Should Beat Gestures Be Learned or Designed? A Bench-marking User Study

Pieter Wolfert[1], Taras Kucherenko[2], Hedvig Kjellström[2], Tony Belpaeme[1]

1 - IDLab, Department of Information Technology at Ghent University – imec, Ghent, Belgium

2 – KTH Royal Institute of Technology Stockholm, Sweden

Read the full paper

*"Machine Learning **generated** gestures **preferred** over **hand crafted** gestures"*

MOTIVATION

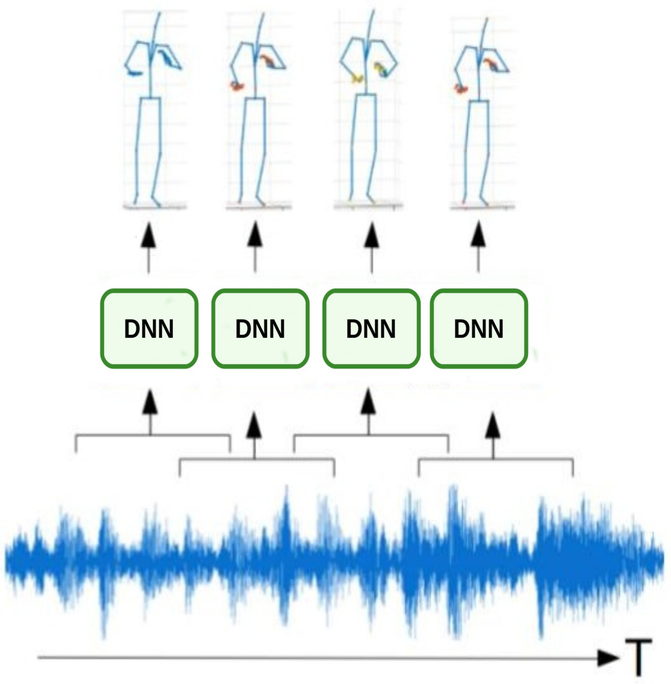
Social robots are often depicted in popular media as being able to communicate like humans do. But this is far from reality, as the current generation of robots only comes with a fixed set of possible nonverbal behaviors. Human-Robot Interaction can be improved by including communicative non-verbal behavior, such as arm gestures. Here we present a user study on generated beat gestures for humanoid agents, which we compare with hand-coded beat gestures.

APPROACH

We compare four types of beat gestures, of which three are hand coded. Our first condition is gestures generated using machine learning. The second condition is random beat gestures, where the gestures follow the pattern of a beat gesture, but are not synced to audio pitch. The third condition is timed beat gestures, in which the gestures are synced with the audio pitch. The fourth condition is noisy gestures, in this condition the hands just move a little.

DATA DRIVEN GESTURES

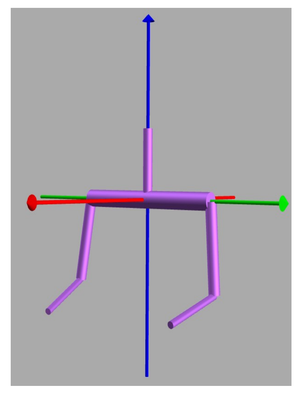
We generated gestures using an encoder-decoder Deep Neural Network, which is trained to map speech to pose[2] . The model is trained on a data set containing an Irish actor who speaks and moves freely (see example above).



The model by Kucherenko et al. learns a representation to take speech audio as input and to generate pose sequences (joint positions).

HAND CRAFTED GESTURES

A 3D model of the human upper body was constructed, using URDF (Unified Robot Description Format)that describes the joints, dimensions and links in XML (Extensible Markup Language). Using this model joint trajectories were simulated and translated to joint positions using inverse kinematics. In the illustration the described joints and links are visualized.

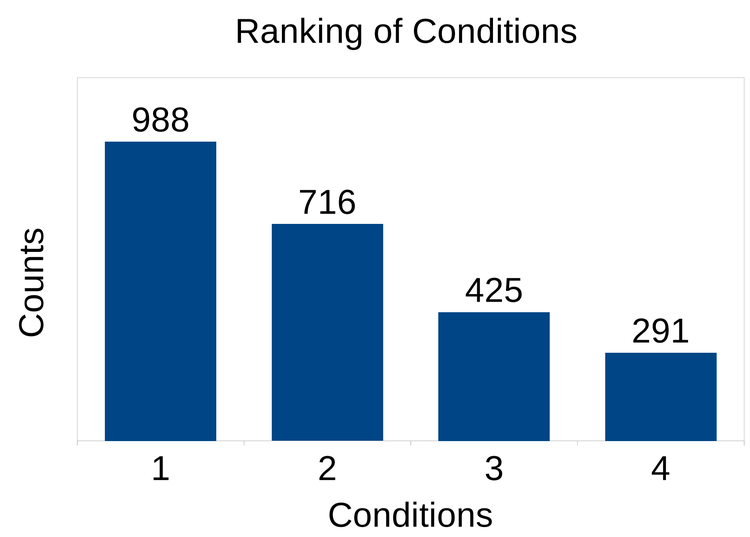


USER-STUDY

A user-study (N=41, average age = 33 years, SD = 9.5 years) was set up, with 10 videos and 4 conditions, which were presented in a pair wise manner to the participants through an Amazon Mechanical Turk survey. Participants were asked to select the most natural gestures.

RESULTS

We found that data-driven generated beat gestures were ranked the highest, followed by random beat gestures. Timed beat gestures and noisy gestures were last on the ranking. A ranking was deduced using the Bradley-Terry-Luce model, and between the pairs a Chi-Square Goodness of Fit was used. The outcome is the number of votes per condition, with the highest number of votes for machine generated beat gestures



DISCUSSION

- We want to emphasize the need for a proper bench marking guideline in gesture generation for humanoid robots, to allow comparative evaluations.
- We found that machine learning based gestures were more appealing than rule based gestures in our specific example.

REFERENCES

1. Ferstl, Y., & McDonnell, R. (2018, November). Investigating the use of recurrent motion modelling for speech gesture generation. In IVA (pp. 93-98).
2. Taras Kucherenko, Dai Hasegawa, Gustav Eje Henter, Naoshi Kaneko, and Hedvig Kjellström. Analyzing input and output representations for speech-driven gesture generation. International Conference on Intelligent Virtual Agents (IVA '19), Paris, July 02–05, 2019