



SCAN ME

Abstract

Gap: There are **no available systems that would allow natural and real-time communication in German Sign Language (DGS)**. The NORA app in Germany relies on written text in German but not in DGS.

Our Contributions: The **DGS-BIGEKO dataset**, a sign language dataset containing a **conversation in a hypothetical emergency scenario** signed by a professional deaf interpreter in German Sign Language (DGS) to support research on bidirectional communications in this scenario.

Related Work

German Sign Language Datasets:

- **DGS Public Corpus** [Konrad et al. (2022)]: One of the largest covering **everyday dialogues**. **50 hours** of video material, **annotated using their double glossing scheme**. Only one conversation about **emergencies** narrating experience but **not maintaining a dialogue protocol**.
- **DGS-Fabeln-1** [Nunnari et al. (2024)]: German text and videos in DGS. Material recorded from different **camera perspectives**. Topic: **German fairy tales** interpreted by a deaf DGS signer.
- **Multilingual TUB corpus** [Avramidis et al., 2025]: more than **525 hours of interpreted DGS** from hearing interpreters but **not topics related to emergency scenarios**.

Sign Language Datasets for Emergency Scenarios:

- **The MSL4Emergency subset from the Myanmar Sign language Corpus** [Myat New et al. (2017)]: It contains 558 recordings with transcriptions in Myanmar glosses and translated Myanmar written text. But **only available in MSL**.
- **The KETI dataset** [Ko et al. (2019)]: The corpus has 14,672 videos about emergencies, signed in Korean Sign Language. It includes 105 sentences produced by 14 deaf signers.

Dataset Acquisition Protocol

Scenario and Dialogue Structure

Hypothetical emergency scenario:

A deaf DGS signer would witness an accident on a German highway and would contact emergency services as the first bystander.

Sentences/Vocabulary Selection:

- **Contact of emergency boards:** They shared anonymous conversations and their experiences during these emergency calls.
- **Persona design:** Synthetic conversations using LeanScope AI©

Dialogue Structure:

- **Rule-based** approach of Questions-Answers.
- **Two sets:** sentences (14 utterances) and vocabulary

Control Center: 112 Emergency Call Center. Thank you for your call. I will now ask you a few questions regarding the emergency. Please note the following instructions: 1. Position yourself centrally in front of your mobile phone camera. 2. Respond as simply and directly as possible. 3. Please answer only the question that is currently being asked.

Have you understand the indications? (Yes or no).

User (Option a): Yes.

User (Option b): No, please repeat it more slowly.

Control Center: What is the emergency? Is the emergency 1. a car accident, 2. a fire, 3. a medical emergency, or 4. something else?

Signer (Option a): Car accident

Signer (Option b): Fire

Signer (Option c): Medical Emergency

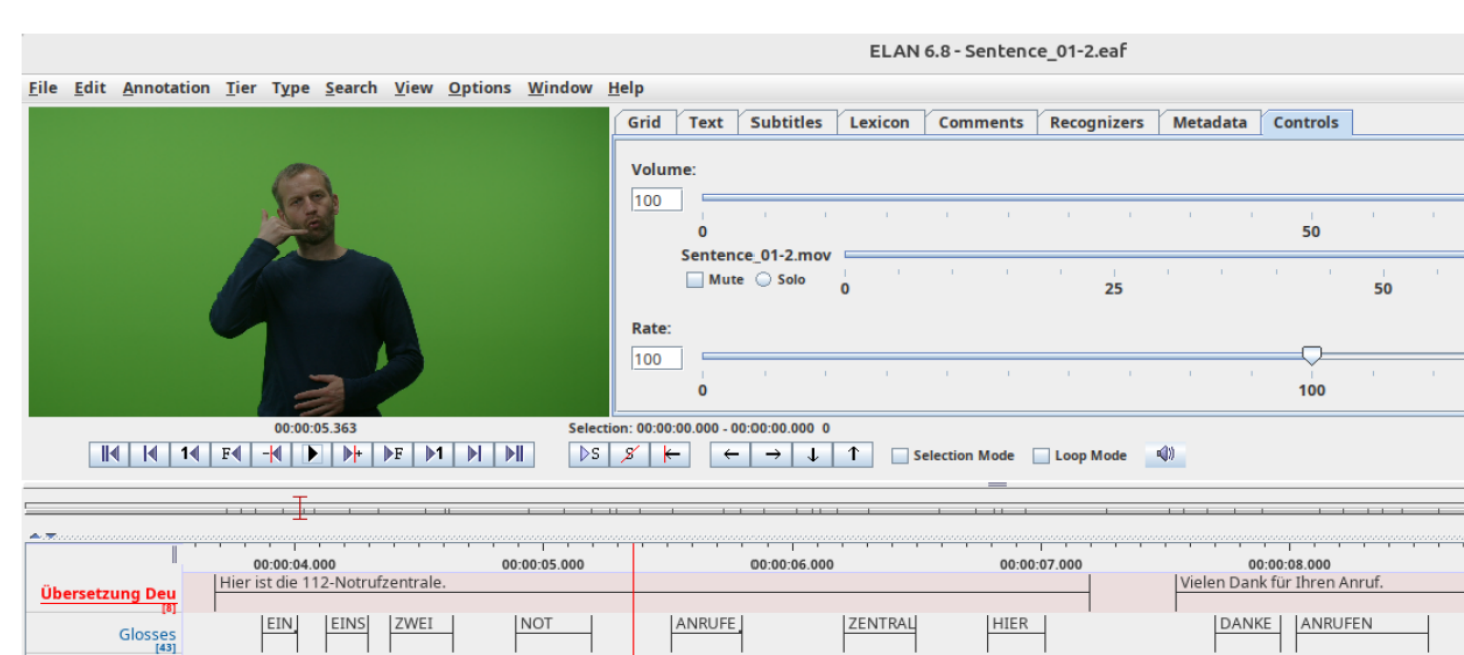
Signer (Option d): Other

...

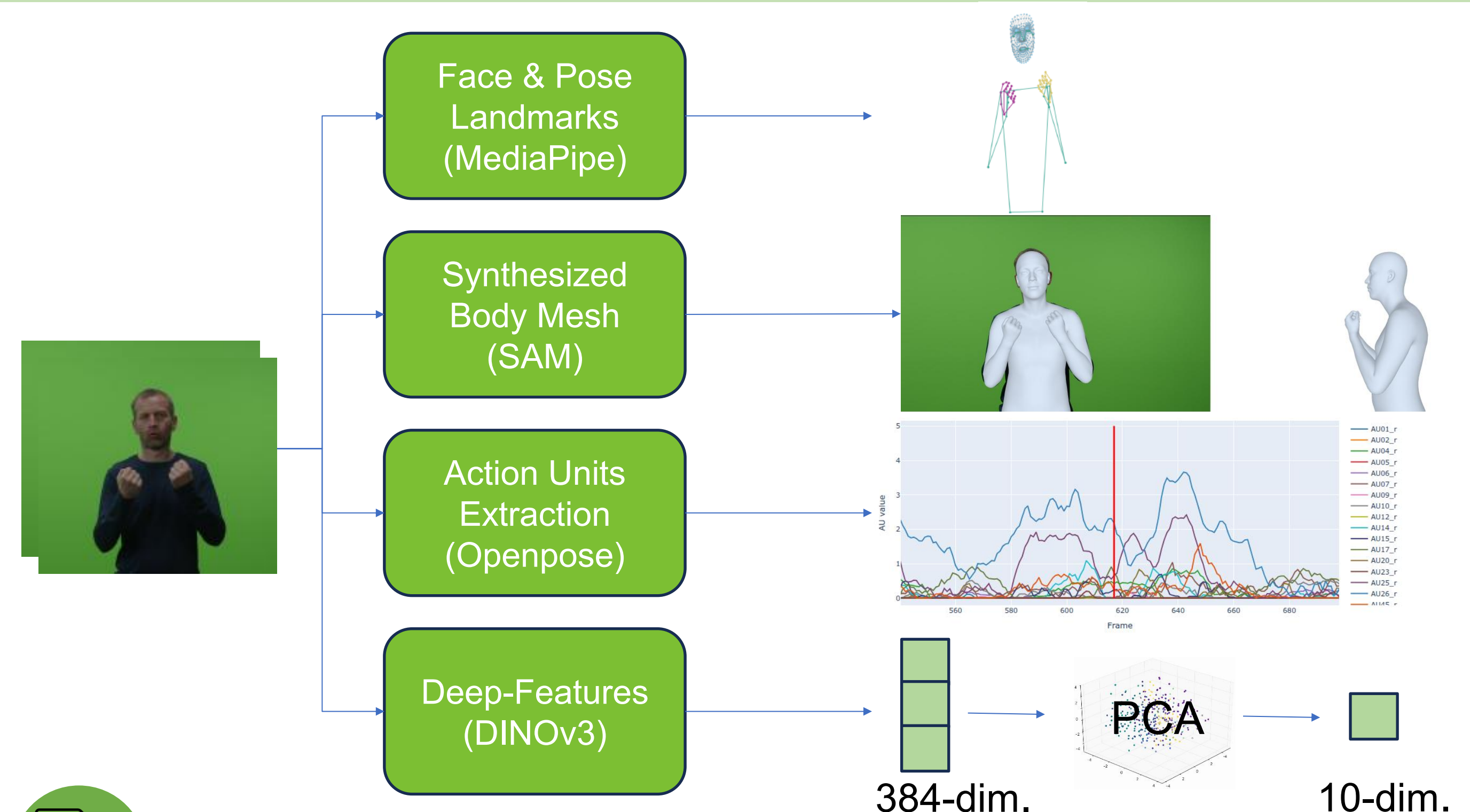
Recording set-up and procedure

Verification & Record: Deaf interpreters received the first version of the dialogue, adapt and record videos (1920x1080) at 50 fps in professional studio by one interpreter. 14 videos of sentences & 14 videos of vocabulary.

Annotation: Annotation by a hearing DGS interpreter following the ID-glossing conventions established by the DGS Corpus, using the ELAN tool.



Features



Corpus Analysis

German Sign Language (DGS)

Sentences: contain **111 different sign types** with an average repetition of $\mu=1.95$, **total of 217 sign tokens annotated**.

Vocabulary: topics related to means of transport in an accident, cities and places, emergencies, number of people, clarification statements, etc. **108 different sign types**.

German Spoken Language

Sentences:

- Sentences contain an average of 4 to 5 words: **Clear and explicit messages**.
- **Reduced vocabulary** but with large **linguistic diversity**.

Metrics for Sentences	Value
# Annotations	60
# Unique Annotations	53
# words	269
# tokens	370
# characters	1,493
Avg. words in sentence	4.20
% hapax legomena	41.64
TTR	56.13
% Short words (<4 characters)	35.69
Avg. Flesh Reading Ease	81.80
Avg. Wiener Sachtextformel	3.48

Table 1: Statistics for the German texts.

Conclusions & Future Work

- ✓ To enrich the available data for technological applications and contexts, **we propose the DGS-BIGEKO dataset. A dialogue containing 14 utterances of questions-answers on hypothetical emergency scenarios.**
- **Future Lines to improve:**
 - **Enlarge the dataset** with more sentences and possible answers that the systems should detect for creating prototypes. Include more subjects.
 - **Evaluate the degree to which some of the DGS sentences** reflect structures associated with **LBG** and possible re-record them.
 - **Consider variability of the language in these critical stressful scenarios.**

References

- [Avramidis et al., 2025] Eleftherios Avramidis, Vera Czehmann, Fabian Deckert, Lorenz Hufe, Aljoscha Lipski, Yuni Amalio Quintero Villalobos, Tae Kwon Rhee, Mengqian Shi, Lennart Stölting, Fabrizio Nunnari, and Sebastian Möller. 2025. The TUB Sign Language Corpus Collection. IVA25, pages 1–7, Berlin, Germany.
- [Nunnari et al. (2024)] Fabrizio Nunnari, Eleftherios Avramidis, Cristina España-Bonet, Marco González, Anna Hennes, and Patrick Gebhard. 2024. DGS-Fabeln-1: A Multi-Angle Parallel Corpus of Fairy Tales between German Sign Language and German Text. In Proceedings LREC-COLING 2024.
- [Konrad et al. (2022)] Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. Public DGS Corpus: Annotation Conventions / Öffentliches DGS-Korpus: Annotationskonventionen.
- [Koller et al. (2015)] Oscar Koller, Jens Forster, and Hermann Ney. 2015. Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems handling Multiple Signers. Computer Vision and Image Understanding, 141:108–125. Pose & Gesture.
- [Myat New et al. (2017)] Hlaing Myat Nwe, Hnin Aye Thant, Hnin Wai Wai Hlaing, Nandar Win Min, Ni Htwe Aung, Swe Zin Moe, and Ye Kyaw Thu. 2017. MSL4Emergency. GitHub. <https://github.com/ye-kyaw-thu/MSL4Emergency>.
- [König et al. (2008)] Lutz König, Susanne König, Reiner Konrad, and Gabriele Langer. 2008. Corpus-based Sign Dictionaries of Technical Terms – Dictionary Projects at the IDGS in Hamburg. In Proceedings of the LREC 2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora, pages 94–100, Marrakech, Morocco. European Language Resources Association (ELRA). [URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/08017.pdf](https://www.sign-lang.uni-hamburg.de/lrec/pub/08017.pdf).
- [Ko et al. (2019)] Sang-Ki Ko, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. 2019. Neural Sign Language Translation Based on Human Keypoint Estimation. Applied Sciences, 9(13):2683

Acknowledgements

This work was funded by the **BIGEKO** project (BMBF, German Ministry for Education and Research, grant number 16SV9094). Additionally, we would like to acknowledge and thank the contribution of yomma GmbH employees for helping with the recordings and translations. We also thank Centigrade GmbH and Charamel GmbH for their support in the guidelines and dialogue flow creation.