The development of canonical proportion continues through 6 years of age

Kasia Hitczenko, Elika Bergelson, Marisa Casillas, Heidi Colleran, Meg Cychosz, Pauline Grosjean, Lisa R. Hamrick, Bridgette L. Kelleher, Camila Scaff, Amanda Seidl, Sarah Walker, and Alejandrina Cristia

Babies' vocalizations becoming increasingly adult-like

Non-canonical vocalizations
(vowel or consonant only)

Canonical vocalizations
(combines vowel and consonant)

Oller et al. (1976); Oller (2000)

Babies’ vocalizations becoming increasingly adult-like

Non-canonical vocalizations
(vowel or consonant only)

Canonical vocalizations
(combines vowel and consonant)

Oller et al. (1976); Oller (2000)

Babies' vocalizations becoming increasingly adult-like

Non-canonical vocalizations
(vowel or consonant only)

Canonical vocalizations
(combines vowel and consonant)

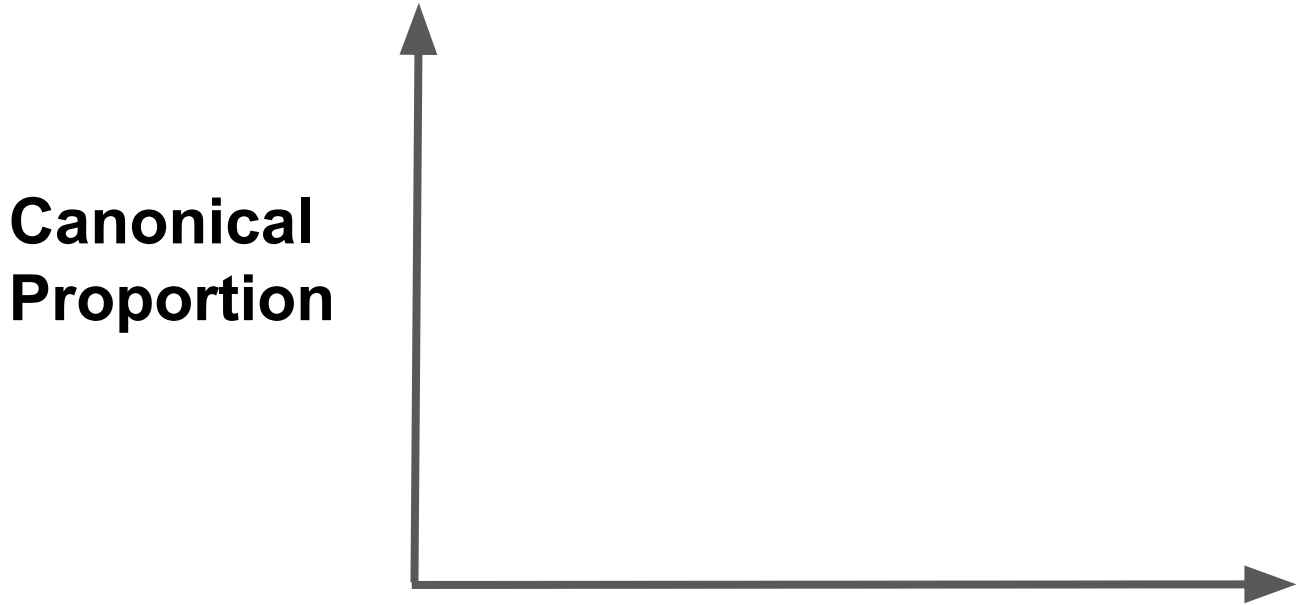
Oller et al. (1976); Oller (2000)

We can measure this development using “canonical proportion”:

$$\text{Canonical proportion} = \frac{\text{Number of canonical vocalizations}}{\text{Number of canonical vocalizations + non-canonical vocalizations}}$$

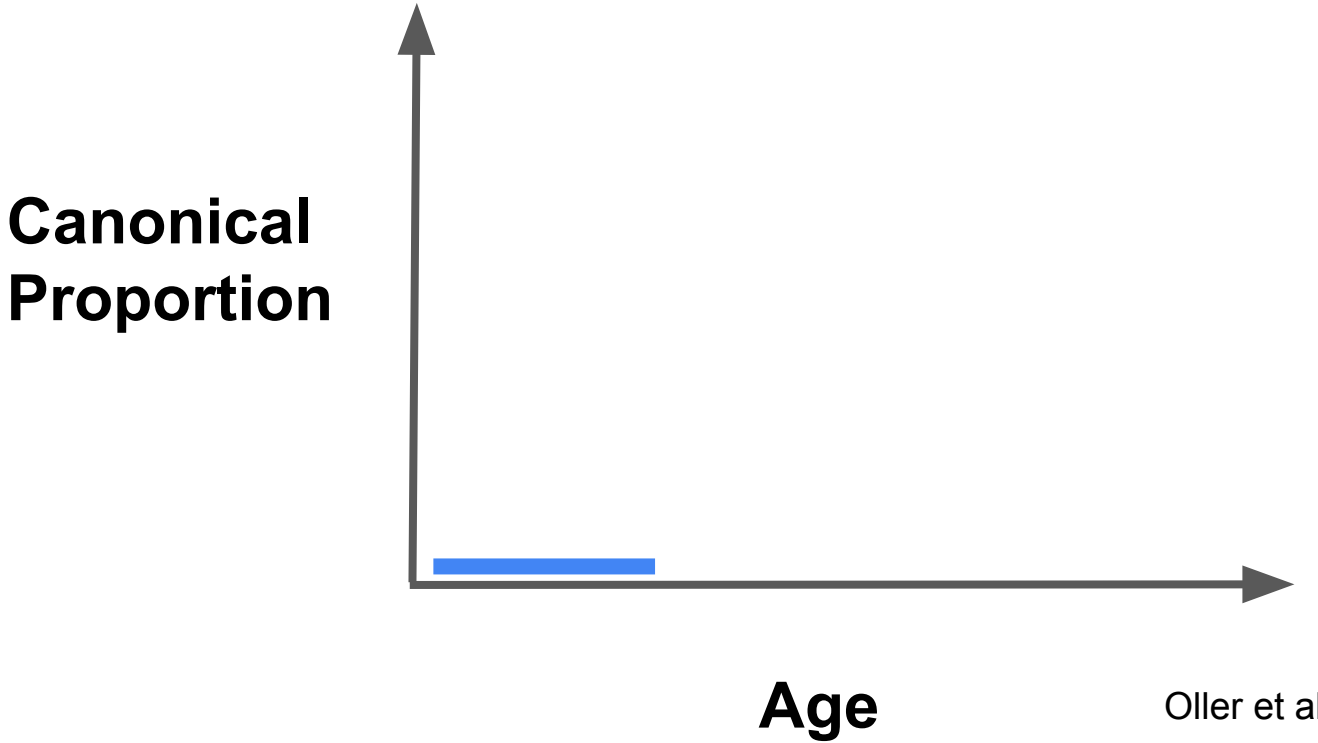
Cychosz et al. (2021); Eilers & Oller (1994);
Lee et al. (2018); Oller & Eilers (1988); Oller (2000)

Under standard accounts of language acquisition, this key aspect of phonological learning is complete by toddlerhood, but this has not been tested



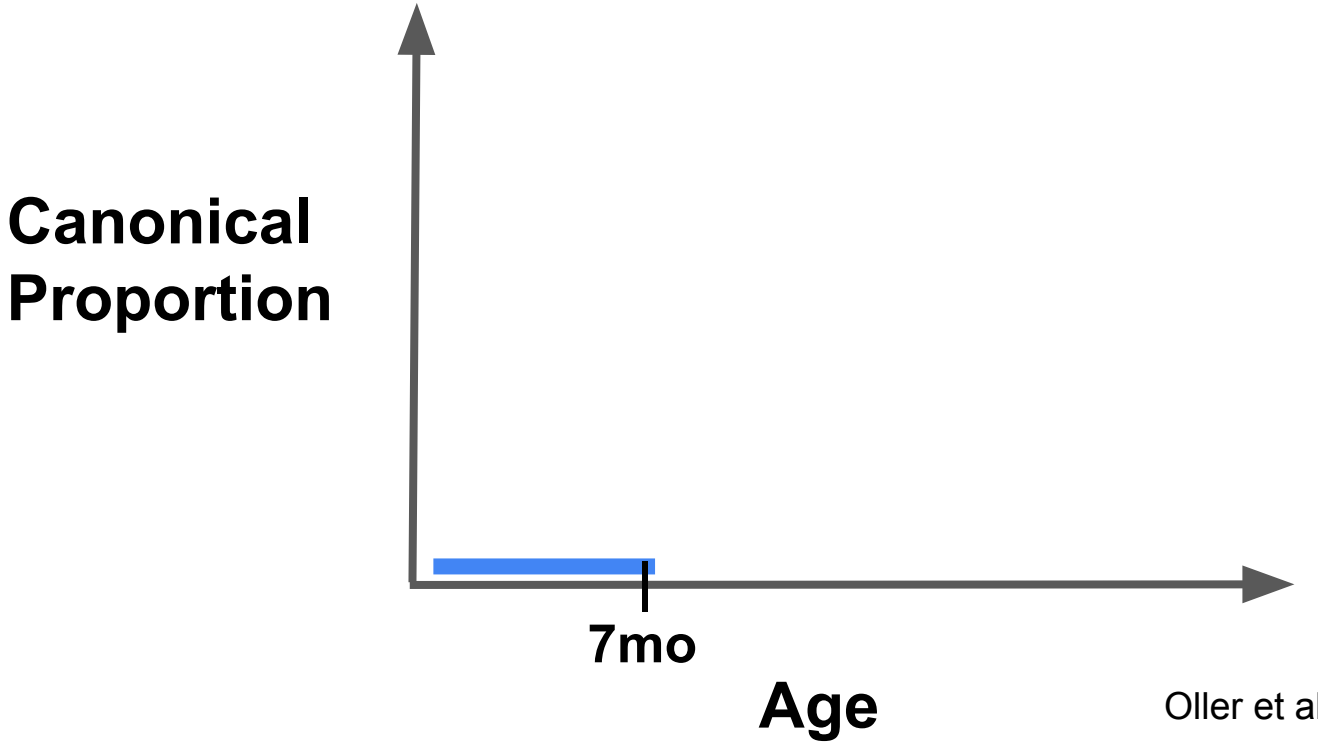
Oller et al. (1976); Oller (2000)

Under standard accounts of language acquisition, this key aspect of phonological learning is complete by toddlerhood, but this has not been tested



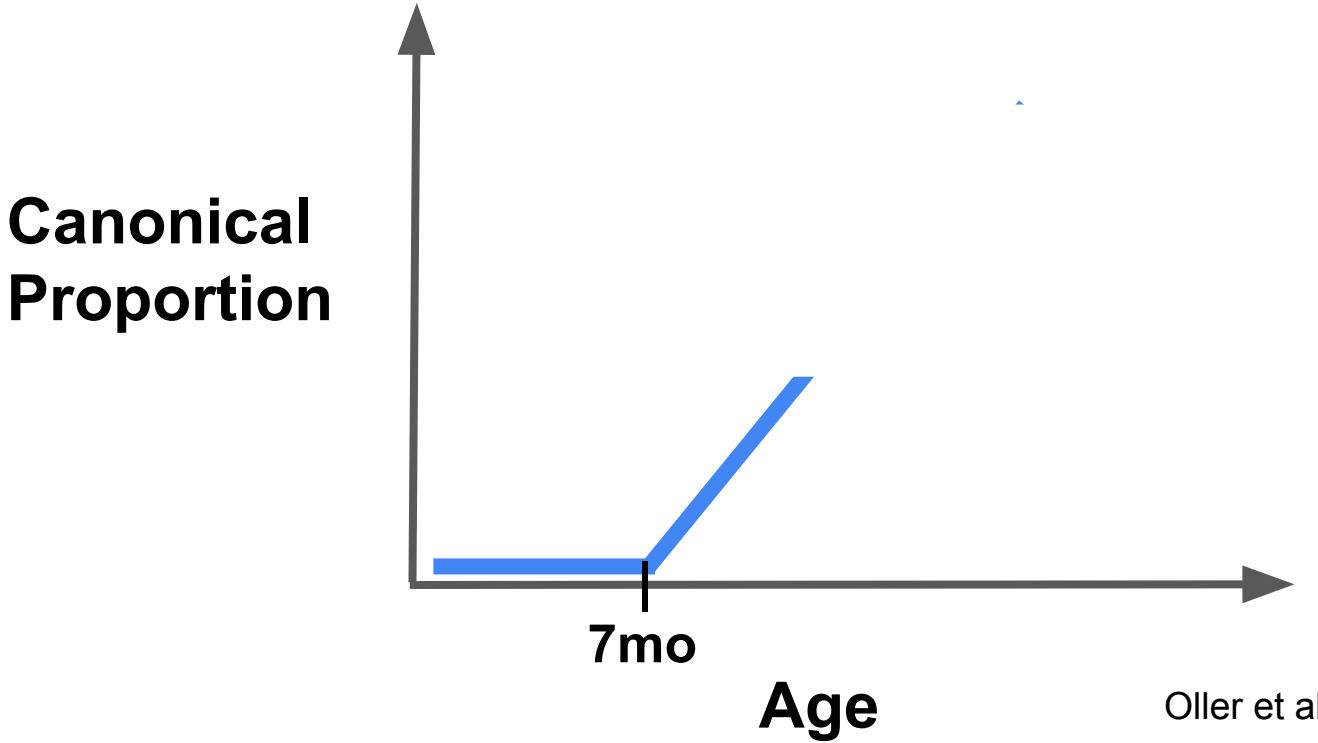
Oller et al. (1976); Oller (2000)

Under standard accounts of language acquisition, this key aspect of phonological learning is complete by toddlerhood, but this has not been tested



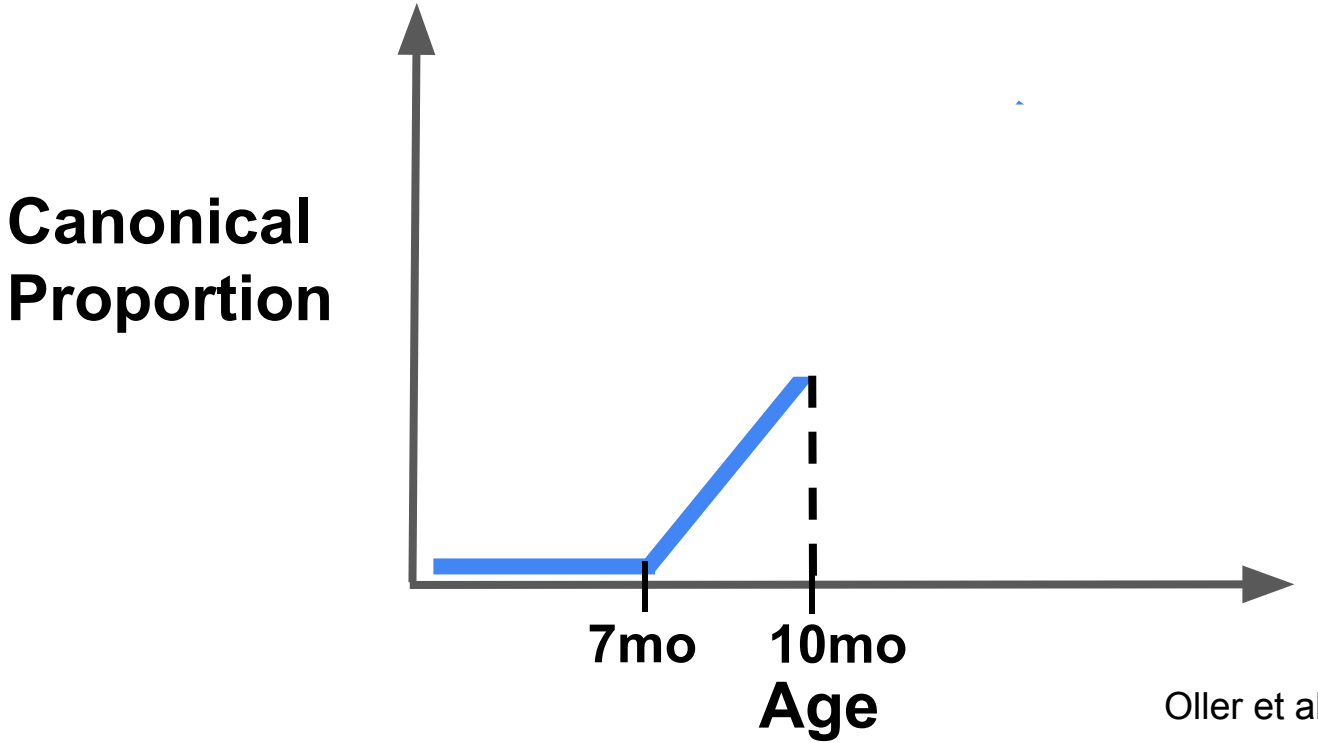
Oller et al. (1976); Oller (2000)

Under standard accounts of language acquisition, this key aspect of phonological learning is complete by toddlerhood, but this has not been tested



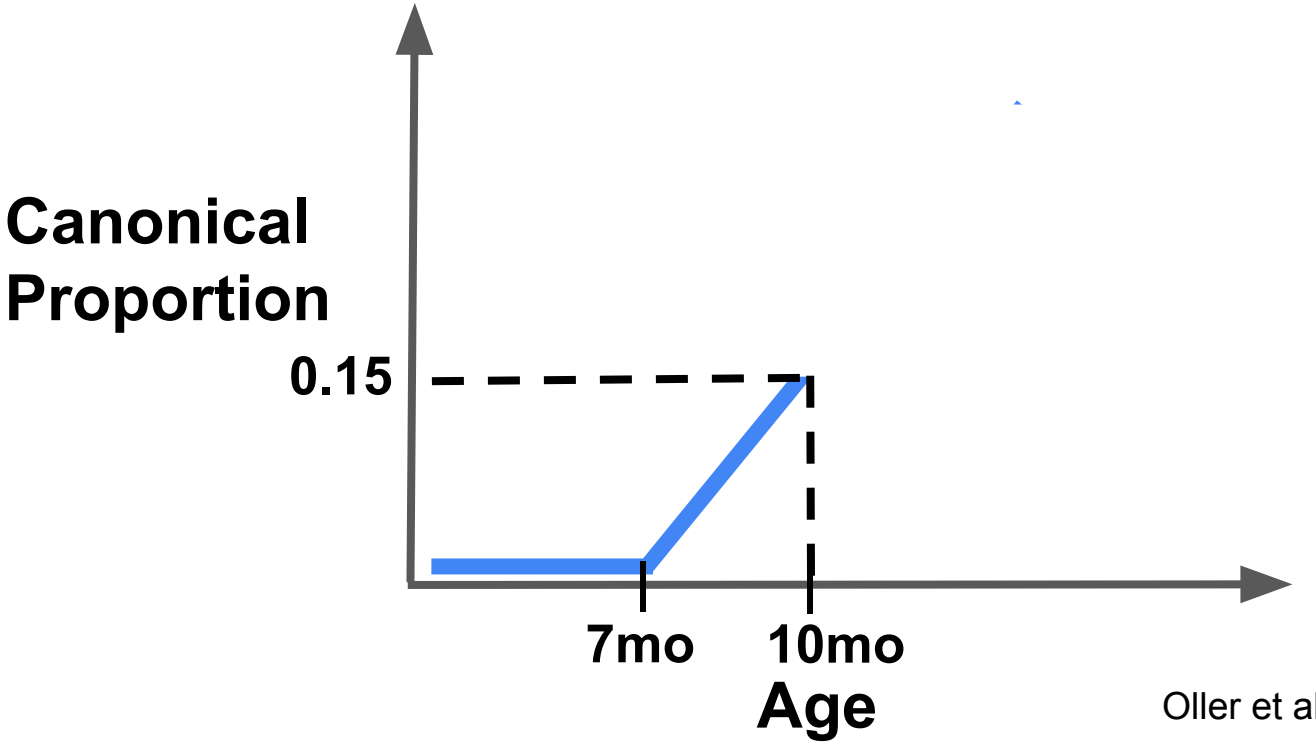
Oller et al. (1976); Oller (2000)

Under standard accounts of language acquisition, this key aspect of phonological learning is complete by toddlerhood, but this has not been tested



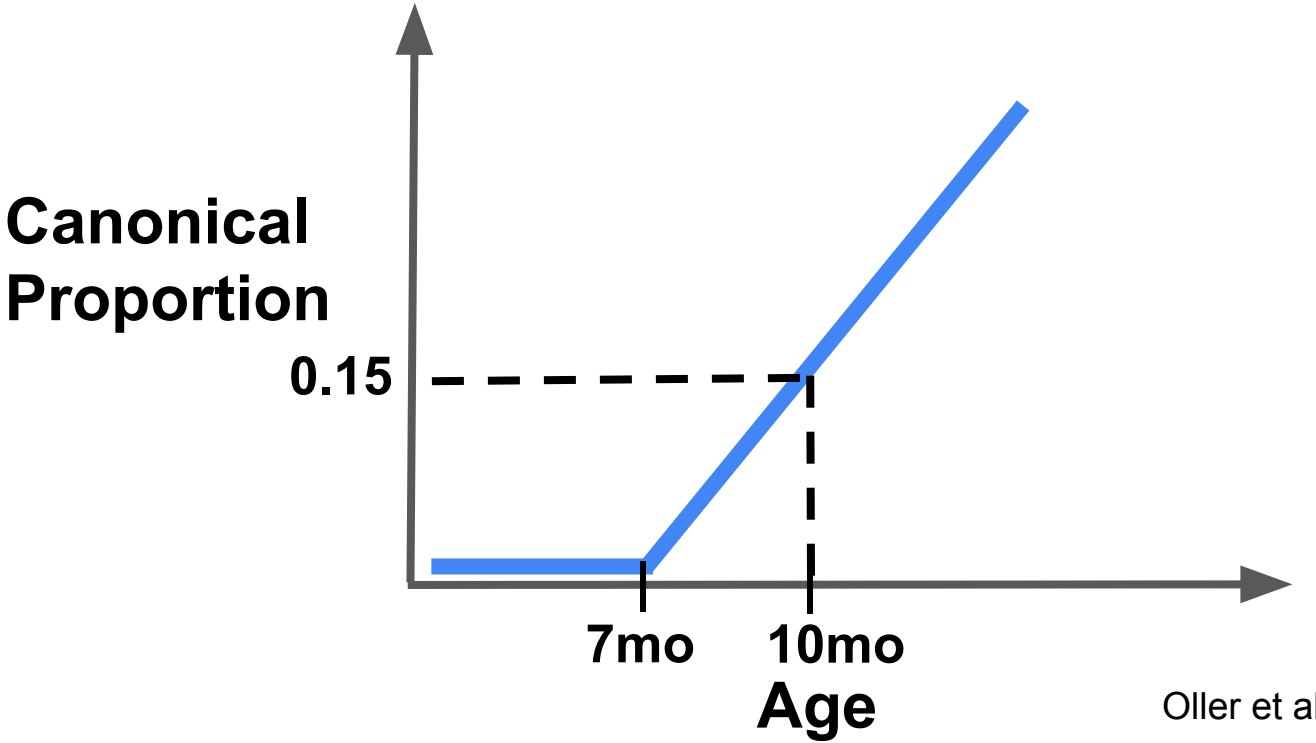
Oller et al. (1976); Oller (2000)

Under standard accounts of language acquisition, this key aspect of phonological learning is complete by toddlerhood, but this has not been tested



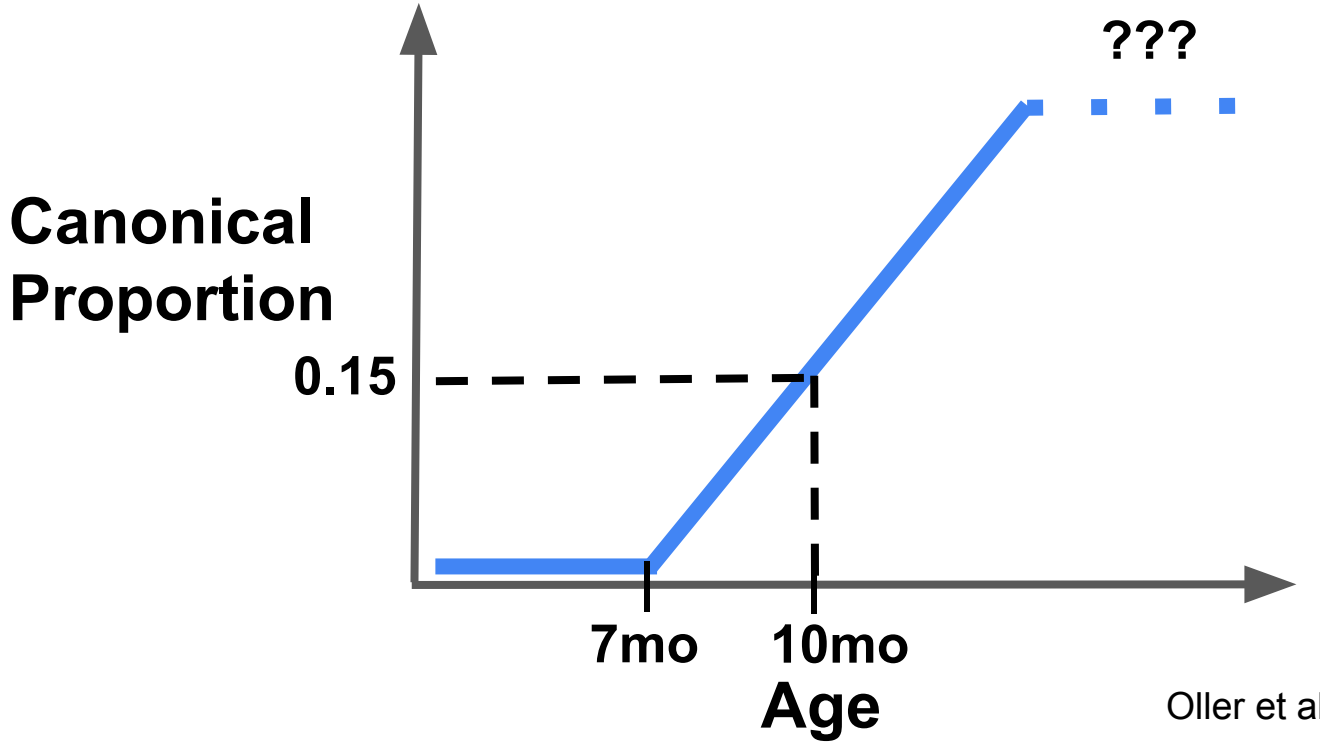
Oller et al. (1976); Oller (2000)

Under standard accounts of language acquisition, this key aspect of phonological learning is complete by toddlerhood, but this has not been tested



Oller et al. (1976); Oller (2000)

Under standard accounts of language acquisition, this key aspect of phonological learning is complete by toddlerhood, but this has not been tested

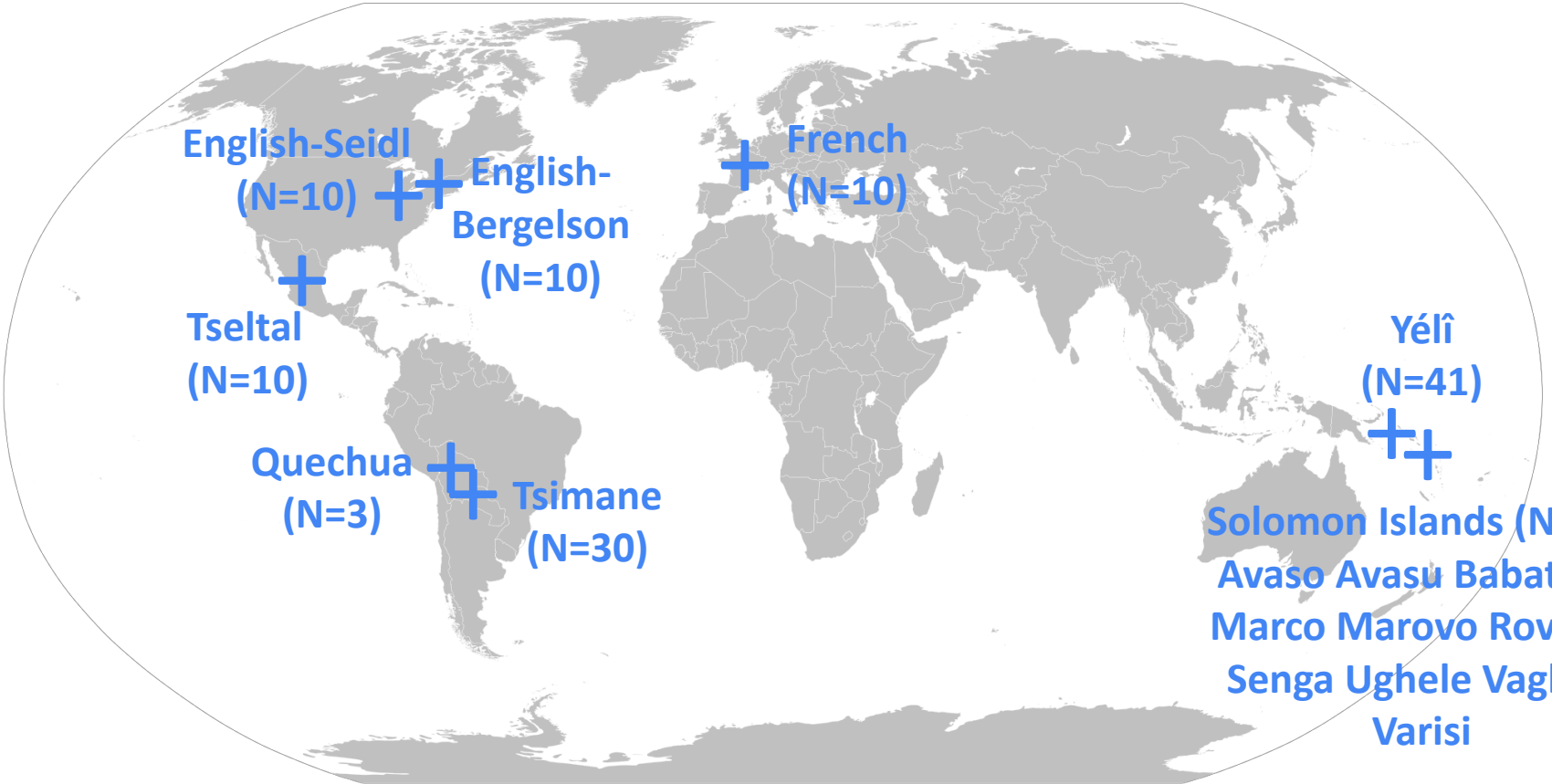


Oller et al. (1976); Oller (2000)

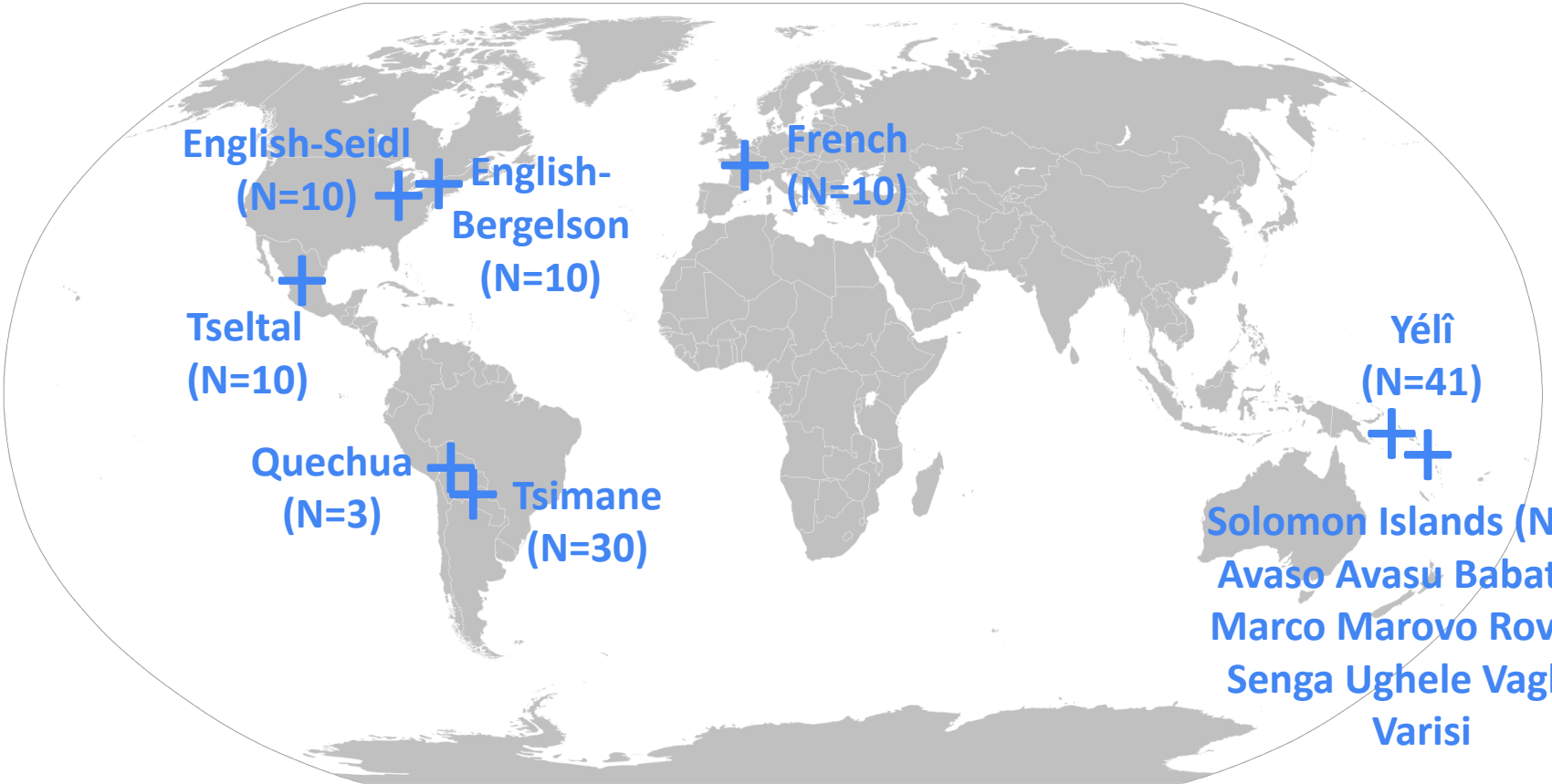
What does the trajectory of canonical proportion look like beyond the ages typically studied?

Does it vary by language?

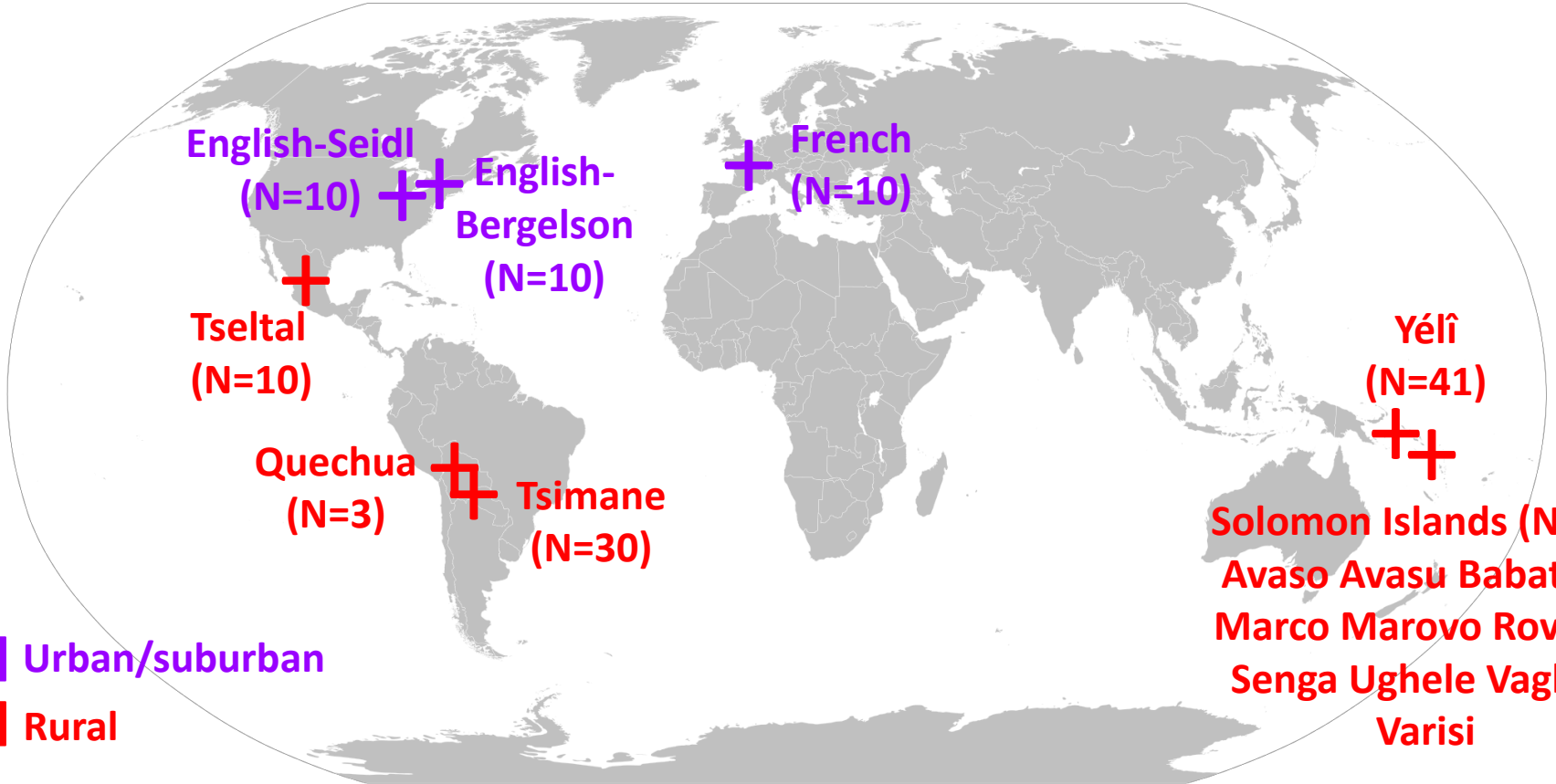
We study a diverse sample of 129 children (0;1-6;0)



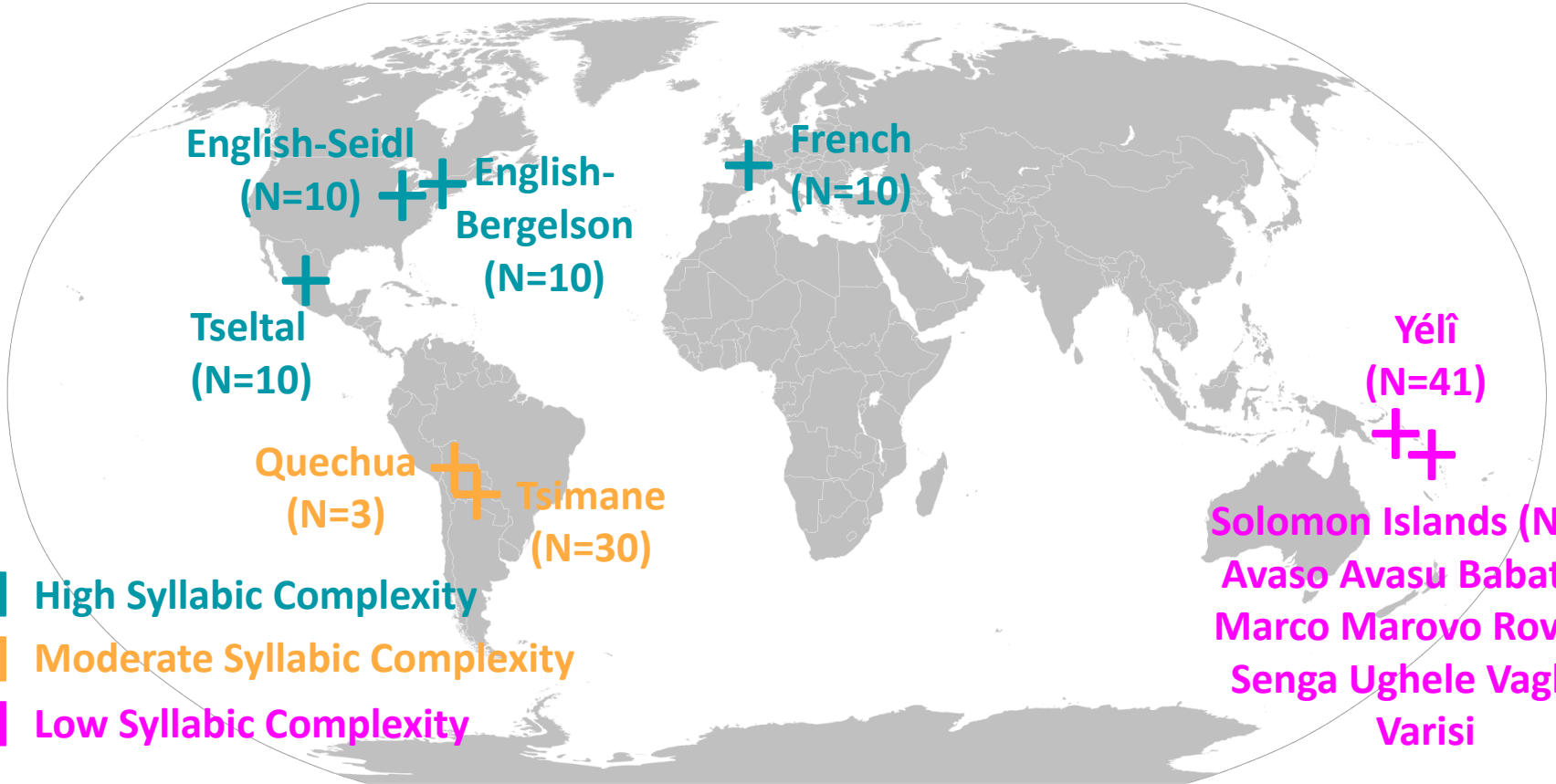
We study a diverse sample of 129 children (0;1-6;0)



We study a diverse sample of 129 children (0;1-6;0)



We study a diverse sample of 129 children (0;1-6;0)



Long-form recordings

Photo by Marisa Casillas

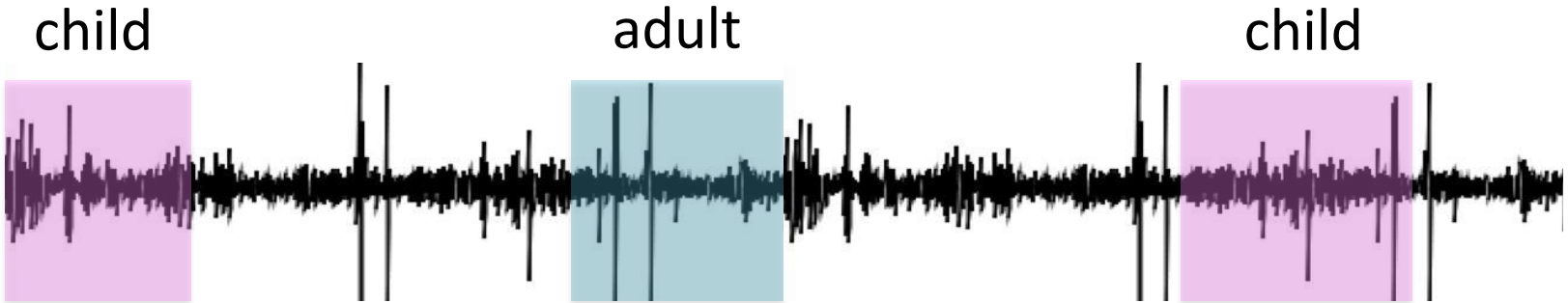
Photo by Heidi Colleran

Bergelson (2017); Semenzin et al. (2021); Cristia (2021); Cychosz (2018); Cassar et al. (2021); Casillas et al. (2017); Scaff et al. (2018); Cristia & Colleran (2018); Cristia & Casillas (2019)

1. Identify child vocalizations automatically

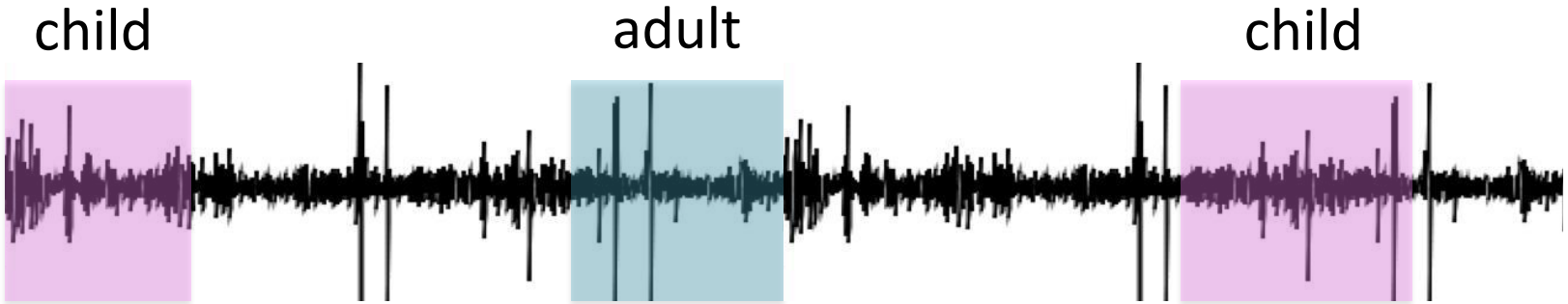
Lavechin et al. (2020)

1. Identify child vocalizations automatically



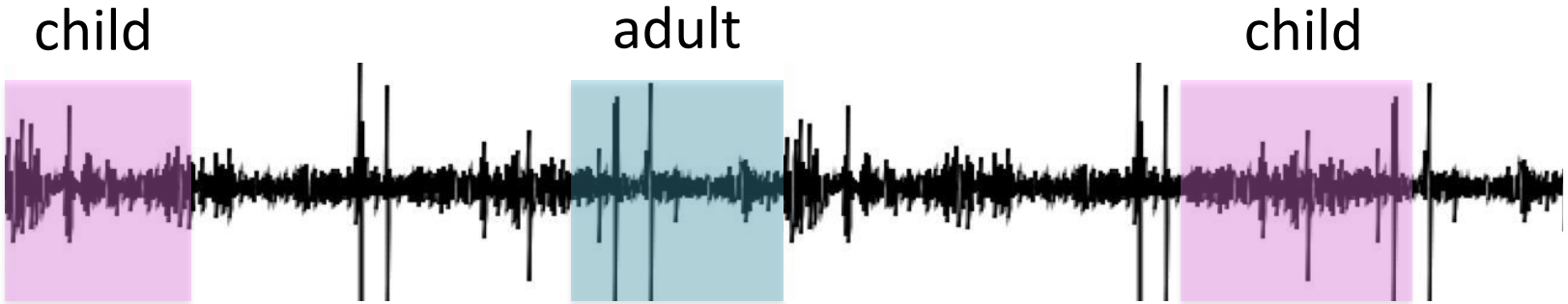
Lavechin et al. (2020)

1. Identify child vocalizations automatically



Lavechin et al. (2020)

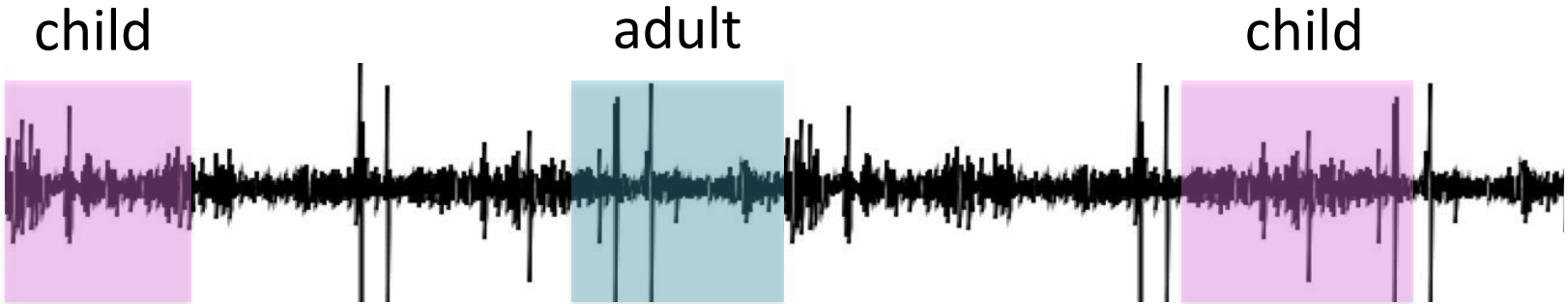
1. Identify child vocalizations automatically



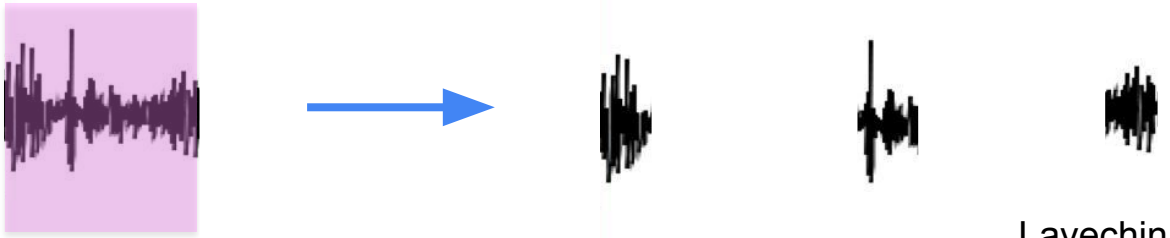
2. Extract and split them into short clips (~500ms)

Lavechin et al. (2020)

1. Identify child vocalizations automatically

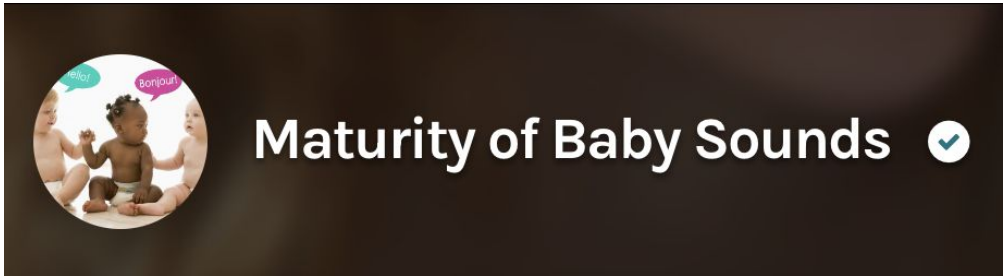


2. Extract and split them into short clips (~500ms)



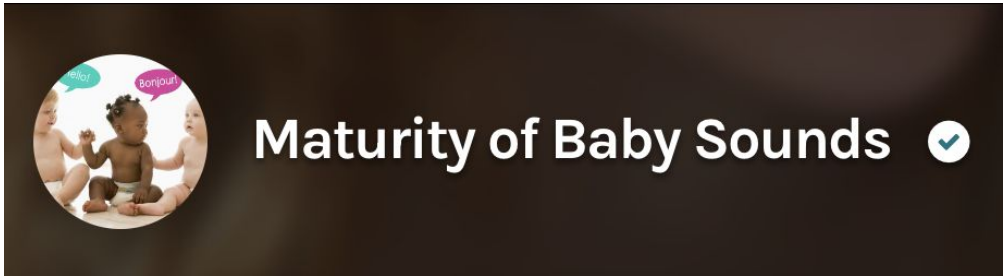
Lavechin et al. (2020)

Citizen science on Zooniverse



Bowyer (2015)

Citizen science on Zooniverse

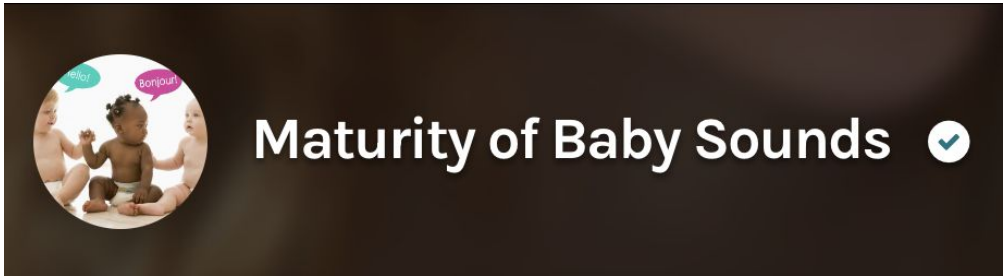


Stats:

- 10,000 individual annotators
- Average: 4,000 labels/day

Bowyer (2015)

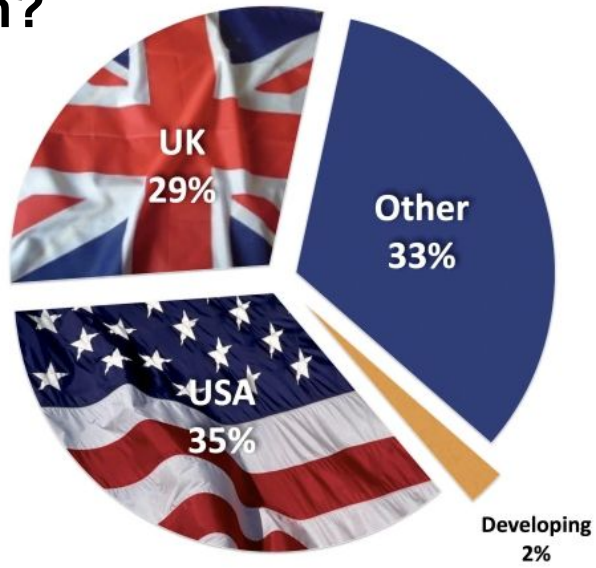
Citizen science on Zooniverse



Stats:

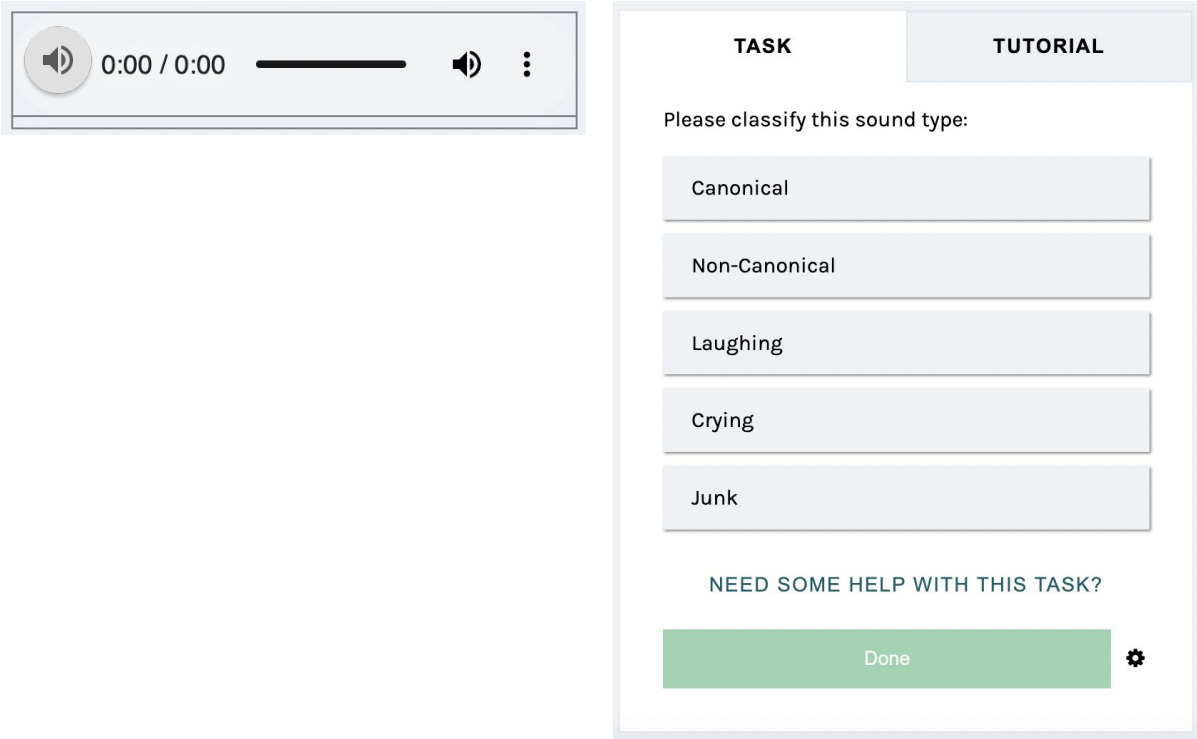
- 10,000 individual annotators
- Average: 4,000 labels/day

Where are citizen scientists from?

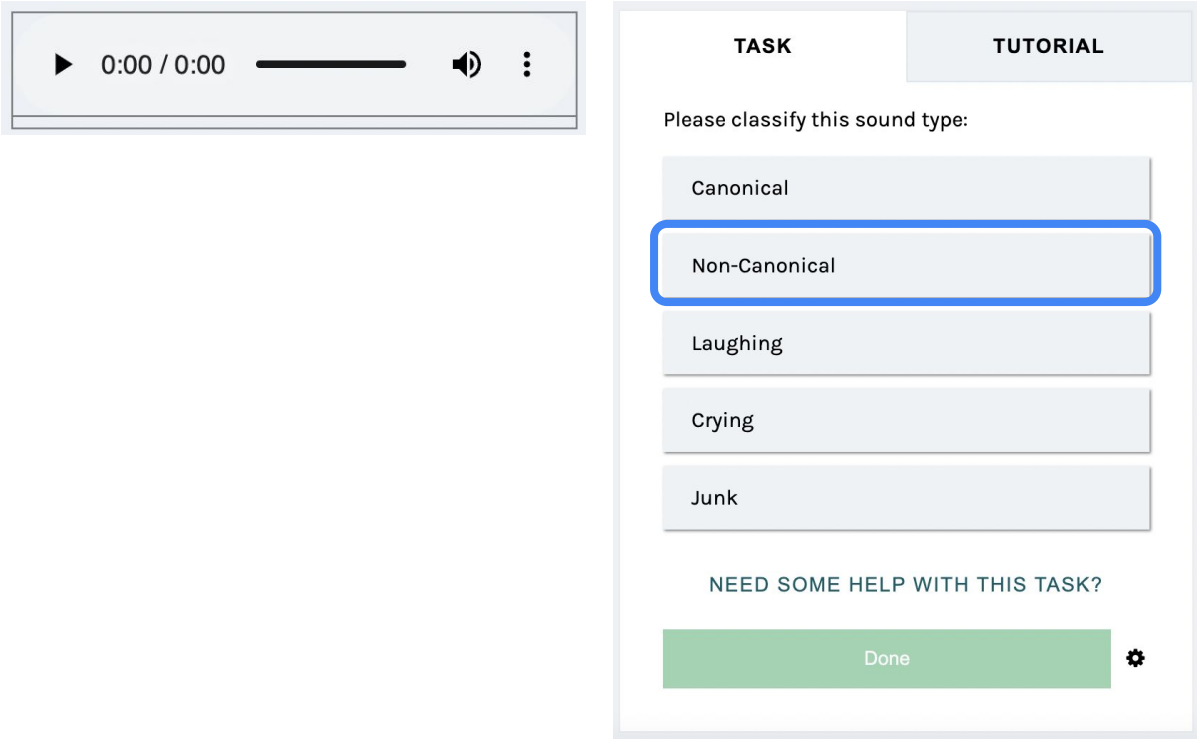


Bowyer (2015)

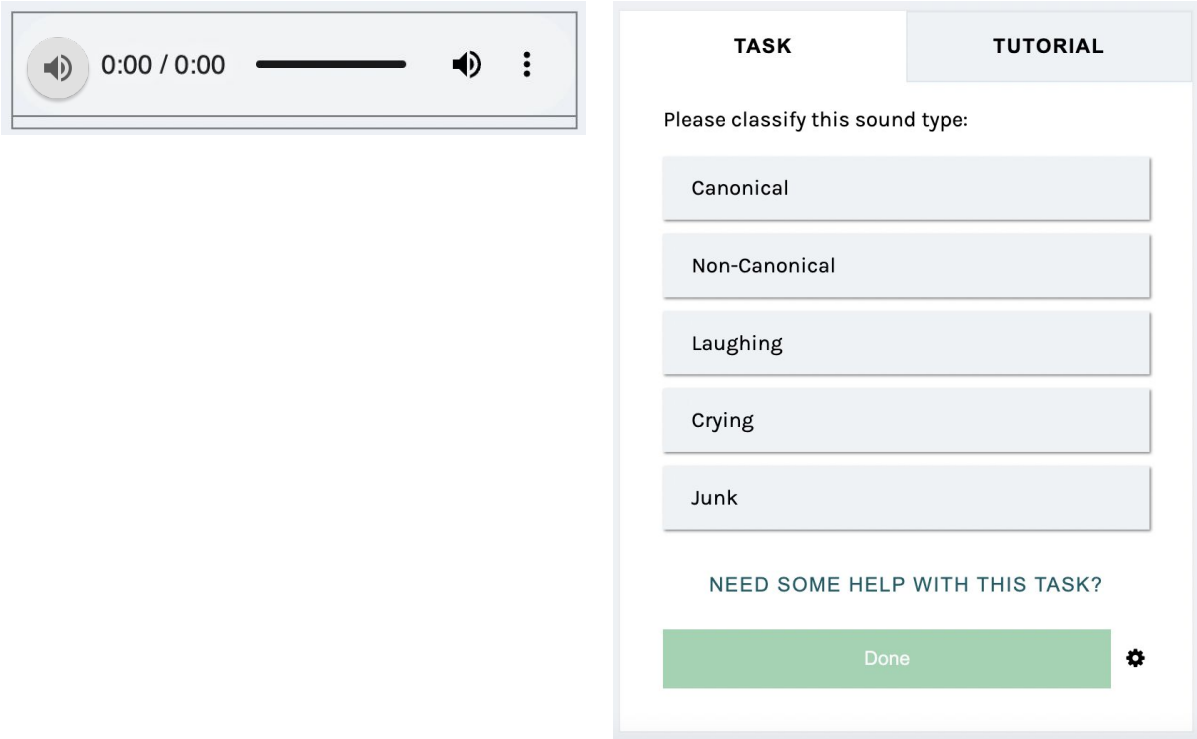
Citizen science annotation on Zooniverse



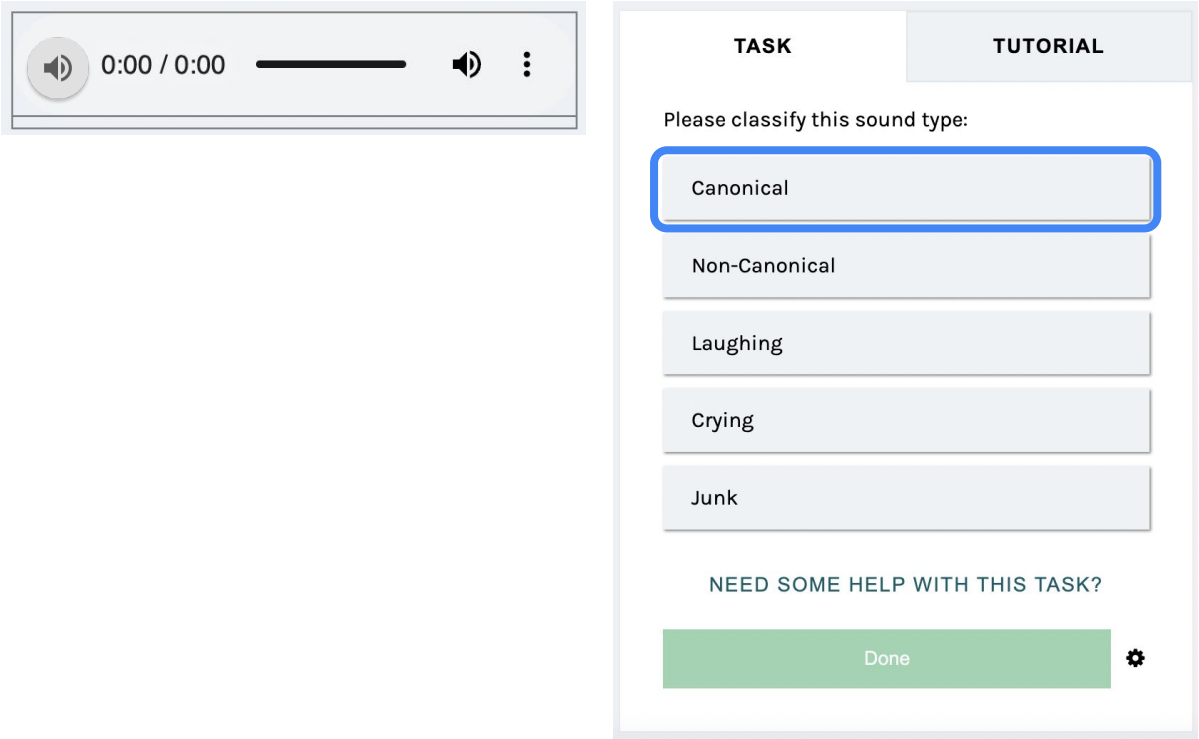
Citizen science annotation on Zooniverse



Citizen science annotation on Zooniverse



Citizen science annotation on Zooniverse

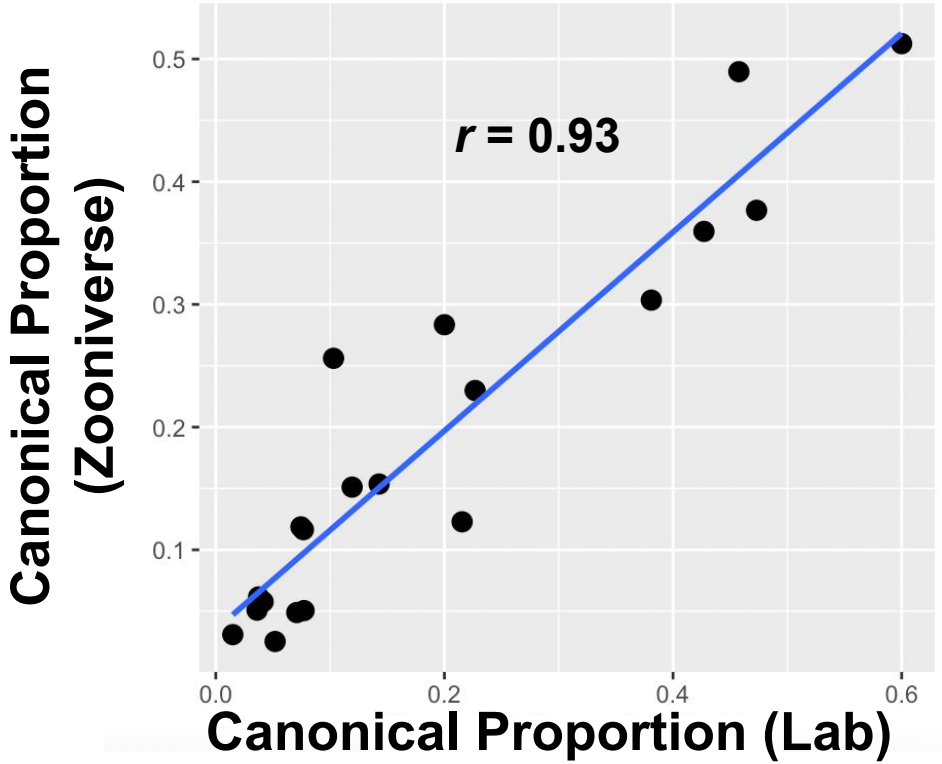


Calculate canonical proportion for each child

$$\text{Canonical proportion} = \frac{\text{Number of canonical vocalizations}}{\text{Number of canonical vocalizations + non-canonical vocalizations}}$$

Cychosz et al. (2021); Eilers & Oller (1994);
Lee et al. (2018); Oller & Eilers (1988); Oller (2000)

Citizen science method is reliable: canonical proportions obtained through this method correlate highly with canonical proportions obtained through lab annotations

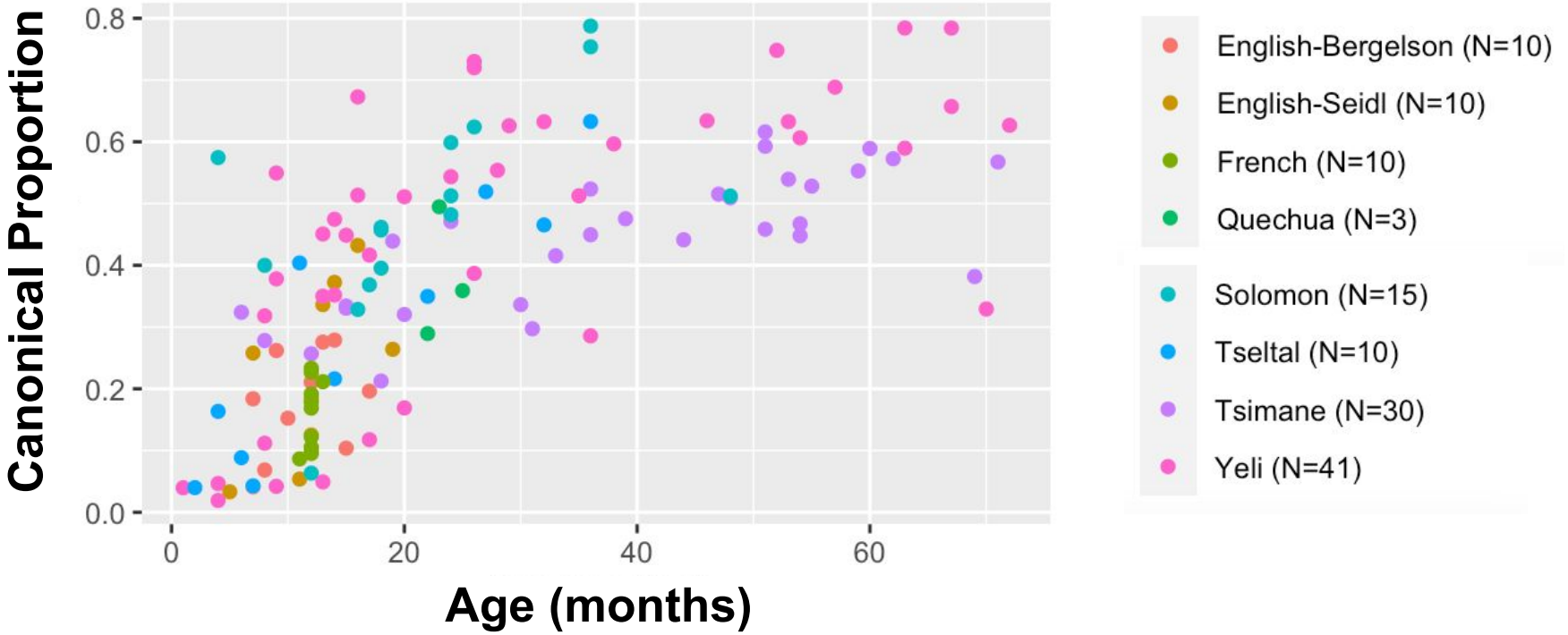


Semenzin et al. (2021)

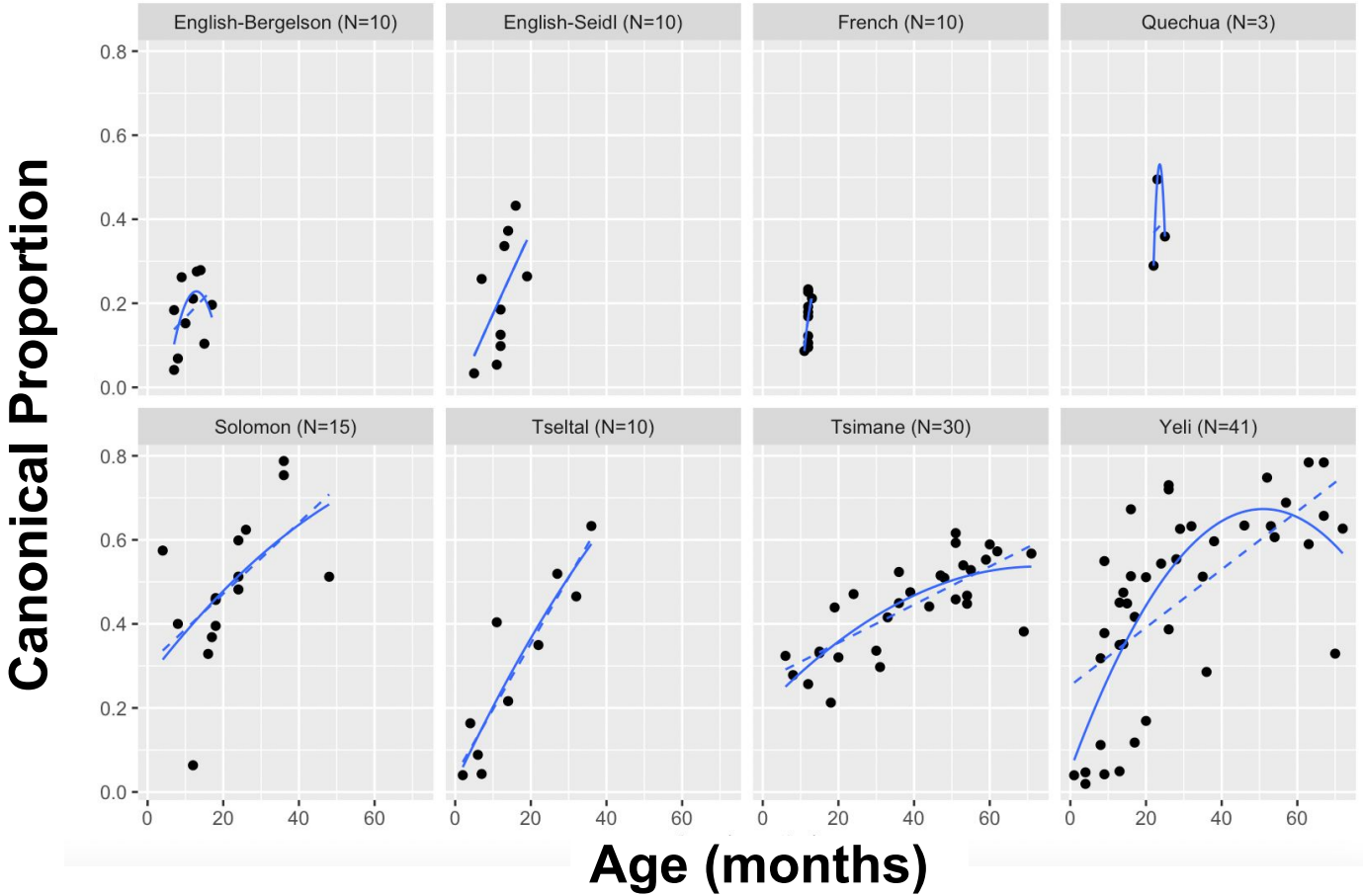
What does the trajectory of canonical proportion look like beyond the ages typically studied?

Does it vary by language?

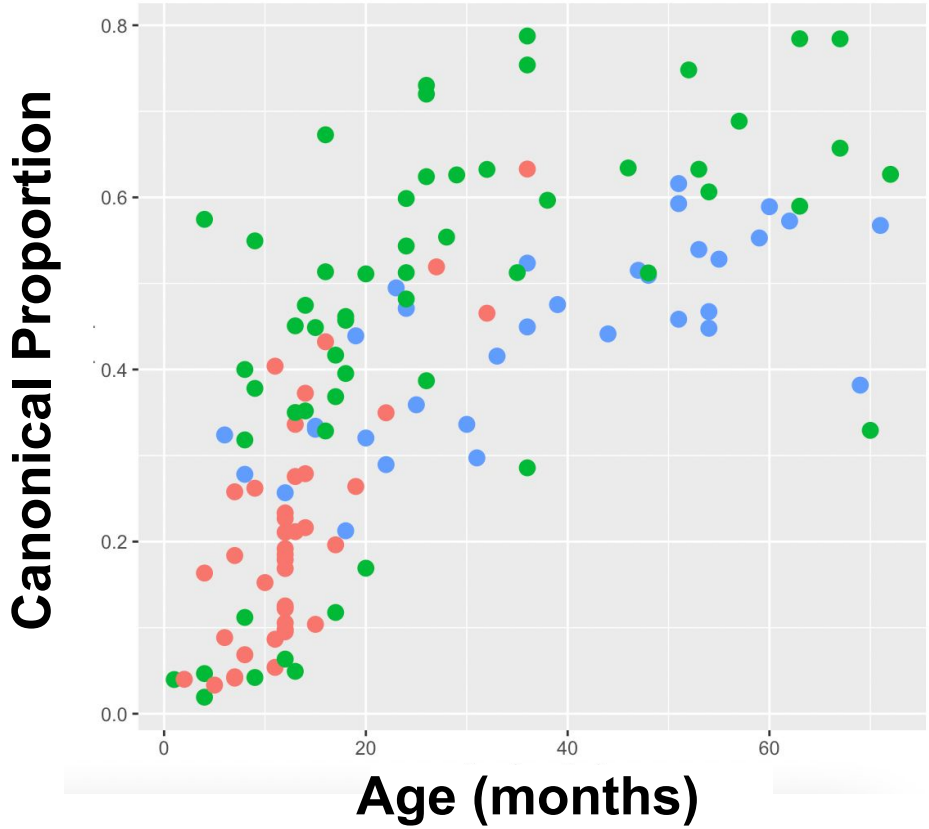
Canonical proportion continues to increase through 6 years of age.



There's some variability by corpus/environment



Potential relationship with the syllabic complexity of the language the child is learning



Syllable complexity

- Low (Yeli, Solomon) "see"
- Moderate (Tsimane, Quechua)
- High (English, French, Tseltal) "striped"

What does the trajectory of canonical proportion look like beyond the ages typically studied?

What does the trajectory of canonical proportion look like beyond the ages typically studied?

Answer: Canonical proportion continues to increase much later than previously thought

What does the trajectory of canonical proportion look like beyond the ages typically studied?

Answer: Canonical proportion continues to increase much later than previously thought

Does it vary by language?

What does the trajectory of canonical proportion look like beyond the ages typically studied?

Answer: Canonical proportion continues to increase much later than previously thought

Does it vary by language?

Answer: Preliminary evidence that it may, but more work needs to be done

What explains the continued increase in canonical proportion?

- Older children's vocalization combinations are thought to be driven by word choice, not by vocalization development
- One (phonetic) explanation: This may capture how children phonologically simplify their early word forms (e.g. omit codas)
- Another (lexical) explanation: Children choose to produce and/or have less phonologically complex vocabulary than adults

Storkel (2004); Hoover et al. (2010); Stokes (2010)

Next steps

- What is the adult/target canonical proportion across the populations we study? When do children reach that target?
- More cross-linguistic data so we can better test the relationship with language properties without confounds
  - Languages with moderate/high complexity beyond 20 months
- Registered report studying what factors influence vocalization development in ~500 children
- Publicly releasing the Zooniverse dataset for others to use

BabbleCor on OSF: Cychosz et al. (2021)

Conclusions

- Children’s phonological development continues much later than toddlerhood
- Children’s vocalization development may be more variable cross-linguistically and cross-culturally than previously thought (but more work needed…)
- Children’s canonical proportion is a candidate measure that can be used across age ranges…
- …and can be measured from long-form recordings without painstaking lab transcription
- This project shows the promise of using coarse semi-automatic methods + citizen science for studying cross-linguistic language acquisition in natural environments

Thank you!

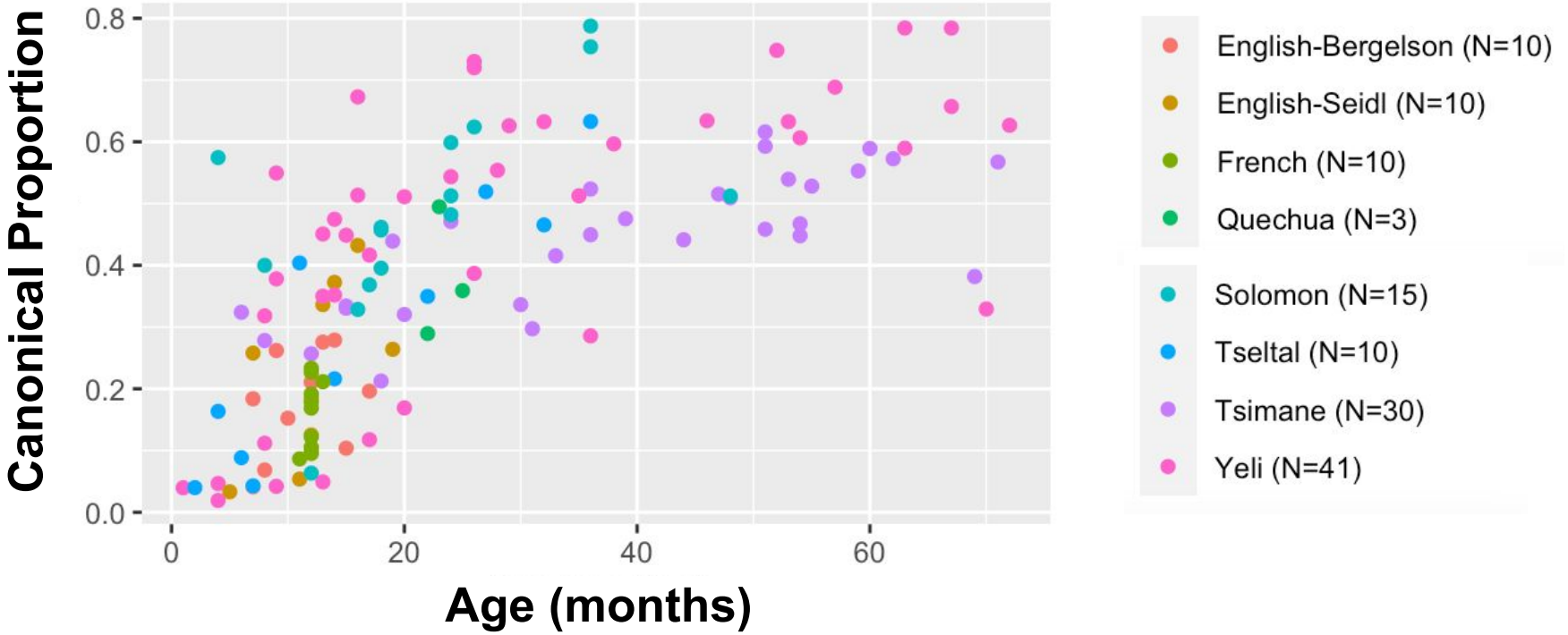
The participants and their families
Coauthors
Nadège Marin
Natalia Kuzminykh
Language Acquisition Across Cultures Team ->

Funding

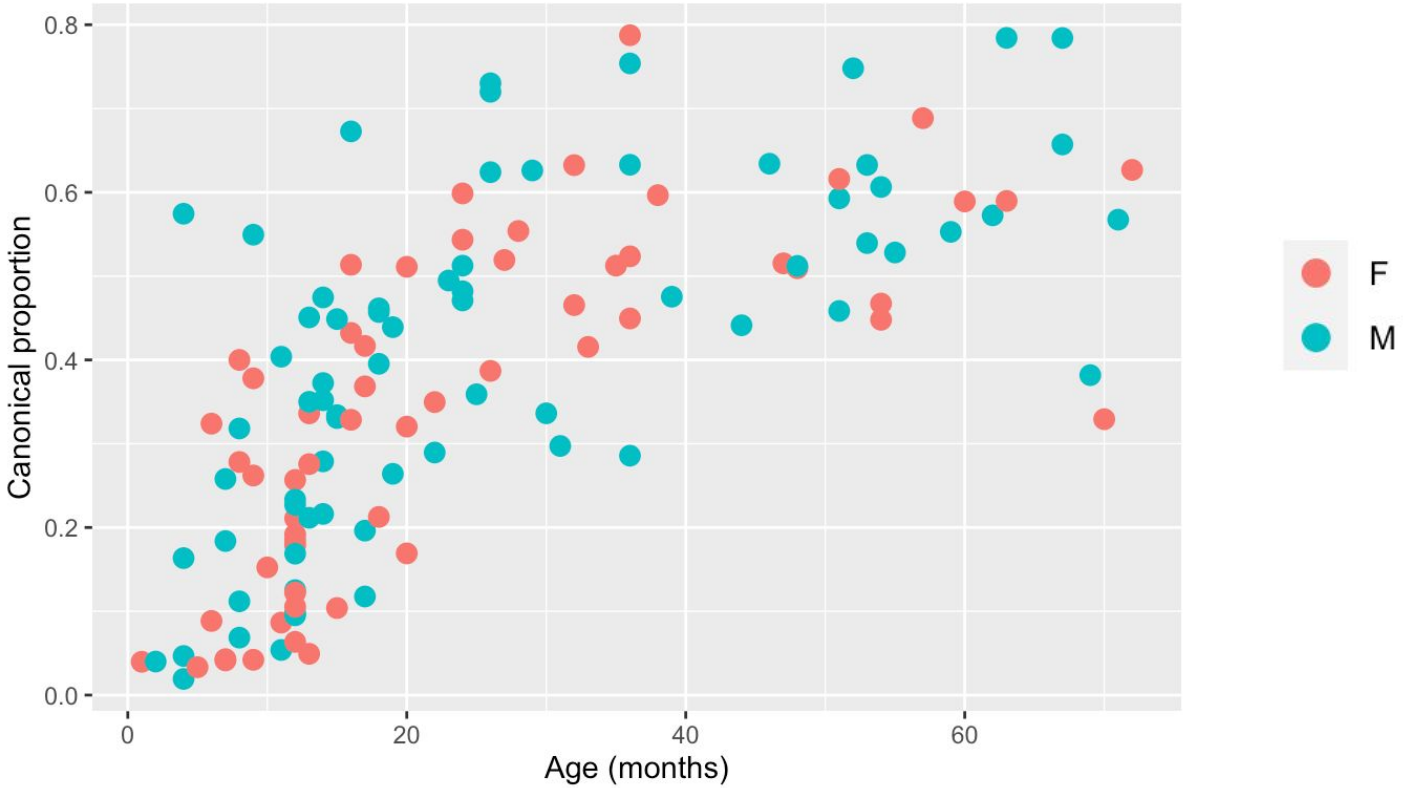
- Agence Nationale de la Recherche
- J. S. McDonnell Foundation Understanding Human Cognition Scholar Award
- European Research Council under the European Union's Horizon 2020 research and innovation programme

Zooniverse Project: Maturity of Baby Sounds
https://www.zooniverse.org/projects/laac-lscp/maturity-of-baby-sounds

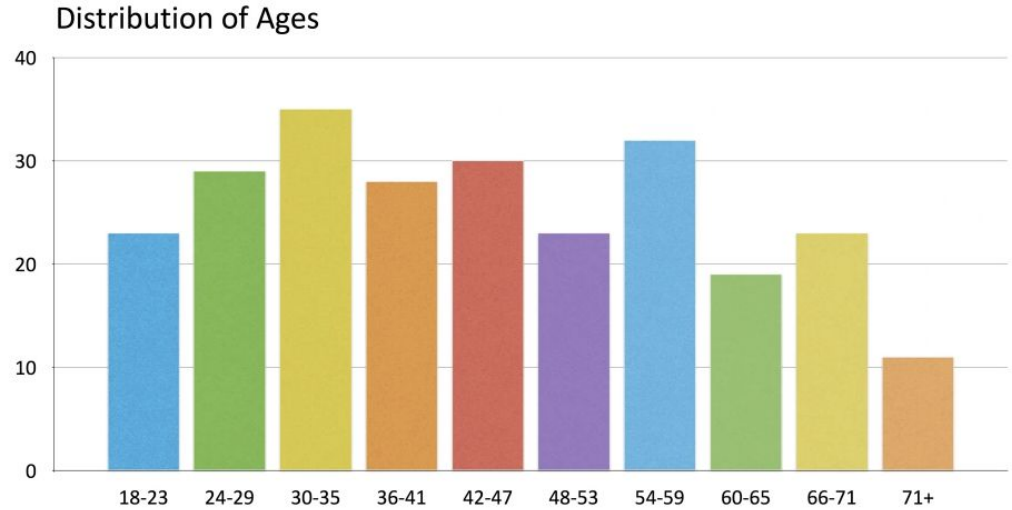
Canonical proportion continues to increase through 6 years of age.



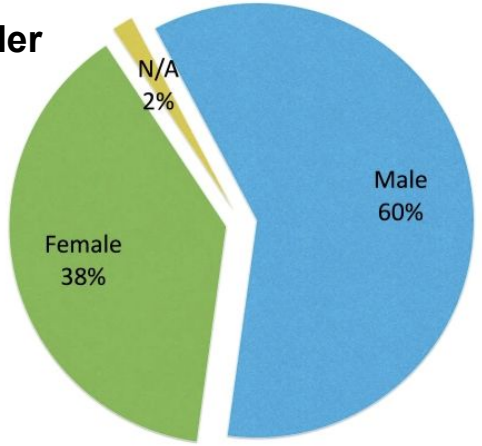
No apparent differences by child sex



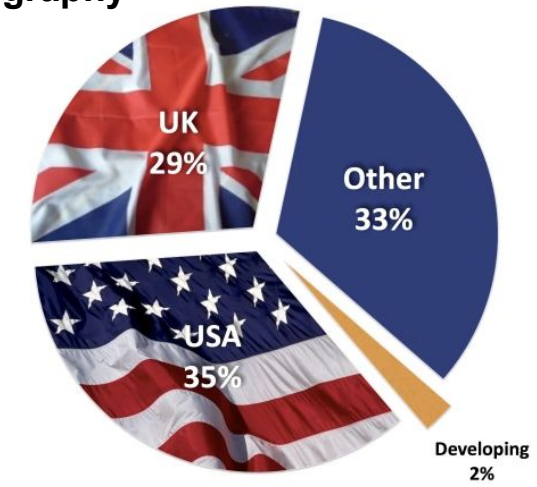
Citizen Scientists



Gender

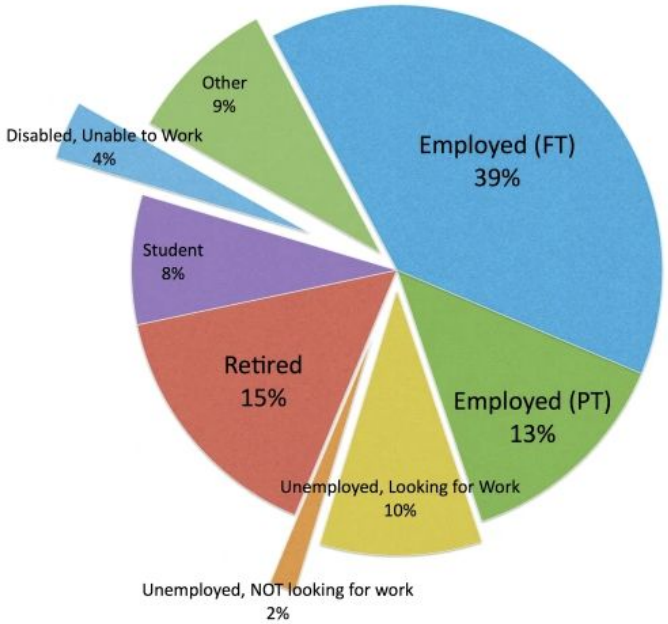


Geography

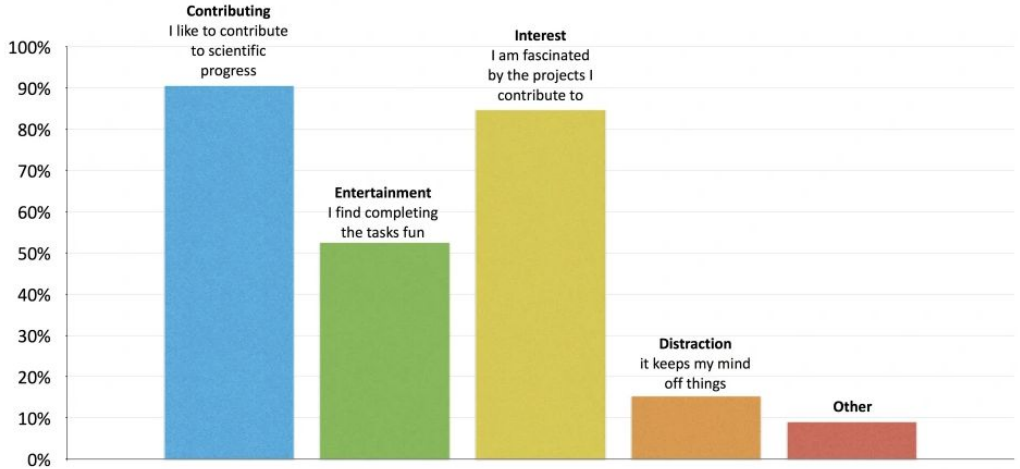


Citizen Scientists

Employment



View of their own participation



Statistical Analyses

- Compare 2 models

```
age.model <- glmer(cp ~ age_mo_round_z + child_gender +
(0+age_mo_round_z|corpus), data = df, family = binomial(), weights =
n_speechlike)

agesqz.model <- glmer(cp ~ age_mo_round_sq_z + age_mo_round_z +
child_gender + (0 + age_mo_round_z + age_mo_round_sq_z|corpus), data = df,
family = binomial(), weights = n_speechlike)
```

Clinical relevance