HEALTHCARE INNOVATION TECHNOLOGY LAB (HITLAB)

Artificial Intelligence-powered algorithm provides insights into speech and language development in infants and children

Ekta Jain, PhD[1]; Vandana Yadav, MSc[1]; Kat Marriott, PhD[1]; Shivangi Das, MSc[1]; Stan Kachnowski, PhD, MPA[1]

[1]Healthcare Innovation and Technology Lab (HITLAB)

ABSTRACT

Speech and language development delays in infants can significantly impact a child's well-being, socioeconomic mobility and future academic success.

The American Speech Language estimates that approximately 10-20% of one to two-year-old children have delayed speech development. These delays begin to arise in their first year of development and have shown to be associated with poor attention, less socialization, and poor literacy levels that the infant is subjected to via their environment. Monitoring of speech and language development in infants and children and its timely assessment is crucial. Presently, children in the US and Canada are screened at 9 and 18-month checkups, respectively and development assessment for autism is mostly conducted at the ~18-month checkup in the US. These evaluation points may be too late. Once flagged, assessment and therapy may take another 6-12 months adding more time to pass before any intervention is begun.

This then widens the gap in language development for children who are already struggling. This is where Babbly's platform addresses this gap. The platform, built using the latest in digital technology, analyzes data in an audio format that captures the communication skills of infants/children.

OBJECTIVE

- Validate the accuracy of the AI-powered algorithm to track communication and cognition during the babbling stages of development in the first 16 months of life (4-16 months age).

STUDY METHODS

Study Design:

- Single-arm validation involving quantitative (demographic surveys) and qualitative (infant voice audio analysis) using Babbly's AI algorithm
- Apart from demographic surveys, a NIDCD (National Institute of Deadness and other Communication Disorders) hearing and communicative development check-list was also administered.
- Manual assessment of infants' audios by three independent annotators, per audio, who are trained speech language pathologists. A total of five annotators contributed to the Study. Agreement of classes between 2 out 3 annotators was considered as the ground truth.
- Findings from the manual/human annotation were compared to that of the AI-powered algorithm. Precision, recall and F1 scores were calculated.

Study Participants:

- Parents/caregivers/guardians of 35 infants were recruited.
- 3-4 audios of a minimum duration of 30 seconds and a maximum duration of 3 minutes, per infant, were collected.
- A total of 101 audios, quality-control approved, were used in the analysis.

The Babbly Algorithm

Babbly's AI-powered algorithm reads in audio clips of baby vocalizations and provides predictions of key early speech milestones.

First, the algorithm splits the provided audio clips into small chunks and passes them to an utterance detection model to classify whether each chunk contains adult voice, baby voice, both, or neither.

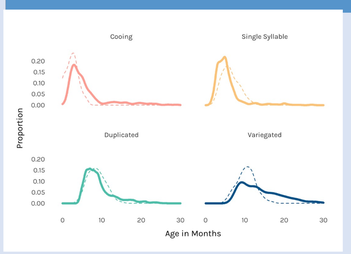
If a baby voice is detected (either on its own or overlapping with adult voice), that audio segment is passed to the speech milestones classification model. This model classifies each baby vocalization as one of the 5 speech classes (explained in the image)

Simultaneously, the detected baby and adult vocalizations is passed to the turn-taking model. In this step, the lengths of the detected adult and baby utterances are calculated, and the number of turns (conversational back-and-forth) between the baby and the adult is established.

The detected baby vocalizations are then compared against established language development norms, compiled across multiple authoritative sources of developmental milestones.

Different Vocalization Sounds defined as Annotation Classes in this Study

Single Syllable Babbling	Duplicated/Reduplicated/ Canonical Babbling	Variegated Babbling	Cooing	Baby - Other	Child Voice
Combination of a consonant and a vowel that is articulated clearly and separated from other vocalizations. (example: Ma, Da, Ne, Mi, Boo)	Combination of a consonant and a vowel, repeated two or more times. (example: Dada,Nene, Dadadadadada, Mememe). The syllables are generally not separated by pauses or breath.	sounds like language but may not have a meaning and includes a combination of different consonants and vowels. (example: Waagooowaa, Dabadiba).	Strings of vowels with occasional consonants (e.g. aaaaaaaamm, gooooooo, mmmmmm, oooo, aaaahhh, iiiiiii), any consonants tend to be velar or bilabial (/g/,/m/).	Any sound by the baby that is not a typical 'language' sound.	Any vocalizations by older children who can clearly speak fluently, but do not yet sound adult-like

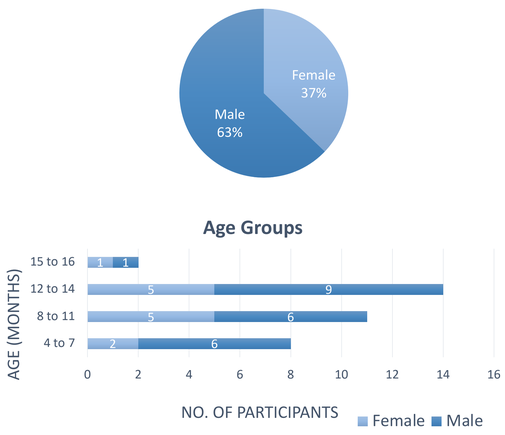


Babbling types detected by the Babbly algorithm (solid lines) vs. Normative data (dashed lines), across age.

STUDY RESULTS

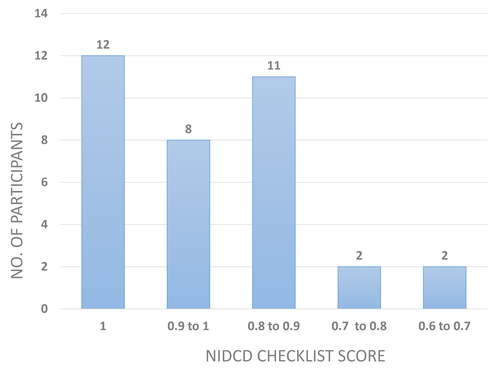
128 audios from 35 infants were collected and annotated. 27 of them were removed from the study due to low agreement between annotators, which was an indication of low quality. 101 audios were used in the study.

Patient Demographics:



The average of all participants' NIDCD checklist scores = 0.90

Caregivers completed the NIDCD Hearing and Communicative Development Checklist. According to the NIDCD questionnaire guidelines, any items checked "no" should be discussed with the child's doctor as they may indicate a delay. For each infant, we calculated the proportion of "yes" responses within the age range relevant to each infant.



These results suggest that the participants in our study were generally on a typical language development trajectory.

Model Results:

Analysis Per Audio

Classification Report: The below table quantifies whether each annotated class was predicted in each audio, whether it was missed by the model, or whether a class was predicted that was not present in the annotations (false positive). The total 'support' value therefore reflects the total number of labels across all audios, and the class-specific 'support' values reflect the total count of occurrences of a given class.

Annotation classes	Precision	Recall	f1-score	Support	True Negatives	True Positives	False Negatives	False Positives
Duplicated/Reduplicated/Canonical Babbling	0.84	0.84	0.84	46.0	50	37	7	7
Single Syllable Babbling	0.62	0.83	0.71	41.0	42	32	7	20
Variegated Babbling	0.69	0.90	0.78	44.0	42	38	4	17
Cooing	0.88	0.88	0.88	91.0	2	77	11	11
Baby - Other	0.91	1.00	0.95	94.0	0	92	0	9
Weighted Averages	**0.82**	**0.91**	**0.86**	**305.0**	**136**	**184**	**29**	**55**

Analysis Per Infant

Classification Report: The analysis was carried out on data summarized for each infant. Instead of calculating the presence/absence of each class in each audio, the below statistics apply to whether each class was present/predicted for each infant. The maximum support value per class is therefore equivalent to the number of infants in the study.

Annotation classes	Precision	Recall	f1-score	Support	True Negatives	True Positives	False Negatives	False Positives
Duplicated/Reduplicated/Canonical Babbling	0.83	0.95	0.89	21.0	8	20	1	4
Single Syllable Babbling	0.78	0.95	0.86	22.0	5	21	1	6
Variegated Babbling	0.70	1.00	0.82	21.0	3	21	0	9
Cooing	0.97	0.94	0.95	32.0	0	30	2	1
Baby - Other	0.97	1.00	0.98	32.0	0	32	0	1
Weighted Averages	**0.87**	**0.97**	**0.91**	**128.0**	**16**	**92**	**4**	**20**

CONCLUSIONS

- Babbly's algorithm detects and classifies babbling milestones with 86% accuracy on audio files of 30-80 seconds, and 91% accuracy when including 2-4 files per infant adding up to 1-4 minutes of audio.
- Babbly's AI provides high-accuracy results from short audio recordings taken by the parents during normal family activities without restriction on location or activity.
- The babbling classes present in each audio are captured by the AI 91% of the time (recall), and 97% of the time if multiple audios are aggregated per infant. The precision per audio is 82% and aggregated per infant is 87%.
- Babbly's algorithm demonstrates high performance across sex and age. Further, no meaningful differences in performance were detected in those splits for any particular babbling class or overall.
- Babbly's algorithm detects the presence of canonical babbling with an accuracy of 89%, making it a valuable tool to track language development in infants and potentially help with early diagnosis.

HITLAB
Healthcare Innovation Technology

ACKNOWLEDGEMENTS

Authors would like to acknowledge the HITLAB research team for study support and implementation; the platform developers for their work and technical support throughout the study, and the study participants.

HITLAB
3960 Broadway, Suite 501
New York, NY 10032
Phone: 212-543-0100
Email: research@hitlab.org