Can a Virtual Human Facilitate Language Learning in a Young Baby?

Extended Abstract

Setareh Nasihati Gilani, David Traum
University of Southern California

Rachel Sortino, Grady Gallagher
Gallaudet University

Kailyn Aaron-lozano, Cryss Padilla
Gallaudet University

Ari Shapiro
University of Southern California

Jason Lamberton
Gallaudet University

Laura-ann Petitto*
Gallaudet University
laura-ann.petitto@gallaudet.edu

KEYWORDS

Empirical studies on social agents/robots; Social impact; Multi-user/multi-agent/robot interaction

ACM Reference Format:
Setareh Nasihati Gilani, David Traum, Rachel Sortino, Grady Gallagher, Kailyn Aaron-lozano, Cryss Padilla, Ari Shapiro, Jason Lamberton, and Laura-ann Petitto. 2019. Can a Virtual Human Facilitate Language Learning in a Young Baby?. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

There is a significant paucity of work on language learning systems for young infants [2, 5, 19] despite the widely understood critical importance that this developmental period has for healthy language and cognitive growth, and related reading and academic success [6, 14]. Deaf babies constitute one vulnerable population as they can experience dramatically reduced or no access to usable linguistic input during this period [18]. This causes potentially devastating impact on children's linguistic, cognitive, and social skills [9, 10, 15, 16, 20]. We introduced an AI system, called RAVE (Robot, AVatar, thermal Enhanced language learning tool), designed specifically for babies within the age range of 6-12 months [8, 17]. RAVE consists of two agents: a virtual human (provides language and socially contingent interactions) and an embodied robot (provides socially engaging physical cues to babies and directs babies' attention to the virtual human). Detailed description of the system's constituent components and dialogue algorithms are presented in [17] and [8].

While RAVE appears to be using the best available research to inform its design, there is still a question of whether it really could be used to facilitate babies' language learning. Is there evidence that the babies' behaviors are influenced and/or facilitated by the avatar's behaviors? As a first step toward the evaluation of the whole system, we focus on babies' behavioral responses toward the avatar which is providing multiple kinds of social and linguistic behaviors. We ask whether the babies' interaction with the system is socially contingent or a random distribution of behaviors from the parties. Can the avatar stimulate babies' production of socially contingent responses, and crucially, nascent language responses?

In this paper, we focus on babies' behavioral responses toward RAVE and especially toward the avatar's different conversational modes. We report preliminary results from an experimental study in order to evaluate the system's performance regarding the above questions, with the ultimate goal of evaluating the potential for AI/Avatar systems to facilitate language learning in young babies.

2 BEHAVIOR TYPES

For the purposes of evaluating the system's ability to engage in socially contingent interaction with the baby, we present our categorization of both the baby and the avatar's behaviors. Avatar conversational modes are as follows.

(1) **Idle Behaviors** are nonlinguistic/nonsigning, and non socially communicating neutral bodily postures. This behavior typically occurs when the avatar is looking at the robot or the baby as a 3rd-party conversationalist.
(2) **Nursery Rhymes** are linguistic stimuli such as the "BOAT-ON-WAVE" nursery rhyme in ASL, with specific rhythmic temporal patterns at the core of all languages' linguistic phonological structure [8].[1,2]
(3) **Social Gestures** include social routines (e.g., HI), conversational fillers (e.g., affirmative head nod), and short lexical phrases such as YES! or THAT (i.e., English's "right!").
(4) **3-Way Behaviors** are avatar's communicative interactions directed to both the baby and the robot like "LOOK-AT-ME."

Below, we categorize the observations reported in [8, 17] of baby's spontaneous behavioral responses towards the avatar. Note that these categories are not mutually exclusive:

(1) **Linguistic Responses** include manual babbling, the production of manual proto sign-phonetic units, proto-signs, and imitations of signs
(2) **Social/Gestural Responses** include pointing, waving, clapping hands or attempts to copy the agents' behaviors, or social referencing to the parent;

*Corresponding Author

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

[1] The formal linguistic notation of natural signed languages, such as ASL, uses glosses showing approximate English translations in capital letters.
[2] While the ASL NR is unique to the deaf culture, a rough semantic neighbor in English would be "Row-Row-Row-Your Boat"

(3) **Sustained Visual Attention** indicates the baby being visually transfixed on the agents for atypically extended periods, defined as greater than one second for this study.

3 HYPOTHESES

As a first step toward evaluating the system, we asked the following main research questions: (1) Do babies attend to the avatar and respond to its communicative behaviors? (2) Can babies with little or no exposure to ASL distinguish among the avatar's different conversational modes; particularly as they appear on a flat TV monitor? (3) Can an avatar stimulate babies' production of socially contingent responses, and crucially, nascent language responses?

Regarding the first question; one possibility is that babies do not see the avatar, or the agents collectively, as interesting social interlocutors or respond to them at all. Another possible outcome is that the infants may enter an agitated mode upon confronting an unknown (or "strange") situation such as the RAVE system [3]. We use the rate of baby's responses to the avatar as a metric to evaluate the overall system's impact in terms of engaging the babies.

The second question asks whether babies can differentiate among the different avatar conversational modes even though it's unlikely that these young babies understand the semantic content of the ASL language productions. This would corroborate the now-classic studies that demonstrates infant's ability to discriminate categorically among classes of linguistic units in different languages based on their contrastive patterning (peaked between ages 6-12 months; [1, 11, 13]). Here, we examine the baby's response rate to the avatar's different conversational modes.

The third question is of scientific interest concerning the mechanisms that drive early language learning: does the avatar's linguistic productions garner the baby's attention, and in particular, garner linguistic responses from the babies? We hypothesize that it is the linguistic patterning that is important in the avatar's productions, not its modality of language production and reception (here, signed; [11, 12]). Specifically, we claim that since we are correctly hitting on just the right temporal patterning in the avatar's productions [13, 14], then all babies would be engaged by the avatar's language productions over other conversational modes. We hypothesize that they would react with more linguistic content when the avatar was in this mode, as compared to its other conversational modes.

4 EVALUATION AND RESULTS

To address the questions about the impact of the avatar behaviors on babies, we designed an experiment whereupon babies interacted with the system in a controlled setting. 4 babies (ages 6-13 months) participated in an intensive case study. Babies were seated on their parent's lap facing the system. Each baby's experimental session lasted until the baby became distracted or fussy in which case we immediately ceased the session. The experiment consisted of several steps: upon arrival, the baby and the parent were greeted and introduced to the robot and then the avatar [7]. Next, a calibration process (a technical requirement of the thermal IR Imaging and Tobii eye tracking systems), followed by the interaction session.

The video-recorded interaction sessions were coded for conversational turns. Analysis was done based on the occurrences of specific behaviors as its the convention in child developmental sciences.

In answer to questions 1 & 2 above, a frequency analysis of baby's behavioral responses (stated in section 2) to avatar behaviors was conducted. Overall, babies responded to more than 60% of avatar's behaviors. In addition, a detailed analysis of babies' different behaviors was done. Results show that overall, 17.9% of babies' behaviors were linguistic, 48% sustained visual attention, 25.4% social/gestural and 36.7% were none of the above categories.

Next, we studied the relationship between the Avatar's behaviors and the baby's response rate. Results show that babies' responses were not equally distributed across different types of Avatar's behaviors. Babies' response rate to avatar's idle behavior was 37%, , 85% to Nursery Rhyme, 75% to Social/Gestural behaviors and 85% to 3-Way behaviors. Note that the distribution of avatar behaviors was also not uniform: 13% of the avatar's behaviors were NR, 13% 3-way, 36% were social, and the remaining 38% were idle.

Regarding the third question, babies responded differently when the avatar was in the linguistic Nursery Rhyme conversational mode versus other modes. The babies produced the largest percentage of linguistic responses to the avatar's Nursery Rhymes (31% to NR vs 10% to Idle, 19% to Social, and 20% to 3-Way). Further, the babies' responses to the avatar's Linguistic Nursery Rhymes involved them to be largely riveted into a state of fixed and Sustained Visual Attention (77%). There appears to be a principled relationship between the avatar's socially contingent communicative turn types and the babies' specific responses. This relationship implies that the avatar was indeed having a linguistic impact on the baby.

5 CONCLUSION

The theoretical question of the present paper was to understand whether a signing virtual human had the potential to facilitate language learning in young babies. To address this, we studied the impact that the agent had on babies' spontaneous behavioral responses, in particular, we asked whether the avatar's linguistic productions in signed language would spontaneously trigger linguistic responses from the babies. We were especially interested if a very young baby would even detect the avatar's different conversational modes, as the avatar was projected onto a flat screen.

We indeed found that babies spontaneously distinguished among avatar conversational modes. Babies produced different categories of behavioral responses to the avatar, and, further, their different behavioral responses were socially contingent (related to) the avatar's different conversational modes. The results indicate that the babies were indeed able to detect the avatar's different conversational modes even though all appeared on a flat screen.

One interesting finding was that babies produced the greatest percentage of linguistic responses to the avatar's linguistic Nursery Rhymes versus other conversational modes even though most of them did not understand ASL and thus could not possibly have been understanding the meanings of the observed language.

ACKNOWLEDGMENTS

This work was supported 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).

REFERENCES

[1] Stephanie A Baker, Roberta Michnick Golinkoff, and Laura-Ann Petitto. 2006. New insights into old puzzles from infants' categorical discrimination of soundless phonetic units. *Language Learning and Development* 2, 3 (2006), 147–162.

[2] Ewa M Golonka, Anita R Bowles, Victor M Frank, Dorna L Richardson, and Suzanne Freynik. 2014. Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer assisted language learning* 27, 1 (2014), 70–105.

[3] David J Greenberg, Donald Hillman, and Dean Grice. 1973. Infant and stranger variables related to stranger anxiety in the first year of life. *Developmental Psychology* 9, 2 (1973), 207.

[4] P Higgins. 1980. *Outsiders in a hearing world*. SAGE Publishing.

[5] Jiyou Jia. 2009. An AI framework to teach English as a foreign language: CSIEC. *AI Magazine* 30, 2 (2009), 59.

[6] Patricia K Kuhl. 2004. Early language acquisition: cracking the speech code. *Nature reviews neuroscience* 5, 11 (2004), 831.

[7] 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, 8-9 (2010), 966–972.

[8] Setareh Nasihati Gilani, David Traum, Arcangelo Merla, Eugenia Hee, Zoey Walker, Barbara Manini, Grady Gallagher, and Laura-Ann Petitto. 2018. Multimodal Dialogue Management for Multiparty Interaction with Infants. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 5–13.

[9] 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 (2007), 1048–1062.

[10] Laura-Ann Petitto. in press. The Impact of Minimal Language Experience on Children During Sensitive Periods of Brain and Early Language Development: Myths Debunked and New Policy Implications. retrieved from http://petitto.net/wp-content/uploads/2014/04/Petitto_Minimal-Language-Experience_Final_Oct-6-2017.pdf.

[11] Laura-Ann Petitto, Melody S Berens, Ioulia Kovelman, Matt H Dubins, K Jasinska, and M Shalinsky. 2012. The "Perceptual Wedge Hypothesis" as the basis for bilingual babies' phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and language* 121, 2 (2012), 130–143.

[12] 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.

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

[14] Laura-Ann Petitto, Clifton Langdon, Adam Stone, Diana 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.

[15] Laura Ann Petitto, Robert J Zatorre, Kristine Gauna, Erwin James Nikelski, Deanna Dostie, and Alan C Evans. 2000. Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proceedings of the National Academy of Sciences* 97, 25 (2000), 13961–13966.

[16] Jenny R Saffran, Ann Senghas, and John C Trueswell. 2001. The acquisition of language by children. *Proceedings of the National Academy of Sciences* 98, 23 (2001), 12874–12875.

[17] Brian Scassellati, Jake Brawer, Katherine Tsui, Setareh Nasihati Gilani, Melissa Malzkuhn, Barbara Manini, Adam Stone, Geo Kartheiser, Arcangelo Merla, Ari Shapiro, David Traum, and Laura-Ann Petitto. 2018. Teaching Language to Deaf Infants with a Robot and a Virtual Human. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 553.

[18] Arielle Spellun and Poorna Kushalnagar. 2018. Sign language for deaf infants: A key intervention for a developmental emergency. *Clinical pediatrics* 57, 14 (2018), 1613–1615.

[19] Glenn Stockwell. 2007. A review of technology choice for teaching language skills and areas in the CALL literature. *ReCALL* 19, 2 (2007), 105–120.

[20] Adam Stone, Laura-Ann Petitto, and Rain Bosworth. 2018. Visual sonority modulates infants' attraction to sign language. *Language Learning and Development* 14, 2 (2018), 130–148.