

Style-Controllable Speech-Driven Gesture Synthesis Using Normalising Flows: Erratum

Simon Alexanderson^{ID}, Gustav Eje Henter^{ID}, Taras Kucherenko^{ID} and Jonas Beskow^{ID}

Division of Speech, Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

Abstract

This page contains an erratum for the article “Style-controllable speech-driven gesture synthesis using normalising flows” presented at EUROGRAPHICS 2020.

After the review process was completed, but shortly before the camera-ready version was due, we discovered an error in the extraction of control features for system MG-V in our paper. The error involves computing finite differences between two different coordinates in the same time frame, instead of the same coordinates across time as intended. The result is that the MG-V system does not control average hand-motion speed, but instead forces the average position of affected joints towards an oblique surface determined by the control. This effect can also be seen in our videos of the MG-V system.

The error does not affect conclusions regarding naturalness nor controllability; only the interpretation of the control exerted by system MG-V differs from what the article states. Consequently, we have elected to not to change the main article from the version approved by the peer review, and instead issue this erratum.

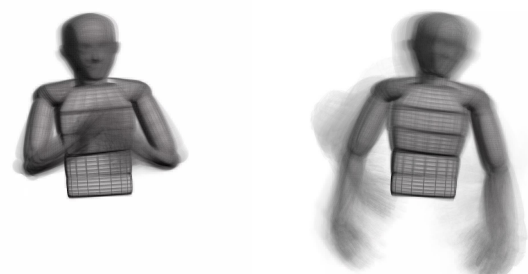
Moreover, we have re-trained the MG-V system with corrected speed-feature extraction. We will refer to this system as MG-VC (‘C’ for “corrected”). Plots of the effect of style control for this system are reproduced in the adjacent Fig. 6. We have also included videos of system MG-VC in the folder `mg-vc` of the supplemental material. The reader is invited to judge for themselves whether the speed control now is more convincing than before.

For the record, the incorrect code used for system MG-V reads as follows:

```
dr=np.linalg.norm(np.diff(rr), axis=1)
dl=np.linalg.norm(np.diff(ll), axis=1)
speed = dr+dl
```

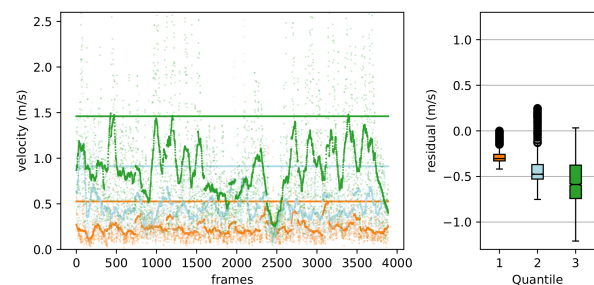
The corrected code (for system MG-VC) is:

```
dr=np.linalg.norm(np.diff(rr, axis=0), axis=1)
dl=np.linalg.norm(np.diff(ll, axis=0), axis=1)
speed = dr+dl
```



(a) Low average speed.

(b) High average speed.



(c) Time series and residual statistics of speed control.

Figure 6: Images illustrating the effect and accuracy of motion speed control for the corrected system MG-VC. It appears that the control has the intended effect of changing motion speed, even if the generated motion speeds do not centre on the intended mean value.