TRAIL Summer Workshop' 25 Project Proposal

Full Name of Team Leader	Liesbet De Vos
Project Title	Recognising HamNoSys handshapes from video and electromyographic signals.
Profile of the Team Leader(s)	Within my thesis, I aim to develop a computational framework for representing and processing signed languages. My knowledge about sign language and sign language representation (including HamNoSys) are particularly valuable for this project.
Abstract	Sign language processing (SLP) focuses on developing artificial intelligence models that recognize, understand, and produce sign language utterances. One of the key challenges within this field is the collection and manual annotation of language data, a time-consuming and expensive endeavor. Previous work has primarily focused on alleviating this burden by studying how signs can be annotated automatically using lexical id-gloss labels. However, such labels are language-dependent and do not provide any information on the sublexical structure of the sign. In contrast, phonetic representation systems such as the Hamburg Notation System (HamNoSys) can be applied to any sign language and utilize a definite set of iconic symbols to describe the shape, orientation, location, and movements of the hands throughout a sign. Although some previous work has focussed on recognising handshape, orientation, location, or movement of signs, this work is limited and often focuses only on handshape in American Sign Language. Currently, no method exists that automatically annotates signs with HamNoSys symbols. **This project takes an initial step at addressing this gap and proposes to develop a trustworthy methodology for extracting HamNoSys handshape symbols from video and electromyographic (EMG) signals. We aim to explore the use of both video and EMG as input signals, assessing their capacities and limitations in sign language handshape recognition. For the recognition component, we will explore the use of neuro-symbolic AI to leverage the structured nature of HamNoSys and integrate it with neural representations. This approach enhances the explainability and transparency of the model, enabling interpretable predictions grounded in linguistic theory.**
Project Objectives	<u>Data acquisition and preprocessing</u> 1. Record isolated sign language handshapes using video + EMG data from multiple participants. 2. Collect videos from existing datasets which are annotated with HamNoSys 3. Preprocess and segment both modalities (e.g., hand tracking for video).

	<u>Develop baseline video handshape recognition model</u> 1. Pretrain the baseline model on videos collected from existing sign language datasets. The goal of the model is to recognise HamNoSys handshapes. 2. Finetune the baseline model on data collected during the workshop 3. Evaluate accuracy of the baseline model on the test set. <u>Develop EMG handshape recognition model</u> 1. Train an EMG model to recognise HamNoSys handshapes 2. Evaluate accuracy of the EMG model on the test set, comparing to performance of the baseline model <u>Research outputs</u> 1. Publish results of the project as a paper, discussing the capabilities and limitations of both input methods 2. Publish developed dataset and code as a brick on the TRAIL factory
Project Dataset	Collection of sign language videos from existing datasets annotated with HamNoSys, including: - The GeoQuery dataset (https://gitlab.unamur.be/beehaif/GeoQuery-LSFB/-/tree/master/Images) - The corpus-based dictionary of polish sign language (https://www.slownikpjm.uw.edu.pl/en) - the dictasign corpus (https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/) Record isolated sign language handshapes using video + EMG data from multiple participants.
Background Information	Sign language processing (SLP) focuses on developing artificial intelligence models that recognize, understand, and produce sign language utterances (Moryossef & Goldberg 2021). One of the key challenges faced within SLP is the representation and collection of language data. In contrast to most spoken languages, no widely accepted writing system exists for signed languages, resulting in low data availability. Available data is stored in video format and annotated using id-glosses (lexical labels) or phonetic/phonological writing systems. Such annotations are crucial both within the field of SLP to develop sign language technologies and within the field of sign language linguistics to perform linguistic analyses. Creating these annotations is time-consuming and expensive as it is performed manually by highly skilled annotators (Ong and Ranganath 2005). As a result, the automation of this annotation task is crucial. Past work has primarily focused on extracting glosses from sign language videos (see Moryossef & Goldberg 2021 for an overview) and less on extracting phonetic/phonological information. The extraction of phonetic details is valuable for several reasons. For

instance, phonetic representation systems are language-independent, while gloss-based annotations differ between sign languages (Arkushin et al. 2023). **Therefore, automatically extracted phonetic representations have a high potential for developing multilingual models that combine data from multiple signed languages.** In addition, phonetic representations contain information about the sublexical structure of signs (i.e., the shape of the hand, orientation, movement, and location involved in the production of the sign), which is not the case for gloss-based representations (Naert et al. 2020). Such information serves linguistic purposes and can be used to drive avatar systems. Finally, previous work has suggested that learning to recognise phonetic representation in combination with glosses might increase performance in sign recognition (Kezar 2023).

This project aims to develop a trustworthy methodology for extracting HamNoSys handshape symbols from video and EMG signals. HamNoSys uses a set of iconic symbols placed in linear order to describe the configurations of signs from any signed language (Hanke 2004). It represents the initial posture of the sign (composed of non-manual features, a handshape, an orientation, and a location) and optional actions or movements that change this initial configuration (Hanke 2004). Within this project, we will primarily focus on extracting information about the handshape, which is part of the initial configuration. Representations of handshapes in HamNoSys combine basic forms with possible diacritics that specify the thumb position or finger bending (Hanke 2004). In total, 33 symbols describe handshapes in HamNoSys, which can be combined using grammatical rules. One key advantage of HamNoSys is its digital integration. It is available as a Unicode font that uses the private usage space of Unicode and can be represented using SiGML, an XML variant of HamNoSys (Elliot et al. 2004). Representations in SiGML can then automatically be rendered into avatar animations using the JASigning software[1]. In addition, Arkushin et al. (2023) propose a transformer model which transforms HamNoSys representations into pose-format.

To the best of our knowledge, the only previous work that focuses on translating sign language videos into HamNoSys format was conducted by Skobov and Lepage (2020), who describe an encoder-decoder method where the encoder produces random (non-existing) signs in HamNoSys format using a tree representation of the HamNoSys grammar. The resulting signs are rendered as avatar-animated signs in video format. The decoder then uses poses extracted from this video to identify the information about the handshape. This approach yielded 22% accuracy on a validation set (Skobov and Lepage 2020). **These results show that the problem of extracting handshapes remains unsolved, especially in the**

[1] http://vh.cmp.uea.ac.uk/index.php/JASigning

	context of real signing videos (rather than avatar-animated ones). While the use of EMG for handshape and gesture recognition has been explored extensively, including within the domain of sign language (Ben Haj Amor et al. 2023), there is no work focussing on identifying HamNoSys handshapes from EMG input signals. Here, we explore whether an EMG model for handshape recognition can improve a baseline handshape recognition model trained on video only. Video input for the baseline model will be pre-processed to extract estimated poses using for instance Mediapipe[2] or OpenPose[3] or hand-shape specific models (see this github-page for an overview). **To extract HamNoSys information from the pose-representations, we would like to explore the use of neuro-symbolic AI to combine the power of neural computer vision models with symbolic AI's capability to ensure generated output is in concordance with the rules of HamNoSys grammar.** Different possibilities exist in terms of neural architectures for sign language recognition, including Sign Language Graph Convolutional Networks (SL-GCN), LSTMs, Transformers, and spatio-temporal graph convolutional neural networks (see Selvaraj 2022 for more information).
Bibliographic References	• Arkushin, R. S., Moryossef, A., & Fried, O. (2023). Ham2pose: Animating sign language notation into pose sequences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21046-21056. • Ben Haj Amor, A., El Ghoul, O., & Jemni, M. (2023). Sign Language Recognition Using the Electromyographic Signal: A Systematic Literature Review. Sensors, 23(19), 8343. https://doi.org/10.3390/s23198343 • Colelough, B. C., & Regli, W. (2025). Neuro-Symbolic AI in 2024: A Systematic Review. *arXiv preprint arXiv:2501.05435*. • Cunnington, Law, Lobo, & Russo. "The role of foundation models in neuro-symbolic learning and reasoning." International Conference on Neural-Symbolic Learning and Reasoning. Cham: Springer Nature Switzerland, 2024. • Dreuw, P., Ney, H., Martinez, G., Crasborn, O. A., Piater, J., Miguel Moya, J., & Wheatley, M. (2010). The signspeak project-bridging the gap between signers and speakers. • Elliott, R., Glauert, J., Jennings, V., & Kennaway, R. (2004). An overview of the SiGML notation and SiGMLSigning software system. *sign-lang@ LREC 2004*, 98-104.

[2] https://ai.google.dev/edge/mediapipe/solutions/guide
[3] https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/?ref=blog.roboflow.com

	• Hanke, T. (2004). HamNoSys–representing sign language data in language resources and language processing contexts. In *sign-lang@ LREC 2004*, 1-6. • Kezar, L., Thomason, J., and Sehyr, Z. (2023). Improving Sign Recognition with Phonology. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 2732–2737. • Lu, H., & Afridi, M. J. (2023). Surveying Neuro-Symbolic Approaches for Reliable Artificial Intelligence of Things. *Journal of Reliable Intelligent Environments*, 9(2), 123–138. • Moryossef, A. & Goldberg, Y. (2021). *Sign Language Processing*. https://sign-language-processing.github.io/ • Naert, L., Larboulette, C., & Gibet, S. (2020). A survey on the animation of signing avatars: From sign representation to utterance synthesis. *Computers & Graphics*, *92*, 76-98. • Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*. • Ong, S. C., & Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *27*(06), 873-891. • Selvaraj, P., Nc, G., Kumar, P., & Khapra, M. M. (2022). OpenHands: Making Sign Language Recognition Accessible with Pose-based Pretrained Models across Languages. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2114-2133. • Skobov, V., & Lepage, Y. (2020). Video-to-HamNoSys Automated Annotation System. *Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages*, 11–16. • Susskind, J., & Arden, O. (2021). Neuro-Symbolic AI: An Emerging Class of AI Workloads and their Characteristics. *arXiv preprint arXiv:2109.06133*. • Wan, Z., Liu, C.-K., Yang, H., Raj, R., Li, C., You, H., Fu, Y., Wan, C., Li, S., Kim, Y., Samajdar, A., Lin, Y. C., Ibrahim, M., Rabaey, J. M., Krishna, T., & Raychowdhury, A. (2024). Towards Cognitive AI. Systems: A Survey and Prospective on Neuro-Symbolic AI. *arXiv preprint arXiv:2401.01040*.
Detailed Work Plan	<u>**Before the workshop (preparation, june - august):**</u> 1. **Literature review**: review state of the art in (hand) pose estimation and neuro-symbolic AI techniques we could use for handshape extraction, as well as state of the art in EMG handshape recognition, especially applied to sign language. 2. **Detailed problem definition:** decide on exact input and output of the model (e.g. which types of HamNoSys handshapes will be included). How will the data be preprocessed, which neural

	models will be used and how will they be combined with symbolic methods for AI? 3. **Dataset definition:** develop a data collection strategy to collect video + EMG data of sign language handshapes from multiple participants **4. Dataset compilation:** compile a dataset of sign language videos from existing resources that are annotated using HamNoSys. 5. **Develop a pre-processing method for video**: develop a methodology to estimate poses from the videos which will be collected during the workshop. Optionally, data from the existing datasets can already be preprocessed before the workshop. **<u>During the workshop (2 weeks):</u>** 1. **Data collection (week 1):** collect video and EMG data of sign language handshapes with multiple participants. 2. **Baseline model development (week 1):** set up backbone of the model: neural models for symbol recognition and symbolic module to combine recognised symbols (e.g. tree structure based on the grammatical rules of HamNoSys). 3. **Baseline training (week 1):** prepare the train/validation/test split and select parameters for the model. Train the baseline video-only model. 4. **Baseline Evaluation (week 1):** evaluate the baseline model on the test set. 5. **EMG model development** 6. **EMG training (week 2):** train the EMG model. 7. **EMG evaluation:** evaluate the EMG model on the test set. 8. **Prepare demonstration and presentation (week 2):** prepare a demo of both models to show at the end of the workshop **<u>After the workshop (september - march):</u>** 1. **Write and submit paper:** the target venue for this paper could be the LREC 2026 workshop on representing and processing sign languages. The paper will discuss the data selection and preprocessing, the architecture used and the results obtained, focussing particularly on the explainability and trustworthiness of the model. 2. **Add brick to trail factory:** the dataset and code used during the project can be released as a brick on the trail factory, for which we will provide clear documentation.
Other Remarks	

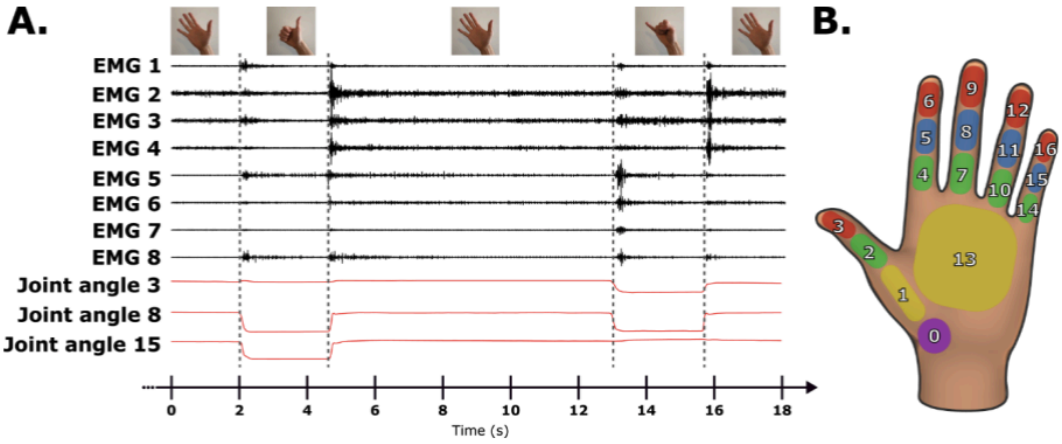

Figure 1. Visualization of the EMG signal in the ULB-MLG EMG and hand gestures dataset (Simar et al. 2024)

Figure 2. Pictures of the MindRove EMG armband available for data acquisition (website : [EMG armband | muscle sensor])