New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children.

Ron Cole[1], Dominic W. Massaro[2], Jacques de Villiers[3], Brian Rundle[3], Khaldoun Shobaki[3], Johan Wouters[3], Michael Cohen[2], Jonas Beskow[2], Patrick Stone[4], Pamela Connors[4], Alice Tarachow[4], Daniel Solcher[5]

[1]Center for Spoken Language Understanding, University of Colorado, Boulder
http://cslu.colorado.edu

[2]Perceptual Science Laboratory, University of California, Santa Cruz
http://mambo.ucsc.edu

[3]Center for Spoken Language Understanding, Oregon Graduate Institute, Portland, Oregon
http://cslu.cse.ogi.edu

[4]Tucker Maxon Oral School, Portland, Oregon
http://www.oraldeafed.org/schools/tmos/index.html

[5]*SBC Communications Inc., San Antonio, Texas*
texdanny@swbell.net

ABSTRACT

This article describes our experiences with an animated conversational agent being used in daily classroom activities with profoundly deaf children at the Tucker Maxon Oral School in Portland Oregon. We first articulate some reasons why animated conversational agents can revolutionize learning and language training by providing a more effective mode of human computer interaction. We then describe the capabilities of our animated agent, Baldi, and the software environment used to design and run interactive media systems. We then describe applications designed by teachers and students that illustrate ways in which students in three different classrooms converse and interact with Baldi. We conclude with a brief look at the next generation of animated conversational agents.

1. INTRODUCTION

Advances in interactive language technologies will eventually revolutionize learning and training. Computer-generated animated characters with expert domain knowledge will interact with learners much like the best teachers do today. These computer-generated characters will have instant access to a virtual world of experts with unique knowledge and skills—speech therapists, historians, museum guides, art critics, scientists, philosophers, therapists, historical figures, priests, etc.

We can only imagine how learning will change when students have access to all available knowledge (and often access to the experts or their simulacrums who created it). Learning will be interactive, individualized, self-paced and infinitely variable. So let's anticipate three scenarios:

Julia's ninth grade conceptual physics class includes a learning module on the nature of time. Julia consults her favorite animated agent, Teacher Molly, and explains her assignment. Saint Augustine appears, introduces himself, and asks: What is time?" He proceeds to describe the puzzle of time. After an interesting discussion, Albert Einstein arrives, and provides some excellent descriptions of the relativity of time, with interesting animation showing the relationship between travel and aging. During the discussion with Dr. Einstein, Dr. Stephen Hawking interrupts to point out the three arrows of time. Excited by these mysteries, Julia asks Teacher Molly how one of the arrows, entropy, can be true in an organized world. Teacher Molly suggests that Julia use the "active worlds" program to simulate a hypothetical world. After several simulations, and some discussions with Teacher Molly (who has the annoying habit of making Julia answer her own questions), Julia understands the tradeoff between entropy and biological evolution. For her science project, Julia incorporates her simulation into a discussion between Albert Einstein and Charles Darwin, which is made available to other students.

Denise, a profoundly deaf child, is working with her favorite Teacher Annie, who is helping her produce phrases like warmth and breath with the "th" sound at the end of the word. Teacher Annie is very patient. She likes to watch Denise say each word, then plays *back a video so Denise can see herself saying it too. When Denise is having trouble saying a word, Teacher Annie sometimes turns mostly transparent, and says the same word, showing how the tongue moves in relation to the teeth and lips. After several tries, Denise still does not have it quite right, so Teacher Annie tells Denise that they will continue their practice the next time they meet and in the meantime Denise should think about this speech segment when she speaks. If progress continues to be slow, Denise will be automatically referred to a human speech therapist, who will have available an organized history of her previous therapy.*

Maria is very nervous about her job interview for a summer internship. Although her grades are good, she gets nervous because her English sometimes fails her and she feels awkward. Her friend Jose directed her to a program he developed during his business class. Maria starts the program, and finds herself in an office facing a young man dressed in a business suit. He is quite friendly, and asks her questions about herself and her family. After a while, he asks her about her classes and her interests, and they have an engaging conversation about her interests, his company and how they might work together. After 20 minutes, the young man tells her that he would be delighted if she would accept a summer internship at his company. Maria is both embarrassed and delighted. She had forgotten about the job interview, and realizes she has completed an entire job interview in English. Maria is now more confident and less anxious about her upcoming interview.

The animated conversational agents in these scenarios represent the eventual maturation of efforts now underway to develop and integrate language technologies to approach human levels of conversational interaction in specific task domains. We cannot predict when research breakthroughs in auditory-visual speech generation and recognition, natural language understanding, discourse and dialogue modeling, and social interaction will enable us to develop lifelike computer characters with sufficient linguistic and social competence to realize these scenarios.

What we can do today is develop animated agents using available language technologies and examine their usefulness in specific learning tasks. Given the synergy between basic and applied research, lessons learned with today's conversational agents will provide a useful testbed to guide research and development of more sophisticated agents in the future.

We believe the development of animated agents is a worthy pursuit because they have awesome potential to improve human computer interaction. This is because talking faces are informative, emotional and personable. We consider each of these in turn.

First, human faces are remarkably informative. We communicate best in face-to-face situations because we are able to combine many sources of information to perceive and understand, even when some of the information is ambiguous or fuzzy. This conclusion is supported by experiments showing that a speaker's face and accompanying gestures, as well as the actual sounds of speech, influence speech perception and understanding [1]. Information in the face is particularly effective when the auditory speech is degraded, because of competing messages, noise, limited bandwidth, or hearing-impairment. For example, in many cases when only about half of the words in a degraded auditory message can be understood, adding visible speech allows comprehension to be almost perfect.

Talking faces are so informative 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. On the other hand, the difference between /ba/ and /pa/ is relatively easy to hear but difficult if not impossible to see.

The visual information presented by a talking face is also robust, in that perceivers are fairly good at using visual cues to aid speech recognition even when they are not looking directly at the talker's lips. Speech reading accuracy is not greatly reduced when the facial image is blurred (because of poor vision, for example), when the face is viewed from above, below, or in profile, or when there is a large distance between the talker and the viewer.

Animated faces can also communicate emotional content, a powerful and independent source of information. In recent keynote talks, Eric Haseltine, Chief Scientist of Disney Interactive, has articulated the importance of emotional content in human computer interaction [2]. He notes that artists and producers at Disney Entertainment design storyboards for animated features of each scene within the production using emotional milestones. Disney has designed a language for composing their scenes and achieving the desired results using emotional archetypes based on forms, colors, movements and expressions. 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. He argues that animated agents can express both emotional and intellectual content and thereby increase the amount of information conveyed. This is a novel and important solution to problems of massive information overload and limited bandwidth on the Internet.

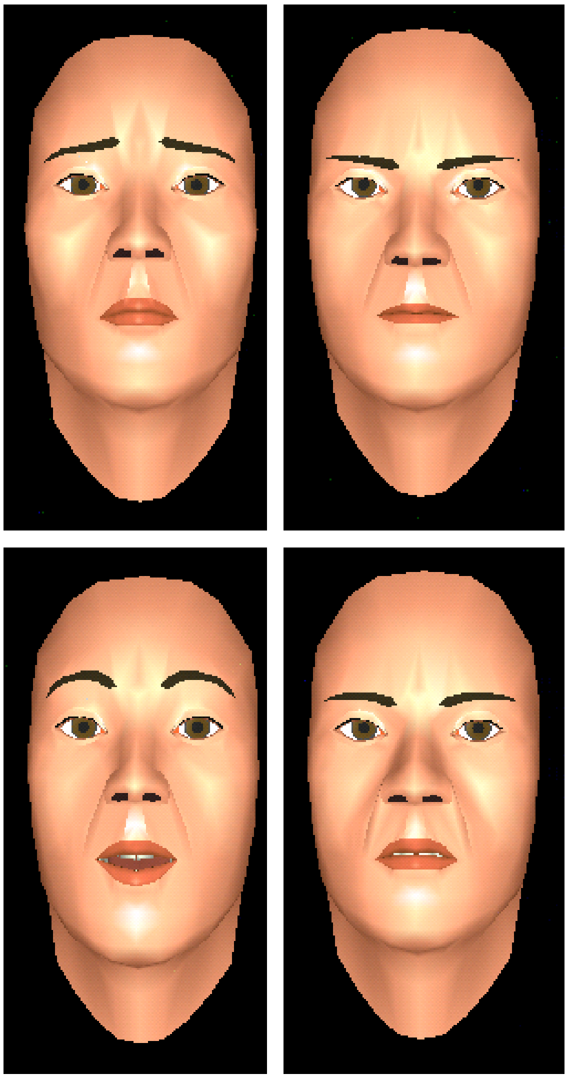

Figure 1: Emotions (top-left: *sad*, top-right: *angry*, bottom-left: *surprised*, bottom-right: *disgusted*).

Third, animated agents bring a personal dimension to human computer interaction. As anticipated from the seminal research of Reeves and Nass [3] 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. Rather than complain about Baldi's speech recognizer or synthesizer, they would note that "Baldi needs to listen better" or "Baldi needs to speak better." 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." In response to "Is he a good teacher?" students replied "Yes, he knows about everything." "I learn from him." "He teaches me how to say words." "He talks slowly and louder like my teacher." "He helps me remember." "I can do it (lessons) many times."

Animated agents can be designed to be attractive and to seem intelligent and knowledgeable, making human computer interaction more natural, meaningful, interesting and fun. 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 these behaviors and incorporating them into conversational agents will provide a more graceful and personable interface.

Finally, we note that animated agents can bring a fanciful and magical experience to human computer interaction. Through the art and technology of animation, animated agents can provide different 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.

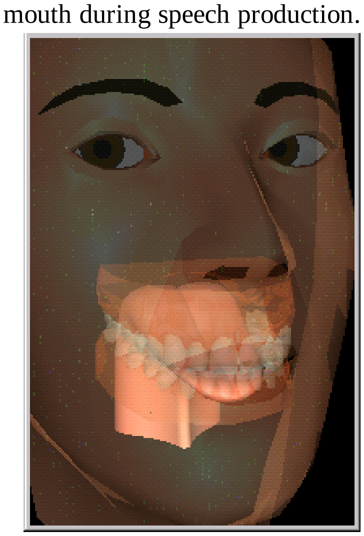

Figure 2: Baldi with visible articulators

2. BALDI AND THE CSLU TOOLKIT

2.1 Baldi

At the Tucker Maxon Oral School (TMOS), teachers, research staff and students use the CSLU Toolkit to design interactive learning experiences for both hearing and profoundly deaf children. The deaf children's hearing is enhanced through cochlear implants, amplification or a combination of both. The students interact with Baldi, the animated conversational agent, through speech, typed input or mouse clicks. Baldi responds to their input using auditory-visual speech synthesis. That is, when Baldi talks, accurate visual speech is presented through facial animation synchronized with auditory speech. Baldi's facial animation can be driven by and synchronized with speech synthesized from text, or speech recorded by a human speaker. By combining natural recorded speech with an image of the teacher or student texture-mapped onto Baldi's 3D polygon surface, Baldi becomes a familiar character. When requested, Baldi's skin can be made transparent or semi-transparent, showing movements of the tongue in relation to the teeth, palette and gums. In addition, Baldi can change facial expressions to display a range of emotions while listening to or producing speech.

Baldi's synthesis program controls a wireframe model, with a control strategy for coarticulation, controls for paralinguistic information and affect in the face, text-to-speech synthesis, and synchronization of auditory and visual speech [1]. Most of the current parameters move vertices (and the polygons formed from these vertices) on the face by geometric functions such as rotation (e.g. jaw rotation) or translation of the vertices in one or more dimensions (e.g., lower and upper lip height, mouth widening). Other parameters work by interpolating between two different face subareas. Many of the face shape parameters such as cheek, neck, and forehead shape, as well as some affect parameters such as smiling use interpolation.

Baldi has been ported Windows (95, 98, NT) and is now available to all educational and governmental institutions free of charge within the CSLU toolkit [4]. Combined with other modules in the toolkit, students and researchers can productively explore problems in visible speech science using an animated talking face.

Compared to the conversational agents in the scenarios above, Baldi and his component language technologies are very primitive. Baldi's visual speech synthesis is actually quite good, since the linguistic content of Baldi's visual speech can be read by not only skilled speechreaders, but also by our students with hearing loss and even those with normal hearing. However, his speech recognition is quite fallible and his conversational skills are quite limited. (His speech recognition will improve considerably in the classroom when new recognizers trained on children's speech are incorporated into the toolkit.) Baldi only engages in structured dialogues under his control. Unlike conversations among individuals, Baldi cannot engage in mixed-initiative dialogues using natural continuous speech. Baldi cannot understand the meaning of an utterance and generate an appropriate response based this interpretation, as humans often do.

Instead, Baldi asks specific questions, and then recognizes (or rejects) specific words or phrases in response to these prompts. Based on the word or phrase that is recognized, Baldi produces a new action or prompt. The speech recognition algorithm is designed to recognize or "spot" a limited set of words or phrases produced by any speaker, and to reject extraneous speech that is not in the recognition vocabulary. When Baldi is driven by recorded speech, the auditory-visual synthesis is highly intelligible and quite natural. When Baldi's auditory speech is driven by the FESTIVAL text-to-speech synthesis (TTS) system [5], Baldi can retrieve textual information from a file or a Web site and produce highly intelligible auditory-visual speech. Although FESTIVAL's auditory synthetic speech is highly intelligible, it sounds artificial compared to a human.

2.2 CSLU Toolkit

Baldi lives and works in the CSLU Toolkit, a comprehensive set of tools and technologies for researching and developing spoken language systems and their underlying technologies [4]. The toolkit integrates the speech recognition, text-to-speech synthesis, and animation technologies that enable Baldi to engage in interactive dialogues. In addition, the toolkit provides the graphical authoring environment that supports development of interactive dialogues that can include media objects such as images and audio files. Since the CSLU Toolkit has been described in several recent articles [e.g., 6, 7] and is available through the CSLU Web site, we limit our discussion to an overview of the main toolkit components used in developing, running and evaluating applications at TMOS.

RAD. The toolkit's Rapid Application Developer (RAD) provides the graphical authoring environment used by students, research staff and teachers to develop interactive media systems that incorporate

Baldi. RAD is easy to use and easy to learn to use. A set of "drag and drop" objects enable users to design interactive dialogues by specifying prompts, recognition vocabularies and actions that followed prompts or user responses. In addition to controlling Baldi's speech and recognition vocabulary at any state during a dialogue, the user can display images, highlight and allow users to click on parts of images, present audio files, and involve Baldi in asking questions about these media objects. A large number of objects have been incorporated into RAD (based on teachers' and students' requests) to make design and evaluation of learning activities easier and more powerful.

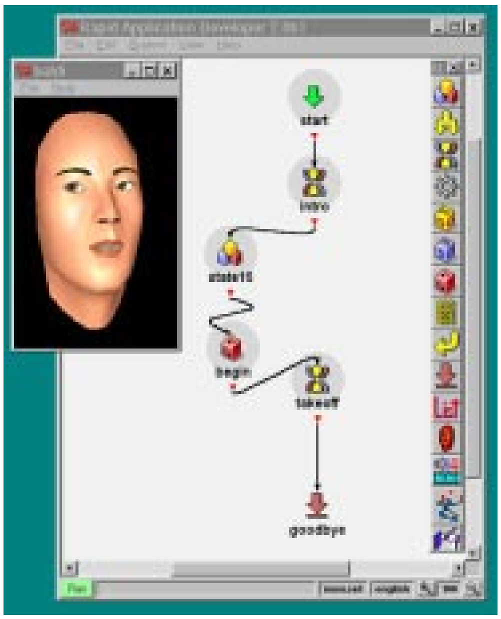

Figure 3: Rapid Application Developer (RAD).

PSL. The Perceptual Science Laboratory provides tools to support research in perception and cognition [8]. PSL provides a user-friendly research environment for designing and conducting multimodal experiments in speech perception, psycholinguistic, and memory. It supports a variety of experimental designs including factorial designs and expanded factorial designs (see [1]) enabling researchers to investigate the manner in which perceivers combine information from different knowledge sources (e.g., speech sounds and facial movements). PSL enables users to manipulate auditory and visual stimuli; design interactive protocols for multi-media data presentation and multi-modal data capture, transcribe and analyze subjects' responses; perform statistical analyses, and summarize and display results. Since PSL tools can be used to teach students to conduct research using the scientific method, offering them new ways to conceptualize problems and investigate the world.

SPAM. Speech Performance Assessment and Measurement (SPAM) is a database program designed to capture and analyze speech data produced during toolkit interactions. SPAM was designed to give an informative picture of deaf students' perception and production within and across toolkit applications. SPAM benefited greatly from the experiences of it's developer Daniel Solcher, an oral deaf person who has been profoundly deaf since the birth, and who overcame many obstacles to become a successful engineer in a hearing world.

3. BALDI GOES TO SCHOOL

3.1 Background (1997-1998)

In October 1997, we were awarded a National Science Foundation Challenge Grant to investigate the use of animated conversational agents for learning and language training with profoundly deaf children. The project is guided by principles of participatory design, in which the end users of the software—the teachers and students—participate in all phases of its design and evaluation. Two of our research staff work on site at TMOS helping teachers and students develop applications. Weekly meetings are held Friday afternoons to review the students' and teachers' experiences and brainstorm on novel ways to apply conversational agents to learning and language training.

Just prior to the start of the project TMOS educators took a short course where they were introduced to Baldi and the CSLU Toolkit. They learned to develop applications using the toolkit's authoring tools, and discussed possible ways of using Baldi in classroom exercises. Five PC platforms (donated by Intel) were then installed in two classrooms and the speech pathologist's office, and Baldi was integrated into daily classroom activities in all subject areas. During the first year of the project over 20 applications were developed and integrated into the daily curriculum in two classrooms [8,9].

The most serious problem encountered during year one was poor speech recognition. The toolkit's speech recognizer, trained on adult speech, performed poorly on children's speech. To address this problem we collected speech data from 1100 children in kindergarten through tenth grade. Recognizers trained on these children's speech reduced error rates on other children's speech by up to 80% compared to performance of the adult-trained recognizers. The kids' speech recognizers are in the new release of the toolkit, described in section 4, and are just being introduced into classroom applications.

In August 1998, TMOS educators, toolkit developers and researchers attended a second short course to learn of toolkit improvements and brainstorm about future classroom applications. On this occasion, we had the pleasure of announcing a donation of 12 additional PC platforms by Intel. One of the main outcomes of the course was the decision to promote a new learning paradigm using the toolkit, in which students design and build their own interactive language systems using Baldi. This represents a paradigm shift from students as users of interactive language systems to students as developers. It was hoped that engaging students in developing applications would expand their understanding of the capabilities and limitations of language technologies and systems, and promote learning, discovery and creative expression of new knowledge through research and development activities needed to build new applications. In the remainder of this section, we summarize experiences in three classrooms at TMOS to this point in time in the 1998-99 school year.

3.2 Classroom Experiences (1998-1999)

In his classroom of students 10-12 years old, teacher George Fortier has been creating interactive media activities with the CSLU toolkit for all classroom subjects. This year, he has transformed his students from users to creators by taking them on a natural course of scientific inquiry.

During the first two months of the 1998-1999 school year, George developed learning and language training applications in which students experienced the range of capabilities available within the toolkit. Next, George revealed to his class how each toolkit object functioned, and worked with the students to develop new learning applications using the graphical authoring tools. Students were then encouraged to work independently in pairs to build similar applications. George continues to introduce the students to new features of the toolkit, but no longer asks them to recreate his applications. Rather, he leaves them with the open-ended task of creating systems that incorporate the new object or feature.

George Fortier is continually exploring new avenues through which his students can express themselves. He finds the toolkit a novel and dynamic medium which not only motivates his students to seek knowledge, but also to show what they have learned in creative, personal ways. In a recent social science activity, George combined content area research on the Iroquois Native Americans with computer skill practice using the toolkit. He showed students how to use the tools to align the animated face with natural speech. He asked them to incorporate this feature into their applications on the Iroquois. Once all groups complete the project, they will vote on the best application, and this will be available from the school Web site.

The 8 to 10 year old students in Alice Davis' class work on the computers individually each day to practice vocabulary and review subject lessons. One day each week they attend a half-hour computer session, in which they learn how to use the toolkit to make their own applications. The instructor usually suggests the computer activity and content themes. For example, one week the class read the story "The Three Billy Goats Gruff." Students had difficulty with the concepts "on the bridge", "over the bridge" and "under the bridge". The teacher decided to use Friday's computer session to develop applications to review these concepts.

On Friday, the students were given paper copies of pictures from the story. As a group, they were asked to place the pictures in the correct order. Once the students had discussed the story and agreed on the correct sequence of events, each child moved to a computer terminal and opened an unfinished template application about the "Three Billy Goats Gruff". This included the first picture of the story and Baldi. The next two dialogue states were on the screen, but empty. After watching how to display a picture and type in a prompt, the students chose their own picture to display next and entered their own text into the blank states. The students were asked to use the prepositions "on", "over", and "under" in their descriptions. Once completed, the dialogues were run and the students made revisions (e.g., moved the picture to a different location on the screen, changed the prompt, or corrected their use of the prepositions.). At this point if the students wished to add to the dialogue, they no longer had a template to fill in. Rather, they used their current application as a model, dragging similar objects to the canvas, entering media and text information, compiling, testing and editing. When the activity was complete, each student shared their dialogue system with the group.

The youngest students using the toolkit are 5 children, ages 6 to 8, who spend half of each school day in the classroom of Kerry Gilley. During the other part of the day, they are in a mainstream classroom. With the help of her students, Kerry creates toolkit applications that reinforce their use of every day vocabulary and sentences. The goal of the applications made during her class is to provide an additional mode through which the students can practice language that is immediate in their lives.

Every Monday, Kerry and her students create a "news" application to which they return during the week for review and vocabulary practice. On Monday each student brings in a news item and related picture to talk about. As each student presents his or her news, the instructor sits at the computer and fills in a dialogue template based upon the child's presentation. After all the students have shared their news, the group scans in the pictures and together run the toolkit dialogues to review the news of the day. Each student converses with Baldi, with the dialogues varying mainly in the content provided by the student. For example:

Baldi: "What is <STUDENT'S NAME> news?"
The picture appears.
User: responds with the news topic (e.g. "Mommy's birthday")
Baldi: "Tell me about that."
User: responds with more information (e.g., "Mommy's birthday is Valentine's Day.")

The students repeat the dialogue at specific times during the week for review.

Presently students in Kerry's class are creating mini-autobiographies using the toolkit. The purpose of the activity is to develop language skills to talk about themselves and to introduce the concept of a timeline by presenting their history in chronological order. For this activity, the children are asked to bring in one photograph for every year of their life. The instructor and assistant scan in the photographs and create an 'autobiography' dialogue template to which the children add content. When the templates are complete, the instructor works with each child to create sentences about the pictures displayed in the dialogue. The outcome of the activity is a dialogue dedicated to and designed in part by each child in which Baldi presents and narrates pictures about each year of the child's life.

3.3 Assessing Outcomes

Procedures are now underway to assess outcomes of using interactive language systems in classroom activities. First, speech production skills of each student are being assessed using a standardized instrument (Assessment of Phonological Processes Revised) administered at the beginning and end of the school year. Second, utterances produced during applications are saved, so that comparisons can be made between utterances produced in the same applications at different times. Third, during the first and second half of the school year, each student is administered the protocol used to collect speech data from 1100 children in grades K-10. This allows to compare speech of the TMOS students with students at different ages in a public school district. In addition to speech production skills, we are testing discrimination of minimal word-pairs. Finally, as part of her Master thesis research, Alice Tarachow is conducting experiments to investigate the influence of the presence or absence of Baldi during applications on both content learning and language skill.

4. THE NEXT GENERATION OF ANIMATED AGENTS

Many of the CSLU Toolkit's principal components have changed significantly since the last beta release. Four modules are of particular interest: text-to-speech system (FESTIVAL), animated agent (Baldi), recognition, and the rapid application developer (RAD).

FESTIVAL. The latest version of FESTIVAL, 1.3.1, has been incorporated into the Toolkit. It features a completely rewritten OGI Residual LPC synthesizer, four American English, two Mexican Spanish, and one British English voice. Festival supports the SABLE standard, an XML based language used to enhance and modify text-to-speech output. In this release Festival is tightly integrated with the Toolkit's TCL core.

BALDI. Baldi has new teeth, a new palate, and two new tongue models. Collision detection leads to realistic deformations of the tongue against the hard palate. The second tongue model was trained on measured ultrasound and electropalatographic data [11]. Baldi can be configured using a simple visual interface that allows modification of physical attributes, including color, emotion, transparency, tongue model, and palate. Face configurations can be named, saved, and recalled. A new set of visemes (facial configurations corresponding to production of phonemes or groups of phonemes) has been added to allow visual synthesis of Spanish speech. A new rendering engine for animating faces has also been developed and will allow for the use of different agents in addition to Baldi.

RECOGNITION and UNDERSTANDING. Both the neural network training and recognition code have been rewritten and are now significantly faster. New adult 8- and 16-kHz recognizers have been trained, along with a number of 16-kHz children's recognizers. An 8-kHz adult Spanish recognizer is also available [12]. A 5000-word speaker-independent continuous speech recognizer has been developed and will soon be incorporated into RAD.

RAD has been massively overhauled in order to incorporate and exploit changes to the underlying Toolkit modules. Some of the major changes are:

- New recognizers, trained using the latest recognition modules, can be added to RAD as simple drop-ins;
- Dialogue states can now be individually configured to use different Baldi, Festival, recognition, and captioning settings;
- Cutting and pasting of dialogue objects and their connections is now possible within RAD;
- Recorded speech can be aligned with animation and saved as part of a dialogue;
- The user interface has been reorganized and clarified;
- Double clicking runs RAD applications;
- A new API to extend RAD's functionality is available. Two packages that use this API ship with the toolkit are: "Tucker-Maxon," which adds support for multimedia desktop interactions and "PSL-Tools," which provides objects that facilitate perceptual experiments.

These advances pave the way for the next generation of animated agents that can engage in more natural conversations in different languages. For example, by combining the new continuous speech recognition capability with natural language understanding using PROFER, the toolkit's robust semantic parser [13], Baldi will be able to carry on natural conversations in specific task domains.

5. 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. We thank Bryan Pellom for his help in putting together this article.

6. REFERENCES

1. Massaro, D., *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*, MIT Press, Cambridge, 1998.
2. Haseltine, E., "The Art of Interfaces," ftp://ftp.wdi.disney.com/pub/incoming/NSFhci.ppt .
3. 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.
4. http://cslu.cse.ogi.edu/toolkit/
5. Black, A., Taylor, P., "Festival Speech Synthesis System: System documentation (1.1.1)," Human Communication Research Centre Technical Report HCRC/TR-83, Edinburgh, 1997.
6. Cole, R., S. Sutton, Y. Yan, P. Vermeulen, M. Fanty, "Accessible technology for interactive systems: A new approach to spoken language research," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA., May, 1998.
7. Sutton, S., Cole, R., et al. "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. http://mambo.ucsc.edu/psl/tools
9. Cole, R., 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.
10. http://cslu.cs.ogi.edu/tm/
11. Stone, M. and Lundberg, A. (1996) "Three-dimensional tongue surface shapes of English consonants and vowels." *Journal of the Acoustical Society of America*, 99, 6, 3728-3737, 1996.
12. B. Serridge, R. Cole, A. Barbosa, N. Munive, and A. Vargas. "Creating a Mexican Spanish version of the CSLU toolkit." *In Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Sydney, Australia, November 1998.

13. Edward C. Kaiser, Michael Johnston, and Peter A. Heeman. "Profer: Predictive, Robust Finite-State Parsing for Spoken Language. *In Proceedings of ICASSP*, Phoenix, Arizona, March 1999.