Speech2Properties2Gestures: Gesture-Property Prediction as a Tool for Generating Representational Gestures from Speech

Taras Kucherenko*; Rajmund Nagy†; Patrik Jonell†; Michael Neff‡; Hedvig Kjellström*; Gustav Eje Henter†

*KTH Royal Institute of Technology, RPL; †KTH Royal Institute of Technology, TMH; ‡University of California, Davis
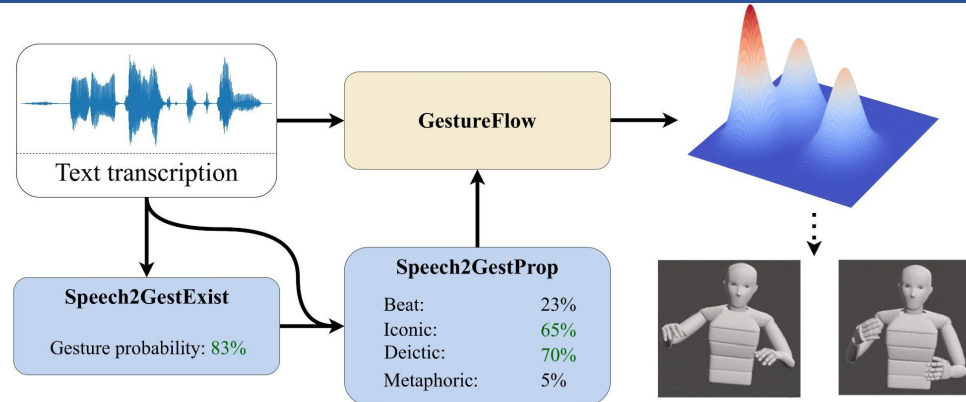
Summary

We outline a three-step gesture generation system that:
1) **predicts when to gesture**
2) **predicts how to gesture**
3) **synthesizes appropriate motion** for the predicted properties.

Problem setting

We evaluate the feasibility of predicting the timing of gestures and various properties of the motion.

Dataset

We utilize the Bielefeld SaGA dataset [1], containing audio/video recordings of 25 participants in a direction-giving scenario with rich **gesture-property annotations.**

Initial results

→ Our experiments show that it is possible to predict various aspects of the gesturing motion.

→ This is surprising, as gestures are highly stochastic and idiosyncratic.

→ Future work will focus on the efficacy of conditioning data-driven systems on these properties.

Label	Gesture category [Macro F$_1$]				Gesture semantics [Macro F$_1$]			
	deictic	beat	iconic	discourse	amount	shape	direction	size
Relative frequency	29.05%	14.47%	72.03%	12.78%	4.7%	13.1%	13.7%	1.9%
RandomGuess	50% ± 2%	50% ± 2%	50% ± 1.5%	50% ± 2%	49% ± 1%	49% ± 2%	49% ± 2%	50% ± 1%
ProposedModel	60% ± 6%	53% ± 6%	63% ± 5%	59% ± 7%	63% ± 8%	65% ± 6%	62% ± 8%	59% ± 9%

project page with the follow up work

Contact

Taras Kucherenko
KTH Royal Institute of Technology,
Stockholm, Sweden
tarask@kth.se

References

1. Andy Lücking, Kirsten Bergman, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2013. Data-based analysis of speech and gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications. Journal on Multimodal User Interfaces 7, 1 (2013), 5–18