GENEA Workshop 2021: The 2nd Workshop on Generation and Evaluation of Non-verbal Behaviour for Embodied Agents

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

Patrik Jonell
pjjonell@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Youngwoo Yoon
youngwoo@etri.re.kr
ETRI & KAIST
Daejeon, Republic of Korea

Pieter Wolfert
pieter.wolfert@ugent.be
IDLab, Ghent University – imec
Ghent, Belgium

Zerrin Yumak
z.yumak@uu.nl
Utrecht University
Utrecht, The Netherlands

Gustav Eje Henter
ghe@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

ABSTRACT

Embodied agents benefit from using non-verbal behavior when communicating with humans. Despite several decades of non-verbal behavior-generation research, there is currently no well-developed benchmarking culture in the field. For example, most researchers do not compare their outcomes with previous work, and if they do, they often do so in their own way which frequently is incompatible with others. With the GENEA Workshop 2021, we aim to bring the community together to discuss key challenges and solutions, and find the most appropriate ways to move the field forward.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Computer systems organization** → Robotics.

KEYWORDS

behavior synthesis, gesture generation, datasets, evaluation

1 WORKSHOP MOTIVATION

Embodied Social AI in the form of conversational virtual humans and social robots are becoming key aspects of human-machine interaction. For several decades, researchers have been proposing methods and models to generate non-verbal behaviors for conversational agents in the form of facial expressions, gestures, gaze and posture [1–3, 7, 9]. The topic has attracted the attention of different communities such as HCI, HRI and 3D graphics/animation, as well as social and behavioral scientists. Yet, these embodied agents are still far from having non-verbal behaviors synthesized on-the-fly in interactive settings, based on factors such as personality, emotions and the social context. Non-verbal behavior synthesis for groups as well as plausibility and controlability of the generated motions remains an important topic of research. In order to advance the field of non-verbal behavior generation, there needs to be clear methods for evaluating and benchmarking the outcomes [4].

This workshop aims to bring together researchers that have different motivations and approaches for non-verbal behavior generation, with the intent to stimulate and advance discussions on how to improve the generation methods and the evaluation of their results. The workshop is a continuation of a successful workshop held at the ACM International Conference on Intelligent Virtual Agents (IVA) 2020; see https://genea-workshop.github.io/2020/. The webpage for the 2021 iteration of the workshop (the one described in this paper) can be found at: https://genea-workshop.github.io/2021/

We aim to address three main questions:

(1) What are the research challenges of generating non-verbal behaviors in different social and emotional contexts, for different individuals, identities and applications?

(2) How can recent advances in machine learning and AI help generate interactive, autonomous, multi-modal and multi-party non-verbal behaviors?

(3) How can we evaluate and benchmark the results of non-verbal behavior synthesis approaches and analyze human perception of synthetically generated non-verbal behaviors?

2 EXPECTED OUTCOMES AND IMPACT

- Facilitating communication and interaction between different communities working on non-verbal behavior synthesis in the field of computer science as well as in connection with social and behavioral sciences.
- Invited talks from experts from within these different communities.
- A deeper understanding of the motivations and challenges in non-verbal behavior-synthesis research.
- Introduction of new data sets, evaluation and benchmarking approaches.
- Presentations of the latest research outputs in non-verbal behavior synthesis and evaluation.
- Group discussions that foster open communication and interaction between participants.

- Diversity and inclusion in all aspects of the workshop (invited talks, PC and organization)

3 WORKSHOP CONTRIBUTIONS

Three contributions were accepted for the workshop, covering a range of topics in HCI. These are summarized below:

- *Probabilistic Human-like Gesture Synthesis from Speech using GRU-based WGAN* [8] by Bowen Wu, Chaoran Liu, Carlos Ishi and Hiroshi Ishiguro. This paper proposes a gesture synthesis from speech. A GRU-based autoregressive model and adversarial training are used to synthesize human-like gesture motion.
- *Crossmodal Clustered Contrastive Learning: Grounding of Spoken Language to Gesture* [5] by Dong Won Lee, Chaitanya Ahuja and Louis-Philippe Morency. This paper introduces a new problem of grounding of speech to gesture. A gesture-aware language embedding is trained through a contrastive loss, and the embedding is proved to be effective for the downstream task of gesture generation.
- *Influence of Movement Energy and Affect Priming on the Perception of Virtual Characters Extroversion and Mood* [6] by Tanja Schneeberger, Fatima Ayman Aly, Daksitha Withanage Don, Katharina Gies, Zita Zeimer, Fabrizio Nunnari and Patrick Gebhard. This paper studies how movement energy and affect priming impact the perceived personality and mood of a virtual character.

4 LIST OF ORGANIZERS

- **Taras Kucherenko** (PhD Student, KTH Royal Institute of Technology, Stockholm, Sweden)
- **Patrik Jonell** (PhD Student, KTH Stockholm, Sweden)
- **Youngwoo Yoon** (Senior Researcher, ETRI, and PhD Student, KAIST, South Korea)
- **Pieter Wolfert** (PhD Student, IDLab Ghent University – imec, Belgium)
- **Zerrin Yumak** (Assistant Professor, Utrecht University, The Netherlands)
- **Gustav Eje Henter** (Assistant Professor, KTH Royal Institute of Technology, Stockholm, Sweden)

5 KEYNOTE SPEAKERS

The workshop also invited two keynote speakers:

- Louis-Philippe Morency (Associate Professor, Carnegie Mellon University, Pittsburgh, USA)
- Hatice Gunes (Associate Professor, University of Cambridge, UK)

These researchers were selected to bring perspectives from both research on language and virtual agents (Louis-Philippe Morency) as well as on robotics (Hatice Gunes).

6 PROGRAM COMMITTEE

The names of the program committee members (in alphabetical order) are the following: Chaitanya Ahuja (CMU); Sean Andrist (Microsoft Research); Jonas Beskow (KTH Royal Institute of Technology); Carlos Busso (University of Texas at Dallas); Oya Celiktutan

(King's College London); Ylva Ferstl (Trinity College Dublin); Carlos Ishi (ATR); Minsu Jang (ETRI); James Kennedy (Disney Research); Stefan Kopp (Bielefeld University); Zofia Malisz (KTH Royal Institute of Technology); Rachel McDonnell (Trinity College Dublin); Michael Neff (University of California, Davis); Catherine Pelachaud (Sorbonne University); Wim Pouw (Radboud University); and Tiago Ribeiro (Soul Machines).

7 CONCLUSIONS

We are confident that the accepted contributions in combination with our keynote speakers will lead the floor for fruitful discussions about the research progress in the field. Our view is that this workshop will bring together researchers who work on gesture generation in avatars and robotics from different angles, which hopefully will lead to cross pollination in terms of methodology and evaluation. This should help advance both non-verbal behavior generation and the methods used for evaluating it.

ACKNOWLEDGMENTS

REFERENCES

[1] Patrik Jonell, Taras Kucherenko, Gustav Eje Henter, and Jonas Beskow. 2020. Let's Face It: Probabilistic multi-modal interlocutor-aware generation of facial gestures in dyadic settings. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. Article 31, 8 pages.

[2] Alex Klein, Zerrin Yumak, Arjen Beij, and A. Frank van der Stappen. 2019. Data-driven gaze animation using recurrent neural networks. In *Proceedings of the ACM SIGGRAPH Conference on Motion, Interaction and Games*. Article 4, 11 pages.

[3] Taras Kucherenko, Dai Hasegawa, Naoshi Kaneko, Gustav Eje Henter, and Hedvig Kjellström. 2021. Moving fast and slow: Analysis of representations and post-processing in speech-driven automatic gesture generation. *International Journal of Human-Computer Interaction* 37, 14 (2021), 1300–1316.

[4] Taras Kucherenko, Patrik Jonell, Youngwoo Yoon, Pieter Wolfert, and Gustav Eje Henter. 2021. A large, crowdsourced evaluation of gesture generation systems on common Data: The GENEA Challenge 2020. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 11–21.

[5] Dong Won Lee, Chaitanya Ahuja, and Louis-Philippe Morency. 2021. Crossmodal clustered contrastive learning: Grounding of spoken language to gesture. In *Proceedings of GENEA 2021: The 2nd Workshop on Generation and Evaluation of Non-verbal Behaviour for Embodied Agents*.

[6] Tanja Schneeberger, Fatima Ayman Aly, Daksitha Withanage Don, Katharina Gies, Zita Zeimer, Fabrizio Nunnari, and Patrick Gebhard. 2021. Influence of movement energy and affect priming on the perception of virtual characters extroversion and mood. In *Proceedings of GENEA 2021: The 2nd Workshop on Generation and Evaluation of Non-verbal Behaviour for Embodied Agents*.

[7] Pieter Wolfert, Taras Kucherenko, Hedvig Kjellström, and Tony Belpaeme. 2019. Should beat gestures be learned or designed?: A benchmarking user study. In *ICDL-EPIROB 2019 Workshop on Naturalistic Non-Verbal and Affective Human-Robot Interactions*.

[8] Bowen Wu, Chaoran Liu, Carlos Ishi, and Hiroshi Ishiguro. 2021. Probabilistic human-like gesture synthesis from speech using GRU-based WGAN. In *Proceedings of GENEA 2021: The 2nd Workshop on Generation and Evaluation of Non-verbal Behaviour for Embodied Agents*.

[9] Youngwoo Yoon, Bok Cha, Joo-Haeng Lee, Minsu Jang, Jaeyeon Lee, Jaehong Kim, and Geehyuk Lee. 2020. Speech gesture generation from the trimodal context of text, audio, and speaker identity. *ACM Transactions on Graphics (TOG)* 39, 6, Article 222 (2020), 16 pages.