

# The SMILE Continuous DSGS Corpus: A Resource for Longitudinal Exploration of Continuous Swiss German Sign Language

## Motivation & Background

- > Sign languages remain **under-resourced**, especially for **longitudinal and learner-focused** data
- > Existing corpora focus on native signers or isolated signs, **limiting broader analysis**
- > **Swiss German Sign Language (DSGS)** lacks a **large-scale, annotated** corpus of **continuous** data suitable for linguistics and sign language processing (SLP)

## Data & Processing

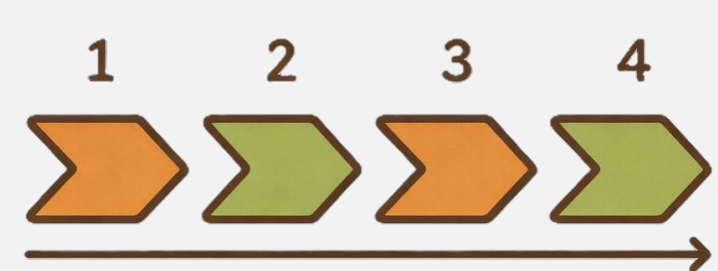
### Participants



10 25

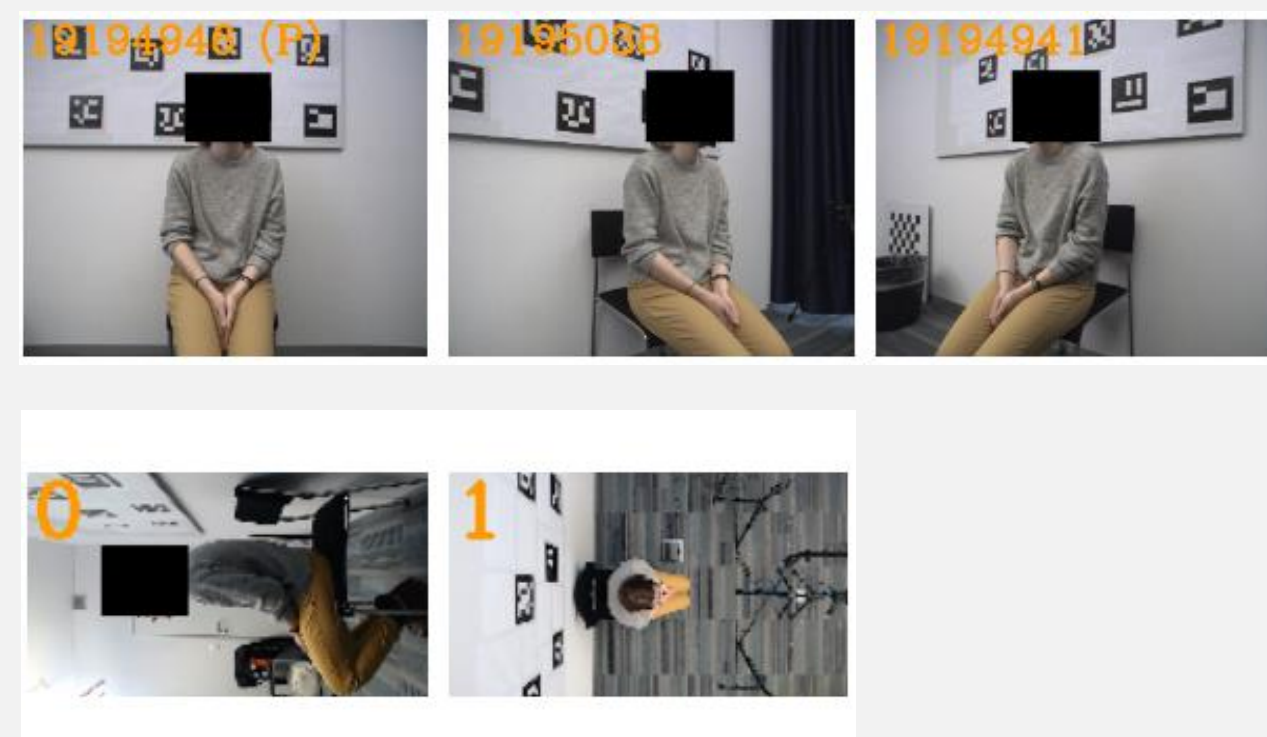
- > metadata
- > informed consent

### Duration



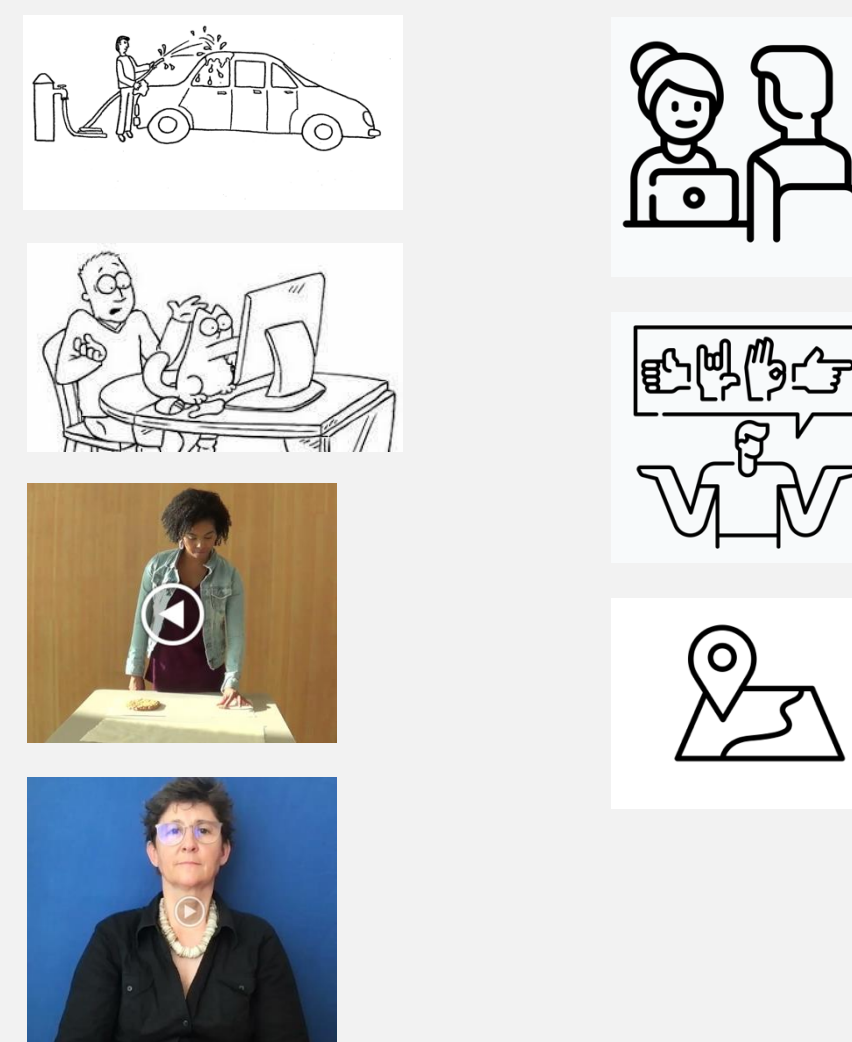
- > 18 months
- > 4 phases
- > longitudinal

### Multi-camera settings



- > controlled setting

### Activities

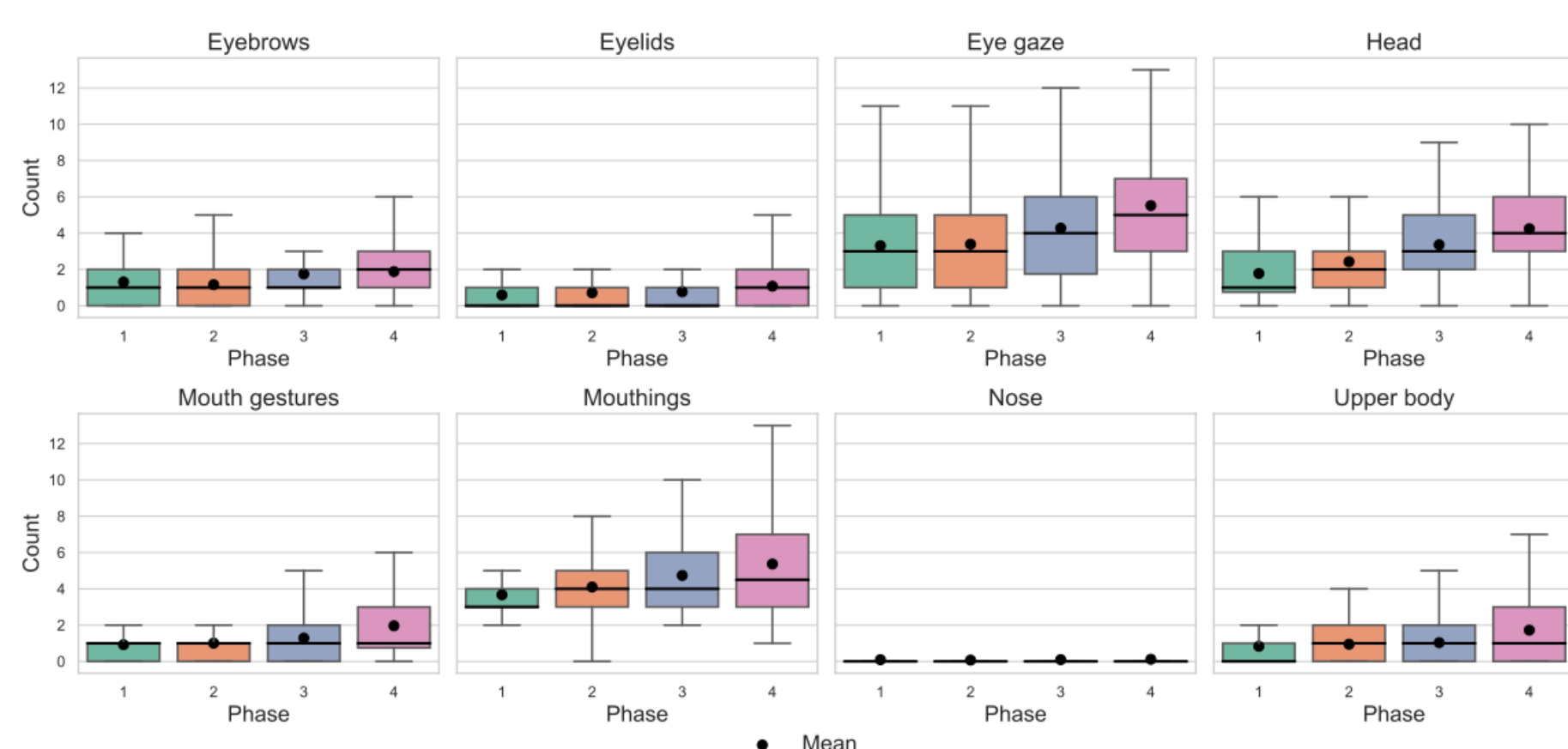


### Annotations

- > Manual components
- > Non-manuals
- > Error types in L2
- > Acceptability in L1 and L2 (Battisti et al. 2024)
- > Annotation agreement
- > Pose processing

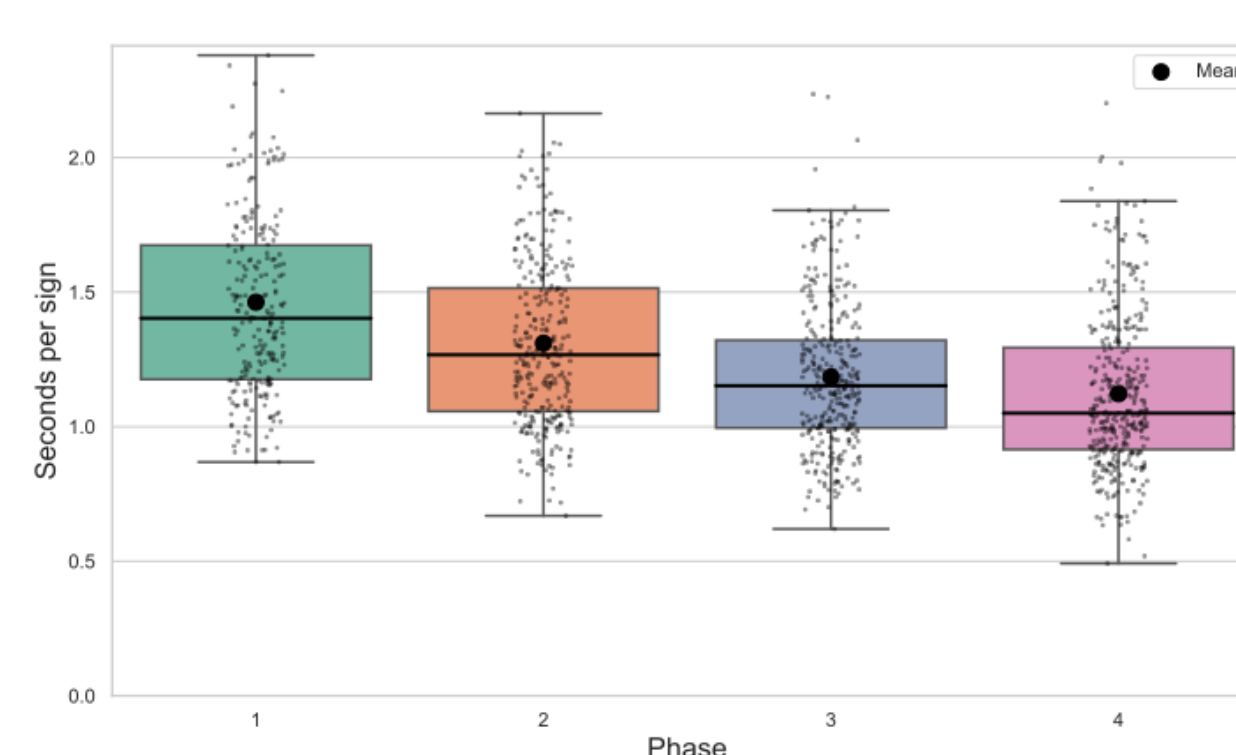
## First analyses: Interlanguage study

### Use of non manual components



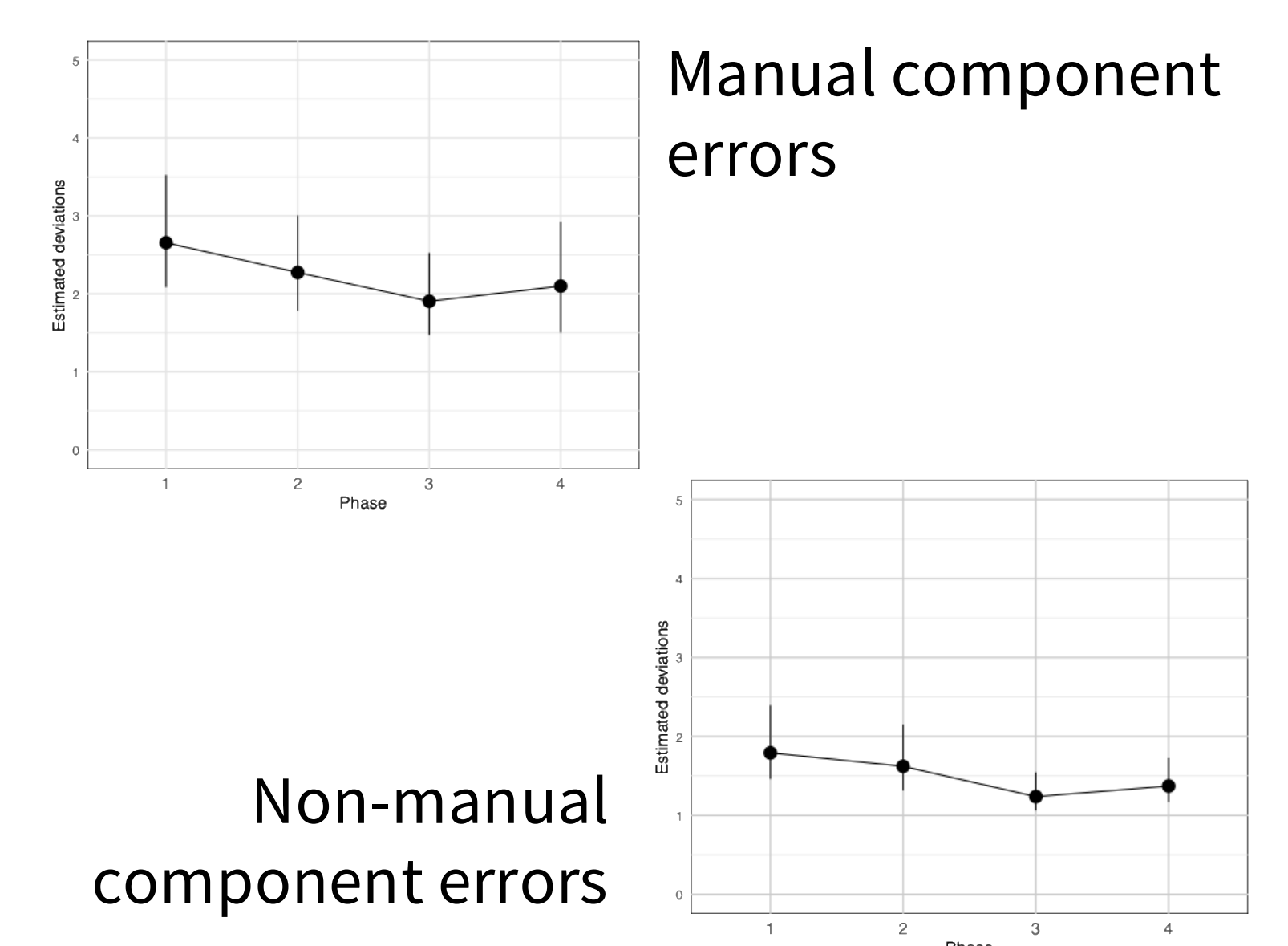
Distribution of individual non-manual components across phases  
> Mouth gestures and eye gaze increase

### Fluency investigation



Average signing time per gloss across phases  
> Median duration per gloss decreases

### Error modeling



Manual component errors

Non-manual component errors

## Conclusion & Outlook

- > **Open source:** publicly available via SwissUBbase for research purposes under CC BY-NC-SA 4.0
- > Valuable data for
  - > modeling **sign language acquisition and interlanguage**
  - > training **sign language recognition systems**
  - > developing **data-driven feedback tools** for learners
- > **Future work:**
  - > extend annotations to more tasks
  - > **support educational and accessibility applications**

