Seeing Who Is Signing and With Which Hand

UNIVERSITY OF BERGEN
Faculty of Humanities

12th Workshop on the Representation and Processing of Sign Languages: Language in Motion (LREC 20, 2026)

Carl Börstell

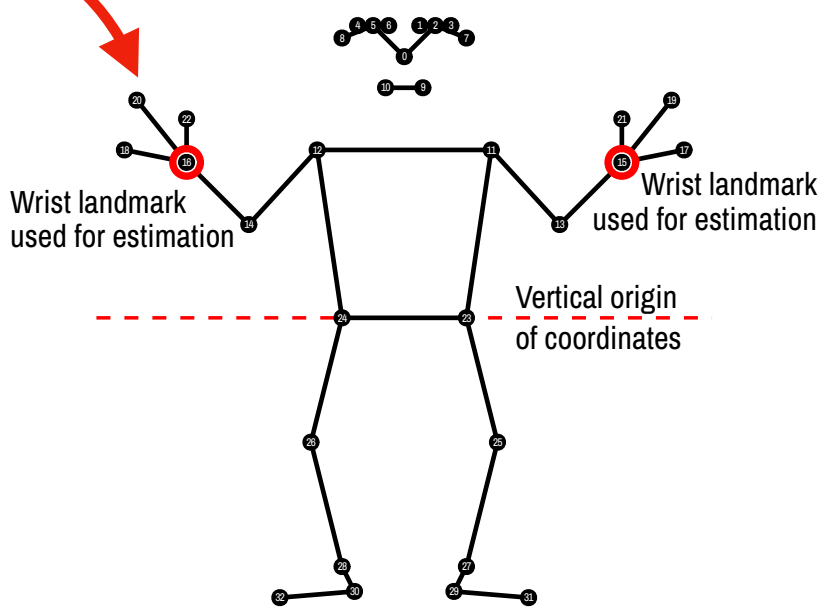
Research questions

Using a computer vision (MediaPipe[1]) analysis of Swedish Sign Language (STS) dyadic, conversational corpus video data, is it possible to accurately classify:

- who the **main signer** is at a given time?
- which **hand** is a signer's **dominant**?

Data & Method

40 dyadic, conversational texts were sampled from the STS Corpus: 80 MPEG video files[2] and their associated ELAN annotation files[3] – about 2 hrs 15 mins of text (≈ 10% of the entire STS Corpus – analyzed with MediaPipe's world landmarks 3D pose estimation: each 5-second segment was classified as having a **main signer** and each signer a **dominant hand**, based on the ***distance*** & ***height*** of hands



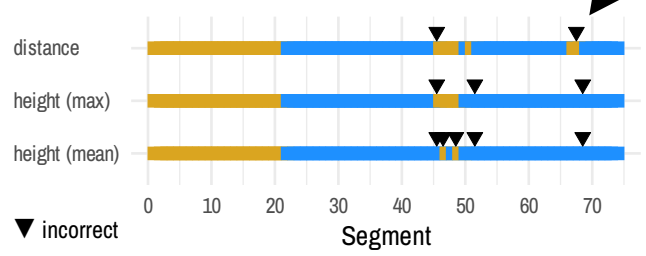
MediaPipe Pose Estimation Landmarks

Results: Who is signing?

Both **distance** and **height** (max & mean) approaches perform well when classifying who the main signer is for each 5-second segment

The optimal approach is a ***combined*** one: majority rule across all methods is applied for each segment, resulting in 96.9% accuracy compared to sign frequency as ground truth

Looking at examples of incorrectly classified segments, they tend to either occur at **turn-taking boundaries**,[4] where main signer naturally changes, *or* when an addressee is moving their hands more or higher for **grooming** or **backchannels/feedback**[5,6]

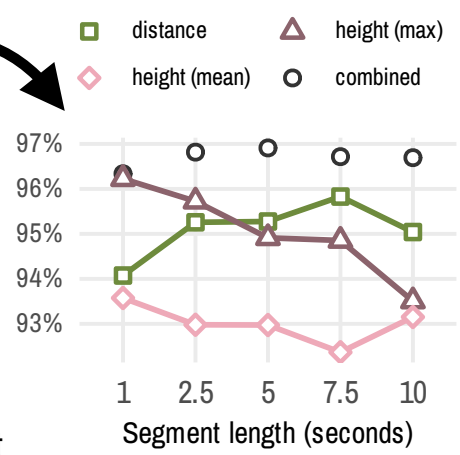


Classification of main signer by segment in one STS Corpus text (SSLC01_020)

Accuracy of main signer classification by method

	Classification	
Method	correct	incorrect
distance	1,573 (95.3%)	78 (4.7%)
height (max)	1,567 (94.9%)	84 (5.1%)
height (mean)	1,535 (93.0%)	116 (7.0%)
combined	1,600 (96.9%)	51 (3.1%)

Varying segment duration shows that approaches favor slightly different lengths, but that the optimal combined approach is the 5-second segment

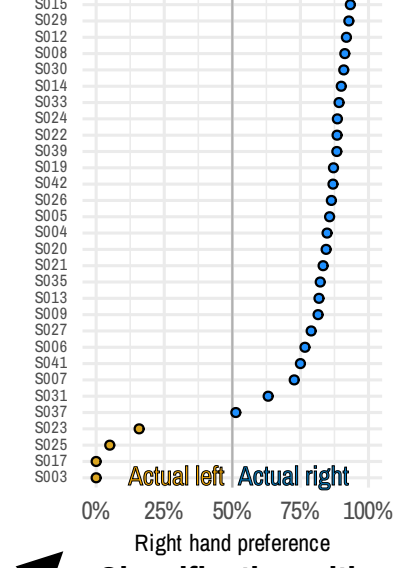


Accuracy of methods with various segment lengths

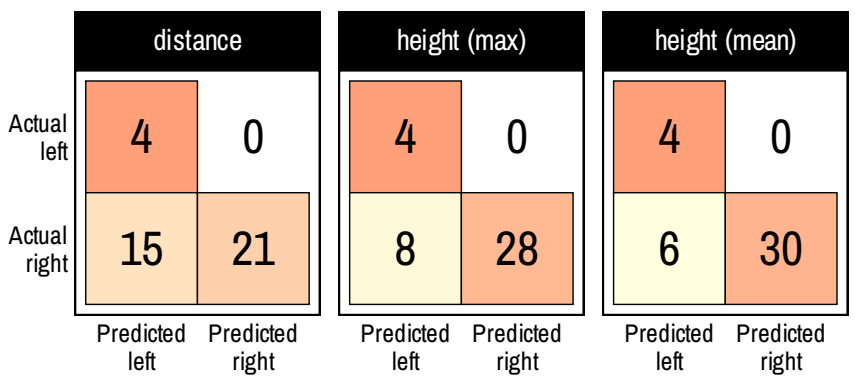
Results: Hand

Classifying each signer's **handedness** is a more challenging task: while all left-handed signers are correctly classified, some right-handed signers are incorrectly labeled left-handed by all methods (distance/height)

But, a two-step approach in which 1) the main signer is identified per segment, and 2) only main signer-relevant segments are used for handedness classification is 100% accurate



Classification with two-step approach



Predicted vs. actual handedness of signers by method

Handedness classification by method as balanced and weighted F-scores

	F-score	
Method	balanced	weighted
distance	.54	.70
height (max)	.69	.84
height (mean)	.74	.88
combined	.71	.86

References

[1] Camillo Lugaresi et al. 2019. MediaPipe: A framework for perceiving and processing reality. In Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019. • [2] Johanna Mesch, Lars Wallin, Anna-Lena Nilsson, and Brita Bergman.2012. Dataset. Swedish Sign Language Corpus project 2009–2011 (v1). Sign Language Section, Dept. of Linguistics, Stockholm University. • [3] Johanna Mesch, Maya Rohdell, and Lars Wallin. 2014. Annotated files for the Swedish Sign Language Corpus (v2). Sign Language Section, Dept. of Linguistics, Stockholm University. • [4] Connie de Vos, Francisco Torreira, and Stephen C. Levinson. 2015. Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. Frontiers in Psychology, 6. • [5] Carl Börstell. 2024. Finding continuers in Swedish Sign Language. Linguistics Vanguard, 10(1):537–548. • [6] Johanna Mesch. 2016. Manual backchannel responses in signers' conversations in Swedish Sign Language. Language & Communication, 50:22–41.