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

NSF INSPIRE IIS-1547178 "The RAVE Revolution for Children with Minimal Language Experience During Sensitive Periods of Brain and Language Development (Petitto, PI)
W.M. Keck Foundation "Seeing the Rhythmic Temporal Beats of Human Language (Petitto, PI)
NSF SBE-1041725 Science of Learning Center, Visual Learning & Visual Learning (Petitto, Co-PI)

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?

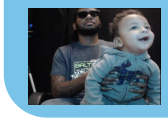
Technical Challenges
(i) Language is not learnable from TV/computer screens[7]
(ii) Babies respond best to "Social Contingency" = conversationally relevant responses when babies are emotionally engaged[4,5,6,8]
Question How to achieve this with AI agents?

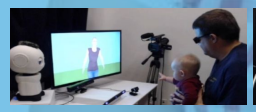
Hypothesis

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

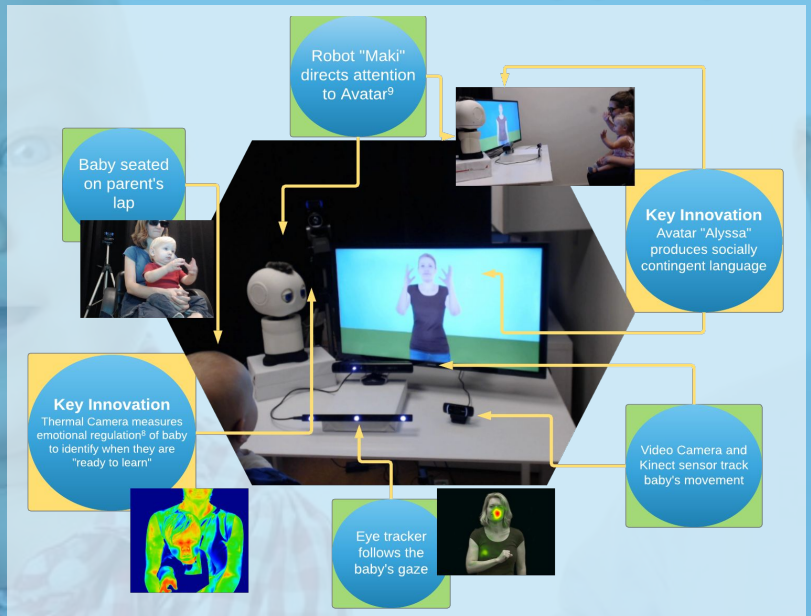


Figure 1. Diagram showing the individual components of the RAVE system

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?

***Corresponding Author** Dr. Laura-Ann Petitto,
Laura-Ann.Petitto@Gallaudet.Edu
Poster presented at Society for Research in Child Development 2019 Annual Meeting, Baltimore, MD