Qeios

Peer Review

Review of: "HoloGest: Decoupled Diffusion and Motion Priors for Generating Holisticly Expressive Co-speech Gestures"

Abdallah Hussein Sham[1]

1. Tallinn University, Estonia

This paper introduces an innovative approach that combines decoupled diffusion models with motion priors to tackle critical issues like unnatural jittering, floating, and sliding observed in co-speech gesture generation. While the approach is promising, several areas require further clarification:

1. The Semantic Alignment (SA) metric described in Equation 8 relies on measuring cosine similarity between gesture and text embeddings. What methods were used to validate this metric in relation to human-perceived alignment?

2. What distinguishing features of DSG (Dynamic Speech Gesture) and EMAGE (Embodied and Multimodal Generative Architectures) should be emphasized to support their selection as baseline methods?

3. Regarding intricate mappings between speech and gestures, how does the current model maintain local semantic coherence without any direct interaction between the models?

4. The experimental results only cover the BEATX dataset. Can you also include findings from other datasets?

5. It would be beneficial to enumerate the key contributions of the study clearly in the introduction.

6. The explanation of Equation 1 in the methods section is missing context in the main text.

7. The manuscript currently overlooks the limitations of this work; an explicit acknowledgment of these limitations is necessary.

doi.org/10.32388/1Z46N2

Declarations

Potential competing interests: No potential competing interests to declare.