ROBOTICS SEMINAR
FRIDAY, March 5, 2021
3:30-4:30 p.m.

Gustav Eje Henter

Assistant Professor in

Intelligent Systems with spec. in Machine Learning
School of Electrical Engineering & Computer
Science, KTH Royal Institute of Technology

Stockholm Sweden

Move over, MSE! – New probabilistic models of motion

Abstract: Data-driven character animation holds great promise for games, film, virtual avatars and social robots. A "virtual AI actor" that moves in response to intuitive, high-level input could turn 3D animators into directors, instead of requiring them to laboriously pose the character for each frame of animation, as is the case today. However, the high bar on visual quality for most character-animation applications has hitherto constrained machine learning to narrow, task-specific solutions, rife with loss-function engineering and custom processing steps, in order to avoid artefacts such as foot sliding and regression to the mean. This talk makes the case that machine learning now has advanced far enough where strong and general (task-agnostic) motion models are possible. These models should furthermore be probabilistic in nature, to avoid excessive averaging in situations where there are multiple ways to move that are consistent with the control input. In response to the above demands, we introduce MoGlow, a new, award-winning deep-learning architecture that leverages normalising flows to create probabilistic models of character motion. Our proposed method has several important advantages: Flow-based models in general are easy to train (since they allow exact likelihood maximisation) and easy to use (since random sampling is straightforward). MoGlow adds a control input allowing the output motion to be controlled without algorithmic latency. We present applications to locomotion with path control for human and bipedal characters, as well as to speech-driven gesture generation with an optional style control. In each application the method produces output with a perceptual quality that is competitive with the state-of-the-art on the respective task, despite the absence of task-specific modelling assumptions. We also cover how our motion-modelling advances can be combined with the latest breakthroughs in synthesising spontaneous-sounding speech, to make a virtual character walk, talk, and gesticulate from text input alone. For a longer introduction showing the models in action, please see the following YouTube videos: Locomotion synthesis: https://youtu.be/pe-YTvavbtA Co-speech gesture generation: https://youtu.be/egf3tjbWBQE Joint synthesis of speech and motion: https://youtu.be/4_Gq9rU_yWg

Brief Bio: Gustav Eje Henter is an assistant professor in machine learning at the Division of Speech, Music and Hearing at KTH Royal Institute of Technology in Stockholm, Sweden. His main research interests are probabilistic modelling frind deep learning for data generation tasks, most prominently speech and motion/animation synthesis. He has an MSc and a PhD from KTH, followed by post-docs in speech synthesis at the Centre for Speech Technology Research at the University of Edinburgh, UK, and in Prof. Junichi Yamagishi's lab at the National Institute of Informatics, Tokyo, Japan, before returning to KTH in 2018.

Host: Matt Travers

For a virtual meeting appointment contact: Stephanie Matvey (smatvey@andrew.cmu.edu)