Hearing Babies Respond to Language's Patterning and Socially-Contingent Interactions with a Signing Avatar: Insights into Human Language Acquisition

Laura-Ann Petitto*, Rachel Sortino, Kailyn Aaron-Lozano, Grady Gallagher, Setareh Nasihati Gilani, David Traum, Arcangelo Merla, Chiara Filippini, Cryss Padilla

Human Language Acquisition

Human babies have peaked sensitivity to specific rhythmic temporal patterns in language within ages 6-12 months. This allows them to segment, categorize, and discern the linguistic stream at the phonological level – a key to early reading success.[1,2,3] Human language acquisition requires social contingency; social interactions (alone) are not enough[4,5,6]

Challenges

Developmental Challenge
Many babies experience minimal language exposure during this critical period, especially deaf babies

Question Can a learning tool be created to augment early language exposure?

Technical Challenges
(i) Language is not learnable from TV/computer screens[7]
(ii) Babies respond best to "Social Contingency" = conversationally relevant responses when babies are emotionally engaged[4,5,6,8]
Question How to achieve this with AI agents?

Hypothesis

Hypothesis Babies will attend to, and perceive, differences among Avatar communicative behaviors on a screen if the Avatar's behaviors contain the precise rhythmic temporal patterns in language to which the baby's brain has peaked sensitivity (6-12 months), irrespective of meaning

Participants

8 Babies ages 7 – 13 months
- 1 deaf, sign-exposed
- 6 hearing, speech exposed
- 1 hearing, sign and speech exposed

Experimental Procedures

- Babies sat on parent's lap in front of RAVE
- Robot directed babies' attention to TV screen[9] where Avatar produced 3 possible types of actions (**Table 1**)
- Thermal IR Imaging camera measured babies' emotional engagement[8] and triggered the START and STOP of Socially Contingent AI dialogue scripts with Baby[4,5,6] (**Figure 1**)
- Experiment lasted an average of 4 minutes
- Babies' full range of behavioral responses to Avatar were analyzed (**Table 2**)

What is RAVE?
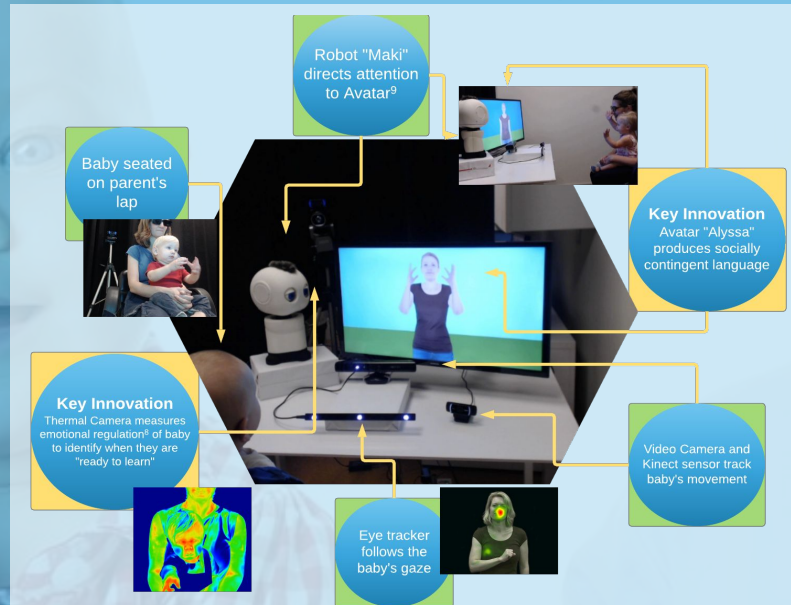### Robot Avatar thermal-Enhanced-thermal language learning tool

Robot "Maki" directs attention to Avatar[9]

Baby seated on parent's lap

Key Innovation Avatar "Alyssa" produces socially contingent language

Key Innovation Thermal Camera measures emotional regulation[8] of baby to identify when they are "ready to learn"

Video Camera and Kinect sensor track baby's movement

Eye tracker follows the baby's gaze

Figure 1. Diagram showing the individual components of the RAVE system

How do I learn language?

Avatar Actions

AVATAR BEHAVIOR	DESCRIPTION
Nursery Rhyme (NR)	Signs American Sign Language (ASL) NR
Social Interaction (S/G)	Waves hello; attention wave; etc.
Idle	Stands still, hands at side, swaying slightly

Table 1. Three categories of behaviors produced by the Avatar's during experiment

Babies' Responses

BABY BEHAVIOR	DESCRIPTION
Linguistic (Ling)	Protosigns; copying Avatar's sign productions; manual babbling
Social/ Gesture (S/G)	Reaching, smiling, raised eyebrows (surprise face)
Sustained Visual Attention (SVA)	Looking at Avatar for ≥ 1 second

Table 2. Three observed categories of behavioral responses produced by babies during experiment

Conclusions & Broad Impact

- Babies demonstrated riveted sustained visual attention more frequently to the Avatar's socially contingent productions (e.g., Linguistic Nursery Rhymes and social gestures), over Avatar's non-contingent idling behaviors (**Figure 2**), even though the Avatar was on a TV screen

- Babies produced the highest rate of linguistic behaviors (e.g., manual babbling, sign copying) during the Avatar's Linguistic Nursery Rhymes (**Figures 2, 3**), even though most babies did not know ASL and did not understand sign meanings

Surprising Science Implications & Early Human Language Acquisition (i) Presence of Language's rhythmic temporal patterning (specifically, patterning at the size of phonetic-syllabic units in all language)[1,2,3] and (ii) Avatar's socially contingent productions (ling. & social gestures over idle) constitute two potent and necessary features of human language acquisition

Broad Implications The study demonstrates the potential for language learning from agents in babies

Want to know more?

Rates of Infant Behavioral Responses to Avatar

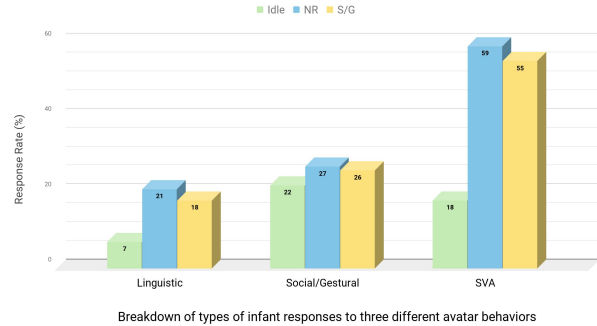Breakdown of types of infant responses to three different avatar behaviors

Figure 2. Rate of babies different responses to Avatar behaviors.
Babies produce different behavioral responses to the Avatar's different types of productions, suggesting that they "see" the differences
Babies produce higher linguistic and sustained visual attention responses when the Avatar is producing Linguistic ASl Nursery Rhymes (even though they do not understand the meanings)

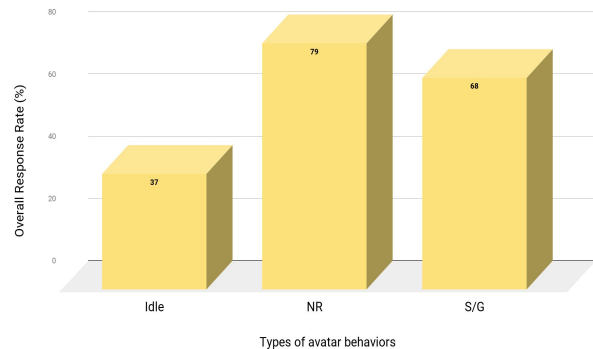**Rate of Infant Responses to Avatar Stimuli**

Types of avatar behaviors

Figure 3. Rate of babies' responses to different avatar behaviors.
Babies respond most to the Avatar when it was producing Linguistic Nursery Rhymes

Laura-Ann Petitto[1]*, Rachel Sortino[1], Kailyn Aaron-Lozano[1], Grady Gallagher[1], Setareh Nasihati Gilani[2], David Traum[2], Arcangelo Merla[3], Chiara Filippini[1,3], Cryss Padilla[1]
1. Gallaudet University, 2. University of Southern California, 3. Universita G. D'Anminizio Chieti-Pescara, Italy

The Petitto Brain and Language Laboratory for Neuroimaging, BL2 at Gallaudet University, Washington, D.C. 20002
Contact: Laura-Ann.Petitto@Gallaudet.Edu
Web page: Petitto.net
More on RAVE: http://petitto.net/whats-new/rave/

References
1. Petitto, L. A., Holowka, S., Sergio, L., & Ostry, D. (2001). Language rhythms in baby hand movements. Nature, 413, 35-36.
2. Petitto, L.A., Holowka, S., Sergio, L.E., Levy, B., & Ostry, D.J. (2004). Baby hands that move to the rhythm of language: Hearing babies acquiring sign languages babble silently on the hands. Cognition, 93, 43-73.
3. Petitto, L.A., Langdon, C., Stone, A., Andriola, D., Kartheiser, G., & Cochran, C. (2016). Visual sign phonology: Insights into human reading and language from a natural soundless phonology. WIREs Cognitive Science. doi: 10.1002/wcs.1404.
4. Nasihati Gilani, S., Traum, D., Sortino, R., Gallagher, Aaron-Lozano, K., G., Padilla, C., Shapiro, A., Lamberton, J., & Petitto, L.A. (2019). Can a virtual human facilitate language learning in a young baby? Extended Abstract of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Montreal, May 2019.
5. Nasihati Gilani, S., Traum, D., Merla, A., Hee, E., Walker, Z., Manini, B., Gallagher, G., & Petitto, L. (2018). Multimodal Dialogue Management for Multiparty Interaction with Infants. In refereed published Paper in the Proceedings of the 20th Association for Computing Machinery/ACM International Conference on Multimodal Interaction, Colorado, October 2018.
6. Scassellati, B., Brawer, J., Tsui, K., Nasihati Gilani, S., Malzkuhn, M., Manini, B., Stone, A., Kartheiser, G., Merla, A., Shapiro, A., Traum, D.,& Petitto, L.A. (2017). Teaching Language to Deaf Infants with a Robot and a Virtual Human. In refereed published Paper in the Proceedings of the 2018 Conference on Human Factors in Computing Systems/CHI (pp.553; 1-553:13). New York, NY, USA: ACM. ISBN 123-4567-24-567/08/06. DOI: http://dx.doi.org/10.475/123_4.
7. Kuhl, P.K., Tsao, F-M., and Liu, H-M. (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, 9096–9101. arXiv:http://www.pnas.org/content/100/15/9096.full.pdf.
8. Merla, A. (2014). Thermal expression of intersubjectivity offers new possibilities to human–machine and technologically mediated interactions. Frontiers in psychology 5, 802.
9. Meltzoff, A.N., Brooks, R., Shon, A.P, and Rao, R.P.N. (2010)."Social" robots are psychological agents for infants: A test of gaze following. Neural networks 23, 8-9, 966–972.