The RAVE: Designing technology-based interactions for deaf infants

Brian Scassellati[1], David Traum[2], , Katherine Tsui[1], , Ari Shapiro[2], Laura-Ann Petitto[3,4]

[1] Social Robotics Lab Yale University, New Haven, CT, USA

[2]Institute for Creative Technologies, University of Southern California, Playa Vista, CA, USA

[3]ITAB - Institute for Advanced Biomedical Technology, and Department of Neurosciences, Imaging and Clinical Sciences – University G. d'Annunzio of Chieti-Pescara, Italy

[4]Next2U s.r.l., Pescara, Italy

[5] Brain and Language Laboratory for Neuroimaging (BL2), Gallaudet University, Washington, DC, USA

[6]NSF Science of Learning Center, Visual Language and Visual Learning, VL2, Gallaudet University, Washington, DC, USA

We identify ways that basic science discoveries in child language development and advances in artificial intelligence, robotics, and human-agent interaction are being brought together to enhance language learning in babies with minimal language exposure. The Robot AVatar thermal-Enhanced learning tool ("RAVE") is an integrated system that makes available multiple components of language in socially contingent and conversational ways within key sensitive periods of early language learning.

Interaction Design: We combine two technologies in a way that minimizes their weaknesses and makes them applicable to this population. Robots have the capability to attract infant attention, but lack the dexterity for producing fluent sign. Virtual humans have the dexterity and flexibility to generate meaningful and fluent sign, but may not be seen by infants as non-interactive television. In a typical interaction with the RAVE system, multimodal sensing (including eye tracking and thermal IR) triggers the robot and virtual human avatar to *start*, *cease*, or *solicit* interaction based on the infant's emotional and social engagement. When an infant's gaze locks with the robot (indicating emotional arousal/attentiveness, and "readiness to learn"), the robot initiates specific rhythmic movements, and, crucially, gaze direction to both maintain interest and direct attention to the virtual character. Robot gaze direction triggers a virtual human to provide nursery rhymes and simple conversation in a visual sign language with speech options.

This system enables fundamental research on questions about language and social development in the following areas:

Robot interaction with infants: Meltzoff et al. (2010) observed that 18 month old babies follow a robot's eye gaze, yet nothing was known about our target group, babies ages 6-12 months. Because deaf babies have acute visual tracking of the human face by 6 months (Petitto et al., 2016), how would their interactions with a minimally-anthropomorphic robot differ? What are the minimum social cues that are necessary to direct language acquisition? Can a physically present robot enable infants to attend to and learn from non-embodied virtual agents?

Virtual Human fluent signing: Virtually all existing signing avatars are difficult for fluent signers to attend to, because of missing elements of signed languages such as obligatory grammatical facial expressions that co-occur with signs and signed sentences. As our approach uses state of the art body and facial scanning and motion capture of fluent signers to create higher quality signing, that it is able to be produced at different speeds and in socially contingent ways, will deaf infants be able to utilize these cues effectively to boost language ability?

Summary: Our work lays a foundation for creating future augmentative communication tools with the potential to provide socially contingent communication for the benefit of all children.