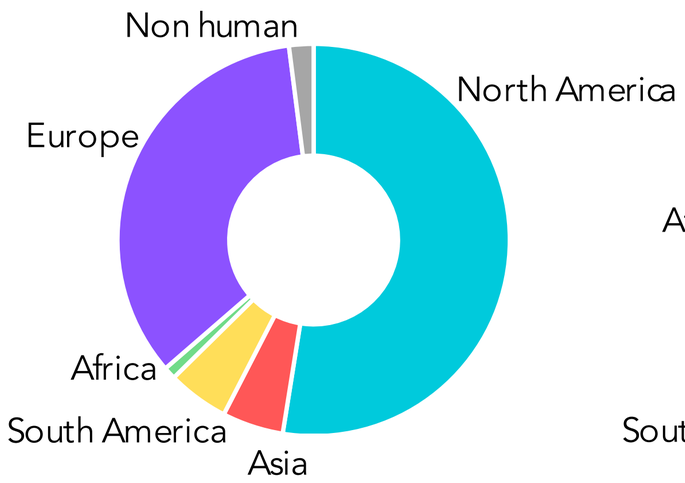
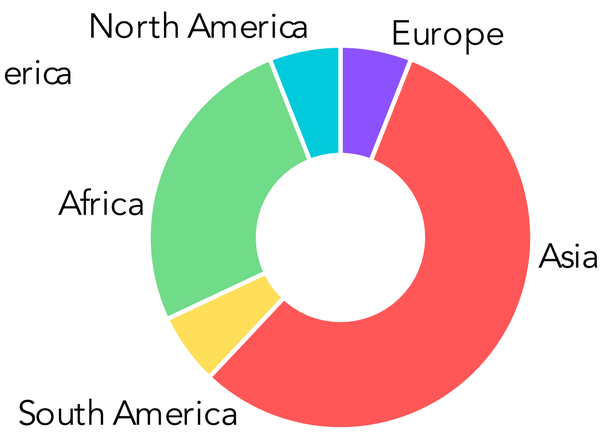
Large-scale audio-recordings to study infant language acquisition

Camila Scaff, Marvin Lavechin & Alejandrina Cristia*
LSCP, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University, *alecristia@gmail.com

Who participates in cognitive development research?



Where do children live?



- Most of developmental data is collected in North America and Europe, only 12% of the world's child population.
- Americans and Europeans are WEIRD (Western, Educated, Industrialized, Rich and Democratic[1])

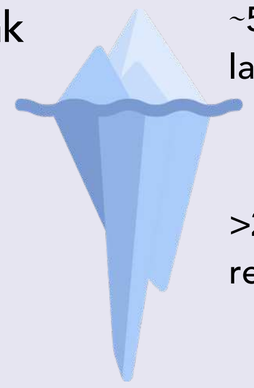
WEIRD
Higher socioeconomic status
Widespread formal education
Fewer children per family
Single caregiver (often mother)

The rest of the world
Lower socioeconomic status
Rarer formal education
More children per family
Multiple caregivers (inc. peers)

Assessing the accuracy of LENA® software

13 Tsimane' children (aged 1y3m to 4y-10m, average 2y10m) wore a LENA® device for 1-2 days.

- 1 minute per hour annotated by a trained listener (who did not speak the language)
- LENA® automatically labels full audio into categories:
  - Key child (CHI, the one wearing the device)
  - Other children (OCH)
  - Male adults (MAL)
  - Female adults (FEM)



~5h of human-labeled data

>270h of recorded audio

Results

LENA®'s **PRECISION** for key child (CHI) is 62%, other speaker categories <45%

Human \ LENA(R)	FEM	CHI	OCH	MAL	NonSpeech
NonSpeech	34 % 22412	26 % 21306	40 % 26551	40 % 5866	91 % 1281645
MAL	7 % 4622	2 % 1557	1 % 534	44 % 6429	1 % 18578
OCH	9 % 5611	7 % 6050	32 % 20880	1 % 205	2 % 29928
CHI	10 % 6710	62 % 49805	17 % 11482	2 % 269	3 % 37164
FEM	40 % 26381	2 % 1988	10 % 6340	13 % 1961	3 % 37726

LENA®'s **RECALL** for key child (CHI) is 36%, other speaker categories <48%

Human \ LENA(R)	FEM	CHI	OCH	MAL	NonSpeech
NonSpeech	51 % 22412	35 % 21306	48 % 26551	59 % 5866	94 % 1281645
MAL	3 % 4622	0 % 1557	0 % 534	20 % 6429	0 % 18578
OCH	9 % 5611	11 % 6050	33 % 20880	2 % 205	2 % 29928
CHI	3 % 6710	47 % 49805	10 % 11482	5 % 269	2 % 37164
FEM	35 % 26381	6 % 1988	9 % 6340	15 % 1961	2 % 37726

Can we debias language acquisition samples through daylong audio recordings?

Some locations where daylong audio-recordings have been gathered, mostly with LENA®

LENA® Foundation products:
- 300 US$/device / >5k US$ license
- Software (trained on American English)

Ongoing research directions

- Evaluating LENA® in further detail: low performance due to language, culture, ...?
- Developing open source re-trainable software alternatives
- Exploring secure hardware alternatives

pyannote

[1] Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. Nature, 466(7302), 29.
[2] Casillas, M., & Cristia, A. (2019). A step-by-step guide to ...long-format speech environment ...) recordings. Collabra, 5(1).
[2] Bredin, H. (2018) https://github.com/pyannote

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR