APPENDIX
Text-Driven 3D Hand Motion Generation from Sign Language Data

Léore Bensabath[1] Mathis Petrovich[2] Gül Varol[1]
[1] LIGM, École des Ponts, IP Paris, Univ Gustave Eiffel, CNRS [2] NVIDIA
leore.bensabath@enpc.fr
https://imagine.enpc.fr/~leore.bensabath/HandMDM

This appendix provides a brief glossary of terminology (Appendix A), an analysis of our BOBSL3DT dataset (Appendix B), implementation details (Appendix C), additional qualitative and quantitative results (Appendix D). We also attach a supplementary video to visualize dynamic motions (see `video.mp4`).

A. Glossary

We list a brief glossary to describe several sign language specific terminology used in the main paper:

- **Gloss:** The written representation of a sign, typically a single word. In linguistic glossing, the transcription follows certain rules (e.g., adding 'PT:' prefix for the pointing sign 'YOU', assigning a unique gloss ID to each sign variant); however, in this paper, we simplify as in [10] and abuse the gloss terminology: the sign-level annotations we use are not careful linguistic glosses, but rather free-form sign-level translations.
- **Phonology tags:** The set of handshapes, movements, locations used to construct the signs. See Sec. 3.1 of the main paper and Appendix C.4 for the specific list of attributes included in our study.
- **Sign variant:** One of the different ways of signing the same word. There may be multiple ways to sign a word due to several reasons. We give explanations and examples in Appendix C.7.

B. BOBSL3DT Dataset Analysis

We analyze our large-scale BOBSL3DT dataset by reporting several statistics (Appendix B.1), and estimating the noise level (Appendix B.2). We also provide more details about our manually cleaned test set verification (Appendix B.3).

B.1. Statistics

Text statistics. In Fig. A.1, we plot the distribution of word count per unique text descriptions. We show two separate histograms for (i) the texts generated from SignBank phonology attributes and (ii) the ones generated from SignBank attributes combined with HMS. The difference between the mean values of the two distributions remains small ($\mu_{\text{[Phonology]}} = 40.2$ and $\mu_{\text{[Phonology+HMS]}} = 41.0$), despite the latter containing more information. Phonology-only descriptions close the gap by containing generic movement words such as 'the hand moves'. In Fig. A.2, we also provide the distribution of the 70 most common word occurrences per unique text. A word cloud of the most common words is shown in Fig. A.3. Measure of 'Textual Lexical Diversity' is 37.8.

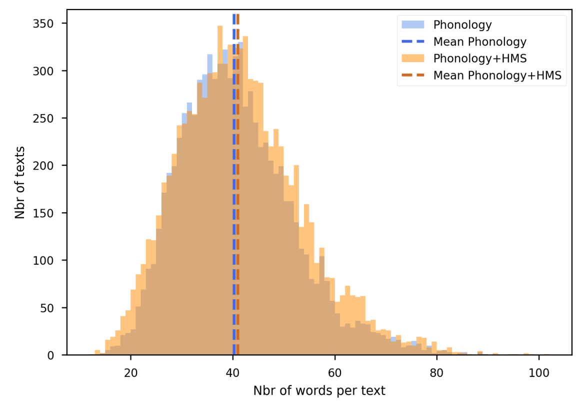


Figure A.1. **Histograms of the number of words per text in BOBSL3DT training set:** We report the distribution of words per unique text, for the descriptions generated from SignBank phonology, and the descriptions generated from SignBank phonology combined with HMS separately.

Motion statistics. In Fig. A.4, we also plot the number of motion samples per unique text description in our LLM(Phonology+HMS) training set. Given that the glosses have a long-tailed distribution in BOBSL, the motions are not uniformly distributed across descriptions either. The dataset includes 17,552 unique descriptions (x-axis), among which 8,201 come from Phonology-only and 9,351 come from the combination with HMS.

We note two details to help interpreting these numbers. (a) Original LLM outputs from SignBank are generated with the dominant/non-dominant hand terminology, which are then replaced with the words right/left when assigning to BOBSL signs depending on the metadata on right-/left-handedness of signers (i.e., a left handed person's dominant hand would be the left hand with mirrored signing). With the original dominant terminology, we have 6,090 (Phonology) and 7,032

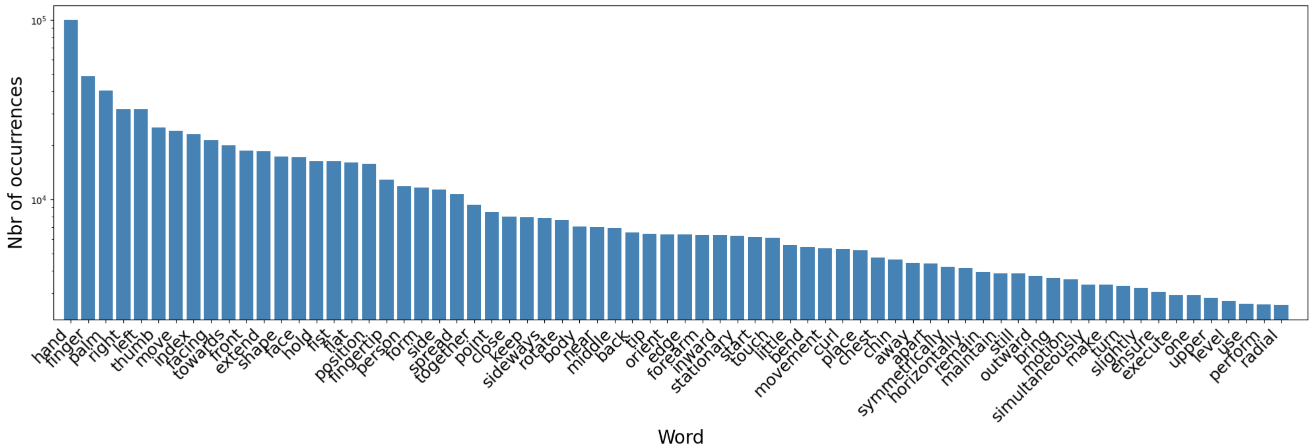


Figure A.2. **Histogram of the number of occurrences of words in the dataset texts:** We report the distribution of the 70 most common words per unique text.

Figure A.3. **Word cloud of the dataset texts:** The point cloud illustrates the frequency of the most common words in the dataset per unique text.

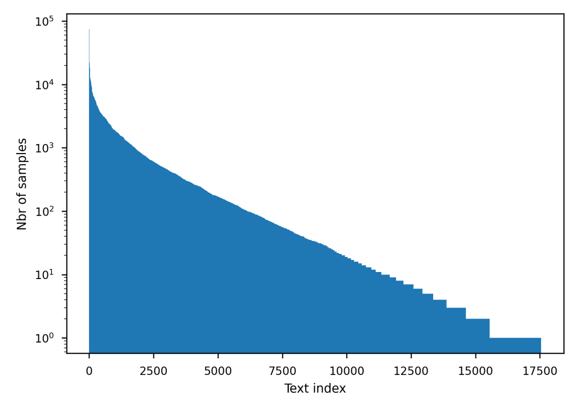


Figure A.4. **Number of motion samples per text description:** We observe a long-tailed distribution on the BOBSL3DT training set. The median is 33 motion samples per text.

(Phonology+HMS) unique text descriptions on SignBank, which then increase to the above numbers after replacing with left/right words during BOBSL assignment. (b) The source of these descriptions are 2,351 glosses. Each gloss is converted to 3 different LLM descriptions (2,351 $\times$ 3 = 7,053). Due to some glosses sharing the exact same set of attributes, we have slightly lower number of unique descriptions than 7,053 (e.g., 6,090 for Phonology portion of the dataset).

In Fig. A.5 and Fig. A.6, we plot the distribution of motion samples per handshape, respectively for non-symmetrical and symmetrical motions. For non-symmetrical motions, we provide two distributions, for right and left hand. The right hand naturally performs a lot more movements than the left, since many of the non-symmetrical signs are single-handed and BOBSL includes mostly right-handed signers (39 right-handed signers and 4 left-handed signers).

B.2. Noise estimation and impact

We quantify the noise on a training subset of 2,525 samples by manually checking whether the motion corresponds to the text. We find that 61% of the motion samples correspond to their

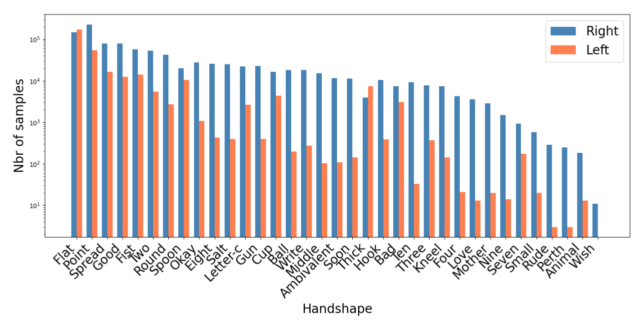


Figure A.5. **Number of motion samples per handshape for non-symmetrical motions:** We show the distributions for the right and left hand separately. The median is 11,669 motions per right handhsape and 370 per left handshape.

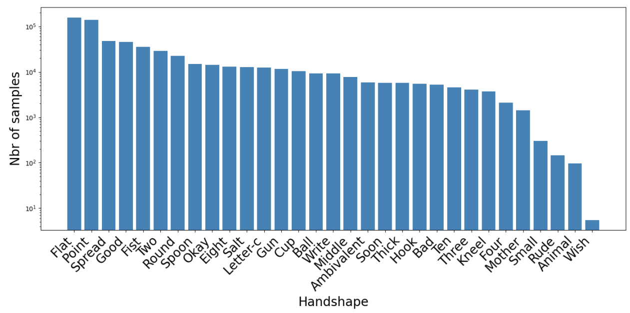


Figure A.6. **Number of motion samples per handshape for symmetrical motions:** Symmetrical motions include motions tagged as *symmetrical*, *alternating* or *parallel*. In these motions, the right and left hand perform the same handshapes, so we include the counts for one hand only. The median is 9,175 motions per handshape.

Model	R@1	R@3
60% PL	15.80	33.62
80% PL	**17.53**	32.18
100% PL	**17.53**	**36.20**

Table A.1. **Pseudo-label** filtering effect: We reduce the dataset noise by discarding samples with pseudo-labelling confidence below a given threshold. Training with the 60% (row 1) and 80% (row 2) less noisy portions of the dataset doesn't show improvement.

descriptions. The main sources of noise are pseudo-glosses (e.g., 60% of 2,624 manually-glossed test set match the pseudo-glosses), and potentially the SignBank variant assignment. We also note that the 3D pose estimates contribute to the level of noise, by not strictly following the video, but our analysis here focuses on the fidelity to the text rather than the motion realism.

Confidence thresholding during pseudo-label assignment reduces the noise induced during this step to some extent (see Sec. 3.1, first paragraph, of the main paper). We experiment with further reduction of the noise by discarding motions with pseudo-labelling confidence bellow a given threshold. A motion confidence is taken as the average confidence over its frames. In Tab. A.1, we use 80% and 60% smaller, less noisy training data by increasing the confidence threshold from 0.60 to 0.74 to 0.82. We do not observe improvements.

B.3. Test set verification

As mentioned in Sec. 4.1 of the main paper, we manually verify the entire test set (seen and unseen) as follows. We first automatically assign a SignBank variant to each BOBSL test motion, by selecting the variant with the highest cosine similarity in the THMR embedding space. Then, for each of these SignBank samples, we collect 4 most similar BOBSL motions, and manually select one of them that matches the SignBank motion, or discard the gloss if we find no match.

We also manually modified the LLM(Phonolgy+HMS) output descriptions for a subset of the test samples, to make sure the text corresponds to the motion. Due to its smaller size, we have done this correction only on the unseen test set.

Note that the entire test set was already manually *glossed* (as opposed to the training set); therefore, we did not need to verify the correctness of matching between the videos and the glosses.

C. Implementation Details

We give additional information about HandMDM training (Appendix C.1), THMR training (Appendix C.2), stitching of HAMER hands to SMPLer-X body (Appendix C.3), SignBank dictionary attributes (Appendix C.4), HandMotionScript (Appendix C.5), the LLM prompt (Appendix C.6), and the assignment between SignBank and BOBSL (Appendix C.7).

C.1. HandMDM training and inference

Similar to [9], HandMDM architecture includes 8 layers, 8 attention heads, its latent dimension is 512 and the intermediate feedforward dimension is 2048. We use a frozen CLIP-ViT-B/32 model to encode the texts. We train the model using a learning rate of 10^{-4} with AdamW optimizer and a batch size of 32. With BOBSL3DT, we train HandMDM for 300 epochs and use a guidance coefficient of $\lambda = 15$ at test time. With ASL-Text and MS-ZSSLR, we train for 2000 epochs and test with $\lambda = 5$. With BOTH57M, we train for 300 epochs and test with $\lambda = 15$. We select our guidance coefficients based on the visual quality of the outputs.

C.2. THMR training

Similar to TMR for bodies [8], we train the THMR networks for hands using a learning rate of 10^{-4} with AdamW optimizer and a batch size of 32. DistilBERT [11] text embeddings are used as input to the text encoder. The transformer encoders and decoders includes 6 layers and 4 attention heads, the latent dimensionality of the embedding space is 256, the intermediate feedforward dimension is 1024, the temperature $\tau = 0.1$, and the contrastive loss weight $\lambda_{NCE} = 0.1$. The network is trained for 100 epochs. As THMR is used for evaluation purposes, to obtain the best possible performance, we use for supervision the 1.3M motion-text samples from our BOBSL3DT training data *prior* to filtering the unseen signs. The input format is the same as HandMDM (see Sec. 3.2).

Performance of THMR on real motions. We report THMR performances on ground truth motions in the first row of Tab. A.3. Note that the unseen set is not unseen with regards to THMR as mentioned above. We refer to Appendix D.2 for comments on the results.

THMR for transfer experiments. For ASL-Text and MS-ZSSLR retrieval score evaluation, we train THMR only on the arm and hand features, discarding the torso. We notice that a significant portion of these datasets include seated signers, where the torso position creates a domain gap with BOBSL3DT containing standing signers. Additionally the 3D reconstruction for seated signers tend to estimate unrealistic torso bending for the ground truth, which also motivated our exclusion of the torso features. Specifically we use the following 6D joint rotations: 2 shoulders, 2 elbows, 2 wrists and 30 for hands, resulting in a total of 216 input features.

BOTH57M full body motions are a lot more dynamic than in sign language datasets, and the textual descriptions for hands only describe finger movements (not arms). Therefore, for the evaluation on BOTH57M, we further reduce the training features of THMR to the 30 hand 6D rotations, resulting in a total of 180 features.

C.3. Stitching HAMER hands to SMPLer-X body

As explained in Sec. 3.1 of the main paper, we use a combination of HAMER [7] and SMPLer-X [2] for estimating 3D bodies with hands from videos. The quality of SMPLer-X hand predictions is not sufficient for our application and we observe consistently more reliable estimates with the hand-specific and more recent approach of HAMER.

In terms of the global location of wrists, we find SMPLer-X to be better since it explicitly estimates the rest of the body, and arm movements define the hand location. However, in terms of the global orientation of wrists, we find HAMER to be more accurate. Therefore, we design our stitching pipeline to incorporate HAMER wrist orientations into the final pose, instead of directly plugging in the hand rotations predicted from HAMER into the SMPL-X body parameters predicted from SMPLer-X. In Fig. A.7, we compare this latter more simple approach with our stitching pipeline results, and observe consistently more reliable wrist estimates when deriving the wrist global orientations from HAMER predictions.

In more detail, HAMER gives us 30 hand joint rotations (local rotations with respect to the parent in the kinematic tree), which we use to replace those estimated by SMPLer-X. For the wrists, we use the global wrist orientations estimated from HAMER which are in world coordinate with the origin at the wrist. We convert these to be local wrist orientations and replace those estimated by SMPLer-X. We first stitch both estimates at the wrists by using SMPLer-X wrist positions and HAMER wrist global orientations. This provides realistic joint positions but can lead to unrealistic wrist rotations. For instance, if SMPLer-X estimates a hand facing down but HAMER estimates a hand facing up, this first stitching results in a local rotation at the wrist of about 180 degrees, which is unrealistic.

Therefore, inspired by [6] and [13], we then perform optimization to adjust the SMPL-X rotations of the arm joints. We optimize the shoulders, elbows and wrists (6 rotations) by minimizing the Euclidean distance between the 3D joints derived from the new rotations and the initial stitched pose. Specifically, the target coordinates are the 3D coordinates of the elbows (2 joints), wrists (2 joints) and hands (30 joints). We initialize the target joint rotations with the SMPL neutral pose for each frame, i.e., with the arms positioned in a T-pose. Similar to [13], we add a temporal smoothing term that minimizes the Euclidean distance between the joint positions from consecutive frames. A regularization term also minimizes the magnitude of the rotations to avoid unrealistic bending. We weight each of the target function terms with the following weights: $\lambda_j = 10$ for the 3D joints Euclidean distance term, $\lambda_{smooth} = 5$ for the temporal smoothing term, $\lambda_{reg} = 0.05$ for the magnitude regularization term.

C.4. SignBank dictionary attributes

We refer to Fig. A.8 and Fig. A.9 to illustrate complete examples for the dictionary attributes. Beyond those mentioned in Sec. 3.1 of the main paper, phonological information also contains whether signs are performed with only one hand, or with both hands. In case of two-handed signs, the phonology also denotes whether the sign is 'symmetric' or 'alternating', and whether one of the hands is static. There may also be information on whether forearm rotations are involved, and whether a handshape change occurs during the motion (in addition to the initial/final handshapes).

To transform these SignBank attributes into our LLM input, we replace linguistic terms in field names and categories with more descriptive English descriptions. Additionally, we provide several descriptions for an attribute where possible, as text augmentation. We randomly select from these attribute descriptions for each gloss when inputting the attributes to the LLM. For instance, for the handshape attributes, 'flat' will be converted into a description among [*'Flat', 'The hand is held flat with the fingers held together.', 'The fingers are extended and together.'*], and 'small' will be converted into a description among [*'The index finger is extended and bent inward. The thumb is extended, parallel to the index finger. The other fingers are curled on the palm.', 'The extended index finger, bent at the palm knuckle, and extended thumb are held parallel to each other.'*]; for the tags, 'double handed' will be converted into the description *'both hands are used, with the same handshape on both hands'* and 'two handed' into *'both hands are used, and their motions aren't symmetrical'*. Specifically, we take part of the handshape descriptions from [1]. The full mapping can be found in `mapping_values.py` in our supplementary files.

C.5. HandMotionScript

We provide a pseudo-code in Algorithm 1 for our HandMotionScript (HMS) pipeline, and qualitative examples in Fig. A.8 and Fig. A.9. We build on PoseScript [3] designed for full human bodies. From PoseScript features, we use the hand **distances** to the relevant body parts, determined by SignBank attributes. If the hand location attribute is not the 'neutral space', but a specific body part, we compute the distances from the hands to this body part, to detect the movement features. If the motion is two-handed, we compute distances between the two hands. If the motion is one-handed, we compute the distances for the signer's dominant hand only. We further break down the distances into the **3 axes (x-y-z)** to provide more detailed information, such as left/right or above/below relationships. As in PoseScript, we convert numerical values into text codes by assigning a range of values for each feature. We adapt some of these ranges for our HMS pipeline. We use the following text codes for the 4 distance features:

Figure A.7. **Stitching HAMER hands to SMPLer-X body:** We observe qualitatively better performance of capturing wrist orientations when using HAMER [7] global wrist orientations (labeled *w/ HAMER wrist*), as opposed to the wrist components of SMPLer-X [2] (labeled *w/ SMPLer-X wrist*). In order to use HAMER global wrist orientations with SMPLer-X body estimations, we stitch SMPL-X hands estimated using HAMER into SMPL-X bodies estimated by SMPLer-X with optimization applied to the arm rotations.

• **Distance:** [*touching, close, medium, spread, wide*]
• **x-coordinate:** [*wide/right, spread/right, medium/right, close/right, touching, close/left, medium/left, spread/left, wide/left*].
• **y-coordinate:** [*wide/below, spread/below, medium/below, close/below, touching, close/above, medium/above, spread-/above, wide/above*].
• **z-coordinate:** [*wide/behind, spread/behind, medium/behind, close/behind, touching, close/in front, medium/in front, spread/in front, wide/in front*].

We also add a new feature to PoseScript to extract the **palm orientations**. Specifically, for each hand, we compute the vector normal to the wrist global orientation. If one of its global coordinate norms is greater than a chosen threshold (0.7), we assign the axis of this coordinate as the palm direction. The sign of the coordinate determines the final orientation. A palm can thus be assigned an orientation among the following possibilities, respectively to the 3 axes: *sideways, down/up, in/out*.

Once the above frame-level features are extracted, we apply the following post-processing in order to shorten them and remove possible noise. We remove every text code with less than 4 consecutive repetitions, and we collapse consecutive repetitions into a single text code. Furthermore, when a distance feature has multiple text codes for multiple axes, we discard the axis-dependent distance features (x/y/z) entirely to prevent too lengthy and noisy descriptions from the LLM. Finally, PoseScript wording uses 'right' and 'left' hands, whereas our raw SignBank descriptions follow 'dominant' and 'non-dominant' terminology (see Appendix B.1). We replace the right/left words with dominant/non-dominant according to the SignBank tags which give information on whether a signer is left-handed or right-handed. The temporal sequence of words are then given to the LLM.

C.6. LLM prompting

We use Gemini 2.5 Pro *Thinking Experimental 01-21* [4] as our LLM. We provide the exact prompt for generating the textual descriptions from the SignBank and HMS attributes in Algorithms 2 to 6. The prompt contains the task instructions and few-shot (six) examples for in-context learning. We also include the prompt version when HMS is not used in our supplementary files (`prompt_phonology.txt` and `prompt_phonology_HMS.txt`). We use both prompts to construct our BOBSL3DT dataset.

LLM output post-processing. For single-handed signs, we remove any mention of the non-dominant hand 'remaining still' from the LLM output. This is because in BOBSL, the non-dominant hand is rarely still (due to the natural signing speed) as opposed to the slow SignBank videos.

Furthermore, once the SignBank LLM-generated texts are assigned to BOBSL motions, we replace the words 'dominant'

Algorithm 1 HandMotionScript pipeline

```
Input: 3D motion world coordinates (n frames)
Input: Global hand orientation (n frames)
Input: SignBank location attributes (SBlocs)
for frame ← 1 to n do
    for SBlocs ← loc do
        if loc is Neutral Space then loc ← other hand
        end if
        if left hand is used then
            Compute distance posecodes: left hand to loc
        end if
        if right hand is used then
            Compute distance posecodes: right hand to loc
        end if
    end for
    if left hand is used then
        Compute orientation posecodes: left hand
    end if
    if right hand is used then
        Compute orientation posecodes: right hand
    end if
end for
Return Distance posecode sequences (n frames)
Return Orientation posecode sequences (n frames)
```

and 'non-dominant' with 'left' or 'right' depending on whether the signer is left-handed or right-handed using the metadata.

C.7. SignBank gloss assignment for BOBSL motions

Each BOBSL video corresponds to one pseudo-gloss, represented as a word (or more rarely a phrase). On the other hand, SignBank is a linguistic dictionary, where each gloss corresponds to a unique sign variant, characterized by one motion. For example, the word 'happy' corresponds to three glosses HAPPY, HAPPYb, HAPPYc. The mapping between words and glosses is complex [10]. A word can map to multiple SignBank glosses, and a gloss can map to multiple words. These may be due to multiple ways of signing the same concept, due to homonyms (same sign motion with different meanings, such as BSL signs for 'battery' and 'uncle'), or due to English synoyms ('happy' and 'content').

SignBank also includes *keywords* for each gloss, related concepts which may share the same signing. For example, the keywords for the gloss 'HAPPY' include 'happiness', 'merry', 'enjoy', 'fun', while they include 'excited', 'jump for joy' for the gloss 'HAPPYc'. When constructing the list of *variants* for a given BOBSL pseudo-gloss word (e.g., 'happy'), we consider all possible keywords (e.g., 'fun', therefore FUN, FUNb, FUNc...), as well as all glosses that map to this word (e.g., 'HAPPYc').

Once we construct a list of candidate gloss motions, for each sign label in BOBSL, we perform k-medoids clustering on the set of its paired BOBSL motions combined with each candidate SignBank variant motion. We initialize the number of clusters as the number of available SignBank variants. For each cluster thus obtained, if the cluster includes: (i) a unique SignBank variant, it is assigned to each of its BOBSL samples; (ii) no SignBank variant, all of its BOBSL samples are filtered out; (iii) more than one SignBank variants, its BOBSL samples are assigned the SignBank variant that maximizes their cosine similarities.

As mentioned in Sec. 3.1 of the main paper, the THMR used for this assignment is trained with random assignment (since initially there is no assignment). But since we only use its motion encoder, we do not observe a big difference in its assignment behavior if we use the second iteration of the THMR model, trained on these better assigned texts. This may be due to a relatively small search space within SignBank candidates. However, note that for our evaluations, we use the latter THMR.

D. Additional results

Here, we complement the results in the main paper by reporting additional metrics evaluating HandMDM ability to generate temporally smooth motions (Appendix D.1), using motion-to-text retrieval for BOBSL3DT-Test (Appendix D.2), evaluating on SignBank as a test set (Appendix D.3), and visualizing LLM inputs and outputs (Appendix D.4). We also provide additional qualitative visualizations (Appendix D.5).

D.1. Temporal smoothness performance of HandMDM

We acknowledge that our motion reconstruction pipeline operates largely on a frame by frame basis, and as the result, this dataset presents some temporal jitter as expected. However, we show in this section that this jitter has little impact on the temporal quality of our motion generation outputs.

In order to evaluate the impact of the training set jitter on HandMDM generation smoothness, we compare smoothness scores from BOBSL3DT ground truth motions and HandMDM generated motions. Besides, we extract smoother subsets of BOBSL3DT, train HandMDM with the subsets and compare how decreasing the jitter in the training set affects the jitter of the predictions.

To this end, we compute for each sample an acceleration score, i.e the average rotational acceleration over all joints and frames. The rotational acceleration is defined as the change in angular velocity between two consecutive frames (in $rad\,/\,s^2$). We compute its 75th and 50th percentiles over the dataset, keep the subsets of samples with an acceleration score below these thresholds, and train HandMDM with each subset.

We report the results in Tab. A.2. We compare both the rotational acceleration scores computed from the SMPL-X pose parameters, and the jitter scores, similar to [12], defined as the acceleration on the 3D joints (in $m\,/\,s^3$). Column 3 and 4 both show the average scores of the training sets used for each model, confirming that the rotational acceleration and the jitter scores are correlated and both decreasing with our subset selection. Row 1 provide the reference scores for the ground truth seen

Input	% train data	GT Train. set		Seen		Unseen	
		Rot. Accel ↓	Jitter ↓	Rot. Accel. ↓	Jitter ↓	Rot. Accel. ↓	Jitter ↓
GT				87	375	77	332
LLM (Phonology+HMS)	100	91	284	7	97	3	47
+ Accel filtering	75	49	229	4	77	2	43
+ Accel filtering	50	31	192	4	68	3	51

Table A.2. **Smoothness metrics:** We assess the impact of the jitter amount in the training set on the smoothness of the generations. The rotational acceleration and jitter scores both show that the generations from our original HandMDM are significantly smoother than the ground truths (Rows 1 and 2). When decreasing the amount of jitter in the training set, the smoothness score improvement of the generations, compared to the initial ground truth jitter, is small for the seen test set, and not significant for the unseen test set.

and unseen test sets. Both metrics support the fact that the smoothness of our LLM (Phonology+HMS) model generated motions (row 2) improves significantly over the smoothness of the ground truth data (rows 1). Besides, rows 2 to 4 illustrate that, even when decreasing the jitter of the training data, the generation smoothness stays fairly stable. This supports the fact that our original model is already robust enough to generate smooth motions from a more jittery training set.

We provide additional qualitative evidence in our supplementary video, that further demonstrates that the jitter present in the training data does not propagate to HandMDM generations. Future work could incorporate full-body motion optimization, as in [5], to further improve the temporal smoothness of BOBSL3DT.

D.2. Motion-to-text retrieval metrics on BOBSL3DT-Test

For completeness, we repeat the last two rows of the Tab. 2 of the main paper, this time by reporting the motion-to-*text* metrics from THMR (as opposed to the motion-to-*motion* metrics used in the main paper for BOBSL3DT). The results in Tab. A.3 support our previous conclusions drawn from the comparison of motion-to-motion scores. We also complement the metrics with diversity and multimodality scores for completeness, and include ground truth evaluation as a reference (first row).

On the unseen test set, performance of the ground-truth motion to retrieve text is overall better than that of generated motions (also because the descriptions are seen by the THMR as mentioned in Appendix C.2). On the seen partition, generations outperform ground-truth motions in terms of text retrieval capability using THMR. A potential explanation is that the generations resemble the *training* set, on which the THMR was optimized on. This phenomenon is similar to the observations in [8] (see Table A.2 of their paper).

D.3. Testing on SignBank motions

Tab. A.4 evaluates the two model variants (corresponding to the last two rows of Tab. 2 of the main paper) by comparing the generated motions to ground truth SignBank motions. Motion-to-motion similarity suffers from a domain gap when comparing them with the shorter and quicker signs of BOBSL. Regardless, the conclusion that using HandMotionScript benefits still holds.

D.4. Qualitative results for the LLM

In Fig. A.8 and Fig. A.9, we provide example inputs and outputs for the conversion from SignBank data to free-form textual descriptions. We first map the linguistic attributes to descriptions with our lookup table (as explained in Appendix C.4), which are then input to the LLM along with our HandMotionScript features.

D.5. Additional qualitative visualizations

We repeat the main paper figures in dynamic motion visualization in Figs. A.10 to A.12, and show more results on BOBSL3DT-Test, this time on the *seen* partition (Fig. A.13).

References

[1] David Brien. *Dictionary of British Sign Language / English; compiled for the British Deaf Association by the Deaf Studies Research Unit, University of Durham.* Faber, 1992. 4

[2] Zhongang Cai, Wanqi Yin, Ailing Zeng, Chen Wei, Qingping Sun, Wang Yanjun, Hui En Pang, Haiyi Mei, Mingyuan Zhang, Lei Zhang, Chen Change Loy, Lei Yang, and Ziwei Liu. SMPLer-X: Scaling up expressive human pose and shape estimation. In *NeurIPS*, 2023. 4, 5

[3] Ginger Delmas, Philippe Weinzaepfel, Thomas Lucas, Francesc Moreno-Noguer, and Grégory Rogez. PoseScript: 3D human poses from natural language. In *ECCV*, 2022. 4

[4] Gemini Team et al. Gemini: A family of highly capable multimodal models. *arXiv:2312.11805*, 2024. 5

[5] Jing Lin, Ailing Zeng, Shunlin Lu, Yuanhao Cai, Ruimao Zhang, Haoqian Wang, and Lei Zhang. Motion-X: A large-scale 3D expressive whole-body human motion dataset. In *NeurIPS*, 2023. 7

[6] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019. 4

[7] Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, and Jitendra Malik. Reconstructing hands in 3D with transformers. In *CVPR*, 2024. 4, 5

[8] Mathis Petrovich, Michael J. Black, and Gül Varol. TMR: Text-to-motion retrieval using contrastive 3D human motion synthesis. In *ICCV*, 2023. 3, 7

[9] Mathis Petrovich, Or Litany, Umar Iqbal, Michael J. Black, Gül Varol, Xue Bin Peng, and Davis Rempe. Multi-track timeline control for text-driven 3D human motion generation. In *CVPRW*, 2024. 3

Input	Seen R@1 ↑		Seen R@3 ↑		Seen FID ↓		Seen Div.←		Seen MultiMod. ↑		Unseen R@1 ↑		Unseen R@3 ↑		Unseen FID ↓		Unseen Div.←		Unseen MultiMod. ↑	
Ground truth	22.58	24.93	40.86	42.80	-		0.94	0.94	-		35.63	**34.48**	**59.77**	**56.32**	-		0.93	0.93	-	
LLM(Phonology)	**52.80**$_{\pm 1.3}$	8.21$_{\pm 0.2}$	**74.06**$_{\pm 0.7}$	16.10$_{\pm 0.4}$	**0.15**	0.23	**0.93**	0.90	0.09	**0.15**	**36.20**$_{\pm 0.6}$	19.54$_{\pm 5.0}$	56.32$_{\pm 2.9}$	39.66$_{\pm 0.6}$	**0.38**	0.46	0.89	0.89	0.10	0.15
LLM(Phonology+HMS)	51.63$_{\pm 0.8}$	**52.29**$_{\pm 0.9}$	72.51$_{\pm 0.4}$	**70.81**$_{\pm 0.9}$	0.16	**0.16**	0.91	**0.91**	0.09	0.09	32.47$_{\pm 0.5}$	30.74$_{\pm 2.9}$	51.72$_{\pm 1.4}$	55.17$_{\pm 1.2}$	0.40	**0.40**	**0.90**	0.89	**0.13**	0.15

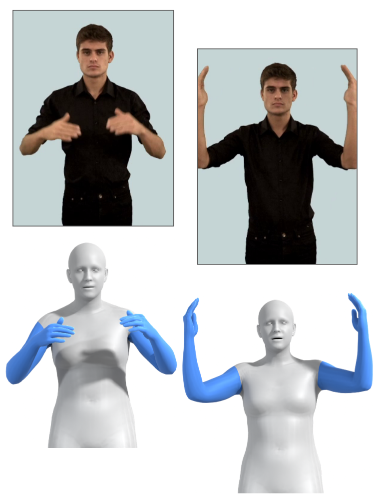
Table A.3. **Motion-to-text metrics:** We complete the results of Tab. 2 by providing the motion-to-*text* retrieval scores for our LLM(Phonology) and LLM(Phonology+HMS) models, as well as the BOBSL3DT-Test ground truth motions. See text for comments.

Input	Seen R@1 ↑		Seen R@03 ↑		Seen FID ↓		Unseen R@1 ↑		Unseen R@03 ↑		Unseen FID ↓	
LLM(Phonology)	**17.56**$_{\pm 0.3}$	5.54$_{\pm 0.4}$	**30.78**$_{\pm 0.6}$	11.74$_{\pm 0.4}$	0.36	**0.42**	**22.70**$_{\pm 1.7}$	12.37$_{\pm 1.6}$	35.92$_{\pm 1.0}$	24.45$_{\pm 2.0}$	0.57	**0.60**
LLM(Phonology+HMS)	16.59$_{\pm 0.4}$	**17.73**$_{\pm 0.1}$	30.40$_{\pm 0.5}$	**32.55**$_{\pm 0.3}$	0.36	0.36	20.98$_{\pm 2.5}$	**21.98**$_{\pm 2.8}$	**41.67**$_{\pm 1.0}$	**34.06**$_{\pm 2.5}$	**0.58**	0.57

Table A.4. **Transfer results on SignBank motions:** We report the two model variants from Tab. 2 of the main paper. The models are trained on BOBSL3DT and tested on motions from *SignBank*. We note a domain gap between co-articulated (fast) motions from BOBSL3DT compared to isolated (slow) motions from SignBank. The notation for the '|' separator is the same as before, i.e., 'LLM(Phonology) | LLM(Phonology+HMS)'.

Phonology from SignBank

Gloss (Sign): SKY
Initial Dominant Handshape: Flat
Initial Subordinate Handshape: Flat
Initial Primary Location: Neutral space
Initial Interacting Dominant Hand Part: Fingertips
Tags: Double handed, Symmetrical



LLM input from phonology

Initial dominant hand shape: Flat
Initial subordinate hand shape: Flat
Final dominant hand shape: Flat
Final subordinate hand shape: Flat
Dominant hand initial location: In front of the person
Dominant hand final location: In front of the person
Location on the dominant hand that the non-dominant hand is facing: Fingertip(s)
Location on the non-dominant hand that the dominant hand is facing: Fingertip(s)
Both hands are used, with same hand shape on both hands
The movement is symmetrical, meaning both hands are doing the same motions at the same time

LLM input from HandMotionScript

DISTANCE BETWEEN HANDS:
Distance from dominant hand to non-dominant hand: [medium, spread, wide, spread]
Distance along x axis from dominant hand to non-dominant hand: [medium, spread, wide, spread]
Distance along y axis from dominant hand to non-dominant hand: [aligned]
Distance along z axis from dominant hand to non-dominant hand: [aligned]
HAND ORIENTATIONS
Palm orientation - dominant hand: [sideways]
Palm orientation - non-dominant hand: [sideways]

LLM output:

1. Both hands are held flat with fingers together, palms facing each other. They move horizontally apart from a medium distance to a wide separation, keeping their fingertips oriented towards one another.

2. The person positions both hands flat, fingers joined, with palms turned inward facing each other. They execute a symmetrical movement, separating the hands horizontally until they are wide apart.

3. Start with both hands flat, fingers together, palms facing inward and fingertips pointing towards each other. Move your hands symmetrically outwards along the horizontal axis to a wide position.

Figure A.8. **Qualitative results for the LLM:** We show an example input-output for the LLM when converting the attributes from the SignBank phonology and from our HandMotionScript into a free-form language description. Note that we show the gloss ('SKY') for reference, but this information is not included in the LLM input. More examples are continued in Fig. A.9.

[10] Charles Raude, K. R. Prajwal, Liliane Momeni, Hannah Bull, Samuel Albanie, Andrew Zisserman, and Gül Varol. A tale of two languages: Large-vocabulary continuous sign language recognition from spoken language supervision. *arXiv*, 2024. 1, 6

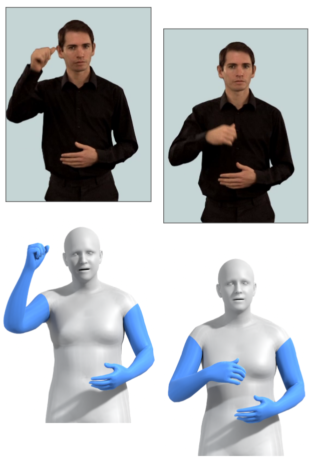
[11] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In *Energy Efficient Machine Learning and Cognitive Computing NeurIPS Workshop*, 2019. 3

[12] Yufu Wang, Yu Sun, Priyanka Patel, Kostas Daniilidis, Michael J Black, and Muhammed Kocabas. Prompthmr: Promptable human mesh recovery. In *CVPR*, 2025. 6

[13] Liang Xu, Xintao Lv, Yichao Yan, Xin Jin, Shuwen Wu, Congsheng Xu, Yifan Liu, Yizhou Zhou, Fengyun Rao, Xingdong Sheng, Yunhui Liu, Wenjun Zeng, and Xiaokang Yang. Inter-x: Towards versatile human-human interaction analysis. In *CVPR*, 2024. 4

Phonology from SignBank

Gloss (Sign): KNOW
Initial Dominant Handshape: Good
Initial Primary Location: Upper face (forehead/temple)
Final Primary Location: Upper face (forehead/temple)
Initial Interacting Dominant Hand Part: Fingertips
Tags: Onehand



LLM input from phonology

Initial dominant hand shape: The thumb is extended from the fist.
Final dominant hand shape: The thumb is extended from the fist.
Initial dominant hand location: Upper face
Final dominant hand location: Upper face
Dominant hand part facing the upper face: Thumb fingertip
Uses the dominant hand only

LLM input from HandMotionScript

DOMINANT HAND DISTANCES:
Distance from dominant hand to head: [close]
Distance along x axis from dominant hand to head: [close, aligned]
Distance along y axis from dominant hand to head: [aligned]
Distance along z axis from dominant hand to head: [close / in front]
HAND ORIENTATIONS:
Palm orientation - dominant hand: [sideways]

LLM output:

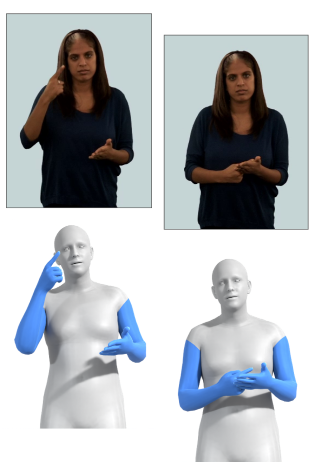
1. The dominant hand, shaped as a thumbs-up with the palm facing sideways, moves slightly horizontally across the upper face, keeping the thumb tip close to the face.

2. The person forms a thumbs-up with their dominant hand, positioning it near their upper face with the palm oriented sideways and the thumb tip pointing towards the face. They then slide the hand slightly sideways across the upper face.

3. Form a thumbs-up with your dominant hand. Place it near your upper face, palm facing sideways and thumb tip close to the skin. Move your hand in a small sideways motion across your upper face.

Phonology from SignBank

Gloss (Sign): DECIDE
Initial Dominant Handshape: Point
Initial Subordinate Handshape: Flat
Initial Primary Location: Upper face (forehead/temple)
Final Primary Location: Neutral space
Final Subordinate Location: Palm
Initial Interacting Dominant Hand Part: Fingertips
Tags: Dominant hand only, two handed



LLM input from phonology

Initial dominant hand shape: The index finger is extended from the fist.
Initial non-dominant hand shape: The fingers are extended and together.
Final dominant hand shape: The index finger is extended from the fist.
Final non-dominant hand shape: The fingers are extended and together.
Initial dominant hand location: Upper face
Final dominant hand location: In front of the person
Dominant hand part facing the upper face during the initial part of the motion: Index fingertip
Location on the dominant hand that the non-dominant hand is facing during the final part of the motion: Index fingertip
Initial location on the non-dominant hand that the dominant hand is facing: Fingertip(s)
Final location on the non-dominant hand that the dominant hand is facing: Palm side
Both hands are used, and their motions aren't symmetrical
Both hands are used, the dominant hand moves, the non-dominant hand is still

LLM input from HandMotionScript

DISTANCE BETWEEN HANDS:
Distance from dominant hand to non-dominant hand: [medium, touching]
DOMINANT HAND DISTANCES:
Distance from dominant hand to head: [close, touching, medium]
HAND ORIENTATIONS:
Palm orientation - dominant hand: [sideways, in]
Palm orientation - non-dominant hand: [up]

LLM output:

1. The non-dominant hand is held flat with the palm facing up. The dominant hand, with the index finger extended, starts by touching the upper face with the fingertip, palm sideways. It then moves away from the face and towards the non-dominant hand, turning inward to touch the non-dominant palm with the index fingertip.

2. Position the non-dominant hand flat, palm oriented upwards, and keep it stationary. The dominant hand, shaped with the index finger pointing out from a fist, initially makes contact with the upper face using the index fingertip, palm facing sideways. Then, the dominant hand travels towards the non-dominant hand, rotating its palm inward, and finishes by touching the palm of the non-dominant hand with the index fingertip.

3. Hold your non-dominant hand still and flat, palm up. Form a fist with your dominant hand and extend the index finger. Touch your upper face with the tip of this index finger, ensuring the palm faces sideways. Move the dominant hand away from your face towards your stationary non-dominant hand, turning the dominant palm to face inward as it moves, concluding the motion when the dominant index fingertip makes contact with the non-dominant palm.-dominant hand.

Figure A.9. **Fig. A.8 continued**

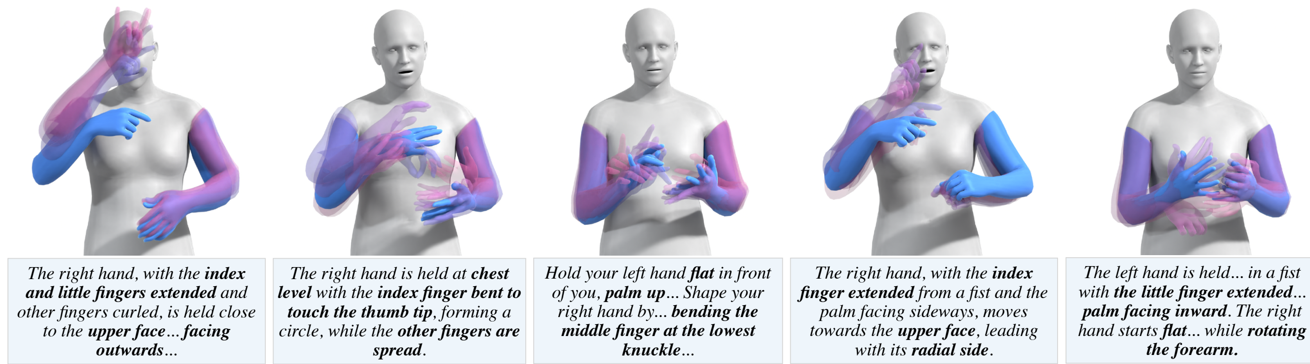


Figure A.10. **Dynamic visualization for Fig. 1 of the main paper:** We display the same examples as in Fig. 1 with a dynamic style including 5 frames evenly sampled. The color coding denotes the temporal evolution, i.e., the last frame with blue, and the first frame with decreased transparency in pink.

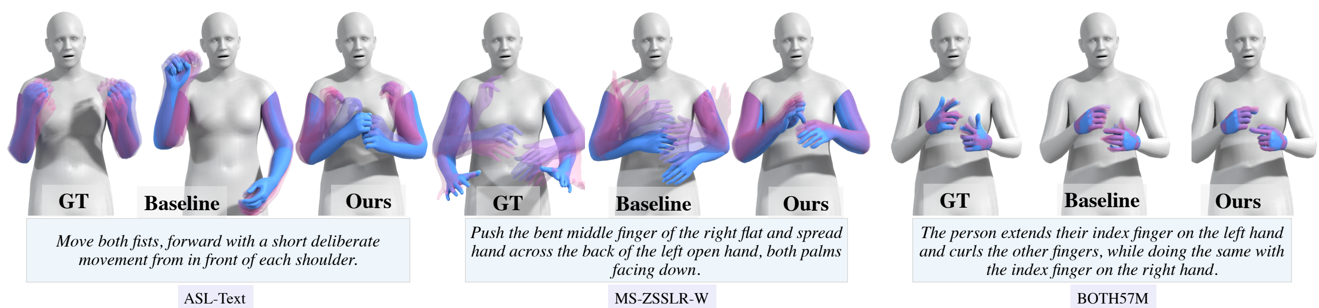


Figure A.11. **Dynamic visualization for Fig. 4 of the main paper.**

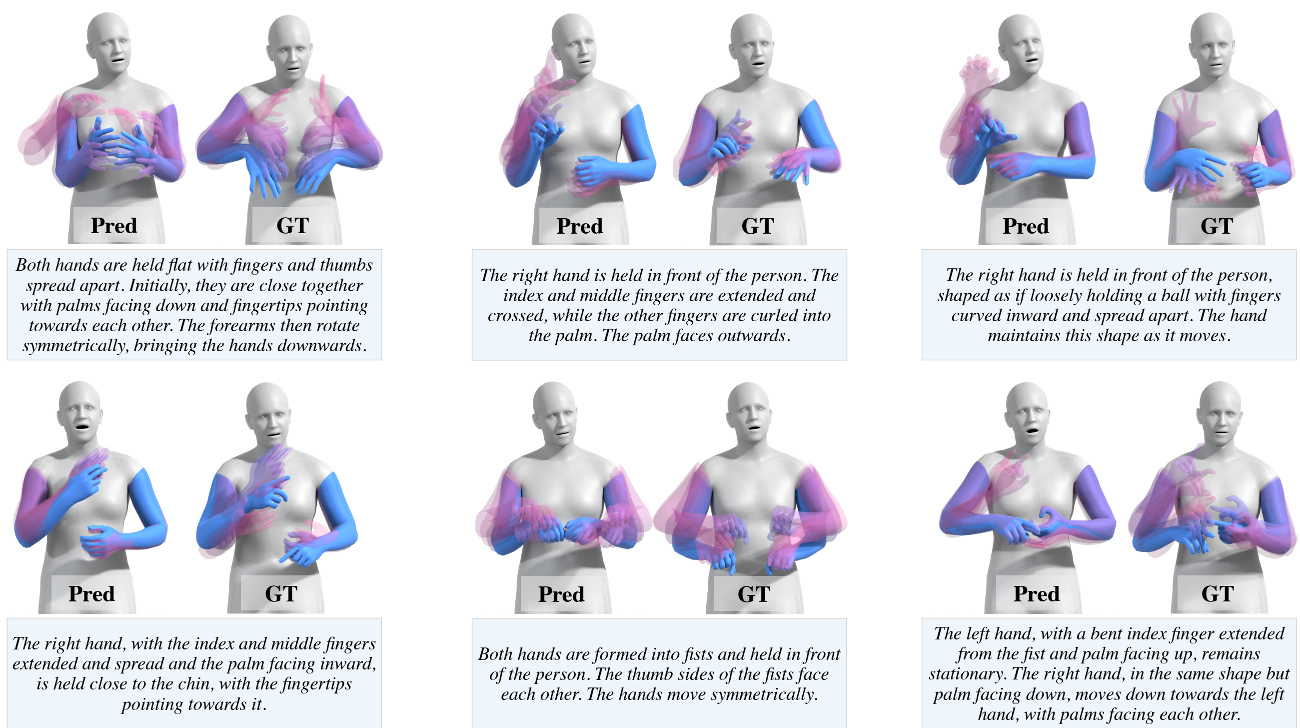


Figure A.12. **Dynamic visualization for Fig. 5 of the main paper.**

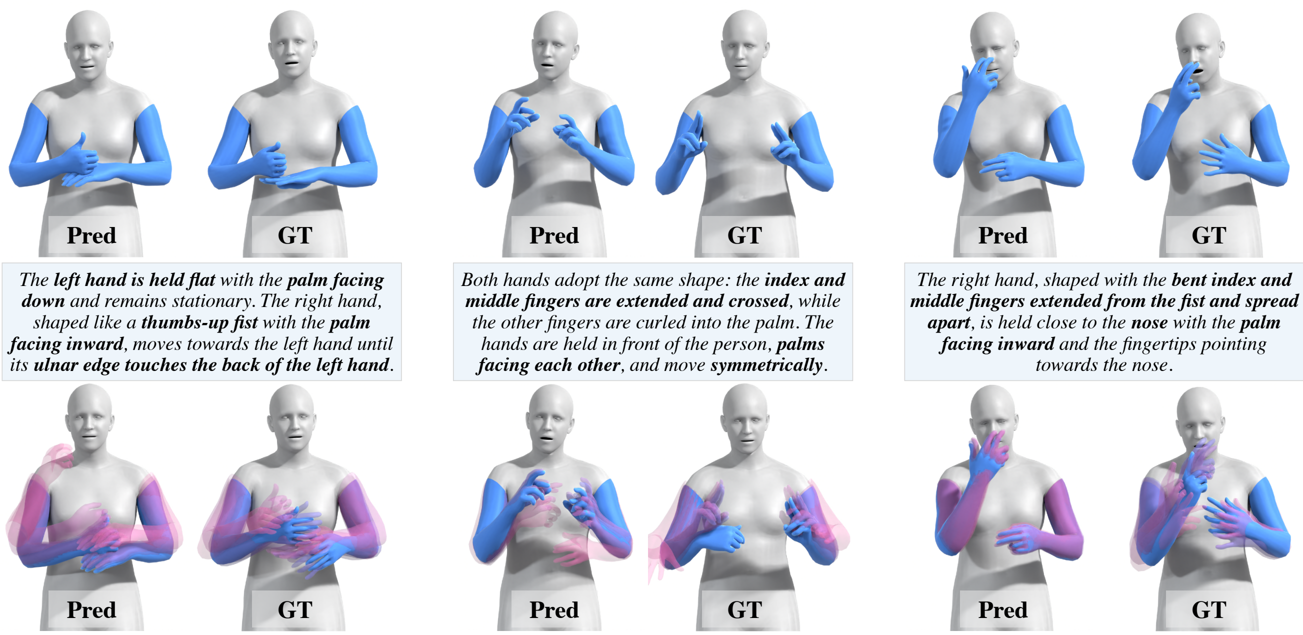


Figure A.13. **Qualitative results on the seen test set of BOBSL3DT-Test:** We complement Fig. 5

Algorithm 2 Prompt for LLM(Phonology + HMS) text generation.

```
You are an assistant whose
    role is to describe the hand motion for a sign in British Sign Language, given a list of information.
Use the information to describe the corresponding motion in 3 different ways.

In the ATTRIBUTES
    section, the given hand shapes concern the initial and final position. Having the information: 'Both
    hands are used, and their hand shapes are different', is not contradictory with the fact that their
    initial and final hand shapes are similar because they can be different in the middle of the motion.
In the DISTANCES sections:
- You are given temporal sequences of the distance between two limbs.
- The x-axis refers to the axis from left to right of the body, the y-axis refers to the axis
    from bottom to top, and the z-axis refers to the axis normal to the body and pointing away from it.
In the HAND ORIENTATIONS section:
- You are given temporal sequences of the palm orientations.

Use all the information in the ATTRIBUTES section unless you find that some points are contradictory.
Infer from the DISTANCES sections the
    directions and change of directions of the movements of the hands and include them in the description.
Infer from the HAND ORIENTATIONS section
    the main hand orientations of the hands during the movement and include them in the description.
Some points
    can be redundant, therefore summarize all the information into concise but exhaustive descriptions.
It is never the case that the dominant hand remains
    stationary. If you can't infer what it is doing from the information, just mention that it moves.

The 3 descriptions should describe the same motion using different wording.
Use a diversified vocabulary from one example to another.

Here are some examples. The comment section of the model answers, when present, is here for your reference
    as indications of the correct practices of the task, it is not to be included in your own output.

{
   {
      "role": "user",
      "content": "Use the information to describe the corresponding
    motion. The comment section of the model answers, when present, is here for your reference
    as indications of the correct practices of the task, it is not to be included in your own output.
      ATTRIBUTES:
      - Initial dominant hand shape: Pointing
      - Initial non-dominant hand shape: Pointing
      - Final dominant hand shape: Pointing
      - Final non-dominant hand shape: Pointing
      - Initial sign location: In front of the person
      - Final sign location: In front of the person
      - Location on the dominant hand that the non-dominant hand is facing: Broad side
      - Location on the non-dominant hand that the dominant hand is facing: Broad side
      - Both hands are used, with same hand shape on both hands
      - The movement is symmetrical, meaning both hands are doing the same motions at the same time
      DISTANCE BETWEEN HANDS:
      - Distance from dominant hand to non-dominant hand: [spread, close]
      - Distance along x axis from dominant hand to non-dominant hand: [spread, close]
      - Distance along y axis from dominant hand to non-dominant hand: [aligned]
      - Distance along z axis from dominant hand to non-dominant hand: [aligned]
      HAND ORIENTATIONS:
      - Palm orientation - dominant hand: [sideways]
      - Palm orientation - non-dominant hand: [sideways]"
   },
   {
      "role": "model",
      "content": {
          "Description 1": "Both hands,
    pointing and facing sideways, palm facing each other, and are moving toward each other horizontally.",
          "Description 2": " The person positions
    both hands so that the index fingers are extended and the hands are pointing sideways, with the
    palms turned inward to face each other. Then, they move their hands horizontally toward one another.",
```

Algorithm 3 Prompt for LLM(Phonology + HMS) text generation. (Algorithm 2 continued)

```
            "Description 3": " Extend both index fingers and place the hands
     with the palm facing each other. Then move your hands horizontally symmetrically toward one another.",
            "Comment": "When the hands are facing sideways, it is always towards one another, not
     towards the outside. Besides, the hands are never crossed when the location is in front of the person.
            They face each other on their broad side, meaning
     not on one of their edges. Frome these 3 points you can deduce that their palms are facing each other.
            When you can deduce this kind information, add it to at least one of the descriptions.
            The horizontal aspect of the motion is deduced
     from the fact that the only direction with a change in the distance between the hands is the x axis."
        }
    }
},
{
    {
        "role": "user",
        "content": "Use the information to describe the corresponding
     motion. The comment section of the model answers, when present, is here for your reference
     as indications of the correct practices of the task, it is not to be included in your own output.
        ATTRIBUTES:
        - Initial dominant hand shape: The index forms a hook.
     The thumb is extended and stuck against the index finger. The other fingers are curled on the palm.
        - Final dominant hand shape: The index forms a hook.
     The thumb is extended and stuck against the index finger. The other fingers are curled on the palm.
        - Dominant hand location: Chest level
        - Dominant hand part facing the chest: Hand edge on the thumb side
        - Uses the dominant hand only
        DOMINANT HAND DISTANCES:
        - Distance from dominant hand to torso: [close]
        - Distance along x axis from dominant hand to torso: [aligned]
        - Distance along y axis from dominant hand to torso: [aligned]
        - Distance along z axis from dominant hand to torso: [close / in front]
        HAND ORIENTATIONS:
        - Palm orientation - dominant hand: [sideways]"
    },
    {
        "role": "model",
        "content": {
            "Description 1": "The dominant hand is held shaped as a hook, with the index
     fingertip touching the thumb. It moves close to the chest, with its radial side facing the chest.",

      "Description 2": "The person holds his dominant hand in a hook shape, bringing the index fingertip to
     meet the thumb. The hand is positioned near the chest, with the thumb-side turned toward the chest.",
            "Description
     3": "Form your dominant hand into a hook. Extend the thumb against the index finger, and make their
     fingertips touch. Bring the hand close to your chest, keeping the thumb-side facing toward your chest."
        }
    }
},
{
    {
        "role": "user",
        "content": "Use the information to describe the corresponding
     motion. The comment section of the model answers, when present, is here for your reference
     as indications of the correct practices of the task, it is not to be included in your own output.
        ATTRIBUTES:
        - Initial dominant hand shape: The fingers are extended and together.
        - Initial non-dominant hand shape: The fingers are extended and together.
        - Final dominant hand shape: Round
        - Final non-dominant hand shape: Round
        - Sign location: Chest level
        - Dominant hand part facing the chest: Fingertip(s)
        - Subordinate hand part facing the chest: Fingertip(s)
        - Both hands are used, with same hand shape on both hands
        - Hand shapes change during the sign
        - The movement is symmetrical, meaning both hands are doing the same motions at the same time
```

Algorithm 4 Prompt for LLM(Phonology + HMS) text generation. (Algorithm 3 continued)

```
        DOMINANT HAND DISTANCES:
        - Distance from dominant hand to torso: [medium, close]
        NON-DOMINANT HAND DISTANCES:
        - Distance from non-dominant hand to torso: [medium, close]
        HAND ORIENTATIONS:
        - Palm orientation - dominant hand: [down]
        - Palm orientation - non-dominant hand: [down]"
    },
    {
        "role": "model",
        "content": {
            "Description 1": "The hands start flat with fingers
     together. They move with the palm facing down toward the chest, while closing into a round shape.",
            "Description 2": "The person begins with both hands flat and fingers held together. As the
     hands move inward toward the chest with palms facing downward, they gradually form a rounded shape.",
            "Description 3": "Start with your hands flat and fingers together. Move
     them toward your chest with palms facing down, and as you do, curve the hands into a round shape.",
            "Comment": ""The fact
     that the sign location is the chest doesn't mean that the hands are at the chest during the whole sign.
            It means that the hands
     are interacting with the chest, or positioned near the chest, at least at some point of the motion.
            Here, from
     the distance information, you can deduce that the motion consists in the hands coming to the chest."
        },
    }
},
{
    {
        "role": "user",
        "content": "Use the information to describe the corresponding
     motion. The comment section of the model answers, when present, is here for your reference
     as indications of the correct practices of the task, it is not to be included in your own output.
        ATTRIBUTES:
        - Initial dominant hand shape: Pointing
        - Initial non-dominant hand shape: The fingers are extended and together.
        - Final dominant hand shape: The hand is held flat with the fingers held together.
        - Final non-dominant hand shape: The fingers are extended and together.
        - Initial dominant hand location: Upper face
        - Final dominant hand location: In front of the person
        - Dominant hand part facing the upper face during the initial part of the motion: Index fingertip
        - Location on
     the dominant hand that the non-dominant hand is facing during the final part of the motion: Back side
        - Initial subordinate hand part facing the upper face: Fingertip(s)
        - Final subordinate hand part facing the upper face: Palm side
        - Forearm rotation
        - Both hands are used, and their motions aren't symmetrical
        - Hand shapes change during the sign
        DOMINANT HAND DISTANCES:
        - Distance from dominant hand to head: [close, touching, close, medium]
        DISTANCE BETWEEN HANDS:
        - Distance from dominant hand to non-dominant hand: [spread, medium, close]
        HAND ORIENTATIONS:
        - Palm orientation - dominant hand: [sideways, in]
        - Palm orientation - non-dominant hand: [up]
    },
    {
        "role": "model",
        "content": {
            "Description
     1": "The dominant touches the head with its extended index finger and the palm facing sideways.
     Then the hands move toward each other, both held flat, with the non-dominant hand turned upward.
     The motion ends with the back of the dominant hand turned towards the palm of the non-dominant hand.",
            "Description 2": "The individual touches her head with the pointed index finger
     of her dominant hand, with the palm facing sideways. Then, she moves both hands toward each other,
     holding them flat, with the dominant hand turning inward and the non-dominant hand turned upward,
     to end the motion with the back of the dominant hand facing the palm of the non-dominant hand.",
```

Algorithm 5 Prompt for LLM(Phonology + HMS) text generation. (Algorithm 4 continued)

```
        "Description 3": "Move your dominant index finger to your head. Then, bring
    both hands toward each other, keeping them flat, with the non-dominant hand palm facing up. Ensure
    to end the movement with the back of your dominant hand facing the palm of your non-dominant hand.",
        "Comment": "The distance sequence from the
    dominant hand to the head indicates that the motion starts near the head and then farther away from it.
        On
    the contrary, the distance sequence between the dominant hand and the non-dominant hand indicates that
    the dominant hand starts far away from the non-dominant hand and gets closer to it during the motion.
        From these 2 points you
    deduce that the dominant hand starts by touching the head, then that both hands move toward each other.
        All the points don't need to be included in each of the descriptions, especially when they seem
    redundant or necessary. For instance, if the dominant hand goes from the head to the non-dominant,
    that the non-dominant hand palm is facing up, and that the dominant hand meets the non-dominant
    hand palm with its back, then the dominant hand will necessarily face in during this motion."
      }
    }
},
{
   {
      "role": "user",
      "content": "Use the information to describe the corresponding
    motion. The comment section of the model answers, when present, is here for your reference
    as indications of the correct practices of the task, it is not to be included in your own output.
      ATTRIBUTES:
      - Initial dominant hand shape: The fingers are held together and tightly bent into a fist.
      - Initial non-dominant hand shape: The fingers are held together and tightly bent into a fist.
      - Final dominant hand shape: The fingers are held together and tightly bent into a fist.
      - Final non-dominant hand shape: The fingers are held together and tightly bent into a fist.
      - Initial sign location: Shoulder
      - Final sign location: Chest level
      - Dominant hand part facing the shoulder during the initial part of the motion: Dorsal side
      - Dominant hand part facing the chest during the final part of the motion: Dorsal side
      - Initial subordinate hand part facing the shoulder: Dorsal side
      - Final subordinate hand part facing the shoulder: Dorsal side
      - Both hands are used, with same hand shape on both hands
      - The movement is symmetrical, meaning both hands are doing the same motions at the same time
      DOMINANT HAND DISTANCES:
      - Distance from dominant hand to non-dominant shoulder: [medium]
      - Distance from dominant hand to dominant shoulder: [close]
      - Distance from dominant hand to torso: [close]
      NON-DOMINANT HAND DISTANCES:
      - Distance from non-dominant hand to dominant shoulder: [medium]
      - Distance from non-dominant hand to non-dominant shoulder: [close]
      - Distance from non-dominant hand to torso: [close]
      HAND ORIENTATIONS:
      - Palm orientation - dominant hand: [in]
      - Palm orientation - non-dominant hand: [in]"
   },
   {
      "role": "model",
      "content": {
        "Description
    1": "Both hands form a fist, and are turned inward. They move near their respective shoulders.",

      "Description 2": "The person bends all his fingers into a fist, turning them inward with the dorsal
    side of the hand facing the shoulders. He then moves his fists close to their respective shoulders.",
        "Description
    3": "Form a fist with both hands, turning them inward. Move your fists close to your shoulders.",
        "Comment": "When the shoulders are mentioned in the location, you
    need to deduce which shoulder each hand is located at, usually by checking which shoulder each hand is
    the closest to - here their respective shoulders. The hand part facing the shoulder sections and the
    palm orientation sections are redundant. In this case, alternate the vocabulary between descriptions."
      }
   }
},
```

Algorithm 6 Prompt for LLM(Phonology + HMS) text generation. (Algorithm 5 continued)

```
{
   {
       "role": "user",
       "content": "Use the information to describe the corresponding
    motion. The comment section of the model answers, when present, is here for your reference
    as indications of the correct practices of the task, it is not to be included in your own output.
       ATTRIBUTES:
       - Initial
    dominant hand shape: The little finger is extended, all the other fingers are curled on the palm
       - Initial non-dominant hand shape: The hand is held flat with the fingers held together.
       -
    Final dominant hand shape: The little finger is extended, all the other fingers are curled on the palm
       - Final non-dominant hand shape: The hand is held flat with the fingers held together.
       - Initial dominant hand location: In front of the person
       - Final dominant hand location: In front of the person
       - Location on the dominant hand that the non-dominant hand is facing: Ulnar side
       - Location on the non-dominant hand that the dominant hand is facing: Palm side
       - Forearm rotation
       - Both hands are used, and their motions aren't symmetrical
       - Both hands are used, the dominant hand moves, the non-dominant hand is still
       DISTANCE BETWEEN HANDS:
       - Distance from dominant hand to non-dominant hand: [close, medium]
       HAND ORIENTATIONS:
       - Palm orientation - dominant hand: [up, out]
       - Palm orientation - non-dominant hand: [up]"
   },
   {
       "role": "model",
       "content": {
           "Description 1": "The non-dominant hand is held flat
    with the palm facing up. The dominant hand, starting with the little finger extended from the fist,
    palm facing up and its ulnar side on the non-dominant hand, moves away from the non-dominant hand.",
           "Description 2": "The
    person holds their non-dominant hand flat with the palm facing upward. Their dominant hand begins in a
    fist with only the little finger extended, palm also facing up, and the little finger edge of the hand
    resting on the non-dominant palm. They then move the dominant hand away from the non-dominant hand.",
           "Description 3": "Hold the non-dominant hand flat with the palm facing upward. Form a
    fist with the dominant hand, extending only the little finger, and position it palm up with the ulnar
    edge resting on the non-dominant palm. Then move the dominant hand away from the non-dominant hand."
       }
   }
}
```