To appear in Special Issue of Educational Technology, 2006.

How Marni Teaches Children to Read

Ron Cole, Barbara Wise, and Sarel van Vuuren
Center for Spoken Language Research
University of Colorado Boulder

Abstract

Our animated conversational agent, named Marni, interacts with children to teach them to read and learn from text within a comprehensive, scientifically-based reading program called *Foundations to Literacy*. Marni is designed to behave like a sensitive and effective reading teacher. We discuss the scientific rationale for the *Foundations to Literacy* program, the technologies and behaviors that make Marni a credible and effective teacher, and the experiences of children in Colorado classrooms who have worked with Marni to learn to read.

Why and how to use Pedagogical Agents

When human computer interfaces are based on the social conventions and expectations that govern our daily interactions with each other, they provide engaging, satisfying and effective user experiences (Johnson, Rickel, & Lester, 2000; Reeves and Nass, 1996; Nass & Brave, 2005). Such programs foster *social agency,* enabling users to interact with the program like they interact with people. Programs that incorporate pedagogical agents, represented by talking heads or human voices, especially inspire social agency in interactive media (Atkinson, 2002; Baylor et al., 2003, 2005; Mayer, 2001; Moreno et al., 2001, Nass & Brave, 2005; Reeves and Nass, 1996). In comparisons of programs with and without talking heads or human voices, children learned more and reported more satisfaction using programs that incorporated virtual humans (e.g., Moreno, 2001, Atkinson, 2002; Baylor et al., 2003, 2005). Students tend to work hard to please a virtual teacher much as they would respond to a real teacher (Lester et al, 1997.) The demonstrated power of individualized instruction with expert human tutors (Bloom, 1984, Cohen et al., 1982) further motivates the use of realistic animated pedagogical agents in tutoring systems to produce effective learning experiences in homes and schools.

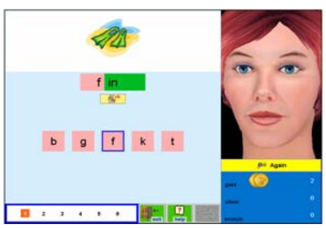
To use agents most powerfully, designers can incorporate suggestions from research about agents concerning speech quality, personality or ethnicity of the agent, or the frequency and verbosity of rewards (e.g., Bouwhuis, 2001; Mayer et al., 2003). Designers can also incorporate what research says about effective human teachers or therapists into the behavior of their agent. For instance, effective expert teachers know a lot about both content and pedagogy specific to the domain they teach, and they scaffold and give appropriate hints to help children find correct answers (Bransford et al., 1999; Ferguson, 1991; Snow, 1989). In our programs we have incorporated both kinds of research to make our agent more powerful. Our 'virtual tutor' Marni gives hints and encouragement to students based on specific errors or error patterns and built-in knowledge about handling these errors. For example, Marni often tells students where to focus so they can self-correct an error (e.g., "You chose 'fat.' Check the vowel sound and look for 'fit'.")

Also, if the student has made several errors on a particular task, and then gets two answers in a row correct, Marni can smile and say “Now you’re getting it!”

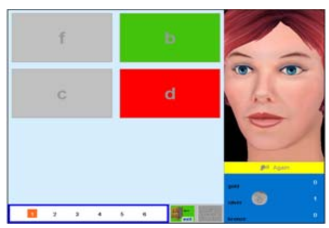
Marni’s Playground: *Foundations to Literacy*

Just like a human teacher, Marni can be effective only with an effective instructional program. *Foundations to Literacy* (*FtL*) is a scientifically- based reading program under development at the University of Colorado (Cole et al., 2003; Wise et al., in press). Most researchers and educators now concur about many aspects of reading instruction, based on the decades of research (McCardle & Chhabra, 2004; Report of National Reading Panel: NRP, 2000; Rayner et al. 2001; Snow, et al, 1998.)

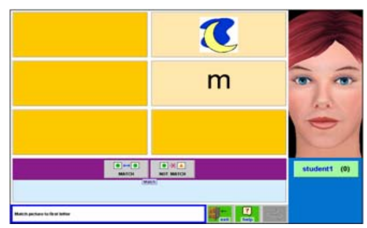
FtL incorporates the five domains and key principles of scientifically-based reading research as summarized in recent meta-analyses of research on preventing and remedying reading difficulties to teach children in kindergarten through 2nd grade to read (NRP, 2000; Rayner et al., 2001). The five domains include phonological awareness, phonics (alphabet, decoding, and spelling), fluency (automaticity with sight words and reading with natural expression), vocabulary, and comprehension. The NRP report also stresses that learning activities be carefully and logically sequenced and structured, and delivered intensively, explicitly, and directly. *Foundations to Literacy* covers them explicitly, systematically and intensively in carefully sequenced ways and it integrates this instruction into engaged reading in books for meaning. All *FtL* exercises aim for *simplicity*, *engagement*, and *empowerment* of success. The experience of success is encouraged by presenting the sequenced items in a “scaffolded” manner (Vygotsky, 1978), ensuring that children start with “comfort level” items from a previously successful level, proceed to supported instructional levels, and end again in “comfort level.” This scaffolding is



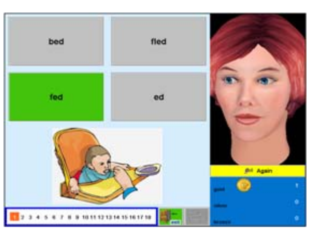
Beginning sounds



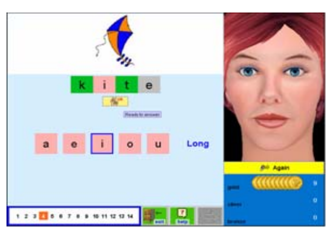
Lower case letters



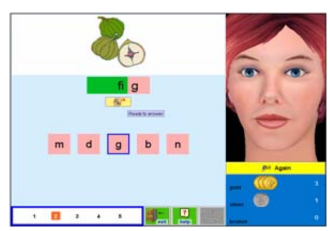
Matching



Word reading



Long vowel sounds



Ending sounds

Sight word spell

Sound blending

Syllable counting

Figure 1. Examples of Foundational Reading Exercises

done seamlessly within the program, as it would be by an expert teacher. The program also incorporates, in each foundational skill domain, a full "cycle of learning" from discovery and practice to competence and transfer (Wise, 2002). After students learn and attain competence with a skill or concept, the program then assigns speeded practice to make the skill or concept become automatic. As foundational reading skills are acquired, the program places students in books designed to apply and transfer these skills to independent reading and writing. Students spend about the same amount of time overall in foundational exercises and interactive books, illustrated in Figures 1 and 2.

FtL monitors each student's performance continuously and adapts the study plan to move the student at his or her own pace through a set of exercises to more advanced exercises and books, or takes the student to lower levels for review and practice. Throughout, a progress graphic allows the student choices about what to do next and gives simple and clear verbal and visual feedback about progress in the program.

Foundational Reading Skills Exercises, shown in Figure 1, teach core competencies needed to recognize words accurately and automatically. Exercise domains include alphabet (letter names and sounds), phonological awareness, letter-to-sound decoding, sight words, spelling (for decodable words and for sight words), and vocabulary. For each knowledge domain, two to four different activities teach and practice the skills and concepts, first to accuracy, and then to automaticity through speeded and time-limited tasks. Each exercise was designed by a team of researchers at CSLR working with teachers and students, using the evidence-based NRP principles described above. Most activities used participatory design methodology, which brings together researchers, programmers, teachers and students to design and test software from the initial stages of development and proceeding through several design-and-test cycles.

Interactive Books, shown in Figure 2, teach students to read fluently and to comprehend text. They enable a wide range of user behaviors and interactions with the pedagogical agent. For instance, Interactive Books allow the story, or any portion of it, to be narrated by the animated character, or they allow the student to click on any individual word or sentence while reading silently or out loud to have it spoken by the agent. They also track the student while reading aloud by moving the cursor to each word as it is spoken. Interactive Books encourage active comprehension by having the student respond to questions posed by the agent, with open-ended recorded questions, by clicking on objects in images or by answering thoughtful multiple choice questions with expansive feedback and hints.

Figure 2. Examples of Interactive Books.

The Technologies that Animate Marni

Our working assumption is that a pedagogical agent that simulates the essential *perceptive* and *generative* behaviors of a sensitive and effective human expert will create engaging and effective learning outcomes. In the systems we have developed, we use 3-D character animation to produce accurate visual speech, and natural facial expression of emotions and movements of the head and face while the agent is talking.

Marni is one of five 3-D models developed at CSLR that produce convincing facial emotions and anatomically correct movements of the lips, tongue and jaw during speech production. These models were developed using CU Animate (Ma et al., 2002) a toolkit designed for research, development, control and real time rendering of 3-D animated characters. Accurate visual speech (Ma et al., 2004a-c) was produced in CU Animate characters using motion capture data collected from markers attached to a person's lips and face while the person is saying words that contain all permissible sequences of adjacent phonemes in English words. The motion capture data for these phoneme sequences are stored in a database and are concatenated (pasted together) to create a representation of the movements of the lips for any English word or sentence. By mapping the motion capture points from concatenated sequences to the vertices of the polygons on the lips and face of the 3-D model, and by applying sophisticated algorithms to assure accurate movements of all associated polygons, the movements of the 3-D model will mimic the movements of a person producing the same speech. This approach produces very natural-looking visual speech, and achieves good visual speech recognition performance in evaluation tests (Ma et al., 2004a-c).

Marni can be made to produce arbitrary movements of the eyes, eyebrows and head using CU Animate Markup Language, CU-AML, an easy-to-use yet flexible and powerful tool for controlling Marni's face movements by marking up text. CU-AML enables designers to control facial expressions and emotions while Marni narrates a text or provides instructions, hints, encouragement or feedback to students in learning tasks. Figure 3 shows the emotions that Marni can simulate while narrating text or interacting with students. Marni's emotions were designed in close collaboration with Dr. Erika Rosenberg, a distinguished researcher in facial expression of emotions, based on the Facial Action Coding scheme developed by Dr. Paul Ekman (Ekman & Rosenberg 1997; 2005).

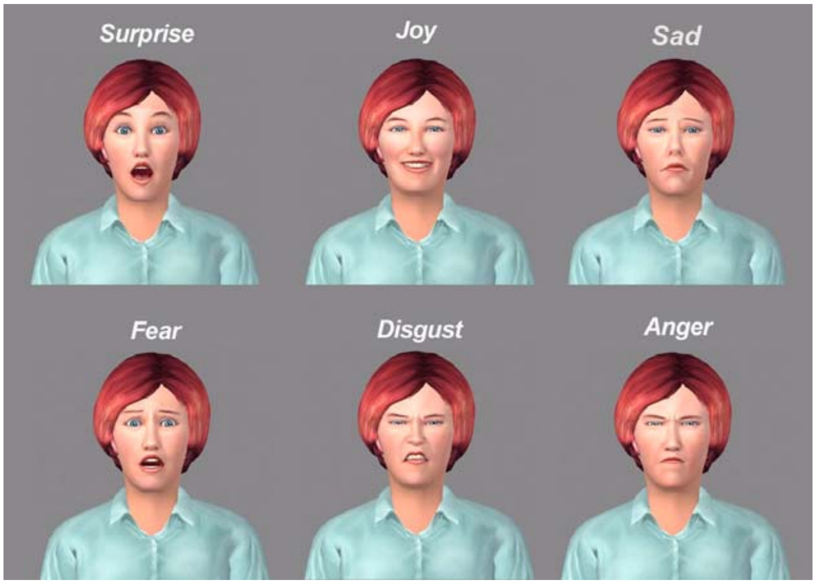


Figure 3. Examples of basic emotions that Marni can model.

We decided to use a natural human "voice talent" for Marni, rather than using text-to-speech synthesis, for several reasons. First, we wanted Marni to have a human personality consistent with the speech patterns of the voice talent. Second, the human voice is a marvelous instrument and we wanted to capture the full range of expression and emotions that a human voice provides. Third, we wanted narration of text to be comparable to a good reader, who uses the discourse structure of the text to emphasize new and important information, and provide appropriate emotional expression. Finally, the use of natural speech has been shown to produce better user experiences and learning outcomes relative to synthetic speech (Mayer et al., 2003).

Computer speech recognition is used currently in *FtL* in two ways; to synchronize the recorded voice with the visual speech produced by the agent, and to track children's speech as they read out loud. The process of making Marni speak is fully automatic—the system automatically synchronizes the recorded auditory signal to the animation given a text string and a recorded voice. First, the text string generates an expected phonetic transcription for the speech. Next, a speech recognizer (Pellom, 2001) aligns the expected sequence of phonetic segments to the speech waveform. The time-aligned phonetic transcription is then used to synchronize the speech to the visual speech targets produced by the animated agent. Speech recognition is also used to provide real time feedback to learners as they read out loud by highlighting each word as it is being read, or moving a cursor to the beginning of the word (Van Vuuren, in preparation).

Marni Works!

Quantitative Outcomes. About 1500 students have used *FtL* in 60 kindergarten, first grade, second grade and a few third classrooms in five urban, suburban, and rural school districts in the state of Colorado. The foundational reading skills of participating students from experimental and control group classrooms were assessed before the intervention began and again at the end of the school year using standardized measures of letter and word recognition. Quantitative analyses indicated that students in the *FtL* condition group in kindergarten and first grades made significantly greater gains in untimed single word reading and letter-identification than their matched controls (Wise, Snyder, Van Vuuren, Schwartz, & Cole, 2004). Our programs have the greatest depth of coverage of foundational skills at pre-reading and early reading levels. We have not yet found similar significant gains over classroom controls with older children, but we are currently working to expand our coverage of skills at these levels and to increase the amount of time reading books, in hopes to extend the effectiveness upwards.

Impressions of Marni. Classroom observations and interviews with teachers revealed strong positive experiences overall. Students worked independently and enjoyed the program, and teachers felt that the program taught what they wanted students to learn, and that they believed that students benefited from it. Some schools embraced the program, and even stopped using all other computer software.

A survey given to 129 kindergarten, 1st and 2nd grade students who used the *FtL* program during the 2004-2005 school years indicated that students enjoyed using the program and perceived Marni as an effective reading teacher. Over 90% of the students gave the highest possible rating scores to questions like “Do you think Marni is a good teacher?” and “How well does Marni help you learn to read?” The interviews revealed that students “bond” with Marni, and feel that she behaves like a real teacher. Over 95% of the students interviewed responded that they wished they could have spent more time on the program.

Summary

Marni, the animated agent in *Foundations to Literacy*, has proved to be an engaging and likeable reading tutor. Students and teachers report liking her and finding the programs useful, and students are using the programs independently, without the support of a human. Future goals of the program are to increase and extend the teaching material to higher levels, to develop new activities designed to assess and train fluent and expressive reading, and to engage students in natural spoken tutorial dialogs with Marni.

Acknowledgments. This work was supported in part by grants from the Coleman Foundation, National Science Foundation, Department of Education and National Institutes of Health: NSF/ITR IIS-0086107, NSF/IERI EIA-012\1201, NICHD/IERI 1R01HD-44276.01. We’d like to thank our colleagues, cited herein, for their contributions.

References

Atkinson, R. K. (2002) Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, 94 (2), 416-427.

Baylor, A. L. & Ryu, J. (2003). Does the presence of image and animation enhance pedagogical agent persona? *Journal of Educational Computing Research, 28(4), 373-395.*

Baylor, A. L. & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*(1).

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-on-one tutoring, Educational Researcher 13, pp. 4-16.

Bouwhuis, D. (August 2001). Multi-agent Environments for Literacy Learning. Paper presented at International Conference on Interactive Literacy Education, Nijmegen, Holland.

Bransford, John D., Brown, Ann L. & Cocking, Rodney R. (Eds.) (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Cohen, P.A., Kulik, J.A., & Kulik, C.L.C. (1982) "Educational outcomes of tutoring: A meta-analysis of findings", American Educational Research Journal, 19, 237-248.

Cole, R. van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S., Wade-Stein, D., Ward, W., Yan, J. (2003) "Perceptive Animated Interfaces: First Steps Toward a New Paradigm for Human Computer Interaction," *in Proceedings of the IEEE: Special Issue on Human Computer Interaction*, 91, (9), pp. 1391-1405.

Ekman, P., & Rosenberg, E. L. (1997; 2005) *What the Face Reveals*. Oxford University Press.

Ferguson, R. F. (1991) "Paying for public education: New evidence on how and why money matters", in Harvard Journal on Legislation, 28, 465-498.

Johnson, W., Rickel, J., & Lester, J, (2000). Animated pedagogical agents: Face to face interaction in interactive learning environments. International Journal of Artificial intelligence in education, 11, 47-78.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Boghal, R. (1997). The persona effect: Affective impact of animated pedagogical agents. In Proceedings of CHI 97 Human factors in computer systems, 359-366. NY: Association for Computing Machinery.

Ma, J., Yan, J., & Cole, R. (2002). CU Animate: Tools for Enabling Conversations with Animated Characters, International Conference on Spoken Language Processing (ICSLP), Denver.

Ma, J., Cole, R., Pellom, B., Ward, W., & Wise, B. (2004a). Accurate Automatic Visible Speech Synthesis of Arbitrary 3D Models Based on Concatenation of Di-Viseme Motion Capture Data. Journal of Computer Animation and Virtual Worlds, Vol. 15(5), pp. 485-500.

Ma, J., & Cole R. (2004b) Animating Visible Speech and Facial Expressions. Visual Computer, Vol. 20(2-3), pp. 86-105.

Ma, J., Cole, R., Pellom, B., Ward, W., & Wise, B. (2004c). Accurate visible speech synthesis based on concatenating variable length motion capture data. *IEEE Transactions on Visualization and Computer Graphics*, 12 (2), 1-11.

Mayer, R. (2001) Multimedia Learning. Cambridge, UK: Cambridge University Press.

Mayer, R. E., Sobko, K. & Mautone, P.D. (2003). Social Cues in Multimedia Learning: Role of Speaker's Voice. Journal of Educational Psychology, 95 (2), 419-425.

McCardle, P. & Chhabra (2004) *The Voice of Evidence in Reading Research*. Baltimore: Paul H. Brookes.

Moreno, R., Mayer, R.E., Spires, H.A., & Lester, J.C., (2001). The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents? Cognition and Instruction, 19(2), 177–213.

Nass C. & Brave S. (2005). Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. MIT Press, Cambridge, MA.

NRP: National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, D.C.: National Institute of Child Health and Human Development.

Pellom, B. (2001) "SONIC: The University of Colorado Continuous Speech Recognizer", Technical Report TR-CSLR-2001-01, CSLR, University of Colorado, March.

Rayner, K., Foorman, B., Perfetti, C., Pesetsky, D., & Seidenberg, M. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*, 31-74.

Reeves, B. & Nass, C. (1996). *The Media Equation:* How people treat computers, television, and new media like real people and places, NY: Cambridge University Press.

Snow, R. (1989). Aptitude-Treatment Interaction as a framework for research on individual differences in learning. In P. Ackerman, R.J. Sternberg, & R. Glaser (ed.), Learning and Individual Differences. New York: W.H. Freeman.

Snow, C., Burns. M., & Griffin (1998). *Preventing Reading Difficulties in Young Children: report of National Research Council (NRC*). Washington, DC: National Academy Press.

Van Vuuren, S., Cole, R., & Ngampatipatpong, N. (In preparation). Providing Feedback to Students while Reading out Loud in Interactive Books. CSLR Tech Report 2006.1.

Vygotsky, L.S. (1978). *Mind in Society*, editor and translator M. Cole, V. John-Steiner, S. Scribner, and E. Soberman. Cambridge, MA: Harvard University Press.

Wise, B. (2002). *Linguistic Remedies for Reading Disabilities*. Boulder, CO: Remedies for Reading Disabilities, Inc.

Wise, B.; Snyder, L., Schwartz, S., Van Vuuren, S., & Cole, R. (June 29, 2004). Interactive Books and Tutors that run “by themselves(?)” in K-2 classrooms. Paper presented at Annual meeting of the Society for the Scientific Study of Reading, Amsterdam.

Wise, B., Cole, R., Van Vuuren, S., Schwartz, S., Snyder, L., Ngampatipatpong, N., Tuantranont, J., & Pellom, B. (In press). Learning to Read with a Virtual Tutor: Foundational exercises and interactive books. In Kinzer, C. & Verhoeven, L. (Eds). *Interactive Literacy Education.* Mahwah, NJ: Lawrence Erlbaum. Available: http://cslr.colorado.edu/beginweb/virtual_tutor/virtual