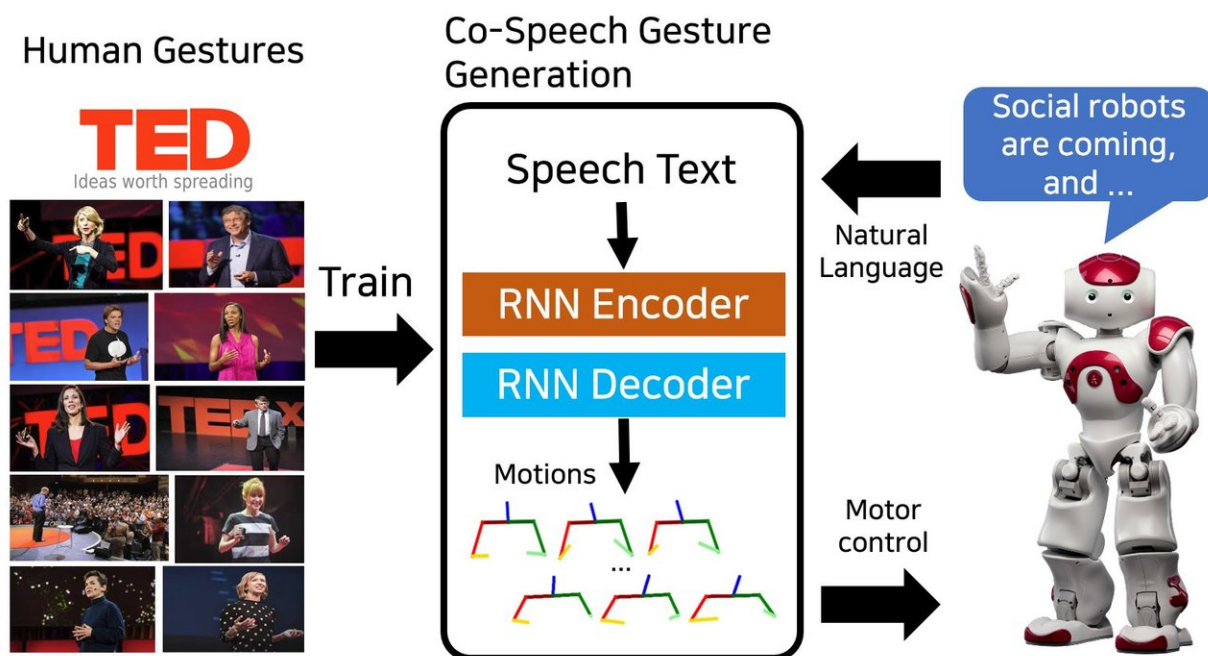


End-to-end learning of co-speech gesture generation for humanoid robots

November 14 2018, by Ingrid Fadelli



The proposed model generates a sequence of upper body poses and is trained on human gestures from TED talks. Credit: Yoon et al.

Researchers at the Electronics and Telecommunications Research Institute (ETRI) in South Korea have recently developed a neural network model that can generate sequences of co-speech gestures. Their model, trained on 52 hours of TED talks, successfully produced human-like gestures that matched speech content.

"Smart devices we are interacting with have evolved from personal computers to mobile phones and smart speakers," Youngwoo Yoon, one of the researchers who carried out the study, told TechXplore. "We think that [social robots](#) could be the next interaction platform. Physical motion is one of key differences between social robots and other smart devices, opening new possibilities for emulating human- or animal-like behaviors, which can increase intimacy."

Co-speech [gestures](#) could greatly improve the quality of interactions between humans and social robots. Most existing robots produce gestures using rule-based speech-gesture association methods. However, these techniques require considerable efforts, as they are based on human expertise and knowledge.

"We wanted to generate natural and human-like social behaviors, especially hand gestures while speaking," Yoon said. "Observing others is a very natural way of learning a new behavior, so we proposed a learning-based gesture [generation model](#) that was trained on a dataset of TED talks."

The model devised by Yoon and his colleagues was trained on a dataset containing 52 hours of video footage from TED talks. After training, the model could generate sequences of human-like gestures and upper body poses to match written speech text.

"Designing the social behaviors of robots is difficult and time consuming because we have to consider contexts, naturalness, the aesthetics of motion, the control space of robots, and a number of other factors," Yoon explained. "Recent end-to-end learning studies have shed light on the potential of artificial intelligences to generate such complex behaviors. After seeing successful applications in autonomous driving and facial motion generation, we decided to apply end-to-end learning to co-speech gesture generation."

The [neural network model](#) developed by Yoon and his colleagues successfully generated several types of gestures, including iconic, metaphoric, deictic, and beat gestures. Moreover, it was able to generate continuous sequences of gestures for speech texts of any length.

The researchers found that their method outperformed baseline methods in creating gestures that resemble those of humans. In a subjective evaluation, 46 people recruited on Amazon Mechanical Turk felt that the gestures it generated were human-like and closely matched the speech content.

"We found that robots can learn social skills," Yoon said. "For the co-speech gesture generation, the model trained on the large-scale dataset is general enough, so the robot can make human-like gestures for any speech. We think this approach can be applied to other social skills, as well as to characters in video games and VR worlds."

The study carried out by Yoon and his colleagues highlighted the potential of end-to-end learning for co-speech gesture generation. In the future, it could be used to enhance human-[robot](#) interactions and could also inspire similar research, as the TED talks dataset they used is publicly available. The researchers are now planning to take their study one step forward, by generating personalized gestures for different robots.

"Robots may have their own personality, like people," Yoon said. "A personalized [gesture](#) generation method could ensure that different robots express themselves with different styles of gestures, according to their persona."

More information: Robots learn social skills: end-to-end learning of co-speech gesture generation for humanoid robots. arXiv:1810.12541 [cs.RO]. arxiv.org/abs/1810.12541

© 2018 Science X Network

Citation: End-to-end learning of co-speech gesture generation for humanoid robots (2018, November 14) retrieved 9 April 2024 from <https://techxplore.com/news/2018-11-end-to-end-co-speech-gesture-humanoid-robots.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.