A Pose-Based Pipeline for Annotation of Headshakes in Sign Language Corpora

Gustaf Gren Nikolaus Riemer Kankkonen
Stockholm University

Contributions

- **Pose-based pipeline for headshake candidate detection** from head yaw trajectories extracted with MediaPipe Holistic.
- **Useful for annotation support:** best LSTM model recovers **87%** of labeled headshake instances while requiring review of only **13%** of frames.
- **Neural models outperform a simple zero-crossing baseline** on DGS headshake-as-negation annotations.
- **Zero-adaptation transfer to Swedish Sign Language is limited,** showing the need for threshold adjustment or further domain adaptation.

Motivation

- Headshakes are important non-manual markers in sign languages, including grammatical negation.
- Large corpora exist, but annotations for headshakes are scarce.
- Manual annotation is slow and expensive.
- **Aim:** surface likely headshake segments for human review

Data

- **Training / test data:** subset of the **DGS Corpus**[2], using annotations from [1]. DGS labels mark **grammatical headshakes in negative clauses**, not all headshakes.
- **Transfer pilot:** ∼50 minutes of Swedish Sign Language Corpus data [3], manually annotated for **all** headshakes.

Models

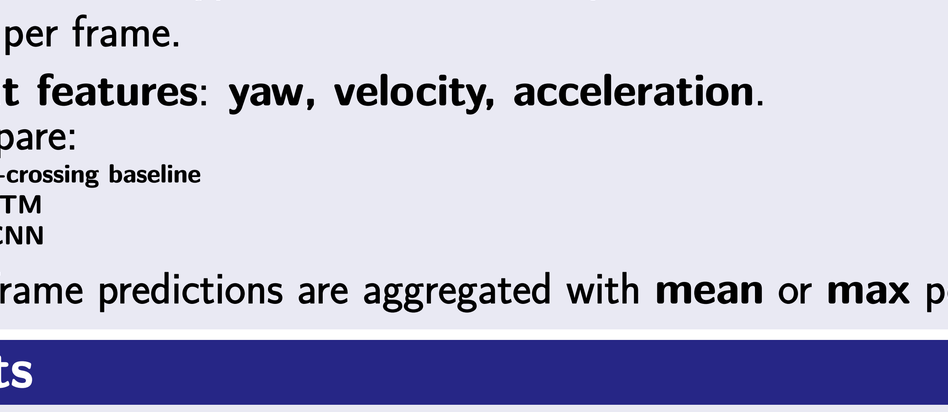
- Extract facial keypoints with **MediaPipe Holistic** and derive head **yaw** per frame.
- **Input features**: **yaw, velocity, acceleration**.
- Compare:
  - Zero-crossing baseline
  - BiLSTM
  - 1D CNN
- Per-frame predictions are aggregated with **mean** or **max** pooling.

Results

	P	R	F_1	F_2
Baseline	0.02	0.59	0.04	0.08
LSTM_{max}	0.20	0.61	**0.31**	0.43
CNN_{max}	0.16	0.53	0.24	0.36
$\text{LSTM}_{\text{mean}}$	**0.21**	**0.63**	**0.31**	**0.45**
CNN_{mean}	0.19	0.50	0.27	0.37

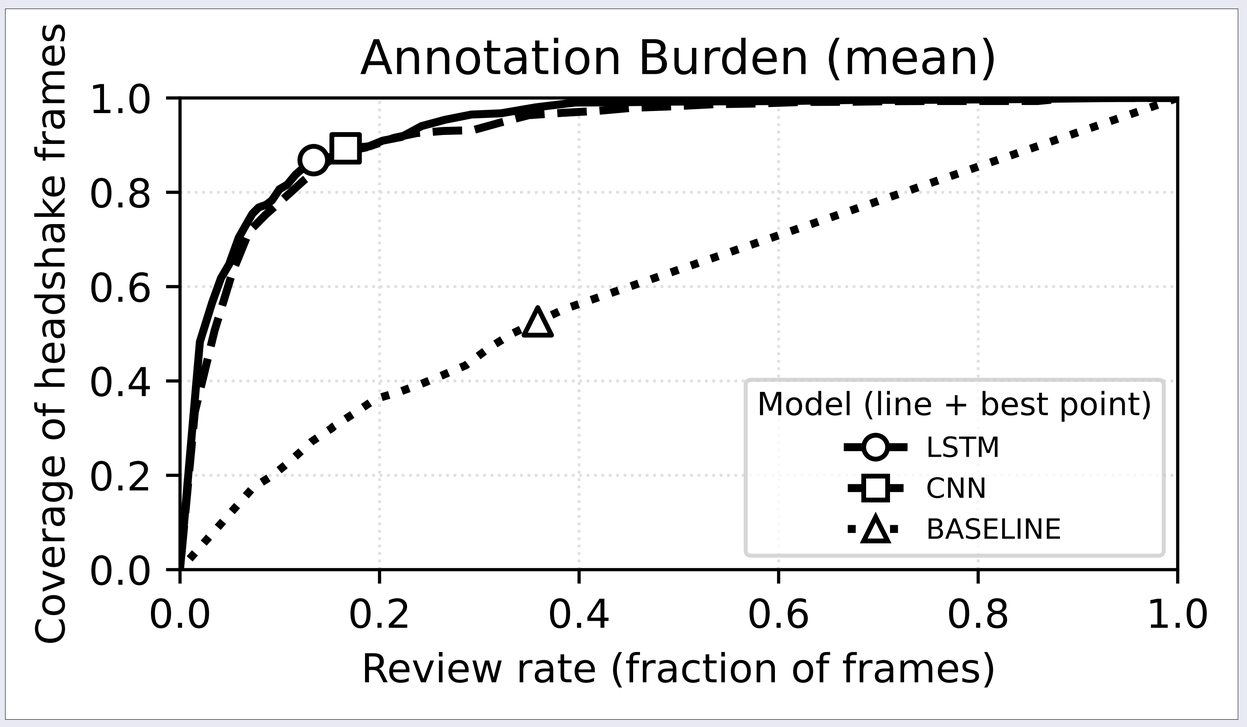
Error Analysis

Many false positives are visually plausible headshakes excluded by the DGS annotation scope:



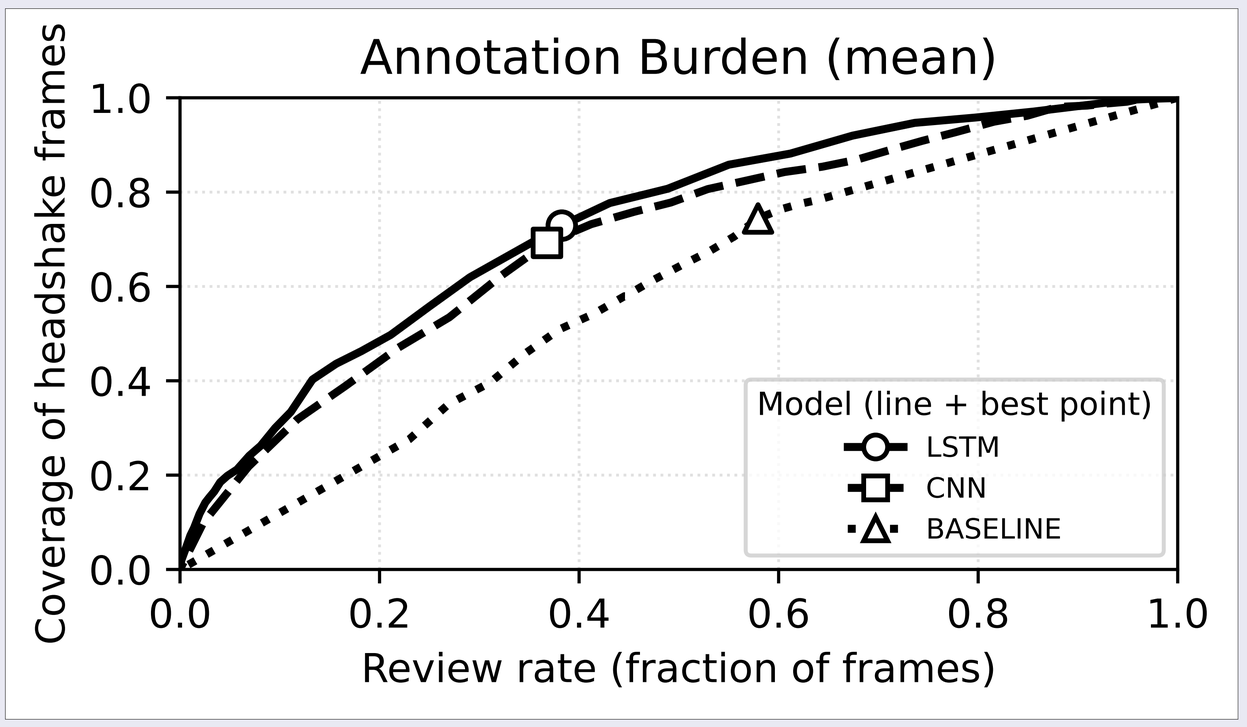
Annotation Burden Reduction

Best performing LSTM only requires 13% review rate to achieve 87% recall, though very similar to CNN:



Zero-shot Transfer to Swedish Sign Language

Performance drops to $F_2 = 0.21$ for both the CNN and LSTM, though annotation burden is still below baseline:



Takeaway

- Pose-based detection is useful for **reducing annotation workload**.
- The main bottleneck is not only modeling, but also **annotation scope**.
- **Human-in-the-loop** approaches, active learning, and manual tuning are promising for future research.

References

[1] Vadim Kimmelman et al. "Kinematics of negative headshake in seven sign languages". Manuscript ready. in preparation.

[2] Reiner Konrad et al. *MEINE DGS – annotiert. Öffentliches Korpus der Deutschen Gebärdensprache, 3. Release MY DGS – annotated. Public Corpus of German Sign Language, 3rd release*. languageresource. Version 3.0. 2020. DOI: 10.25592/dgs.corpus-3.0. URL: https://doi.org/10.25592/dgs.corpus-3.0.

[3] Zrajm Öqvist et al. "STS-korpus: A Sign Language Web Corpus Tool for Teaching and Public Use". English. In: *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*. Ed. by Eleni Efthimiou et al. Marseille, France: European Language Resources Association (ELRA), May 2020, pp. 177–180. ISBN: 979-10-95546-54-2. URL: https://aclanthology.org/2020.signlang-1.29.