



MambaTalk: Efficient Holistic Gesture Synthesis with Selective State Space Models

Zunnan Xu, Yukang Lin, Haonan Han, Sicheng Yang, Ronghui Li, Yachao Zhang[†], Xiu Li[†]
Tsinghua University



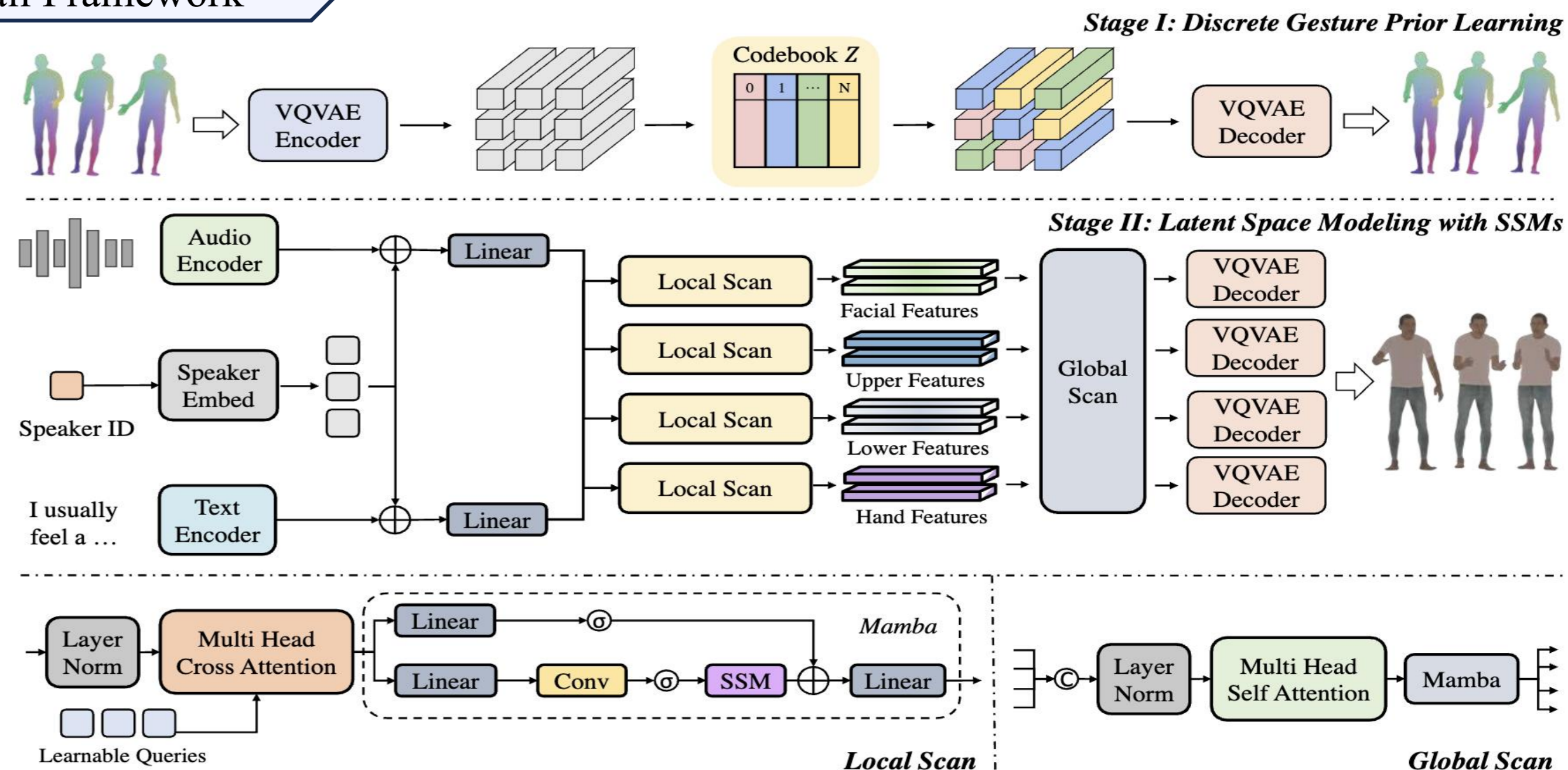
Introduction

In this study, we address the challenge of generating long and diverse gesture sequences with low latency in human-computer interaction. We introduce MambaTalk, a novel framework that leverages state space models to enhance gesture diversity and rhythm through multimodal integration, outperforming state-of-the-art models in gesture synthesis quality.

Contributions

- (1) We explore the potential of state space models (SSMs) in gesture synthesis, addressing the challenges of high computational complexity and unnatural jittering in generated gestures.
- (2) We introduce MambaTalk, a framework integrating hybrid scanning modules to refine latent space representations for gesture synthesis, capturing distinct movement patterns across various body parts.
- (3) Experiments demonstrate that our method surpasses state-of-the-art models in performance, achieving diverse and rhythmic gestures with low latency.

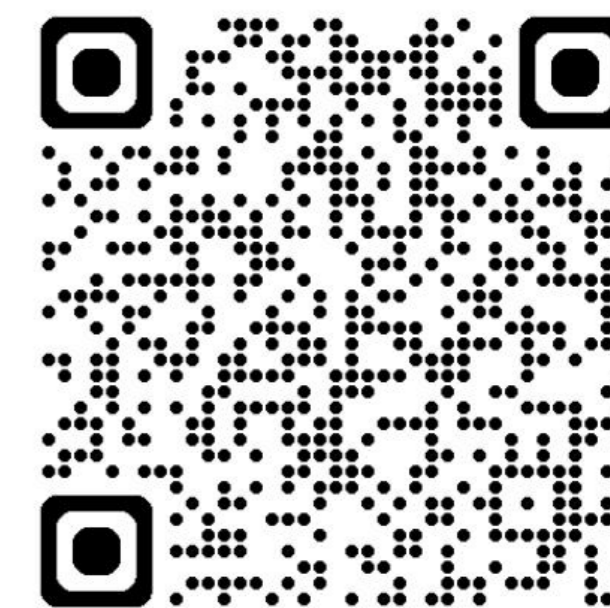
Overall Framework



Results on BEAT/BEATX

- (1) MambaTalk surpasses state-of-the-art models in gesture synthesis tasks, achieving a 14.3% increase in Beat Constancy and competitive Fréchet Gesture Distance scores, indicating superior synchronization and naturalness in generated gestures.
- (2) On the BEAT dataset, MambaTalk demonstrates a significant reduction in Fréchet Gesture Distance to 51.3 and enhances Semantic-Relevant Gesture Recall to 0.256, outperforming the baseline CaMN and other methods.

Scan me



Please scan the QR for our full paper.