REFERENCES

[1] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2001. Beat: the behavior expression animation toolkit. In *Annual Conference on Computer Graphics and Interactive Techniques*.

[2] Chung-Cheng Chiu and Stacy Marsella. 2011. How to train your avatar: A data driven approach to gesture generation. In *Proc. International Workshop on Intelligent Virtual Agents*. 127–140.

[3] Chung-Cheng Chiu, Louis-Philippe Morency, and Stacy Marsella. 2015. Predicting co-verbal gestures: a deep and temporal modeling approach. In *International Conference on Intelligent Virtual Agents*.

[4] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: encoder–decoder approaches. *Syntax, Semantics and Structure in Statistical Translation* (2014), 103.

[5] Dai Hasegawa, Naoshi Kaneko, Shinichi Shirakawa, Hiroshi Sakuta, and Kazuhiko Sumi. 2018. Evaluation of speech-to-gesture generation using bi-directional LSTM network. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, 79–86.

[6] Chien-Ming Huang and Bilge Mutlu. 2012. Robot behavior toolkit: generating effective social behaviors for robots. In *ACM/IEEE International Conference on Human Robot Interaction*.

[7] Mark L. Knapp, Judith A. Hall, and Terrence G. Horgan. 2013. *Nonverbal Communication in Human Interaction*. Wadsworth, Cengage Learning.

[8] David Matsumoto, Mark G. Frank, and Hyi Sung Hwang. 2013. *Nonverbal Communication: Science and Applications*. Sage.

[9] David McNeill. 1992. *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago press.

[10] Victor Ng-Thow-Hing, Pengcheng Luo, and Sandra Okita. 2010. Synchronized gesture and speech production for humanoid robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[11] Najmeh Sadoughi and Carlos Busso. 2017. Speech-driven animation with meaningful behaviors. *arXiv preprint arXiv:1708.01640* (2017).

[12] Kenta Takeuchi, Souichirou Kubota, Keisuke Suzuki, Dai Hasegawa, and Hiroshi Sakuta. 2017. Creating a gesture-speech dataset for speech-based automatic gesture generation. In *International Conference on Human-Computer Interaction*. Springer, 198–202.

[13] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research* 11, Dec (2010), 3371–3408.