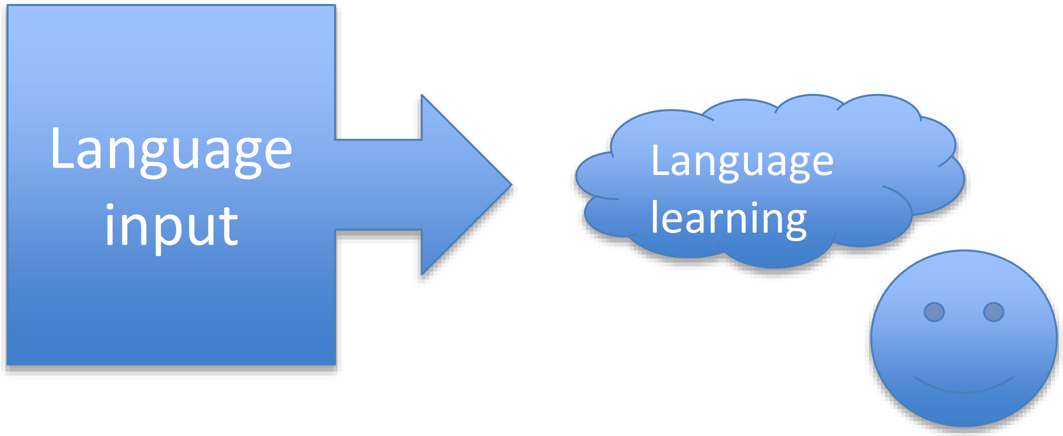
6~~-17~~-month-olds' Noun Input: Human and Automated Corpus Analyses

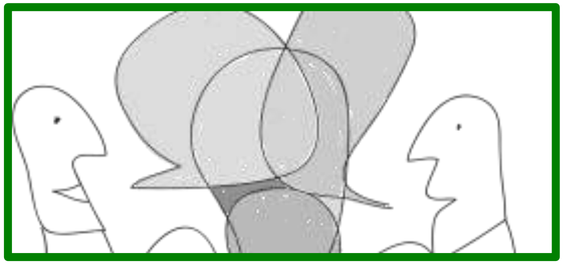
Sharath Skoorathota
Shaelise Morton
Andrei Amatuni
Elika Bergelson
University of Rochester/Duke University

Overview

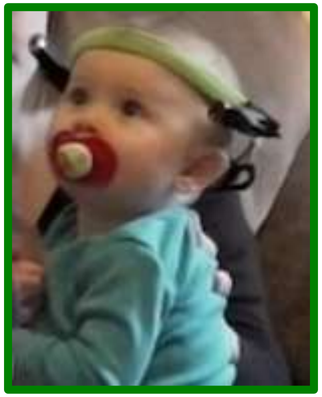
- SEEDLingS Corpus & Annotation description
- Describing Variability: getting the groundwork
- Human vs. LENA^{TM} speaker tags



Within each infant (n=44), measure:



1) Linguistic Context
(audio recordings)



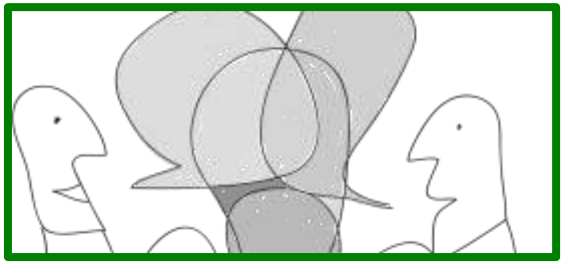
2) Visual Context
(on-head cameras +
context camera)

SEEDLingS

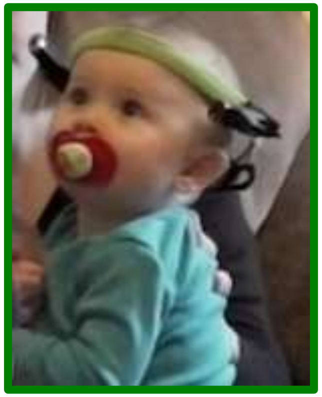
Study of Environmental Effects on Developing LINGuistic Skills

How does infants' environment give rise to word learning?

Within each infant (n=44), measure:

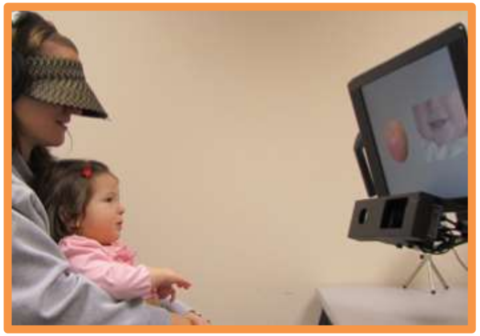


1) Linguistic Context
(audio recordings)



2) Visual Context
(on-head cameras + context camera)

Shared with
data repositories
(CHILDES-Homebank,
Databrary)



3) Word Comprehension
(eyetracking in the lab)

A) Common words
B) Child-specific words
C) Newly taught words

6 months 18 months

4) Longitudinally

A) Motor Development
B) Onset of Pointing
C) Demographics
D) MCHAT
E) Maternal Vocabulary

SEEDLingS Sample

- Monthly: 1 hr. video, all-day audio (~10 waking hours)
- **44** infants from Rochester, NY area
  - 47% female
  - 43% of moms stay-at-home full time at study-start
  - 50% of moms have at least Master’s; high vocab (PVT)
  - Moms are 23-42 y.o., average = 33.3 y.o.
  - Most families have 1 or more older siblings
  - Mix of urban, suburban, and rural families.
- Current Dataset: 6-14mo.: 100% 15-17mo.: 70-95%

Annotation

- Mark each object word (~concrete noun) directed to **baby** in daylong audio (~10 waking hrs.) and 1 hr. video

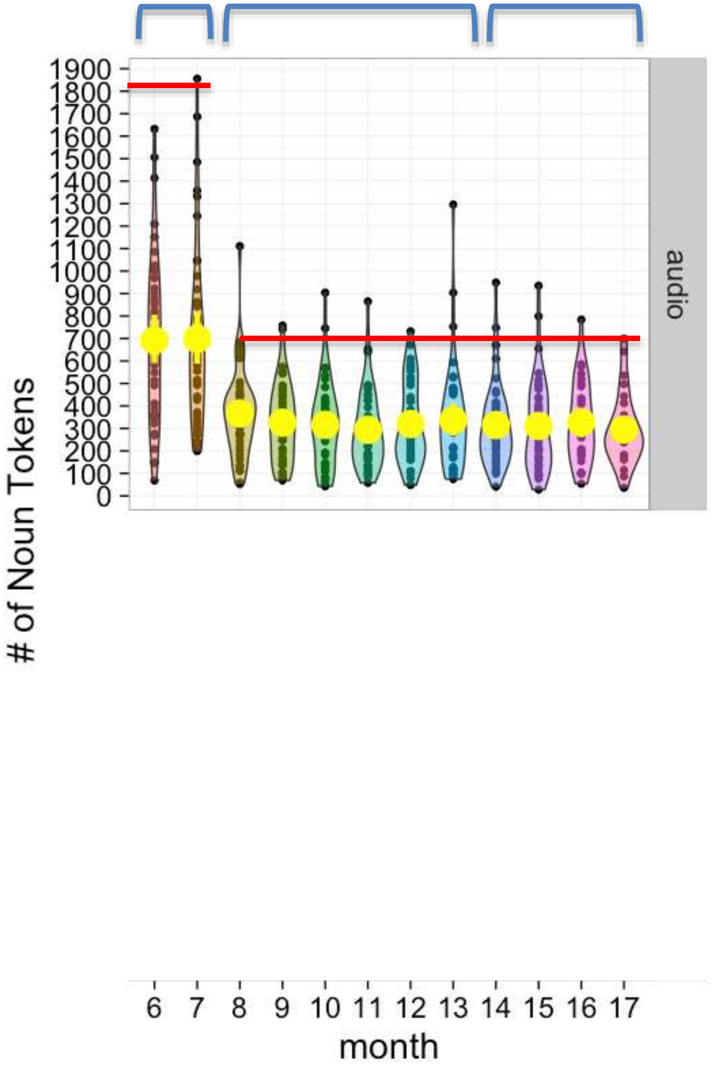
  (Operationalizing this is hard, ask me later)

- For each object, annotate:
  - **Object word**, as said (e.g. teethies, ball)
  - **Utterance type** (declarative, question, imperative, reading, singing, short-phrase)
  - **Object presence**: does it seem like object is present and attended to
  - **Speaker** (3 letter code for each speaker in a file/family)
  - Convert word to lemma, i.e. '**basic level**' of word
    - teeth, toothie, tooths, tooth -> tooth

- Total corpus:
  - >500 hour-long videos
  - >500 day-long audio recordings
    - Subsampled: full recordings month 6/7, 3-5 hours months 8-17
    - Used LENA 'Adult Word Count' and 'Child Vocalization Count' Average for Subsampling

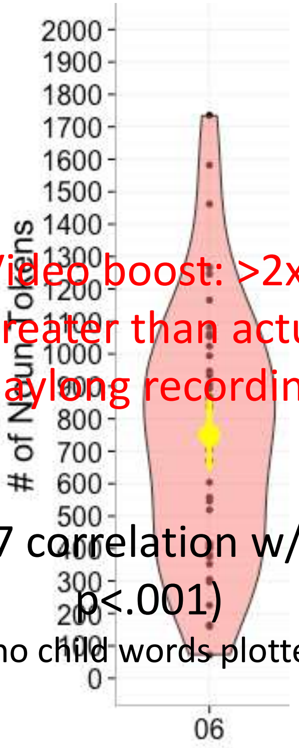
How many object words do 6–17 m.o. hear ?

10+ 4 hrs. 3 hrs.



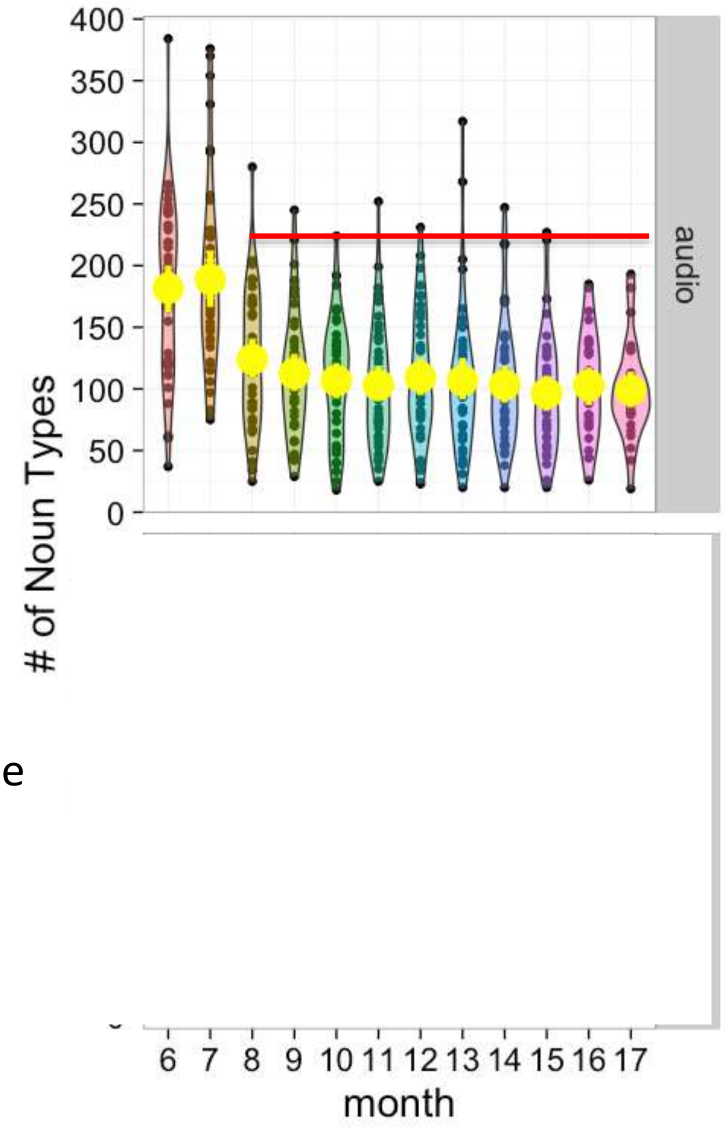
~700 tokens/day
~180 tokens/video-hour

~ 1 Type: 3 Tokens



Video boost: >2x greater than actual daylong recording

.7 correlation w/age
p<.001)
(no child words plotted)

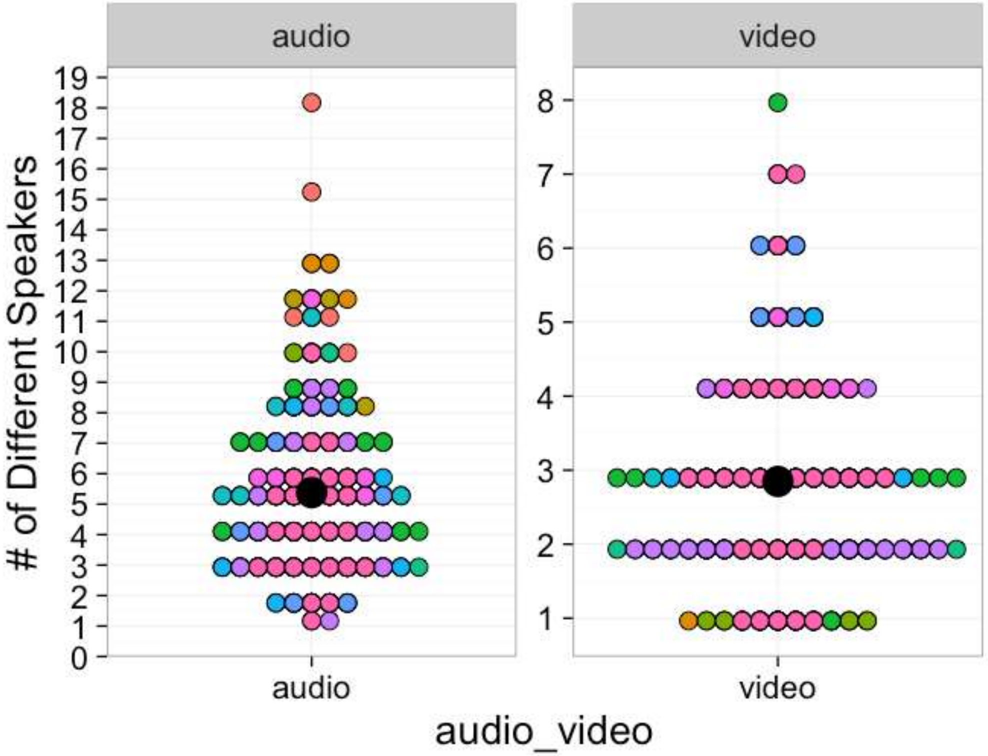


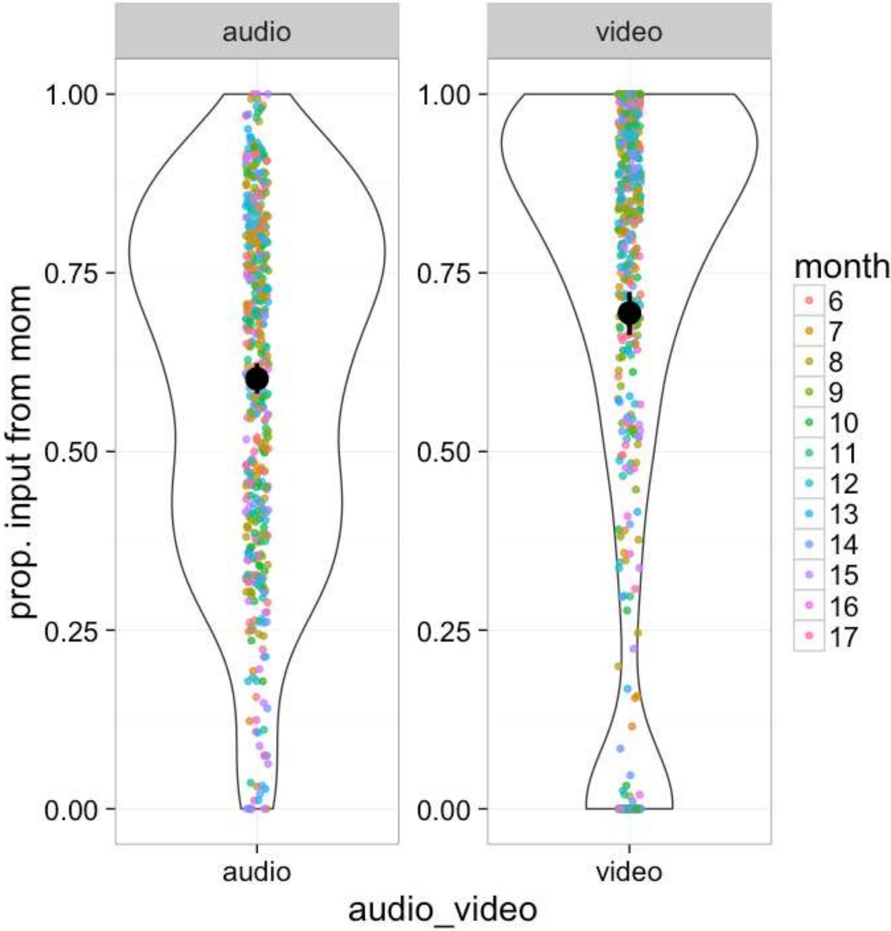
~180 types/day (more in denser hours)
~ 65 types/video-hour

What’s the speaker variability in the input?

~5 speakers/day
~3 speakers/video-hour

~63% input from mom

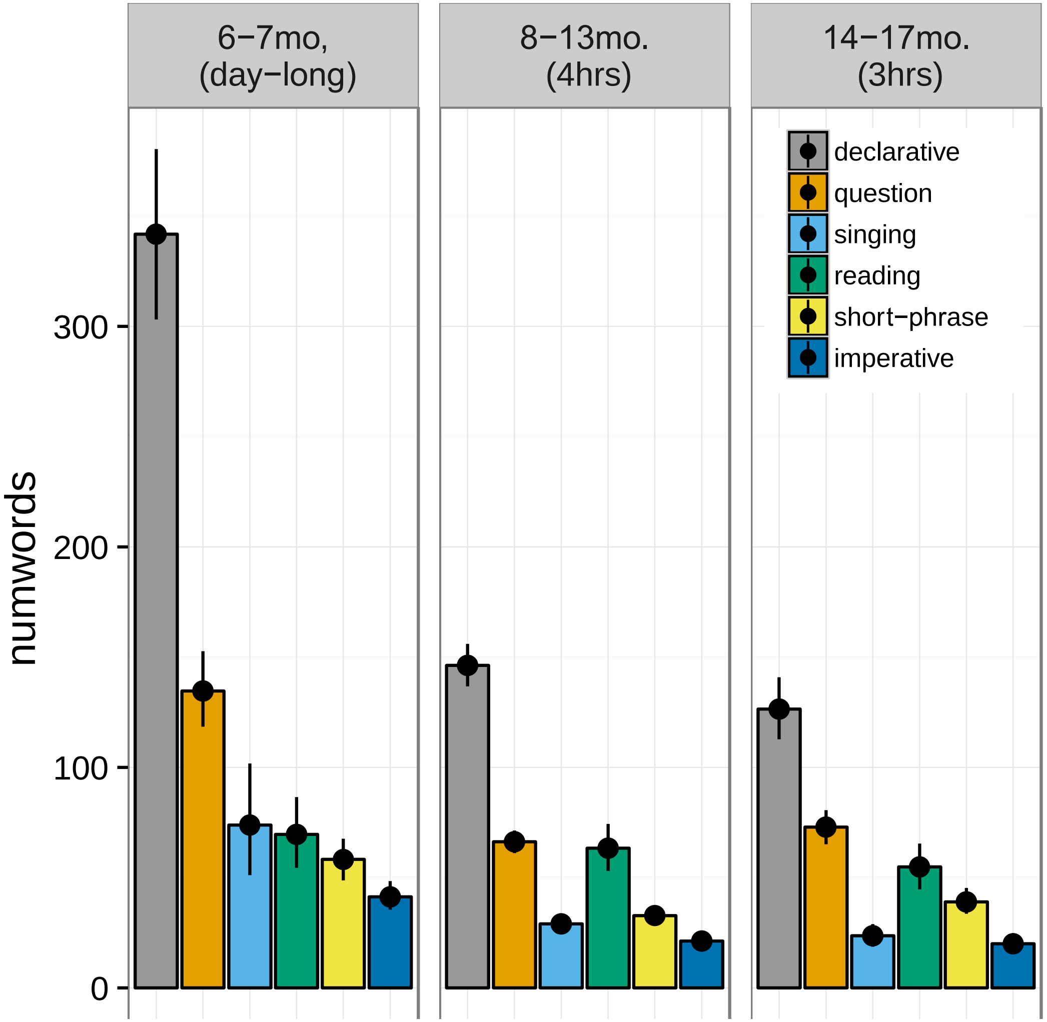




~>79% input from parents (not pictured)
[sampling bias due to privacy concerns]

How are object words distributed across utterance types?

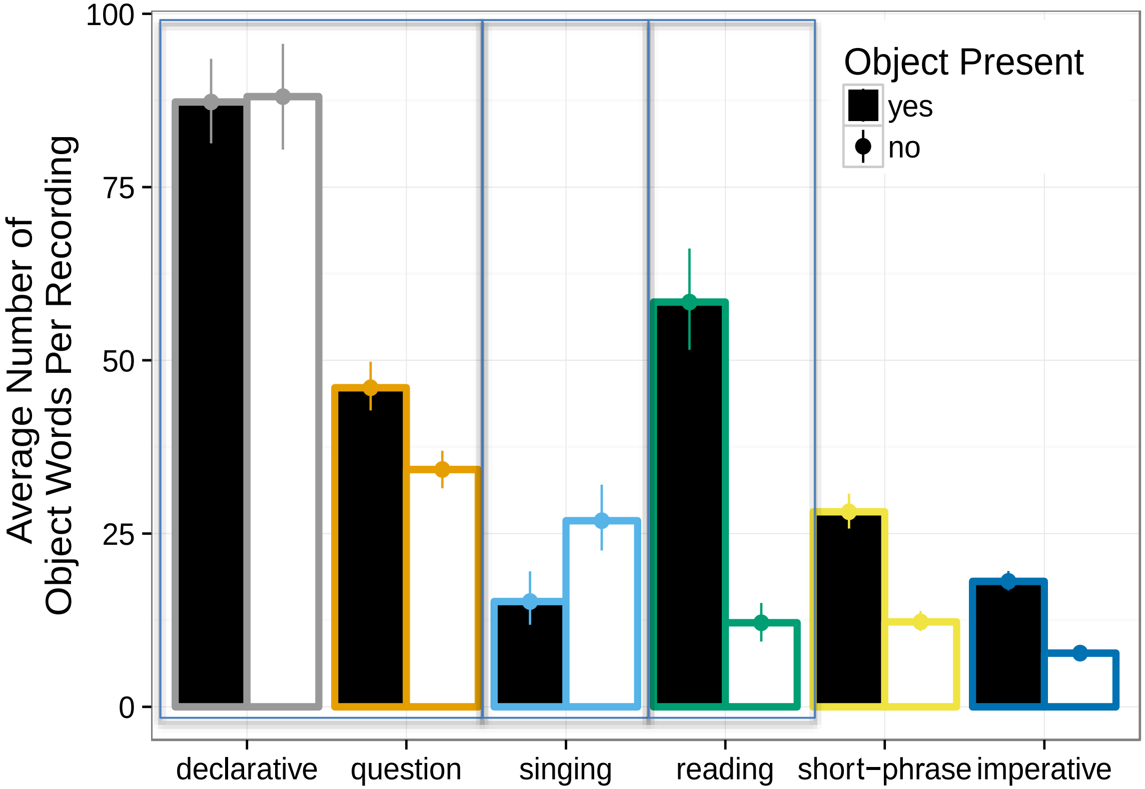
(audio recordings only; child's own utterances removed)



- Declaratives: ~45% of input
- Questions: ~20%
- Short-Phrase: ~10%
- Imperatives: ~5%
- Reading & Singing:
  - time-of-day effects?

Are objects equally 'present' across utterance types?

(audio recordings only; child's own utterances removed)

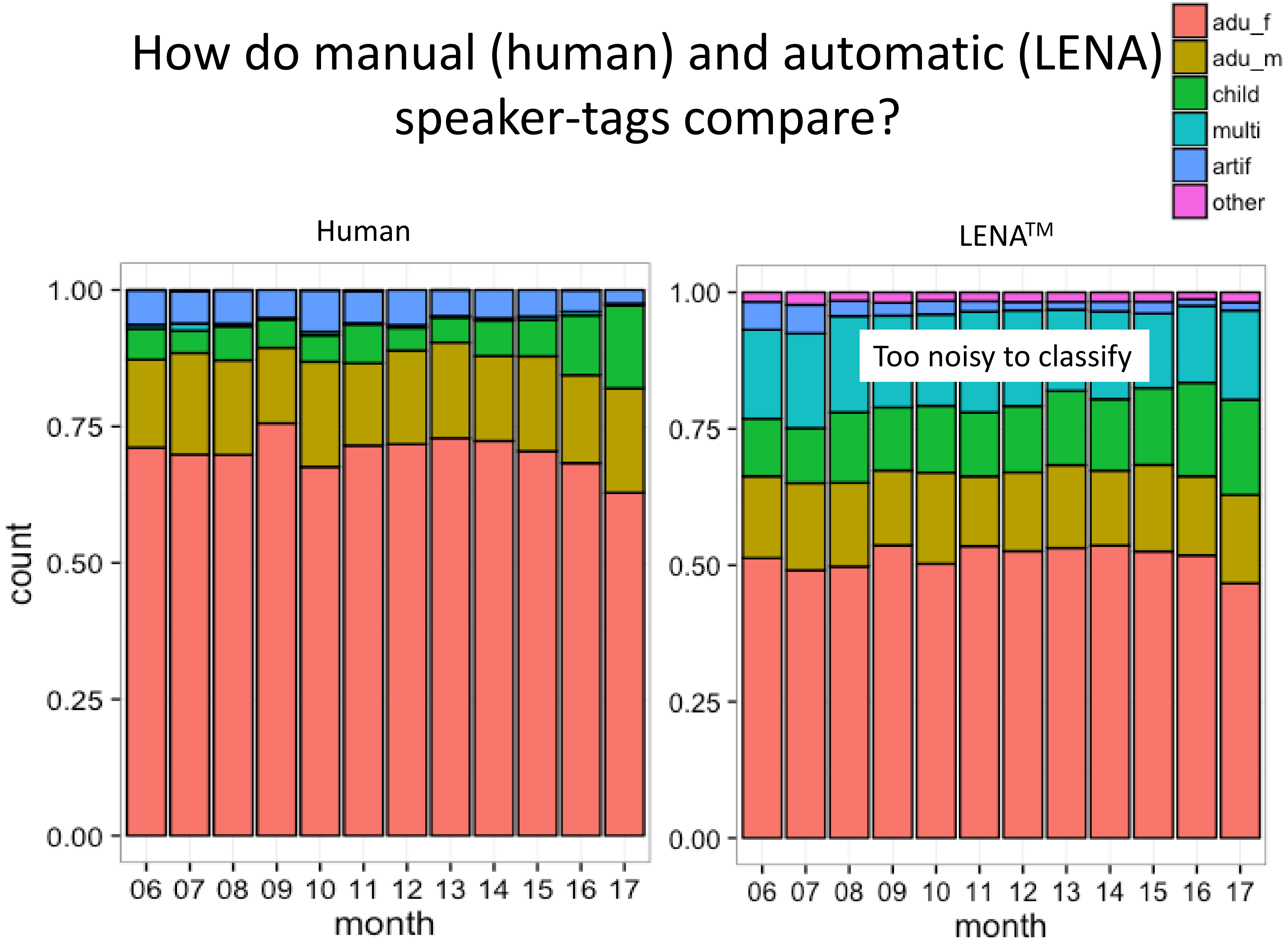


- Most common utterance-types have relatively low object presence
- Singing has least object-presence -> least 'learnable'?
- Reading benefit may relate to concept **and** word co-availability (not just sheer # types/tokens)

How do manual (human) and automatic (LENA) speaker-tags compare?

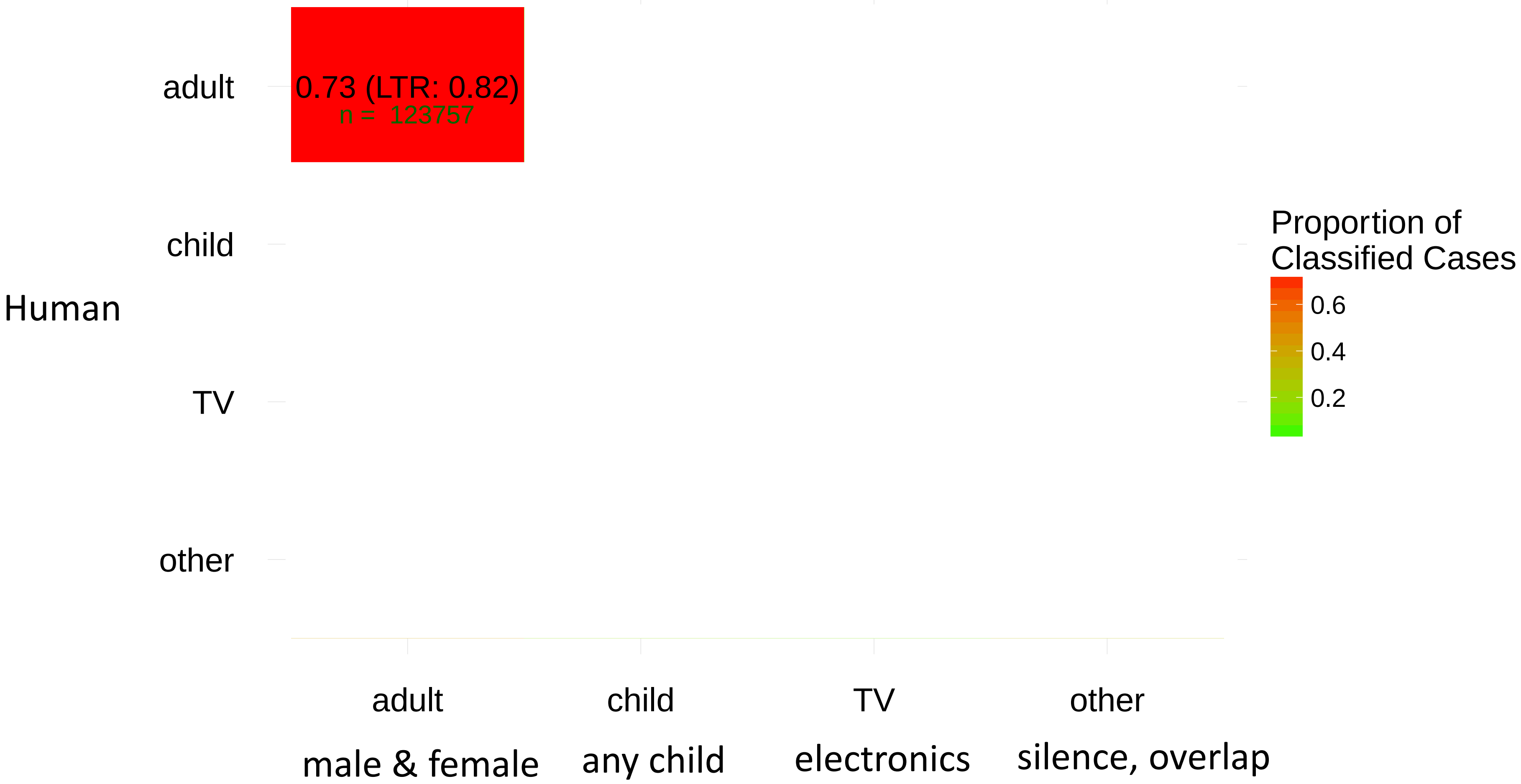
- LENA black-box algorithm spits out:
  - ‘Utterance-level’ segmentation
  - speaker-tag in 14 categories
    - e.g. Female Adult Near, Overlapping Sound Far
- Human coding spits out:
  - Speaker-tag at the individual level
    - e.g. Aunt Mary, Neighbor’s Kid
  - Tags are placed on LENA-segmented line
- Speaker-tag categories can be compared

How do manual (human) and automatic (LENA) speaker-tags compare?



Too noisy to classify

Confusion Matrix: LENA vs. Human Speaker-Tags



LENA

- Our accuracy prop's correlate with Lena Technical Report values (τ=.64, p<.001)
- Overall LENA vs. Human correlation across all 200K words more modest (τ=.34, p<.001)

LENA’s “Key Child” Category by Human Annotator Category

- <14mo., LENA’s ‘Key Child’ tag most often actually ‘adult female’ by human coders
    - Algorithm doesn’t modulate by age
- By **17** months, significant agreement between LENA and Human annotation of key-child talker
- Room for error improvement for both human and algorithm; limitations of method

Conclusions

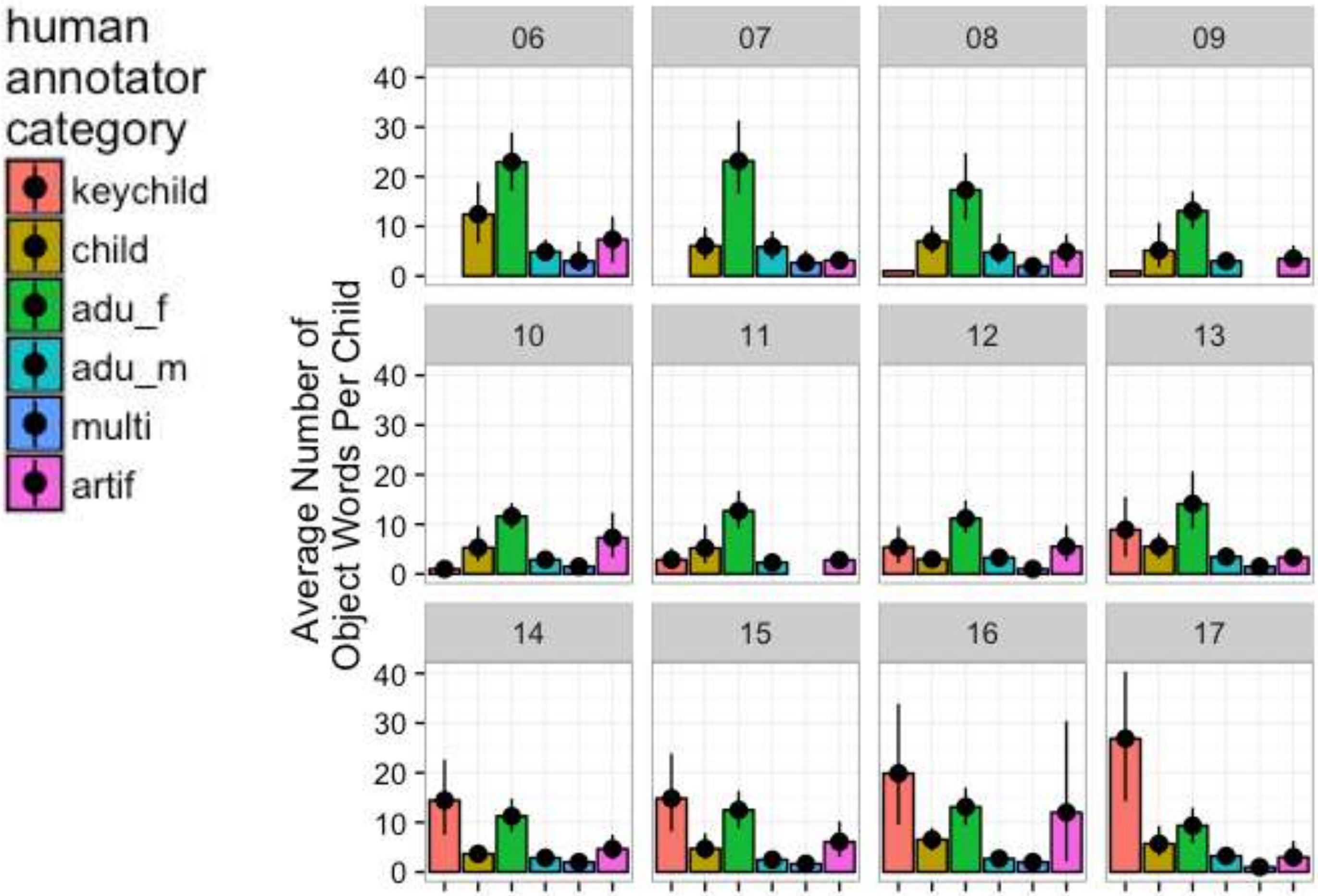
- **Quantity**: Infants hear ~700 object words tokens a day
- **Talker Variability**: Most come from 1-2 speakers
- **Utterance Types**: Declaratives & Questions dominate
- **Object Presence**: varies by utterance-types
- LENA algorithm ‘sufficiency’ depends on goals
  - Good flashlight for further human coding
- Relevant variables can (**must**?) be linked to in-lab comprehension and at-home production to test theories of language development

Thanks!

- SEEDLingS Staff: Sharath Koorathota, Shaelise Morton, Andrei Amatuni, Josh Schneider, Shannon Dailey & small army of RAs! (see our website)
- NIH Early Independence Award
- Dick Aslin, U. of Rochester Brain & CogSci
- Our 44 SEEDLingS and their families!

Backup slides

LENA's "Key Child" Category by Human Annotator Category

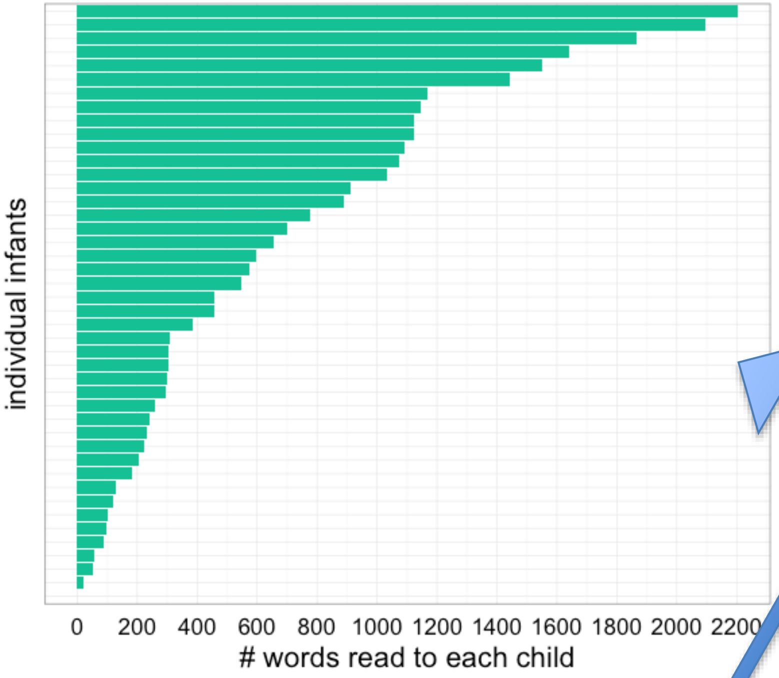


- <14 months, LENA's 'key child' is mostly adult females
- By 17 months, accurate key-child classification
- Room for error improvement for both human and algorithm

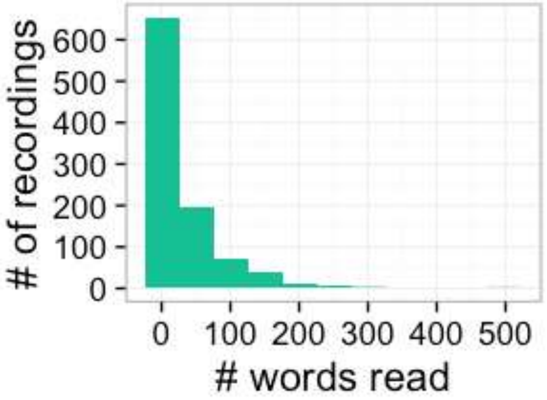
Why study babies longitudinally?

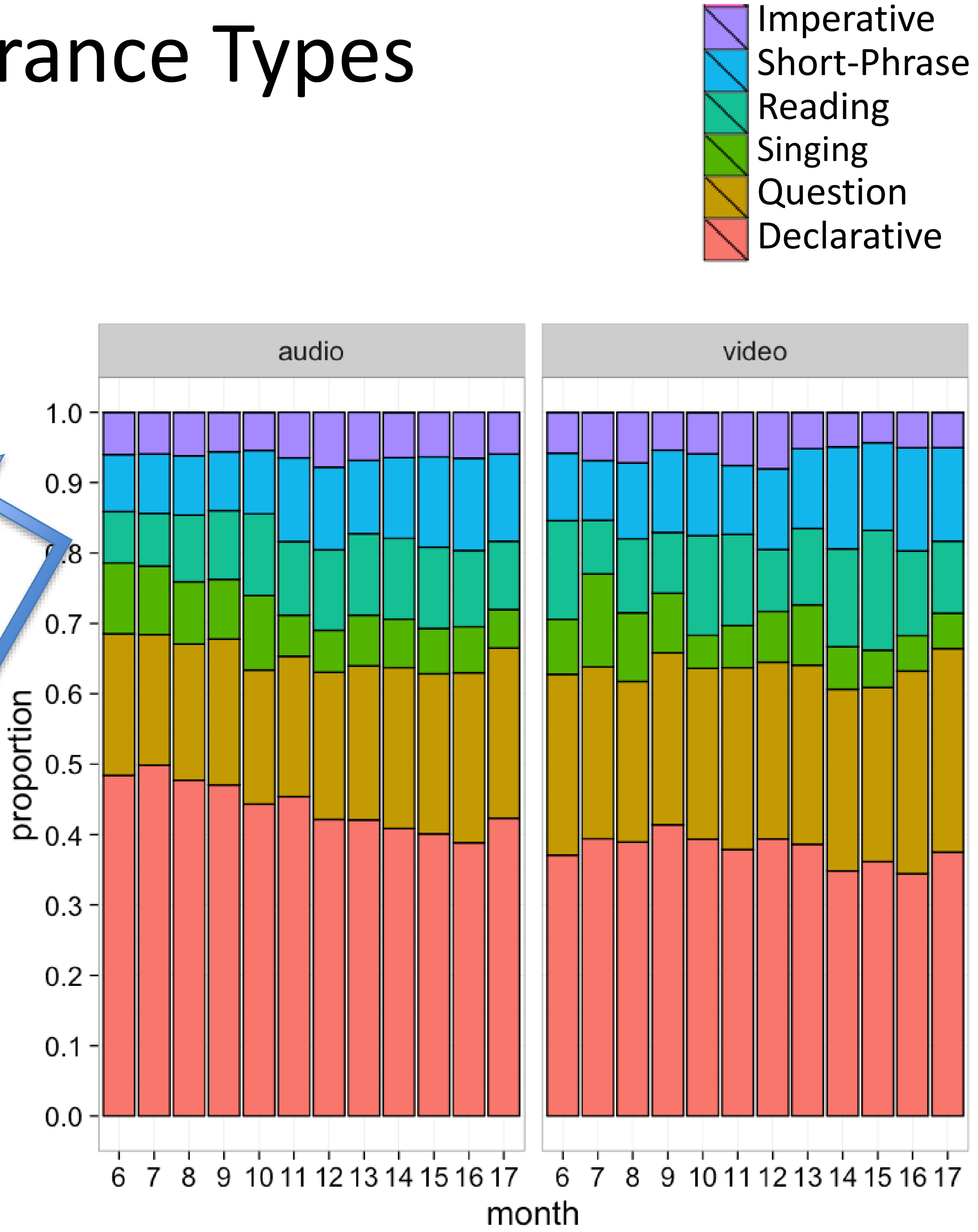
- Because you’re masochistic and very patient
- To control for within-child variability
  - ‘fussy’ babies in the lab
  - age vs. family variability in the home
- To look at individual differences
- To inch closer to causality
  - Or at least, have stronger input ➔ output tests

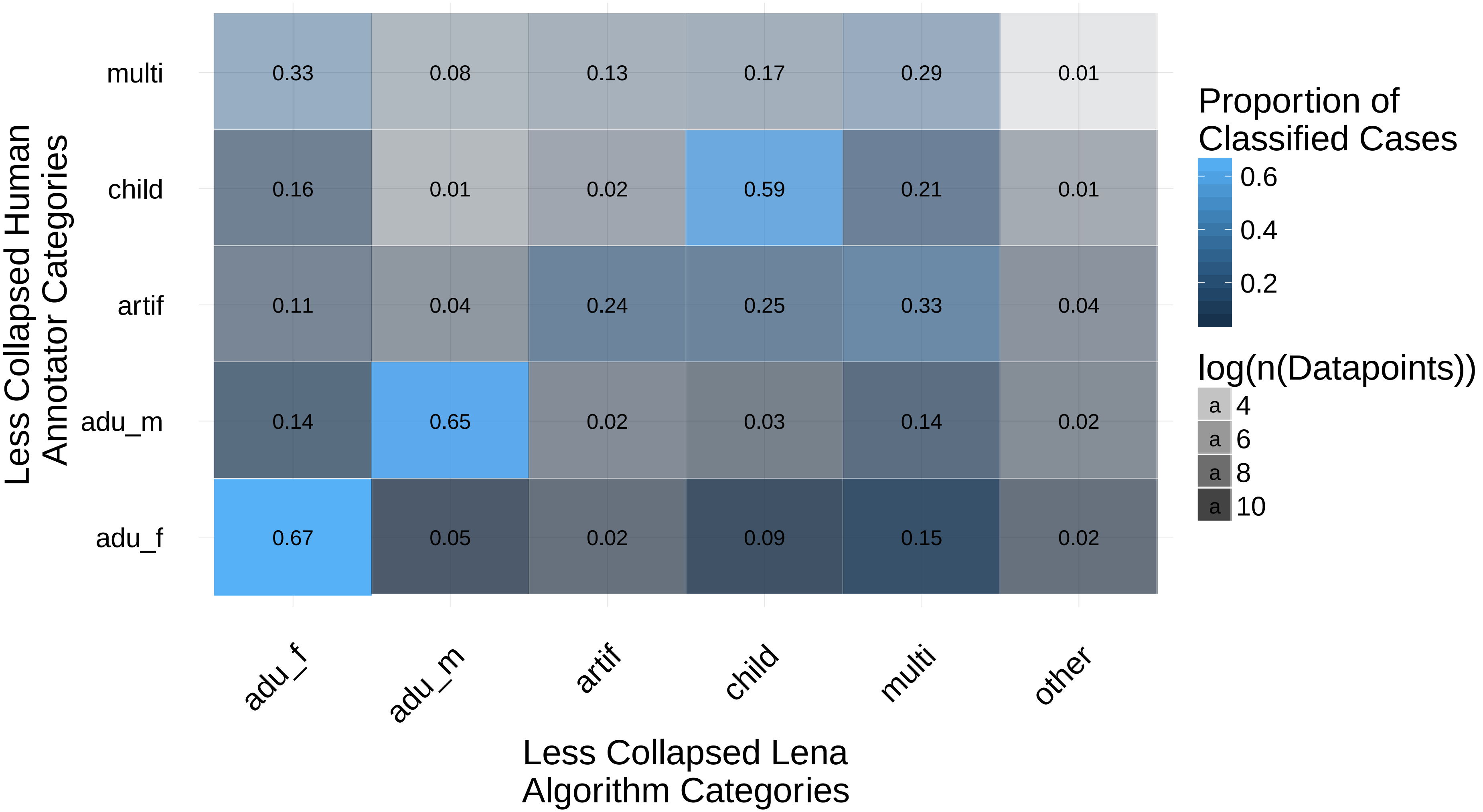
Utterance Types



(all months combined)






Less Collapsed Human
Annotator Categories
multi
child
artif
adu_m
adu_f
0.33
0.08
0.13
0.17
0.29
0.01
0.16
0.01
0.02
0.59
0.21
0.01
0.11
0.04
0.24
0.25
0.33
0.04
0.14
0.65
0.02
0.03
0.14
0.02
0.67
0.05
0.02
0.09
0.15
0.02
adu_f
adu_m
artif
child
multi
other
Less Collapsed Lena
Algorithm Categories
Proportion of
Classified Cases
0.6
0.4
0.2
log(n(Datapoints))
a 4
a 6
a 8
a 10