Large-Scale Multi-Character Interaction Synthesis Supplementary Material

Ziyi Chang
ziyi.chang@durham.ac.uk
Durham University
Durham, United Kingdom

He Wang
he_wang@ucl.ac.uk
University College London
London, United Kingdom

George Alex Koulieris
georgios.a.koulieris@durham.ac.uk
Durham University
Durham, United Kingdom

Hubert P. H. Shum*
hubert.shum@durham.ac.uk
Durham University
Durham, United Kingdom

1 TWO-CHARACTER GROUP DIVISION CHOICE

We clarify that our two-character interaction setup is a modeling choice driven by two factors: (1) Unlike prior work that loosely couples single-character motions, we focus on close and continuous interactions (CCI) with frequent, prolonged contact (e.g., dancing), which rarely involve more than two people. (2) Data availability is a major constraint. Capturing two-person CCI is already rare, and, to our knowledge, no datasets exist for multi-person CCI.

Our model is not inherently limited to two-person interactions. To demonstrate this, we adapt our method using Multi-Track Timeline Control [Petrovich et al. 2024], allowing a character to be in multiple groups (e.g., A-B, A-C) with minimal modifications: diffusion model taking two interaction groups as input and further constraining distances among three individuals. Results are shown in Table 1:

Table 1: **Allowing three-character division choice.**

	TS	HD
Two-character division	0.071	1.963
Allowing three-character division	0.079	1.888

This confirms our method's high extensibility with minimal modifications, enabling diverse scenarios and demonstrating potential for broader multi-character interaction synthesis.

2 EVALUATION METRICS

2.1 User Study

We provide a small user study where participants were asked to rank the quality of results generated by our method against InterGen and InterGen†. 94.12% prefers ours to the other two, 5.88% prefers InterGen†, and InterGen is unfavorable due to heavy character overlap.

2.2 Quantitative Metircs

Quantitative metrics remain an open challenge. Following common practice, we measure variance as the diversity metric for dancing, with and without planning. Table 2 shows the diversity results.

*Corresponding author.

Table 2: **Diversity as a quantitative metric.**

	Coordinatable Space	Planning	Diversity
InterGen (dancing)	NA	NA	4.938
Ours	✓	×	5.001
Ours	✓	✓	5.010

We attribute the slightly increased diversity to our coordinatable space and planning for synthesizing transitions that are absent in the original two-character data.

Essentially, without directly comparable ground truth, only self-contained metrics can be calculated or measured via quality assessment [Zhou et al. 2023]. Our TS metric has already integrated speed information by considering acceleration changes, which are crucial for physical plausibility and perceptual quality. Biomechanically, smooth acceleration results from the gradual modulation of neural signals regulating muscle force, aligning with Newton's second law.

3 HYPER-PARAMETER SETTINGS

Hyper-parameters related to the reinforcement learning part have been shown in Table 3.

Table 3: **Hyperparameters**

Hyperparameter	Value
batch size	16
gamma	0.98
epsilon start	0.08
epsilon end	0.01
epsilon decay	200
lr	0.0005

DQN strategy is used to train the transition planning network. We used the official diffusion model checkpoints and trained on an NVIDIA RTX 2080 Ti.

4 DISCUSSION ON BIASES, RISKS, AND POTENTIAL MITIGATION

Potential biases in motion synthesis can stem from unrepresentative datasets. Some datasets may fail to capture the inherently rich variability of human motions and interactions performed by people from different cultures, genders, and ages. The generated motions and interactions usually concentrate on dominant patterns while under-representing minorities. This limitation not only risks reinforcing stereotypical portrayals and cultural insensitivity but also raises privacy concerns, as models trained on data with skewed distributions may expose sensitive personal information from motion patterns such as genders and ages. Moreover, the under-representation of minority groups can exacerbate the issue, resulting in a loop where certain gestures or movements are insufficiently recognized or reproduced, thereby perpetuating social biases and limiting the generalizability of synthetic models.

To mitigate these challenges, several solutions can be potentially applied. For example, data augmentation for interactions [Li et al. 2024] expands and diversifies training data to better represent the full spectrum of motions and interactions. Another promising solution is diffusion minority sampling [Um et al. 2024]. Sampling strategies have been designed to promote minority patterns contained in the training data to receive adequate emphasis during the generation process. These techniques aim to enhance the fairness of motion synthesis systems, ultimately fostering more inclusive and ethically responsible applications in areas ranging from entertainment and virtual reality to rehabilitation and robotics.

5 ABLATION STUDIES ON METHOD SCALING

We performed an ablation study by removing the proposed components to showcase the scaling ability. Table 4 shows the results.

Table 4: **An ablation study on method scaling.**

	Coordinatable Space	Planning	TS	HD
a	×	×	0.071	0.564
b	✓	×	0.202	3.422
Full	✓	✓	0.075	3.372

6 ABLATION STUDIES ON DISTANCE THRESHOLD

We performed an ablation study by changing the distance threshold to demonstrate its effect. The results are shown in Table 5.

7 VISUALIZATION ON THE EFFECTS OF TRANSITION SMOOTHNESS METRIC

We provide an example as an illustration of the effects of the transition smoothness metric. Generally speaking, higher values indicate larger discontinuities like abrupt movements, which is indicated by the orange color in Figure 1.

Table 5: **an ablation study on the distance threshold.**

Distance threshold	TS	HD
0.5	0.059	1.723
1.0	0.057	1.841
1.5	0.067	1.901
2.0	0.071	1.963
2.5	0.085	2.003
3.0	0.092	2.355

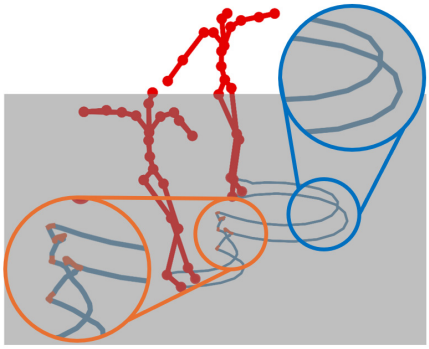

Figure 1: **An illustrative figure for the effects of transition smoothness metric. Discontinuities in trajectories are highlighted in orange color.**

REFERENCES

Baiyi Li, Edmond SL Ho, Hubert PH Shum, and He Wang. 2024. Two-person interaction augmentation with skeleton priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1900–1910.

Mathis Petrovich, Or Litany, Umar Iqbal, Michael J Black, Gul Varol, Xue Bin Peng, and Davis Rempe. 2024. Multi-track timeline control for text-driven 3D human motion generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1911–1921.

Soobin Um, Suhyeon Lee, and Jong Chul Ye. 2024. Don't Play Favorites: Minority Guidance for Diffusion Models. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=3NmO9lY4Jn

Kanglei Zhou, Yue Ma, Hubert PH Shum, and Xiaohui Liang. 2023. Hierarchical graph convolutional networks for action quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 12 (2023), 7749–7763.