IMPERIAL

Signs as Tokens: A Retrieval-Enhanced Multilingual Sign Language Generator

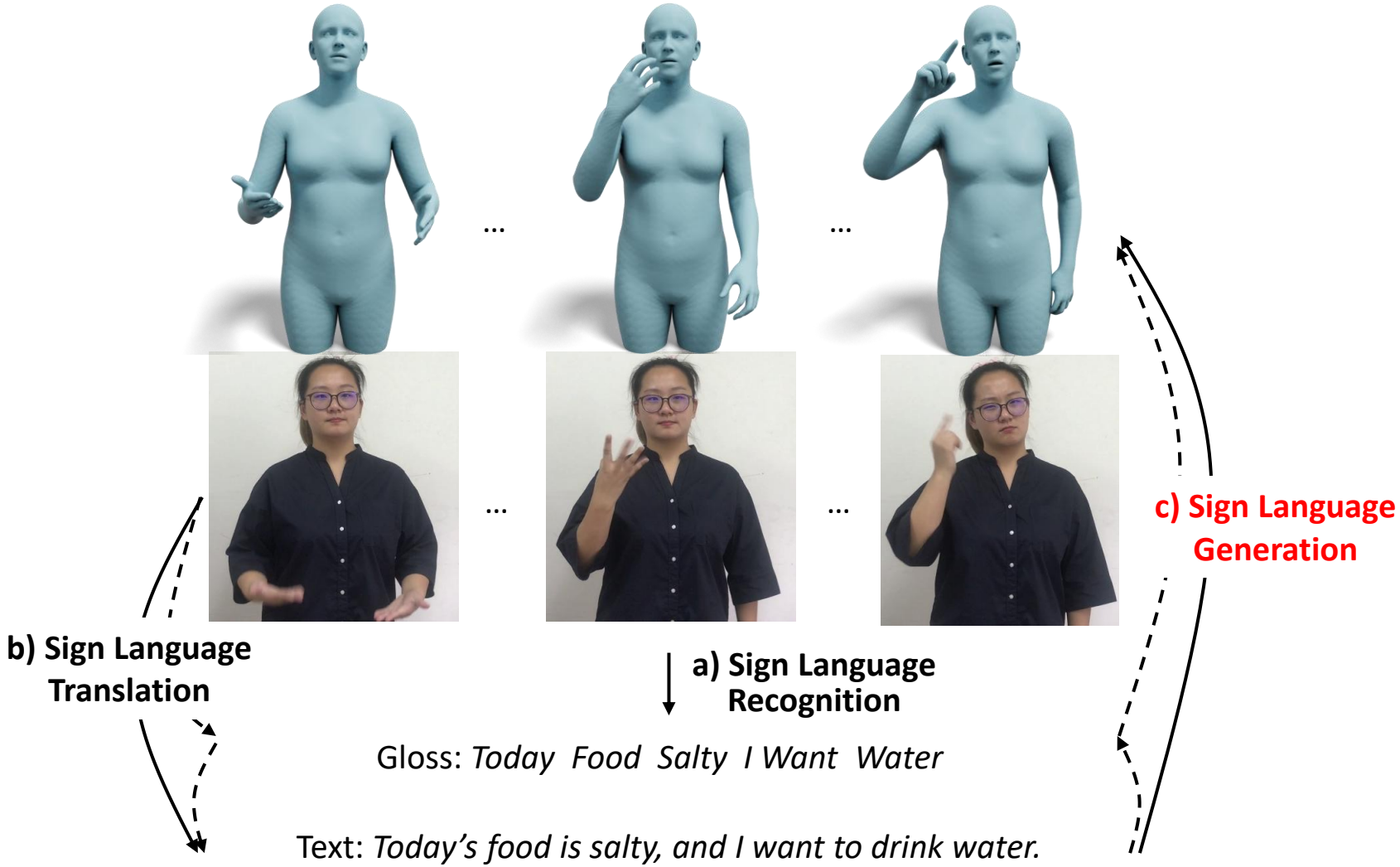
Ronglai Zuo, Rolandos Alexandros Potamias, Evangelos Ververas, Jiankang Deng, Stefanos Zafeiriou

Dr. Ronglai Zuo
Research Associate, Imperial College London
13 September 2025

Sign Language

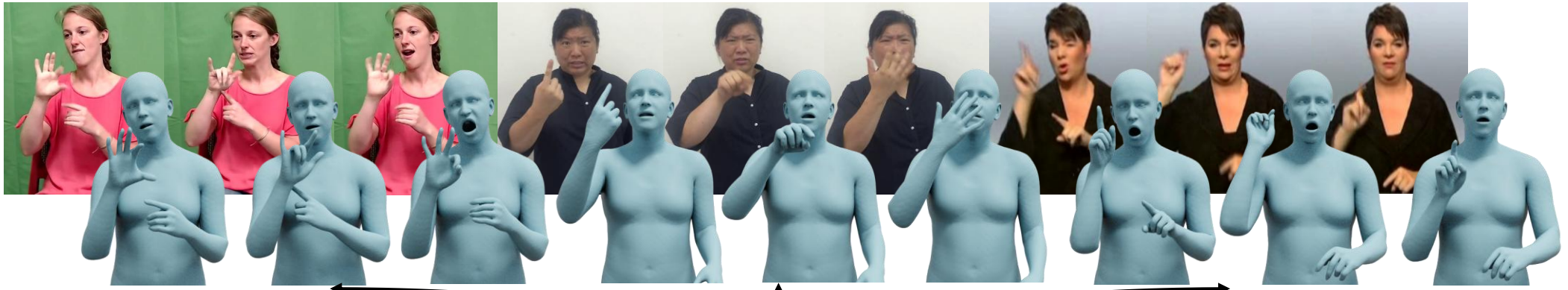
From World Federation of the Deaf, 2021

Sign Language Processing
Three Core Problems



Signs as Tokens (SOKE) [Zuo et al., ICCV25]

A Language Model-Based Solution

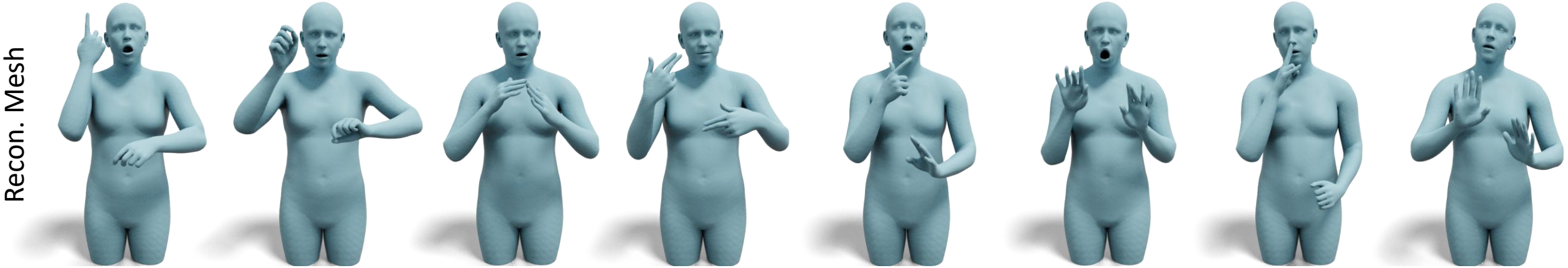


A Generalizable Pose Fitting Pipeline

OSX [Lin et al., CVPR'23] & WiLoR [Potamias et al., CVPR'25]

Ground Truth

Recon. Mesh



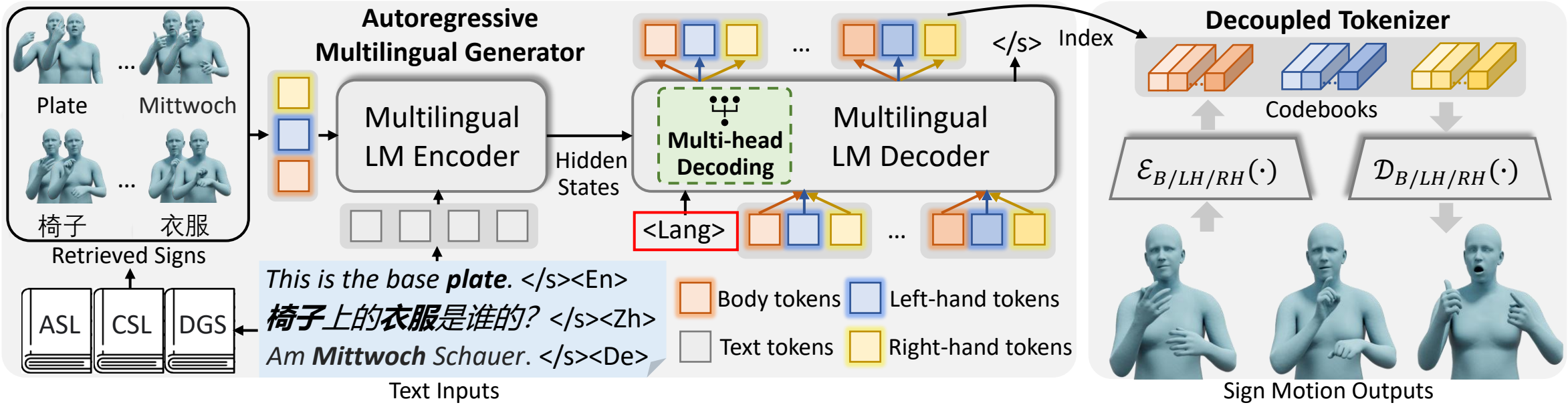
A Generalizable Pose Fitting Pipeline

OSX [Lin et al., CVPR'23] & WiLoR [Potamias et al., CVPR'25]

Method	Body↓	Left Hand↓	Right Hand↓
FrankMoCap [50]	78.07	20.47	19.62
PIXIE [14]	60.11	25.02	22.42
PyMAF-X [74]	68.61	21.46	19.19
SMPLify-X [45]	56.07	22.23	18.83
SGNify [15]	55.63	19.22	17.50
OSX [36]	47.32	18.34	18.12
NSA [3]	**46.42**	<u>16.17</u>	<u>15.23</u>
Ours	<u>46.73</u>	**10.55**	**8.94**

Excellent **zero-shot** performance on SGNify mocap dataset [Forte et al., CVPR23].

Framework Overview

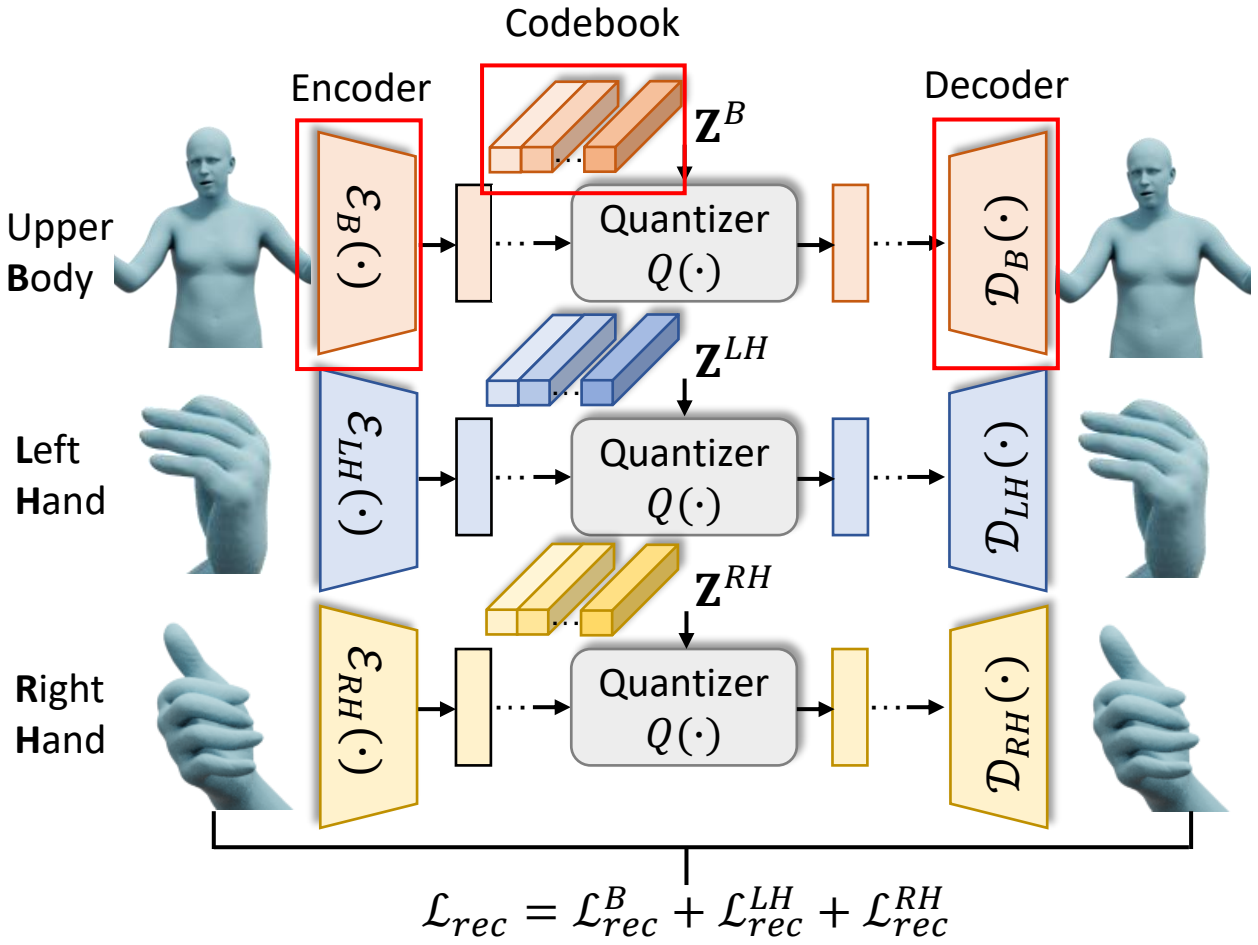


Base LM: mBART-large (0.6B)

Support multilingual SLG in one model via language tokens

Decoupled Tokenizer

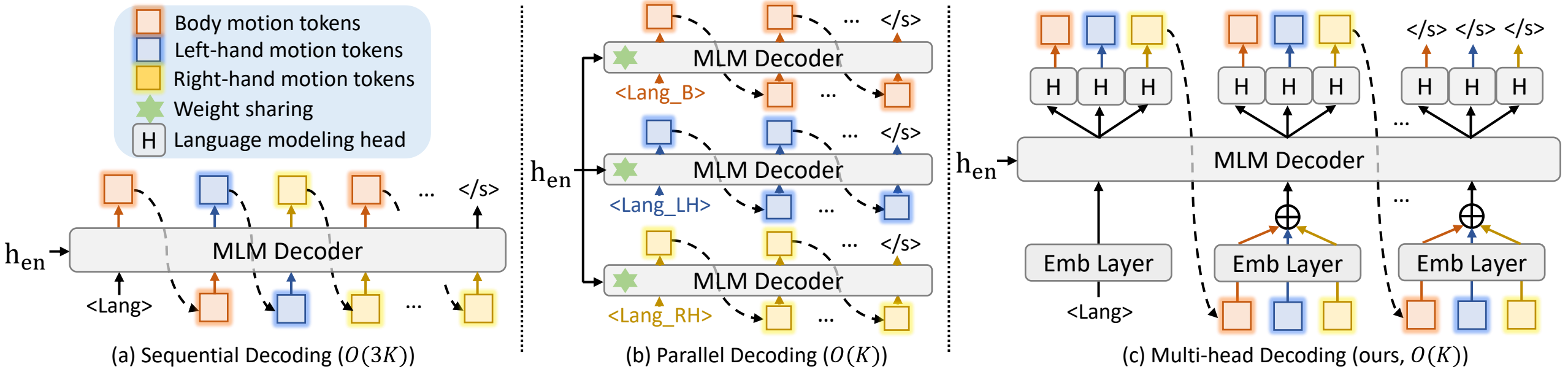
Discretize sign motions



Use $\boldsymbol{3M}$ tokens to represent $\boldsymbol{M^3}$ signs.

Multi-Head Decoding

Predict part-wise tokens at one step



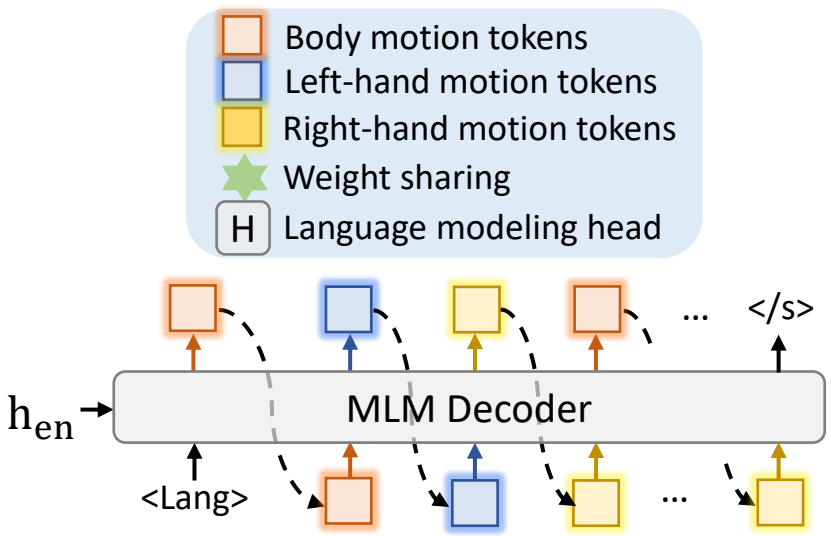
(a) Sequential Decoding ($O(3K)$)

(b) Parallel Decoding ($O(K)$)

(c) Multi-head Decoding (ours, $O(K)$)

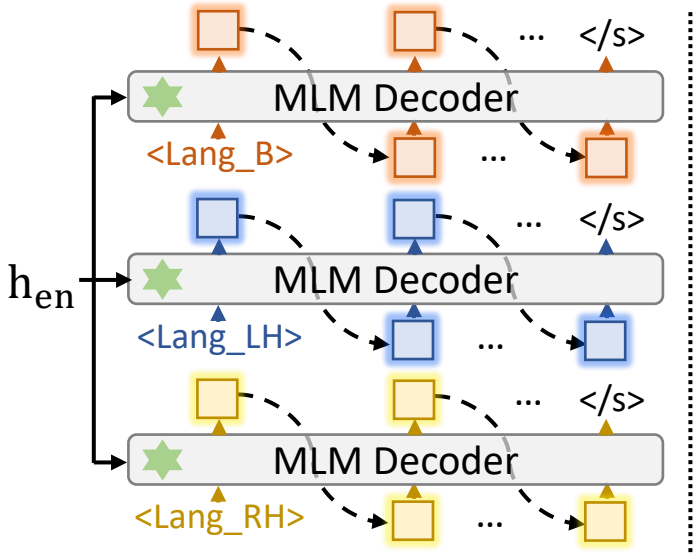
Multi-Head Decoding

Predict part-wise tokens at one step



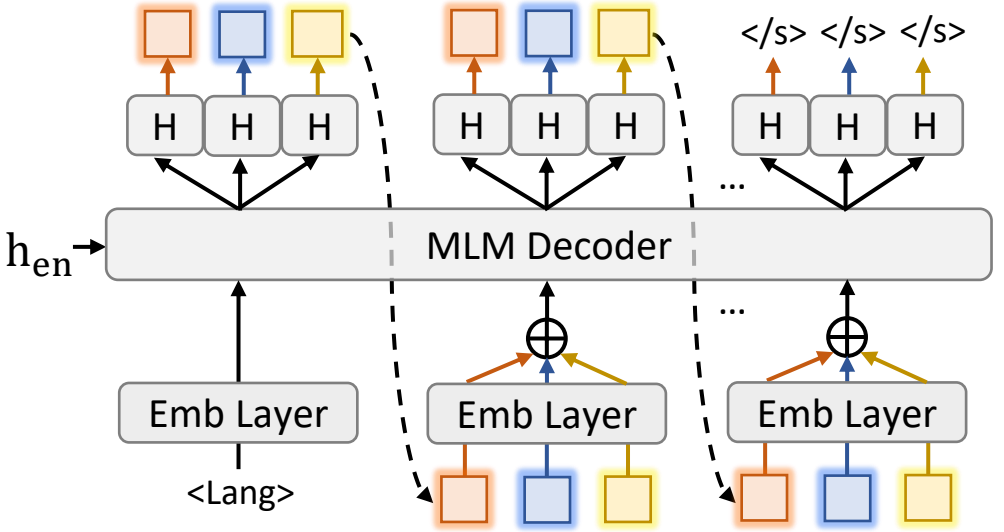
(a) Sequential Decoding ($O(3K)$)

3x decoding steps



(b) Parallel Decoding ($O(K)$)

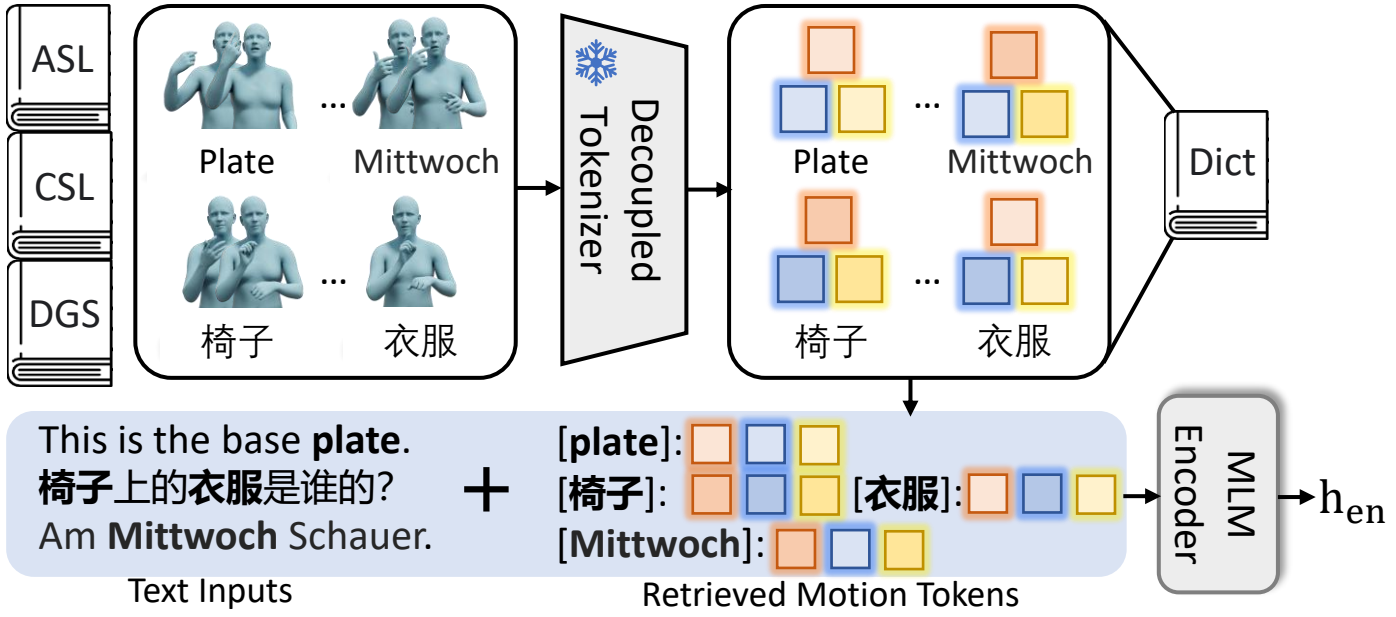
No interactions among body parts



(c) Multi-head Decoding (ours, $O(K)$)

Fewer steps and efficient interactions

Retrieval-Enhanced Generation



Retrieve external dictionaries to provide accurate signs.

Qualitative Results

Maybe someone else is better.

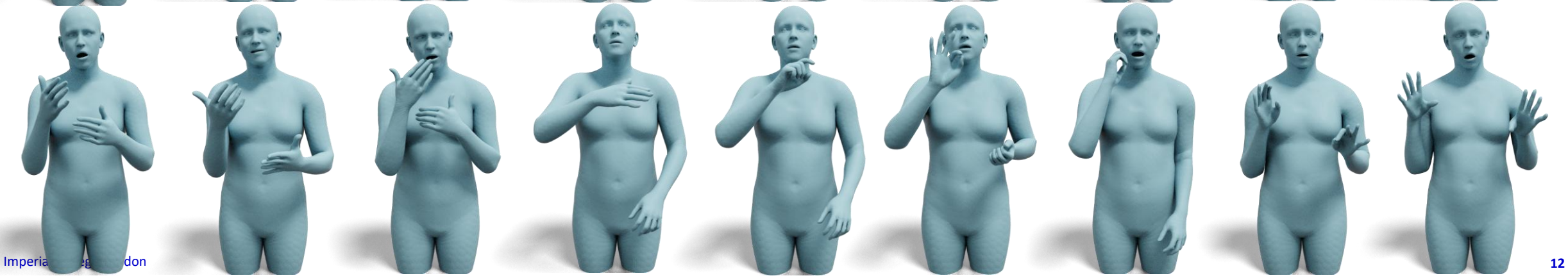
卫生间在哪里？ (Where is the toilet?)

Liebe zuschauer guten abend. (Dear viewers, good evening.)

Ground Truth

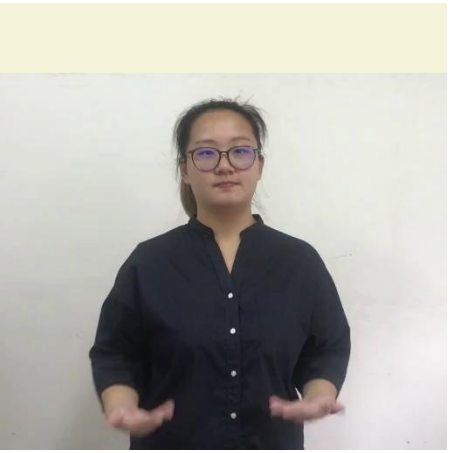
S-MotionGPT

SOKE (Ours)



Video Demos

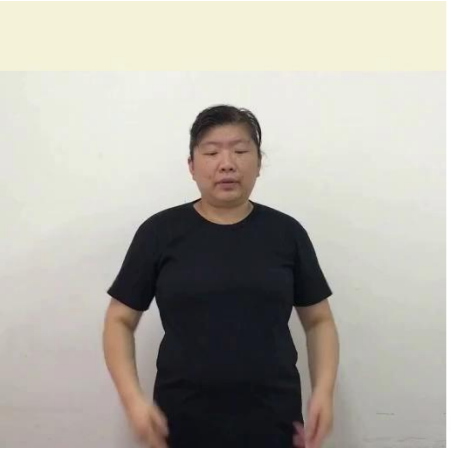
Ground Truth | S-MotionGPT | **SOKE (ours)**



Text: *今天的菜好咸。* *(Today's food is too salty.)*

Ground Truth | S-MotionGPT | **SOKE (ours)**

Text: *我爸爸当年差点考进大学。* *(My dad almost got into college.)*



Text: *今天我生病了。* *(I am sick today.)*

Text: *明天考试完了就放假。* *(We will have a holiday after the exam finishes tomorrow.)*

IMPERIAL

Thank you!

Signs as Tokens: A Retrieval-Enhanced Multilingual Sign Language Generator
13 September 2025