Toward Natural Child-Machine Communication: Spoken Language Interactions in Preschoolers
"Bridging linguistic, social and technological challenges"

FACULTY: Elaine Andersen, Dani Byrd, Joann Farver, Margaret McLaughlin, Shrikanth Narayanan

The Vision: Our vision is a world in which there is ubiquitous and natural access to information worldwide. Recent advances in information and communication technologies have not been paralleled by similar advances that ensure access to information by everyone, regardless of their age, skills, abilities and preferences. We believe that machines that facilitate interaction, using natural modalities of human communication, can be the most effective in gracefully bridging the gap between information systems and their users. Our vision is to advance science in this direction by developing technology that accommodates the broadest range of human abilities, requirements and preferences. ***We wish to address the problem of natural child-machine communication specifically targeting preschool age (aged 2.5-4) and early elementary school children (aged 4-6).*** While the state of the art in sensory interfaces to information (e.g., machine speech recognition & synthesis), is still not perfect for the adult population, the task of enabling and measuring the effect of sensory technologies for children poses even greater challenges. Children are a crucial segment of the society that will benefit from advances in multimedia information and communication technologies that educate and entertain.

The specific issue: With limited fine motor skills and without the ability to read or write (well or at all), young children are one of the primary potential beneficiaries of computers that use *conversational* interfaces, for example in multimedia games and computer instructional materials, including those designed to promote prosocial behavior. Computer systems interacting with children need to be tailored for these users so that they will understand child intent and so that the child will have a positive and successful experience with the system. But current automatic speech recognition (ASR) and natural language processing (NLP) systems are neither designed for nor successful with preschool children. One reason is that the acoustic characteristics of the speech of preschoolers have not been adequately described. A second reason is that these children are still learning linguistic rules of social and conversational interaction. Their concepts of social structure are still solidifying and are different from those of adults. This means that their behavior in interacting with a computer as an interlocutor is also different from the behavior of adults. Moreover, there are likely to be significant differences in this regard between three- and five-year olds, between boys and girls, and among children of different socio-economic, ethnic and linguistic backgrounds. Although ASR/NLP systems that can accommodate these differences will be of great value to both education and entertainment, to date there has been very little research in this area.

The project: The goal is to establish a highly interdisciplinary research program focused on *Communication, Technology, and Children* to explore realistic child-machine interactions, with the aim of developing and improving spoken language and multimedia technologies for children. Specific initial goals aim to design, collect, and analyze pilot experimental data of young children interacting with machines using natural modalities of speech and gestures, as well as more traditional modalities (e.g., mouse, keyboard, & joystick). Such data are indispensable in helping us refine our research hypotheses and lending credibility to seeking extramural funds to launch a full fledged research program. To solve this complex problem, we have brought together a group of faculty and students representing five programs across three schools: Linguistics, Psychology and Neuroscience in LAS; Electrical Engineering Systems in the School of Engineering, and Communication in the Annenberg School for Communication. This team has the knowledge and skills needed to realistically approach the problem, including expertise in child language acquisition and cognitive development, sociolinguistics, phonetics (speech science), and media effects on prosocial development, as well as computer science and signal processing. USC provides the appropriate setting for this research to move forward, by integrating the SAIL (Speech Analysis & Interpretation Laboratory) laboratory led by Professor Narayanan, the Discourse Performance Laboratory led by Professor Andersen, the Phonetics Laboratory led by Professor Byrd**,** and the research teams led by Professors McLaughlin and Farver.

Research questions

1. Understanding how children interact with machines using natural communication modalities; the investigation and quantification of variability along dimensions of age, gender, socio-economic, ethnic and linguistic backgrounds.

Questions include: How does the spontaneous speech of preschool children compare with that of older children and adults? How do discourse patterns in child-machine communication compare to human-human (child-adult, child-child) communication? What is the effect of cognitive load of task on spoken language patterns? What is the effect of machine recognition and understanding errors on the interaction? Do explicit acoustic and linguistic cues emerge to indicate politeness, frustration, attention etc that can be automatically tracked? How do active child-machine spoken interactions compare to "passive" and "quasi-active" interactions with broadcast media (TV/movies)? Is the child more comfortable using conventional computer interaction modalities such as mouse and keyboard or just naturally talking and gesturing to the machine? Can the child combine multiple communication modalities?

2. Developing spoken language and multimedia technologies tailored to young children, particularly technologies that are automatically adaptable to individual children.

Questions include: What are the acoustic characteristics of young children's speech (at phoneme, word, phrase/utterance, discourse levels)? What are the challenges for automatic speech recognition (ASR) of young children's speech? How does ASR performance compare to human recognition of children's speech? Can the child user's behavior be modeled so as to enable an adaptive and customized experience? What are the effects of oral, visual and multimodal cues in presentation? What is the role of 'anthropomorphism' in the interface design (the role of social functions)? What sort of social status do we need to assign to the interface for maximizing its effectiveness?

3. Developing applications that can enhance the educational experiences of preschoolers, related to cognitive, social and linguistic (including emergent literacy) development.

Questions include: What are the implications for developing successful educational applications (including reading and literacy assessment and development)? How can we use computer technology to structure optimal center-based early learning environments to support children's emergent literacy? How do we design and tailor computer instruction as part of an emergent literacy curriculum that meets the needs of children from diverse ethnic, linguistic, and social class backgrounds? Can we use the computer to encourage pro-social learning? How does children's previous media usage (including educational and entertainment television, video games, computers, and Internet) affect their interaction with computers?