POSE AUGMENTATION: MIRROR THE RIGHT WAY

Jonathan Windle[1] Sarah Taylor[1] David Greenwood[1] Iain Matthews[1]

INTRODUCTION

A common challenge faced during speech-to-motion generation is acquiring sufficient data.

The ability to effectively double the amount of data using a lateral mirroring technique is attractive.

We introduce an identity embedding and compare naïve mirroring augmentation using the same speaker identity against using a new, ***virtual*** identity.

We evaluate the use of this *virtual* identity technique against the same quantity of real data.

EXPERIMENTAL SETUP

We train an audio speech-to-motion LSTM-based model on different splits of data using various augmentation settings defined as follows:

All Data: We use all available training data with no additional augmentation.

Half Data: We randomly subsample the training data to reduce the number of samples by approximately 50%.

Mirrored Virtual Identity: We augment the *Half Data* training set by laterally mirroring the pose at each frame and assigning a **new, *virtual*** identity label to the mirrored data.

Mirrored Same Identity: We augment the *Half Data* training set by laterally mirroring the pose at each frame and assigning the mirrored data the **same** identity label as the original speaker data.

All Data Mirrored Virtual Identity: All available training data plus the laterally mirrored augmentation with the *Virtual Identity* setting.

RESULTS

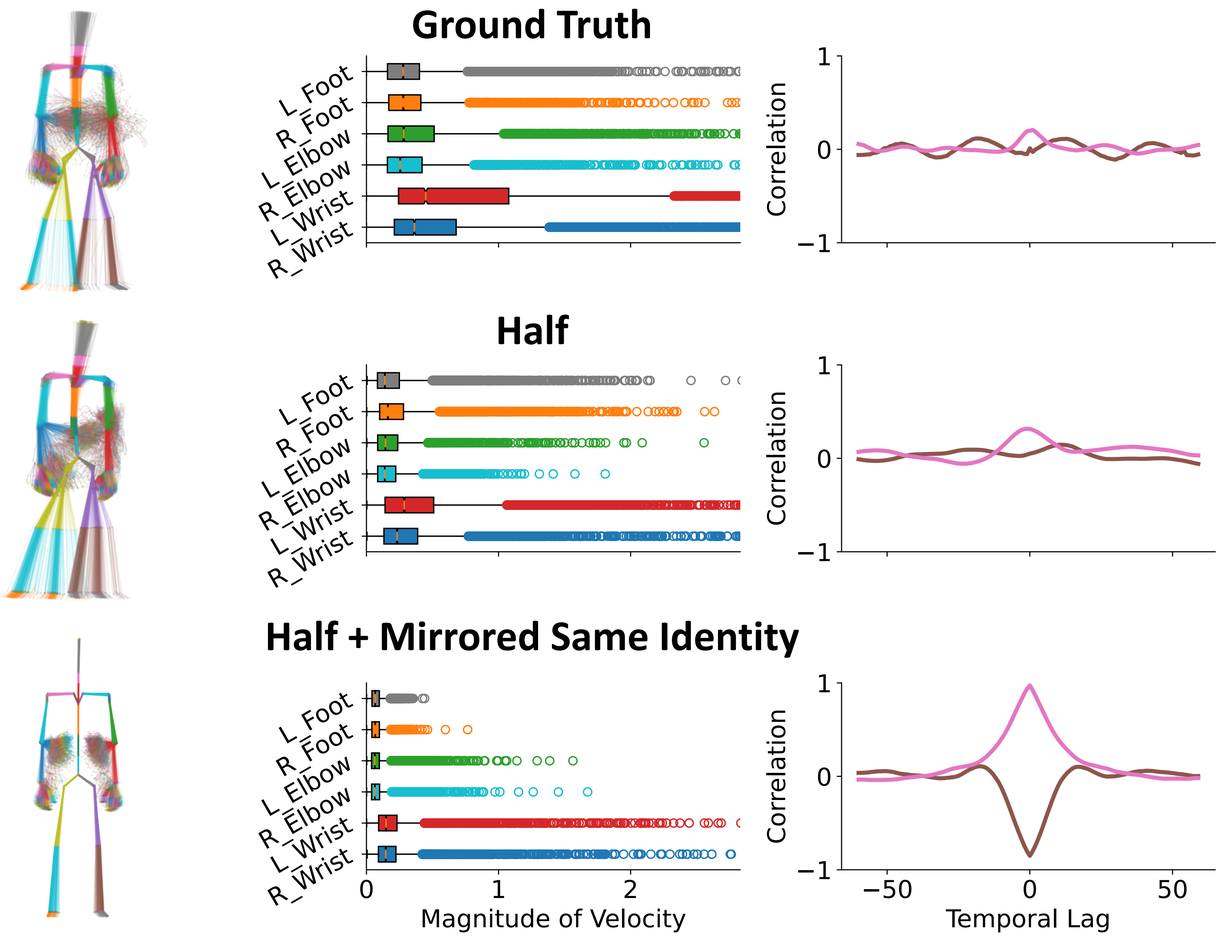

Figure 1: Comparison of a single speaker's generated motion showing the detrimental impact of including mirrored motion under the same identity.
Column (a) is the orthographic projection of poses at every second showing high positional symmetry when mirrored under the same identity.
Column (b) is the distribution of velocity magnitude; this is how far a joint moves between two frames showing symmetrical amounts of movement on both sides. The cross-correlation lags between the onset of left wrist motion given the right wrist motion in column (c) show high temporal symmetry when including mirrored motion under the same identity.

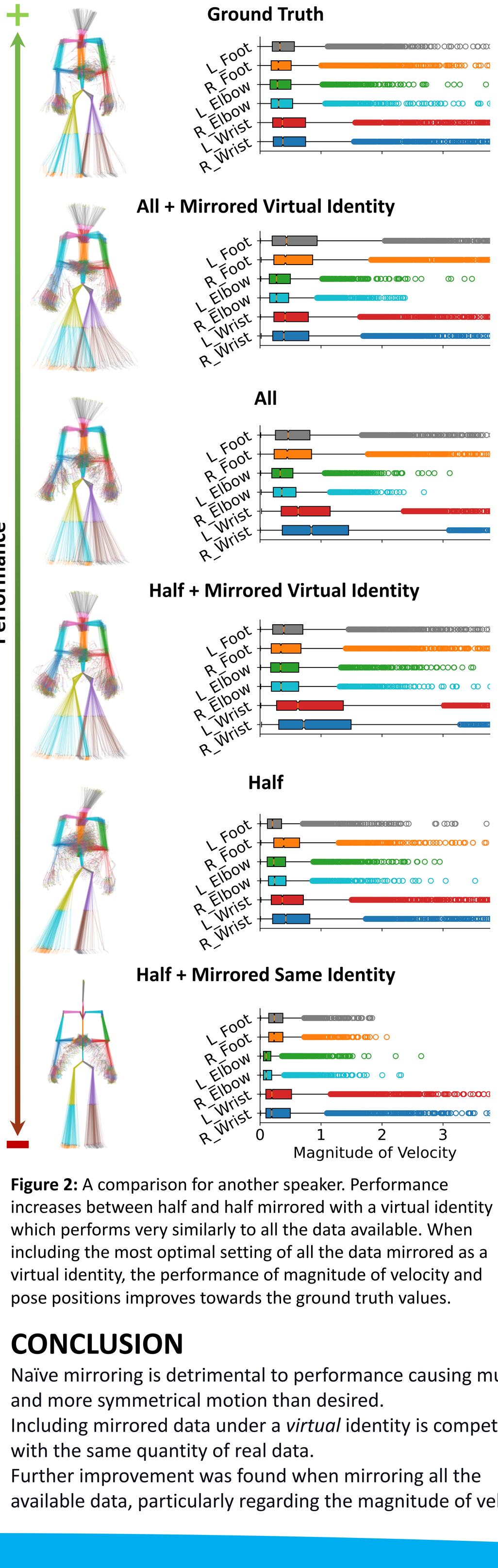

Figure 2: A comparison for another speaker. Performance increases between half and half mirrored with a virtual identity which performs very similarly to all the data available. When including the most optimal setting of all the data mirrored as a virtual identity, the performance of magnitude of velocity and pose positions improves towards the ground truth values.

CONCLUSION

Naïve mirroring is detrimental to performance causing muted and more symmetrical motion than desired.

Including mirrored data under a *virtual* identity is competitive with the same quantity of real data.

Further improvement was found when mirroring all the available data, particularly regarding the magnitude of velocity.

University of East Anglia

[1] University of East Anglia