Teaching Language to Deaf Infants with a Robot and a Virtual Human

Brian Scassellati[1], Jake Brawer[1], Katherine Tsui[1], Setareh Nasihati Gilani[2], Melissa Malzkuhn[3], Barbara Manini[3], Adam Stone[3], Geo Kartheiser[3], Arcangelo Merla[4], Ari Shapiro[2], David Traum[2], Laura-Ann Petitto[3]

[1]Dept. of Computer Science
Yale University
New Haven, CT, USA
firstname.lastname@yale.edu

[2]Institute for Creative Technologies
University of Southern California
Los Angeles, CA, USA
lastname@ict.usc.edu

[3]Dept. of Psychology
Gallaudet University
Washington, DC, USA
firstname.lastname@gallaudet.edu

[4]Dept. of Neuroscience & Imaging Sciences
University G. d'Annunzio
Chieti, Italy
firstname.lastname@unich.it

ABSTRACT

Children with insufficient exposure to language during critical developmental periods in infancy are at risk for cognitive, language, and social deficits [55]. This is especially difficult for deaf infants, as more than 90% are born to hearing parents with little sign language experience [48]. We created an integrated multi-agent system involving a robot and virtual human designed to augment language exposure for 6-12 month old infants. Human-machine design for infants is challenging, as most screen-based media are unlikely to support learning in infants [33]. While presently, robots are incapable of the dexterity and expressiveness required for signing, even if it existed, developmental questions remain about the capacity for language from artificial agents to engage infants. Here we engineered the robot and avatar to provide visual language to effect socially contingent human conversational exchange. We demonstrate the successful engagement of our technology through case studies of deaf and hearing infants.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

Social robots; virtual humans; sign language; assistive technology; language development

CHI'18, April 21–26, 2018, Montreal, Canada

ISBN 123-4567-24-567/08/06. . . $15.00

DOI: http://dx.doi.org/10.475/123_4

INTRODUCTION

Children have a tremendous ability to acquire language from observation of and engagement in social interactions. This ability is particularly salient during the first year of life, when infants begin to display the capacity to both understand and generate language through vocal or manual babbling [56, 54, 53, 51]. Children who do not have sufficient exposure to language during this critical period can be at risk for delays in cognitive, linguistic, and social skills that can last for years [65].

While children may have minimal language input for a variety of reasons, children born deaf or with severe hearing loss are particularly at risk and inaccessible to traditional interventions [52]. More than 90% of deaf infants are born to hearing parents with little sign language experience; these parents face the challenge of learning to sign at a rate that matches or exceeds the rate at which their child requires language input [48]. While there are some technology-based interventions that can help restore or support hearing in some children (such as cochlear implants), these technologies often cannot be deployed until after 18-24 months of age and can result in continued language deficits following more than a year of minimal language input [49].

In this paper, we describe the design of a novel technology-based intervention for language learning for deaf infants. We base our selection of both the age range and the specific language stimuli on foundational research conducted as a part of the present study using functional near infrared spectroscopy (fNIRS) of infants across different early-life developmental periods while processing different types of language stimuli [52, 55]. These studies yielded evidence that select brain sites and systems underlying human language learning are sensitive to specific rhythmic temporal patterns central to phonological structure in all world languages (spoken and signed). Further, these neural systems are most negatively impacted by minimal language input specifically during the ages 6 to 12 months [55, 73]. Crucially, an infant's exposure to language stimuli during this period with select phonetic-syllabic rhythmic temporal patterning (frequencies) most powerfully engages these neural sites and systems to support later healthy language, phonological, reading and cognitive growth [55]. We then built American Sign Language (ASL) nursery rhymes with these specific phonetic-syllabic rhythmic temporal patterns and used them as our linguistic stimuli to match infants' biological sensitivity within this precise developmental period.

Providing linguistic stimuli to 6-12 month old deaf infants presents a unique set of design challenges. Our system must be capable of dealing with the developing perceptual, cognitive, and social responsiveness of 6-12 month old infants. Most of the standard application design methodologies, which involve posing detailed questions to the user population, have extremely limited usefulness with this population. The system must also not rely on any type of auditory cueing that is typical for language-based applications.

Because social interaction and context are essential to language learning in children [33, 71, 74], social agents are a natural starting point for our design. The choice of social agent, however, exposes the central design challenge for this work. We might choose a virtual agent, a character on a screen, to provide linguistic input to the infant. Virtual characters have both the manual dexterity and the expressiveness of posture and facial expression to produce sign language at a limited, but reasonable fidelity [68, 28]. However, even exceptionally well-designed educational material presented to infants on a screen have resulted in only minimal learning gains by those infants [61]. Phoneme distinction [33], word referent mapping [32, 31], and lexical category retention [70] all showed only minimal gains when using a screen-based intervention.

We could instead choose to use a physically-present robotic agent. Infants as young as six months of age have been shown to respond socially to physically-present robots, selectively following the robot's gaze and engaging the robot with social overtures [45, 7]. However, these robots lack both the physical dexterity and expressiveness to mimic even a limited set of signs. Typically, robots often have only 2-3 degrees of freedom in their hands because of space and cost requirements, far less than is required for most signed languages [30]. Expressiveness displayed through posture and facial expression is similarly limited, with only a few systems approaching anything close to what visual languages require [3]. Even robots that have been designed specifically to produce signed language [72, 41] or to act as tutors for signed languages [29, 78] often fail to have the full range of manual dexterity and expressiveness required. Most significantly, all of the robots developed to date for signing lack the temporal patterning and fluidity of motion that is a critical part of infant sensitivities and none were designed to be appropriate for infants due to their size and/or appearance.

The solution that we present here seeks to use both a screen-based virtual human and a physically-present robot to engage infants in triadic interactions (see Figure 1). Our hypothesis is that the pair of agents will provide both the dexterous, expressive language production (via the virtual human) and the socially-engaging physical cues that trigger learning (via the robot). The following section reviews the related work on infant-technology interactions, on socially-assistive robots for child tutoring, and the use of virtual avatars for language instruction. We then describe the system design, implementation, and evaluation via three case studies that highlight aspects of our design.

Figure 1. Typical setup of the complete system from the infant's point of view. Here the robot and avatar (left and right) can be seen with the eyetracker (bottom center) and thermal IR camera (top center).

RELATED WORK

Infants and Technology

Prior work explored the pedagogical and social functions of screen-based and physically-present technologies for infants. Anderson and Pempek [6] posited that infants and young children learn less well from a demonstration on a television or video program than from similar real-life experiences. This "video deficit effect" has been widely replicated and found to exist in several aspects of language development including phoneme distinction [33], word referent mapping [32, 31], and lexical category retention [70]. However, DeLoache et al. [15] found that young children (age 30 months) were better able to learn new words from watching a video when its content was of two people's social interaction than when a single person appeared to directly address the child, emphasizing the important role social interaction plays in infant language development.

In terms of physically-present technologies, Meltzoff et al. [45] found that toddlers (at approximately 18 months of age) were more likely to follow the gaze of a robot that had been seen interacting in socially communicative ways with an adult. Similarly, Arrita et al. [7], using a looking-time paradigm to gauge expectation, found that infants expected robots to be spoken to like a human, but only if the robot had been seen previously to interact with a human interlocutor. Together these findings suggest that infants can attribute social agency to robots.

Socially Assistive Robots

Robots that provide social and cognitive support, rather than engage in direct physical manipulation, are called socially assistive robots [16]. This type of robot is often a desktop-mounted system that has a small number of degrees of freedom for expressive movement but is incapable of locomotion or manipulation. These robots have been used in a variety of tutoring and educational domains, including providing emotional support to children undergoing blood draws [20], teaching children how to deal with bullies [36], tutoring basic mathematic problems [60], and supporting social skills tutoring for children with autism spectrum disorder[66]. Language instruction has been particularly successful with older children (typically 4-6 year olds) with these robotic systems; Spanish-speaking children struggling to learn English in public schools showed 2 standard deviations of improvement following a 5-session intervention with a robot tutor [38]. The present work unites a multidisciplinary team to extend beyond previous work by considering both a much younger population and a specific population that has not been previously addressed (deaf infants).

Virtual Agents

Virtual Humans [63] combine an animated humanoid avatar body with a behavior control mechanism, to create a human-like animation interface to computing. Virtual humans have been successfully used for a number of purposes including education [62, 19, 24], and practicing human interaction activities [64, 17, 58]. There have been several recent efforts to create signing virtual humans (e.g., [57, 79, 23]), however so far they have focused on only manual sign-shape, rather than the full complement of necessary synchronized behaviors for grammatical production, including facial expressions [28, 25]. Virtual humans have been successfully used with young children for educational purposes [75, 26, 76], however, to our knowledge, not previously with deaf infants.

OVERALL SYSTEM DESIGN

As we knew of no other interactive technologies developed for this population, we followed an integrated, iterative design methodology. We tested prototypes and partially-operational systems with infants as we iteratively improved our hardware designs, software systems, and interaction goals. A sketch of the individual system components and their relative positioning can be found in Figure 2.

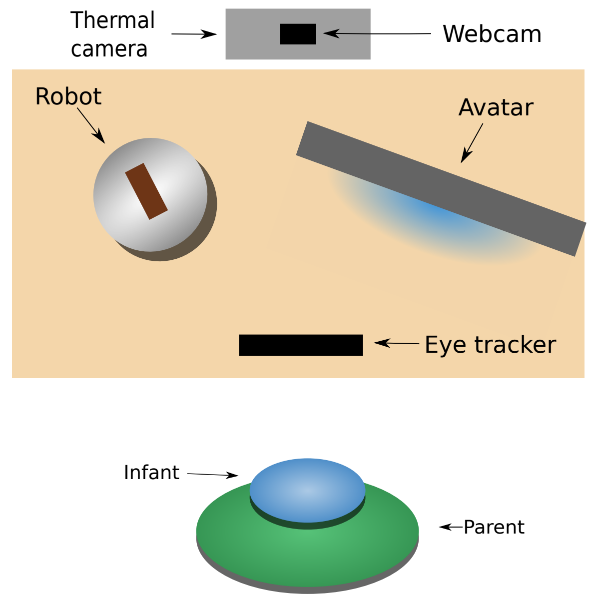

Figure 2. An overhead view of the experimental setup. Infants were seated on the parent's lap during interactions with the system.

The overarching aim of the joint avatar-robot system is to provide language exposure during a critical period for language development, thereby mitigating developmental delays due to minimal language experience in deaf infants. To this end, the design of the system was guided by two goals: 1) to provide socially contextual language through the interactions of a robot-avatar-child triad, and 2) to maintain child attention for reasonable age-appropriate periods of time. The system behavior should be contingent on the behavioral and arousal state of the infant. Appropriate social interaction scaffolds the development of social [21] and higher cognitive functions [44], and social contingency in particular is hypothesized to be critical to language acquisition [52]. Maintaining attention is important because if the infants are not sufficiently engaged then language learning is impossible to accomplish.

The following sections describe our design process for the robot, the virtual human, the perception system, and the interaction controller. We discuss not only the final product that we implemented, but some of the design guidelines that we uncovered during our iterative design process. We conclude with sections on the evaluation of the complete system using a set of case studies drawn from interactions with infants throughout our iterative design process.

ROBOT DESIGN

The robot's primary purpose is to capture and direct the infant's attention to the avatar. The robot used here is a modified version of Maki v1.4 by Hello Robo [50], which is a 3D printable, open source robot (Figure 3). Maki's compatibility with the design guidelines we identified as relevant for infant-robot interaction made it a strong candidate as our base robotic platform. Described below, these guidelines were drawn from research fields including animation, developmental psychology, and human-robot interaction.

Design Guideline 1: Attractive

The robot's appearance needs to be salient enough to capture a deaf infant's attention without impeding its ability to transfer the infant's attention to the avatar when appropriate. To support this, we chose an anthropomorphic appearance with accentuated infantile characteristics. Standing 13.5-inches high on a small torso, we enlarged the eyes and minimized the other facial features. Infant-like appearance has been posited to be particularly salient to all humans [42], and has been empirically demonstrated to be highly salient to children as young as 3 years old [9].

The robot relies on the infant's ability to follow gaze. Infants as young at 6 months are capable of following gaze direction, although this skill is more accurate in infants aged 12 months and older [10]. Through means of joint attention episodes, the robot influenced the infant's attention. The robot has an articulated head (pan left/right, tilt up/down), articulated eyes (pan left/right, tilt up/down) and eyelids (open/close) to enable naturalistic directional shifts of attention. Once the robot had captured the infant's attention, it turned its head and eyes to look at the avatar. We believe that the robot having physical eyes and eyelids (as opposed to screen animated facial features) are important for infants to consider it as a social and communicative agent.

Design Guideline 2: Simplified

As the primary purpose of the robot is to direct attention to the avatar, we seek to minimize morphological features and behaviors that might distract the infant. We chose to hide the large, dark circular "ears" of the robot as these high contrast variations were particularly distracting to younger infants. By eliminating the color contrast difference, the robot's eyes remained the most salient feature, especially when the head turned toward the avatar. While we also created a version of the robot that added a pair of 2 degree-of-freedom eyebrows, we chose not to use the added expressiveness that these features provided because of the additional salience that they added.

It is imperative for the robot to act as a supporting actor to the avatar, like a sidekick to a hero. (This relationship in human-robot interaction was suggested initially by Vázquez [80].) We limited the behavioral repertoire of the robot to a few core behaviors: directing attention through joint head and eye movement, responding to overtures from infant and avatar, showing surprise, and engaging the infant with a game of "peek-a-boo". To maintain novelty, we opted to provide small variations to these core behaviors rather than attempting to provide a richer set of behavioral options.

Design Guideline 3: Identifiable

As eye gaze was the robot's primary method of interaction and communication with the infant, it was imperative that the robot's gaze direction and head orientation must be easily identifiable to the infant at all times. While human-robot gaze has been studied extensively (see Admoni & Scassellati [1] for a review), none of this work specifically has targetted infants. In human-to-human interactions, infant gaze has been studied extensively. Prior to age 14 months, infants largely look at an adult's head orientation as an attention following mechanism [13]. Wilson et al. [82] suggested that the change in profile and nose angle from center might be cues that humans use to determine head orientation [34].

We made three primary modifications to the base Maki robot to accentuate head orientation and gaze direction. First, we increased the color contrast of the eyes in order to render them more visible. Second, we added a strip of faux-fur material to the top of the robot's head to give the appearance of a stripe of hair. As the robot's head was nearly spherical and the other directional features (nose and mouth) were only minimally represented, this feature provided a key mechanism for discriminating head orientation. Third, inspired by principles of studio animation, we exaggerated the robot's movements and behaviors to exploit the full range of physical motion that the robot was capable of performing. While we did not test the effects of these modifications individually with infants, each provided our own (adult) staff a noticeable increase in performance in estimating head orientation and gaze direction.

Design Guideline 4: Agentic

In order to sustain the idea that the robot is a social and communicative agent, the robot must further demonstrate intention, respond appropriately and in a timely manner to interactions towards it for an appropriate duration, and demonstrate passive life-like movement when resting in between interaction periods. We constructed a simple behavioral repertoire for the robot that supported the illusion of agency. This included both goal-directed actions (e.g., looking toward a target location), expressive actions (e.g., showing "surprise" by opening the eyelids wide and drawing the head slightly back), and idle behaviors that were active at random intervals whenever other activities were not being performed (e.g., blinking and small shifts in gaze position). The length of all of these behaviors were limited to be no more than a few seconds both to accommodate the relatively attenuated attention span of infants and to allow more rapid responsiveness to actions by the infant or avatar.

Finally, we found during the course of our design and evaluation that transitivity was maintained with the robot. As Brooks and Meltzoff noted, when known social agents (human experimenters) treated the robot as a social agent, then toddlers would tend to attribute agency to the robot [45]. We used a familiarization protocol during which the researcher was instructed to act as if the robot and the virtual agent were social agents by greeting them individually, making eye contact, responding to social gaze from the system, and engaging the system in a simple social exchange.

Design Guideline 5: Safe

While the robot was generally kept out of reach of the infant, we considered physical safety during the design of the system. We eliminated potential pinch points in the robot hardware, used a physical design that was large enough to be difficult for an infant to grasp and lift while still maintaining a lightweight frame, and limited the speed of movement to limit possible accidental contact. Electronics are maintained out of reach and the robot can be wiped down with disinfectant wipes between children.

AVATAR DESIGN

The virtual human's primary purpose is to provide visual language stimulus to the infant, and, along with the robot, to engage the infant in a contingent social interaction. Thus, some of the design guidelines are similar to those for the robot, while others are complementary. Like the robot, the avatar should be attractive and agentic. Rather than striving for a simplified appearance, the goal of providing high fidelity sign language stimulus pushed in the opposite direction - a very human-like avatar that is capable of performing fluent visual

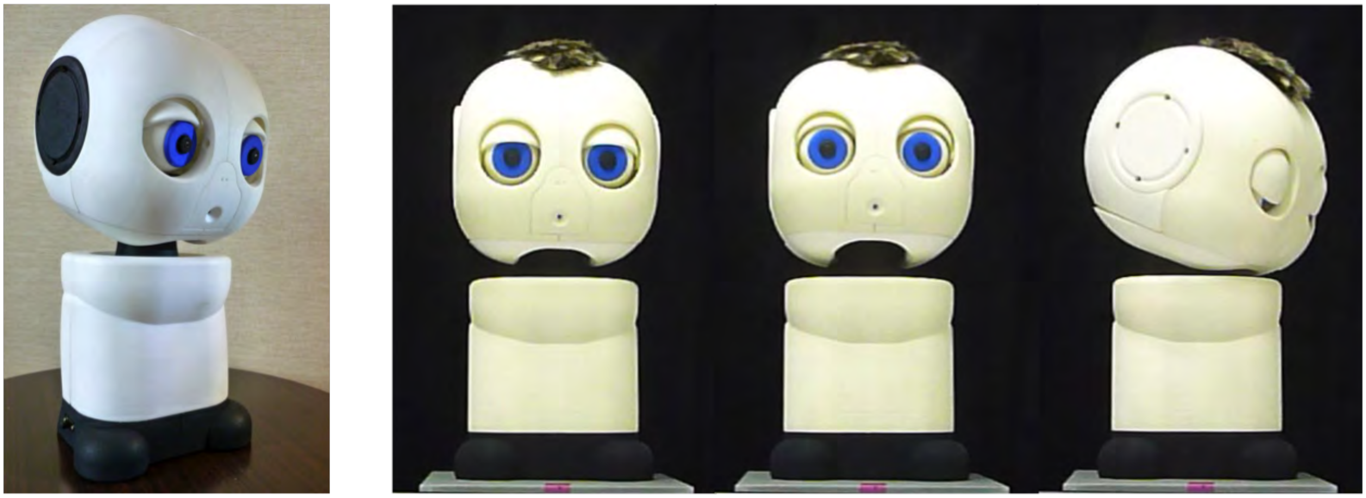

Figure 3. The robot used in this study (shown in three figures to the right) is based on the open-source Maki platform from Hello Robot (shown at left). Modifications have been made to accentuate infant-like features, to reduce distractions, to increase the saliency of the directionality of head positions, and to support judgments of animacy and agency.

sign language samples. Safety is not a concern, given the agent is in the virtual world (the concern is rather with the conventional display hardware, rather than the avatar itself).

To meet the goals of attracting the infant and being able to perform realistic human-like behavior, we decided to base the avatar on scans of a young deaf female native signer of American Sign Language (Figure 4). This choice also allows use of motion capture of the same individual, which increases the ability to target the motion to the avatar. We briefly describe the process for creating the avatar model, behaviors, animation, and control of the avatar.

As above, specific rhythmic temporal patterns underlying phonetic-syllabic organization in all languages served as the structural template on which the specific ASL nursery rhymes used in the present studies were built [55, 54, 53, 52].

3D model construction

The avatar was constructed by capturing a native signer inside a photogrammetry cage using 25 megapixel DSLR (Digital Single Lens Reflex) cameras (Figure 4). The 3D body model was then reconstructed using photogrammetry software [2] to create a 3D body model. A virtual skeleton was then added to the 3D model to allow for articulation and deformation using linear blend skinning. A set of facial scans was also captured using a lighting cage [14] and used as reference imagery. A set of joints were added to the avatar's 3D face to allow for deformation and movement, as well as 3D models for the eyes, teeth and tongue.

Motion capture

Motion data were built with the collaborative resources of Gallaudet University's Motion Capture Laboratory (M. Malzkuhn and J. Lamberton) and the University of Southern California's Institute for Creative Technologies (A. Shapiro) produced through full body capture via a camera-based motion capture system (VICON). The raw motion data was then re-targeted [18] onto a skeleton that matched the topology of the avatar model. Facial animation was manually keyframed by an animator. A variety of nursery rhymes, conversational fillers (e.g., "yes"), short utterances ("What's that?"), and idling poses were captured and processed.

Realtime animation and control

The 3D avatar was animated and controlled through the use of a real-time character animation system [69]. The animation system includes control mechanisms for playing prerecorded animations, gazing, and head movements (such as nodding and shaking).

PERCEPTION MODULE

In order to enable an ecological interaction between the infant and the artificial agents, a perceptual system was constructed based on thermal infrared (IR) imaging, eye tracking, and machine vision. The details of this system are beyond the scope of this paper, but we include here a sufficient description such that the evaluation of the overall system can be understood. The perceptual system is able to detect an infant's key behaviors and physiological states, classify them on the basis of an integrated theoretical model, and trigger robot and avatar behavior in response to infant behavior.

Thermal IR imaging allows the system to detect subtle changes in the infant's internal state, which are significant for discriminating when the infant is engaged with the interaction. Facial thermal patterns depend on subcutaneous vessels transporting blood heat, and these vessels regulate blood flow via local vascular resistance (vasodilation and vasoconstriction) and arterial pressure [5, 46]. Therefore, by recording the dynamics of the facial cutaneous temperature, it is possible to assess autonomic nervous system activity and infer the subject's emotional state [47, 22, 43, 46].

We choose the nasal tip as the salient region of interest (ROI) for assessing the psychophysiological activity of the infants,

Figure 4. The generation of the 3D avatar model. Left: The participant inside a photogrammetry cage. Right top: The 3D avatar model generated from photogrammetry capture. Right bottom: Light Stage facial scans used for reference for avatar construction.

because of its strict neurovascular relationship with adrenergic activity associated with expression of emotional states [22, 43]. The nose tip's average temperature was extracted from each frame thus obtaining a temperature signal in real time. The dynamics of the temperature was used to classify the arousal state of the infants by assuming that a decrease of temperature is linked with a sympathetic-like response (associated with distress and disengagement) whereas its increase is due to a parasympathetic prevalence on the subject autonomic state (related with interest and social engagement)[43].

Thermal IR imaging is performed by means of a digital IR thermal camera FLIR A655sc (640 x 480 microbolometer FPA, NETD: $<$ 30 mK @ 30 °C, sampling rate: 50 Hz). In order to preserve the ecology of the recording in challenging situations like experiments with infants and toddlers, the facial ROIs need to be automatically recognized and tracked in all the frames of the thermal video. The original solution developed includes three different processes: i) automatic recognition of facial landmarks in the visible domain using the Open Face library [4]; ii) frame-to-frame tracking of the ROIs in the visible domain by referencing them with respect to the facial landmarks; iii) co-registration of the visible ROIs with their corresponding ROIs in the thermal videos.

Infant gaze was also used as an assessment measure for triggering behavioral responses from the robot and avatar. A Tobii Pro X3-120 eye tracker measured the eye gaze and duration as an indicator of an infant's focus of interest. Eye movement data were collected every 17 ms (for an effective sampling rate of 60 Hz), and identified by a time stamp and (x,y) coordinates. The eye tracker was controlled by a customized Python algorithm perceiving the position of the infant gaze on the experimental setting, i.e. if the infant was looking at the robot, at the avatar screen, if it was looking between the two, or looking somewhere else.

INTERACTION DESIGN

The interaction design follows the Information State approach to dialogue management [77], in which a set of typed information variables are monitored and updated and a policy is used to decide on behaviors given the current values. Unlike most interactive dialogue systems, the states do not refer to aspects of natural language dialogue context, but rather the perceptually informed status of the infant. The goals of the system are to use the robot and avatar to engage the infant's attention, initiate visual language episodes when the infant is engaged with the social agents, and react contingently to the infant's initiatives to either of the agents. The infant's visual attention state (robot, avatar, or neither) was primarily derived from the eye tracker output. The infant's engagement was modeled as one of four states (from very positive to very negative) based on classification of the thermal signals [11].

The main idea is for the robot to attract the infant's attention and, assuming the infant is engaged, direct the infant's attention to the avatar, who will provide a nursery rhyme in American Sign Language to the infant. When the infant is not attentive, the agents provided social routines in order to re-engage the infant, or interacted with each other before redirection of the interaction to the infant.

In the first integration of robot and avatar output (used for case study 1, below), the perceptual components were not yet trained on appropriate data and available for real-time use. We thus used a "Wizard of Oz" interface to allow human controllers to select specific behaviors for the robot and avatar, following a rough protocol based on the above guidelines. This process allowed full human perceptual abilities and decision making to guide selection of behaviors, but allowed evaluation of whether the pair of agents and set of behaviors could be used to engage the infant as desired.

The next version (used for case study 2, below) was fully automated, including a dialogue policy written in Python, that was informed by the thermal imaging and eye-tracking perception, as well as physical states of the robot and avatar. Each combination of information states led to behaviors by robot and avatar (sometimes just waiting, sometimes complex, synchronized behaviors, such as the robot directing attention to the virtual human who then starts the rhyme).

As observed in the case studies, infant behavior, not just attentional focus and arousal, are important in establishing contingency - in particular communicative and social behaviors toward the agents must be reacted to in order to establish contingency. This represents a challenge, because the state of the art of recognition of infant's posture and communicative behaviors is not as well developed as that for adults, and most ready-to-use perception systems are not trained on infant data. In order to address this, we have added additional visual perception components, initially to record data, with a goal of later customizing recognition components based on this data. For the third phase (used in case study 3), we introduced a semi-automated behavioral policy that relies on a human observer to signal specific complex perceptions (beyond arousal and attention), but uses the automated controller to make avatar and robot behavior decisions based on the updated state.

EVALUATION

The development of this system took place over three years. We tested the system, or parts of the system, with infants in three pilot sessions designed to evaluate the overall efficacy of the system as well as additional experiments focused on specific scientific investigations which are not reported in this paper. Overall, we have seen more than 68 infants with some variant of this system for this and related studies, but report here on three case studies drawn from a total of 36 infants recorded during our system integration and evaluation sessions. The design was revised incrementally after each integration session before the next testing phase. Because of the difficulty in recruiting participants who were within a limited age range (6-12 months), were part of a limited population (deaf infants), and were available during the times when testing was planned, we at times tested many of the components and early versions of the system on hearing infants with no sign exposure or hearing infants with sign exposure (typically, through deaf parents). Evaluation with these hearing infants was essential for our other scientific goals, but they were not the population this system was specifically designed to support.

The focus of this paper is the design of the unique dual-agent system for social exchange with a previously unstudied population in human-computer interaction. To evaluate this design, we consider a few case studies, drawn from different points in time, that demonstrate how the system evolved. Controlled experimental evaluation of the system (and subsystems) have also been conducted for specific research questions focused on language development and the application of the physiological perception module; these results will be presented in publications focused in other communities. Some of the infants described below took part in those other studies, but our discussion here is focused on the qualitative performance of the system and not on specific empirical questions.

Case Study 1: Free-play, Wizard-of-Oz

The first case study was drawn from a sample of infants recorded during August of 2016. Following a controlled experimental protocol that focused on the ability of the robot to guide infant attention to the avatar's screen, a free play session was introduced. The experimental protocol used pre-scripted behavioral sequences from the robot and virtual agent. Our goal with the free play session was to understand if infants found the system to be engaging when both the robot and the avatar performed in socially contingent manners towards the infant and also towards each other, even though the system could not yet achieve this result autonomously.

Infant "Albert" (Figure 5), a deaf sign-exposed male, aged 13 months, was the first deaf infant to interact with this system. During the free play session, we controlled the robot and avatar using a "Wizard of Oz"-based open-loop controller. The behaviors for the robot and the avatar were operated by two human "wizards," and a third human "conductor" coordinated the timing and execution of the robot and the avatar's behaviors together. It was the conductor that created the illusion of social contingency based on the infant's direction of attention and emotional state. The conductor was responsible for providing additional brief periods of rest if the infant became overstimulated or agitated during the free play session, and ultimately controlled the duration of the free play session and its conclusion.

Albert was very engaged with both the robot and the avatar. The longest period of uninterrupted engagement (based on gaze toward robot or avatar) lasted roughly one minute. In between periods of engagement with the system, Albert turned to his mother (deaf) in several social reference episodes; he tapped his mother hand and signed "mother" at one point to capture her attention on the avatar and/or the robot. He never showed signs of distress during the procedure. During the engagement period, Albert pointed at the robot on two separate occasions.

Our experience with Albert demonstrated that the language generated by the avatar was effective for children at the up-

Figure 5. Case study 1 infant "Albert" pointing. Here the infant, seated on his mother's lap, as seen from the system's point of view.

Figure 6. Composite image of case study 2 infant "Bella". Here the infant can be seen from multiple angles (bottom left and right) interacting with the autonomous system (top).

per limit of our age range who already had exposure to sign language. We also saw successful engagement by both robot and avatar, successful direction of attention by the robot to the avatar, and attempts to share the robot with his mother (via declarative pointing). This interaction demonstrated to our team that this system could be successful even with older infants (typically known for showing briefer tolerance in controlled experimental paradigms) and that we may continue to expect only short interactions. This also demonstrated that the inadvertent noises made by the robot (motor noise) were not critical to the ability of the system to engage an infant, and supported the use of design guidelines drawn from hearing infants. Based on this case, and others obtained during this same testing sample, we opted to begin to incorporate autonomous response and eliminate our use of the wizards.

Case Study 2: Freeplay, introducing limited autonomy and perception

By February 2017, we had integrated limited parts of the perception system to allow for an interaction guided by perception, but still triggered by a hidden human operator. This iteration of the system used arousal and gaze information collected from a thermal camera and eye tracking camera, respectively, to trigger specific behaviors in the avatar and robot. The behavioral repertoire of the robot and avatar were relatively simple. The robot had one behavior designed to engage the infant (a "peek-a-boo" style engagement in which the robot showed a startle response when the infant looked at it) and one behavior designed to direct attention toward the avatar. The avatar would engage in sign-language nursery rhymes following the robot gaze-direction behavior and would be quietly idle at other times.

Infant "Bella" (Figure 6), a hearing, sign-exposed female aged 8 months, was among the first infants to interact with this autonomous system. The entire session lasted approximately 4 minutes and 20 seconds before Bella completely disengaged. The longest period of uninterrupted engagement lasted roughly three minutes, three times longer than the most engaged infant from the freeplay sessions. While Bella visually tracked both the robot and the avatar, she displayed a strong preference for the robot and produced numerous robot-directed pointing gestures. Remarkably, at one point she appeared to copy the robot's 'startle' behavior – pitching her head down with closed eyes, followed by rapid upward pitch while concurrently opening eyes in a startle response – directly after it had been produced by the robot.

This interaction was highly successful in many ways; the semi-autonomous system enabled even longer engagement and interaction than we had been able to achieve with a strictly human-controlled system, the coordination between the perception system, robot, and avatar allowed for smooth transitions of the infant's attention from one target to the other, and we even saw an example of robot-to-infant copying. However, this interaction also pointed out two significant deficits of our system. First, there were numerous attempts by Bella to engage the system in ways that we had not anticipated, and the system therefore failed to respond to them. Bella at times clapped her hands together in excitement to try to generate some kind of response or engagement with the robot, but it failed to respond. Her copying of the robot's startle behavior was also a missed opportunity for our system. Second, when Bella was not engaged with the system, both the avatar and the robot were still and passive. This resulted in a break in the illusion of agency for both agents, as both appeared to be less social as they sat still and unmoving.

Case Study 3: Greater autonomy with tetradic interaction

To address the deficits that we saw in the interactions with Bella, we made substantial improvements to both the interaction design (to account for these "still" periods) and to the perception system. During periods where the infant is distracted or otherwise not attending to the system, we trigger interactions between the robot and the virtual human. This maintains the animacy of both agents while providing social interaction exemplars to the infant that might serve as points of engagement.

Infant "Celia" (Figure 7), a hearing, sign-exposed female, aged 11 months, was one of a cohort of 23 pilot-study infants that interacted with this updated system during June, 2017. For the

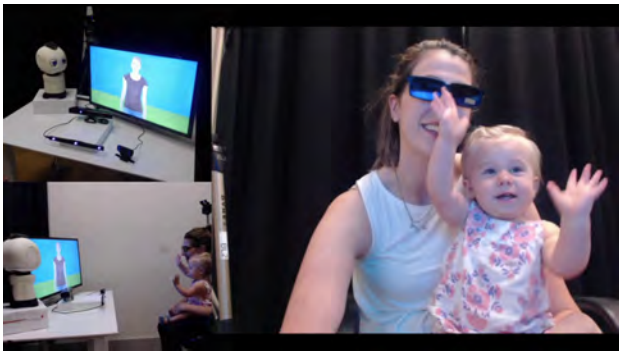

Figure 7. Composite image of case study 3 infant "Celia". Here the infant (right and bottom left) can be seen producing a waving-arm gesture towards the system (top left).

first time, we also allowed the parent to participate in the infant-robot-avatar interaction half way through the session in order to observe how this system might be used *in vivo*. The session lasted approximately five and a half minutes with the longest period of uninterrupted engagement lasting roughly a minute and a half. Celia visually tracked both the robot and the avatar, and made a number of robot-directed manual productions including pointing gestures and attempts to produce parts of what the virtual agent had signed (as interpreted by sign-fluent observers). During the parent-engaged interactive phase of the session, Celia's engagement appeared to increase, producing many instances of socially communicative gestures, including hand-clapping, waving and pointing. Celia was not unique in reproducing the virtual human's signing, and this was observed even in some non-sign-exposed infants. While Celia did not attempt to copy the robot's behavior, other members of the cohort did.

Perhaps the most interesting thing that we observed with Celia was the production of the virtual avatar's signs. This attempt was notable not only because it was in an interaction that lasted only a few minutes, but also because it was directed not back at the virtual agent but at the (non-signing) robot.

This most recent cohort has also made clear additional future challenges for this system. Recognition of infant behavioral attempts to engage (such as hand clapping, flapping, and copying) still remains a future goal. This is challenging primarily because most of our commonly used tracking systems (such as structured lighting systems like the Kinect) are based on body models that fail to track infants due to their differences in body morphology from adult standards. Infants being held in the arms of an adult also provide challenges due to occlusion and visual interference. We also plan to implement a more complex physiological model of infant attentiveness based on the thermal imaging system. Finally, discriminating communicative attempts (linguistic manual babbling) automatically from other random exploratory movement (nonlinguistic motor movements) remains an open challenge.

DESIGN LESSONS AND LIMITATIONS

While our efforts in this project were focused on building a specific tool for supplementing language for deaf infants, some of our design process has touched on issues that are more generally applicable. In this section, we highlight three design lessons that emerged from this work and discuss their broader application to child-agent design and the design of agents as pedagogical tools.

First, our work supports and affirms the importance of physical embodiment. While our protocols were not designed to explicitly compare the impact of an embodied system to a screen-only system, we based our design on the considerable evidence showing that infants do not learn from screen-based technologies [33, 12]. Infants in our case studies were able to successfully engage with our paired robot and avatar system and some of the infants displayed instances of copying (a limited form of social learning). While anecdotal, we view our results as being aligned with prior studies that demonstrate the benefits of physically embodied systems for teaching language to older children [39], as well as the overall benefits of physical embodiment for improving learning gains in adults [40], enhancing social performance and perception [81], and ensuring compliance with challenging pedagogical and therapeutic tasks [8]. Our observations are encouraging, but future efforts must focus on repeated exposures over longer periods of time which would possibly allow for the demonstration of language learning.

Second, our work supports the potential benefits of incorporating parents into agent-infant pedagogical interactions. The introduction of a parent-engaged interactive session with our third case study supports multiple results in robot-child interaction that demonstrate the usefulness of involving parents in pedagogical engagements. In particular, this work draws from observations in producing therapeutic agents that teach cognitive and social skills to children with autism spectrum disorder [67]. In clinical domains, designing interactions that provide enhanced agent-child interactions have limited value, as teaching a child to interact with a robot may or may not generalize to human-human interactions. Instead, modern agent-based interventions focus on using the agent to explicitly support child-adult interactions [27, 59]. In the future, allowing more direct support between infant and parent may provide a more rapid and generalizeable method for supporting language learning.

Finally, our work expands and reinforces the use of paired agents to provide controlled, semi-scripted interactions for the benefit of a child observer. Having two agents allowed our system to produce scripted call-and-response interactions between agents, to demonstrate positive social responses toward the social overtures of the other agent, to use social mechanisms to signal to the child that both avatar and robot were social agents, and to capitalize on capabilities that only one of the agents possessed. This technique was used previously in interactions involving groups of 2 robots and 1-3 older children (4-6 years old) by Scassellati's group [37, 35], and in this work closely resembles the framing used by Vázquez [80]. The computational methods for supporting these multi-agent pedagogical designs and the types of interactions that can be achieved merit further exploration.

CONCLUSION

This paper describes the design of a unique dual-agent system that uses a physical robot and a virtual human to engage 6-12 month old deaf infants in linguistic interactions. Our system was bolstered by a perception system capable of estimating infant attention and engagement through thermal imaging and eye tracking. We documented our experiences in designing for a unique population (deaf infants), and summarized the lessons and guidelines that we established over an iterative design process. Our design was informed by experimental sessions spread over three years, highlighted here by three case studies. This system has been successful at soliciting infant attention, directing attention to the linguistic content, and keeping the infant engaged for developmentally appropriate lengths of time. We also observed instances of infants copying robot behavior, of infants producing signs displayed by the avatar, and of infants producing signs to the non-signing robot agent that they had observed the virtual human perform. These initial experiences give us hope that longer-term exposure to a system based on this work may be able to impact long-term learning in this unique population.

ACKNOWLEDGMENTS

We are grateful to the families who took part in the evaluation of this system. This work was supported primarily by the W.M. Keck Foundation ("Seeing the Rhythmic Temporal Beats of Human Language," PI: Petitto) and NSF (IIS-1547178, "The RAVE Revolution for Children with Minimal Language Experience During Sensitive Periods of Brain and Language Development," PI: Petitto). Additional funding for robot design and evaluation was derived from an NSF Expedition in Computing award (#1139078, "Socially Assistive Robotics," PI: Scassellati). Additional funding for components of this project derived from the NSF, Science of Learning Center Grant SBE 1041725, specifically, Petitto (PI): NSF-SLC Project Funding of her study entitled, "The impact of early visual language experience on visual attention and Visual Sign Phonological processing in young deaf emergent readers using early reading apps: a combined eye-tracking and fNIRS Brain imaging investigation."

REFERENCES

1. Henny Admoni and Brian Scassellati. 2017. Social Eye Gaze in Human-Robot Interaction: A Review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.
2. LLC AgiSoft and Russia St Petersburg. 2014. Agisoft photoscan. *Professional Edition* (2014).
3. Neziha Akalin, Pinar Uluer, and Hatice Kose. 2014. Non-verbal communication with a social robot peer: Towards robot assisted interactive sign language tutoring. In *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*. IEEE, 1122–1127.
4. Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. 2016. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science* (2016).
5. Michael Anbar. 2003. Physiological, clinical and psychological applications of dynamic infrared imaging. In *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, Vol. 2. IEEE, 1121–1124.
6. Daniel R Anderson and Tiffany A Pempek. 2005. Television and very young children. *American Behavioral Scientist* 48, 5 (2005), 505–522.
7. Akiko Arita, Kazuo Hiraki, Takayuki Kanda, and Hiroshi Ishiguro. 2005. Can we talk to robots? Ten-month-old infants expected interactive humanoid robots to be talked to by persons. *Cognition* 95, 3 (2005), B49–B57.
8. Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3, 1 (2011), 41–52.
9. Marta Borgi, Irene Cogliati-Dezza, Victoria Brelsford, Kerstin Meints, and Francesca Cirulli. 2014. Baby schema in human and animal faces induces cuteness perception and gaze allocation in children. *Frontiers in Psychology* 5 (may 2014). DOI: http://dx.doi.org/10.3389/fpsyg.2014.00411
10. A Borji, D Parks, and L Itti. 2014. Complementary effects of gaze duration and early saliency in guiding fixations. (2014).
11. Daniela Cardone and Arcangelo Merla. 2017. New Frontiers for Applications of Thermal Infrared Imaging Devices: Computational Psychopshysiology in the Neurosciences. *Sensors* 17, 5 (2017), 1042.
12. Dimitri A Christakis. 2009. The effects of infant media usage: what do we know and what should we learn? *Acta Paediatrica* 98, 1 (2009), 8–16.
13. Valerie Corkum and Chris Moore. 1995. Development of joint visual attention in infants. (1995).
14. Paul Debevec. 2012. The light stages and their applications to photoreal digital actors. *SIGGRAPH Asia* 2, 4 (2012).
15. Judy S DeLoache, Cynthia Chiong, Kathleen Sherman, Nadia Islam, Mieke Vanderborght, Georgene L Troseth, Gabrielle A Strouse, and Katherine OâĂŹDoherty. 2010. Do babies learn from baby media? *Psychological Science* 21, 11 (2010), 1570–1574.
16. David Feil-Seifer and Maja Mataric. 2011. Socially Assistive Robotics. *IEEE Robotics & Automation Magazine* 18, 1 (mar 2011), 24–31. DOI: http://dx.doi.org/10.1109/mra.2010.940150
17. Sudeep Gandhe, David DeVault, Antonio Roque, Bilyana Martinovski, Ron Artstein, Anton Leuski, Jillian Gerten, and David Traum. 2008. From domain specification to virtual humans: An integrated approach to authoring tactical questioning characters. In *Ninth Annual Conference of the International Speech Communication Association*.

18. Michael Gleicher. 1998. Retargetting Motion to New Characters. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '98)*. ACM, New York, NY, USA, 33–42. DOI: http://dx.doi.org/10.1145/280814.280820
19. Arthur C Graesser, Patrick Chipman, Brian C Haynes, and Andrew Olney. 2005. AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education* 48, 4 (2005), 612–618.
20. Jillian Greczek and Maja Mataric. 2015. Toward Personalized Pain Anxiety Reduction for Children. In *AAAI 2015 Fall Symposium-Artificial Intelligence for Human-Robot Interaction*. 74–76.
21. Tobias Grossmann, Sarah Lloyd-Fox, and Mark H. Johnson. 2013. Brain responses reveal young infants' sensitivity to when a social partner follows their gaze. *Developmental Cognitive Neuroscience* 6 (oct 2013), 155–161. DOI: http://dx.doi.org/10.1016/j.dcn.2013.09.004
22. Stephanos Ioannou, Sjoerd Ebisch, Tiziana Aureli, Daniela Bafunno, Helene Alexi Ioannides, Daniela Cardone, Barbara Manini, Gian Luca Romani, Vittorio Gallese, and Arcangelo Merla. 2013. The autonomic signature of guilt in children: a thermal infrared imaging study. *PloS one* 8, 11 (2013), e79440.
23. Kabil Jaballah and Mohamed Jemni. 2013. A Review on 3D signing avatars: Benefits, uses and challenges. *International Journal of Multimedia Data Engineering and Management (IJMDEM)* 4, 1 (2013), 21–45.
24. W Lewis Johnson, Carole Beal, Anna Fowles-Winkler, Ursula Lauper, Stacy Marsella, Shrikanth Narayanan, Dimitra Papachristou, and Hannes Vilhjálmsson. 2004. Tactical language training system: An interim report. In *Intelligent Tutoring Systems*. Springer, 336–345.
25. Hernisa Kacorri, Pengfei Lu, and Matt Huenerfauth. 2013. Evaluating facial expressions in American Sign Language animations for accessible online information. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 510–519.
26. Colleen Kehoe, Justine Cassell, Susan Goldman, James Dai, Ian Gouldstone, Shaunna MacLeod, Traci O'Day, Anna Pandolfo, Kimiko Ryokai, and Austin Wang. 2004. Sam goes to school: story listening systems in the classroom. In *Proceedings of the 6th international conference on Learning sciences*. International Society of the Learning Sciences, 613–613.
27. Elizabeth S Kim, Rhea Paul, Frederick Shic, and Brian Scassellati. 2012. Bridging the Research Gap: Making HRI Useful to Individuals with Autism. *Journal of Human-Robot Interaction* 1, 1 (2012).
28. Michael Kipp, Alexis Heloir, and Quan Nguyen. 2011. Sign language avatars: Animation and comprehensibility. In *Intelligent Virtual Agents*. Springer, 113–126.
29. Hatice Kose, Neziha Akalin, and Pinar Uluer. 2014. Socially interactive robotic platforms as sign language tutors. *International Journal of Humanoid Robotics* 11, 01 (2014), 1450003.
30. Hatice Kose, Rabia Yorganci, Esra H Algan, and Dag S Syrdal. 2012. Evaluation of the robot assisted sign language tutoring using video-based studies. *International Journal of Social Robotics* 4, 3 (2012), 273–283.
31. Marina Krcmar. 2011. Word Learning in Very Young Children From Infant-Directed DVDs. *Journal of Communication* 61, 4 (2011), 780–794.
32. Marina Krcmar, Bernard Grela, and Kirsten Lin. 2007. Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology* 10, 1 (2007), 41–63.
33. Patricia K Kuhl, Feng-Ming Tsao, and Huei-Mei Liu. 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences* 100, 15 (2003), 9096–9101.
34. Stephen RH Langton, Helen Honeyman, and Emma Tessler. 2004. The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception & psychophysics* 66, 5 (2004), 752–771.
35. Iolanda Leite, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati. 2015. Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 75–82.
36. Iolanda Leite, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati. 2017a. Narratives with Robots: The Impact of Interaction Context and Individual Differences on Story Recall and Emotional Understanding. *Frontiers in Robotics and AI* 4 (jul 2017). DOI: http://dx.doi.org/10.3389/frobt.2017.00029
37. Iolanda Leite, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati. 2017b. narratives with robots: The impact of interaction context and individual Differences on story recall and emotional Understanding. *Frontiers in Robotics and AI* 4 (2017), 29.
38. Daniel Leyzberg, Samuel Spaulding, and Brian Scassellati. 2014a. Personalizing Robot Tutors to Individuals' Learning Differences. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction (HRI '14)*. ACM, New York, NY, USA, 423–430. DOI: http://dx.doi.org/10.1145/2559636.2559671

39. Daniel Leyzberg, Samuel Spaulding, and Brian Scassellati. 2014b. Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 423–430.

40. Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Cognitive Science Society*, Vol. 34.

41. Sheng-Yen Lo and Han-Pang Huang. 2016. Realization of sign language motion using a dual-arm/hand humanoid robot. *Intelligent Service Robotics* 9, 4 (2016), 333–345.

42. Konrad Lorenz. 1943. Die angeborenen formen möglicher erfahrung. *Ethology* 5, 2 (1943), 235–409.

43. Barbara Manini, Daniela Cardone, Sjoerd JH Ebisch, Daniela Bafunno, Tiziana Aureli, and Arcangelo Merla. 2013. Mom feels what her child feels: thermal signatures of vicarious autonomic response while watching children in a stressful situation. *Frontiers in human neuroscience* 7 (2013).

44. Vesna Marinović, Stefanie Hoehl, and Sabina Pauen. 2014. Neural correlates of human–animal distinction: An ERP-study on early categorical differentiation with 4- and 7-month-old infants and adults. *Neuropsychologia* 60 (jul 2014), 60–76. DOI:`http://dx.doi.org/10.1016/j.neuropsychologia.2014.05.013`

45. Andrew N Meltzoff, Rechele Brooks, Aaron P Shon, and Rajesh PN Rao. 2010. "Social" robots are psychological agents for infants: A test of gaze following. *Neural networks* 23, 8 (2010), 966–972.

46. Arcangelo Merla. 2014. Thermal expression of intersubjectivity offers new possibilities to human–machine and technologically mediated interactions. *Frontiers in psychology* 5 (2014).

47. Arcangelo Merla and Gian Luca Romani. 2007. Thermal signatures of emotional arousal: a functional infrared imaging study. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, 247–249.

48. National Institute on Deafness and Other Communication Disorders. 2015. (15 December 2015). Retrieved August 29, 2017 from `https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing`.

49. Johanna Grant Nicholas and Ann E. Geers. 2007. Will They Catch Up? The Role of Age at Cochlear Implantation in the Spoken Language Development of Children With Severe to Profound Hearing Loss. *Journal of Speech Language and Hearing Research* 50, 4 (aug 2007), 1048. DOI: `http://dx.doi.org/10.1044/1092-4388(2007/073)`

50. Tim Payne. 2013. MAKI - A 3D Printable Humanoid Robot. (26 March 2013). Retrieved August 29, 2017 from `https://www.kickstarter.com/projects/391398742/maki-a-3d-printable-humanoid-robot`.

51. Laura-Ann Petitto. 2005. How the brain begets language. *The cambridge companion to chomsky* (2005), 84–101.

52. Laura Ann Petitto. in press. Impact of Minimal Language Experience on Children During Sensitive Periods of Brain and Early Language Development: Myths Debunked and New Policy Implications.. In *The Science of Learning*, S. Guerriero P. Kuhl, S-S. Lim (Ed.). The Organization for Economic Co-Operation and Development (OECD).

53. Laura Ann Petitto, Siobhan Holowka, Lauren E Sergio, Bronna Levy, and David J Ostry. 2004. Baby hands that move to the rhythm of language: hearing babies acquiring sign languages babble silently on the hands. *Cognition* 93, 1 (2004), 43–73.

54. Laura Ann Petitto, Siobhan Holowka, Lauren E Sergio, and David Ostry. 2001. Language rhythms in baby hand movements. *Nature* 413, 6851 (2001), 35–36.

55. Laura Ann Petitto, Clifton Langdon, Adam Stone, Diane Andriola, Geo Kartheiser, and Casey Cochran. 2016. Visual sign phonology: insights into human reading and language from a natural soundless phonology. *Wiley Interdisciplinary Reviews: Cognitive Science* 7, 6 (2016), 366–381.

56. Laura Ann Petitto and Paula F Marentette. 1991. Babbling in the manual mode: Evidence for the ontogeny of language. *Science* 251, 5000 (1991), 1493.

57. Farzad Pezeshkpour, Ian Marshall, Ralph Elliott, and J Andrew Bangham. 1999. Development of a legible deaf-signing virtual human. In *Multimedia Computing and Systems, 1999. IEEE International Conference on*, Vol. 1. IEEE, 333–338.

58. Brian Plüss, David DeVault, and David Traum. 2011. Toward Rapid Development of Multi-Party Virtual Human Negotiation Scenarios. In *Proceedings of SemDial 2011, the 15th Workshop on the Semantics and Pragmatics of Dialogue*. 10.

59. Sarah M Rabbitt, Alan E Kazdin, and Brian Scassellati. 2015. Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical psychology review* 35 (2015), 35–46.

60. Aditi Ramachandran, Alexandru Litoiu, and Brian Scassellati. 2016. Shaping Productive Help-Seeking Behavior During Robot-Child Tutoring Interactions. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Piscataway, NJ, USA, 247–254. `http://dl.acm.org/citation.cfm?id=2906831.2906875`

61. Rebekah A Richert, Michael B Robb, and Erin I Smith. 2011. Media as social partners: The social nature of young childrenâĂŹs learning from screen media. *Child Development* 82, 1 (2011), 82–95.

62. Jeff Rickel and W. Lewis Johnson. 1999a. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence* 13 (1999), 343–382.

63. Jeff Rickel and W Lewis Johnson. 1999b. Virtual humans for team training in virtual reality. In *Proceedings of the ninth international conference on artificial intelligence in education*, Vol. 578. 585.

64. Jeff Rickel, Stacy Marsella, Jonathan Gratch, Randall Hill, David Traum, and William Swartout. 2002. Toward a new generation of Virtual Humans for Interactive Experiences. *IEEE Intelligent Systems* 17 (2002), 32–38.

65. J. R. Saffran, A. Senghas, and J. C. Trueswell. 2001. The acquisition of language by children. *Proceedings of the National Academy of Sciences* 98, 23 (oct 2001), 12874–12875. DOI: http://dx.doi.org/10.1073/pnas.231498898

66. Brian Scassellati, Henny Admoni, and Maja Matarić. 2012a. Robots for Use in Autism Research. *Annual Review of Biomedical Engineering* 14, 1 (aug 2012), 275–294. DOI:http://dx.doi.org/10.1146/annurev-bioeng-071811-150036

67. Brian Scassellati, Henny Admoni, and Maja Matarić. 2012b. Robots for use in autism research. *Annual review of biomedical engineering* 14 (2012), 275–294.

68. Jerry Schnepp, Rosalee Wolfe, John McDonald, and Jorge Toro. 2013. Generating Co-occurring Facial Nonmanual Signals in Synthesized American Sign Language. (2013).

69. Ari Shapiro. 2011. Building a character animation system. *Motion in Games* (2011), 98–109.

70. Clare E Sims and Eliana Colunga. 2012. Language Development in the Age of Baby Media: What We Know and What Needs to be Done. 2 (2012), 384–396.

71. Jeffrey L Sokolov. 1993. A local contingency analysis of the fine-tuning hypothesis. *Developmental psychology* 29, 6 (1993), 1008.

72. Michelle Starr. 2014. Toshiba's new robot can speak in sign language. (October 2014). http://www.cnet.com/news/toshibas-new-robot-can-speak-in-sign-language/ [Online; posted 8-October-2014].

73. Adam Stone, Laura-Ann Petitto, and Rain Bosworth. 2017. Visual sonority modulates infants' attraction to sign language. *Language Learning and Development* (dec 2017), 1–19. DOI: http://dx.doi.org/10.1080/15475441.2017.1404468

74. Catherine S Tamis-LeMonda, Marc H Bornstein, and Lisa Baumwell. 2001. Maternal responsiveness and children's achievement of language milestones. *Child development* 72, 3 (2001), 748–767.

75. Andrea Tartaro and Justine Cassell. 2006. Authorable virtual peers for autism spectrum disorders. In *Proceedings of the Combined workshop on Language-Enabled Educational Technology and Development and Evaluation for Robust Spoken Dialogue Systems at the 17th European Conference on Artificial Intellegence*.

76. David Traum, Priti Aggarwal, Ron Artstein, Susan Foutz, Jillian Gerten, Athanasios Katsamanis, Anton Leuski, Dan Noren, and William Swartout. 2012. Ada and Grace: Direct interaction with museum visitors. In *Intelligent Virtual Agents*. Springer, 245–251.

77. David R Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In *Current and new directions in discourse and dialogue*. Springer, 325–353.

78. Pınar Uluer, Neziha Akalın, and Hatice Köse. 2015. A new robotic platform for sign language tutoring. *International Journal of Social Robotics* 7, 5 (2015), 571–585.

79. Lynette van Zijl and Jaco Fourie. 2007. The development of a generic signing avatar. In *Proceedings of the IASTED International Conference on Graphics and Visualization in Engineering, GVE*, Vol. 7. 95–100.

80. Marynel Vázquez, Aaron Steinfeld, Scott E Hudson, and Jodi Forlizzi. 2014. Spatial and other social engagement cues in a child-robot interaction: Effects of a sidekick. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 391–398.

81. Joshua Wainer, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. 2006. The role of physical embodiment in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*. IEEE, 117–122.

82. Hugh R Wilson, Frances Wilkinson, Li-Ming Lin, and Maja Castillo. 2000. Perception of head orientation. *Vision research* 40, 5 (2000), 459–472.