

Gustav Eje Henter

CURRICULUM VITÆ

Division of Speech, Music and Hearing
KTH Royal Institute of Technology
Lindstedtsvägen 24
SE-100 44 Stockholm, SWEDEN
☎ +46 70 526 0185
✉ ghe@kth.se
🏠 people.kth.se/~ghe

Speech & motion technology • Probabilistic modelling • Machine learning

Employment

- 2020– **Assistant Professor**, *KTH Royal Institute of Technology*, Stockholm, Sweden.
Assistant Professor in intelligent systems with specialisation in machine learning at the Division of Speech, Music and Hearing (TMH) in the School of Electrical Engineering and Computer Science.
- 2018–2019 **Post-doctoral researcher**, *KTH Royal Institute of Technology*, Stockholm, Sweden.
Post-doctoral position at the Division of Speech, Music and Hearing (TMH) in the School of Electrical Engineering and Computer Science.
- 2016–2018 **Project researcher**, *National Institute of Informatics*, Tokyo, Japan.
Post-doctoral position at the Yamagishi Laboratory in the Digital Content and Media Sciences Research Division.
- 2014–2016 **Research fellow**, *The University of Edinburgh*, UK.
Post-doctoral position at the Centre for Speech Technology Research (CSTR) in the School of Informatics.
- 2013–2014 **Marie Curie research fellow**, *The University of Edinburgh*, UK.
12-month Marie Curie INSPIRE ITN Experienced Researcher post-doctoral position at the Centre for Speech Technology Research (CSTR) in the School of Informatics.
(Please see the Research section for information regarding professional activities.)

Education

- 2007–2013 **Ph.D. in Electrical Engineering (Telecommunications)**, *KTH Royal Institute of Technology*, Stockholm, Sweden.
Research and studies at the Communication Theory group (formerly the Sound and Image Processing Laboratory) in the School of Electrical Engineering.

Ph.D. dissertation “*Probabilistic sequence models with speech and language applications*.” Supervisors Prof. W. Bastiaan Kleijn and Prof. Arne Leijon. Opponent Prof. Dr. Gernot Kubin, Signal Processing and Speech Communication Laboratory, TU Graz, Austria.
- Jan–June 2011 **Visiting Ph.D. student**, *Victoria University of Wellington*, New Zealand.
Collaborative research in the Communications and Signal Processing (CaSP) Group at the School of Engineering and Computer Science.
- 2001–2007 **M.Sc. (Civilingenjör) in Engineering Physics**, *KTH Royal Institute of Technology*, Stockholm, Sweden.
Combined M.Sc. and B.Sc. studies. Specialisation in applied mathematics, with focus on machine learning, mathematical statistics, and numerical methods.

M.Sc. thesis “*Mathematical techniques for applied demand estimation and profit maximization*.” Supervisor Jani Ekonoja, Actuva AB. Examiner Assoc. Prof. Jan Grandell, Department of Mathematical Statistics, KTH.

- 2000–2001 **Exchange student in Engineering Physics**, *UBC (University of British Columbia)*, Vancouver, BC, Canada.
Two semester studies at the Faculty of Applied Science, equivalent to the second year at KTH.
- 1999–2000 **M.Sc. (Civilingenjör) in Engineering Physics**, *KTH Royal Institute of Technology*, Stockholm, Sweden.
First year of combined M.Sc. and B.Sc. studies.

Research

- 2020– **WASP Starting Grant in AI/MLX**, *Knut and Alice Wallenberg Foundation*.
Individual research grant for new Assistant Professors in machine learning.

Principal investigator. Leading, performing, and publishing research on machine learning and statistical modelling for data generation, in particular for speech synthesis and motion generation.
- 2018–2020 **EACare: Embodied Agent to support elderly mental wellbeing**, *Swedish Foundation for Strategic Research Smart Systems Grant*.
Multi-site research project developing a communicative robot for elderly people with dementia.

Workpackage lead on WP2 System Integration and Evaluation, focussing on data collection. Performed and published research on synthesising conversational speech and motion such as co-speech gestures.
- 2017–2018 **Harnessing Latent Variation in DNN-Based Speech Synthesis**, *KAKENHI Wakate B (Japan Society for the Promotion of Science Grant-in-Aid for Young Scientists)*.
Individual research grant for applying latent-variable theory to improve speech synthesis.

Performed research on unsupervised learning of speech-synthesis expression control. Manuscript available on arXiv.
- Apr–Sep 2017 **DIACEX (Dissecting Foreign Accent)**, *Spanish Ministry of Economy and Competitiveness (MINECO) Grant*.
Multi-site research project on cues to foreign accent and their effect on communication.

Performed and published research on synthesising foreign-accented speech from native-speech recordings.
- 2016–2017 **uDialogue (User Generated Dialogue Systems)**, *JST CREST Funding Programme*.
Multi-site research project on adaptable speech synthesis and dialogue systems.

Performed and published research on supervised speaker identity adaptation and unsupervised learning of expression control in speech synthesis.
- 2014–2016 **NST (Natural Speech Technology)**, *UK EPSRC Programme Grant*.
Multi-site research project to advance the state-of-the-art in speech technology, including speech recognition and synthesis.

Performed and published research on statistical modelling of acoustics and duration in speech synthesis, as well as speech synthesis evaluation methodologies.

2013–2014 **INSPIRE (Investigating Speech Processing In Realistic Environments)**, *EU FP7 Marie Curie Initial Training Network*.

Multi-site, multi-disciplinary training network studying speech communication in real-world conditions.

Performed and published research on the effects of mathematical assumptions in speech synthesis. Recorded two speech databases, one of which has been released to date.

2010–2013 **LISTA (The Listening Talker)**, *EU FP7 Future and Emerging Technologies project*.

Multi-site research project on enhancing the intelligibility of natural and synthetic speech in adverse conditions.

Performed and published research on statistical sequence modelling for data generation and automated speech-intelligibility assessment. Wrote and reviewed deliverables.

2007–2009 **ACORNS (Acquisition of Communication and Recognition Skills)**, *EU FP6 Future and Emerging Technologies project*.

Multi-site research project on language acquisition based on human infants' language learning.

Performed and published research on unsupervised pattern discovery using information theory. Wrote deliverables.

Teaching

2018–2020 **Teaching assistant**, *DT2119 Speech and Speaker Recognition (7.5 credits)*.

M.Sc.-level course. Answering student questions and grading student assignments.

2020 **Course responsible**, *FDT3317 Speech Synthesis from Beginning to End-to-End (7.5 credits)*.

Ph.D.-level course. Course development, giving lectures, leading seminars, creating and grading assignment. Jointly course responsible with Dr. Zofia Malisz.

2019 **Guest lecturer**, *EQ2341 Pattern Recognition and Machine Learning (7.5 credits)*.

M.Sc.-level course. Giving a lecture.

2018–2019 **Teaching assistant**, *DT1130 Spectral Transforms (7.5 credits)*.

B.Sc.-level course. Giving tutorial sessions and grading exams.

2018–2019 **Project supervisor**, *DT2140 Multimodal Interaction and Interfaces (7.5 credits)*.

M.Sc.-level project course. Supervised two student projects.

2007–2012 **Teaching assistant**, *EN2200/EN2202 Pattern Recognition (6.0/7.5 credits)*.

M.Sc.-level course. Wrote course book chapters, gave lectures and exercises, and took part in designing and grading exams. Responsible for the course project (tasks, code, supervision, grading). Consistently strong course evaluation results – see appendix.

Supervision

2021 **M.Sc. student supervision.**

Gustavo Teodoro Döhler Beck, preliminary thesis title “*Controllability and interpretability of generated speech through WaveGlow and low-level speech features*”. Principal supervisor Ulme Wennberg.

2020– **Ph.D. student supervision.**

Shivam Mehta, thesis topic: End-to-end speech synthesis. Principal supervisor Prof. Jonas Beskow.

2020– **Ph.D. student supervision.**

Ulme Wennberg, thesis topic to be decided. Principal supervisor Prof. Jonas Beskow.

- 2020– **Ph.D. student supervision.**
Patrik Jonell, thesis topic: Using social and physiological signals for user adaptation in conversational agents. Principal supervisor Prof. Jonas Beskow.
- 2020 **M.Sc. student supervision.**
Harsha Holenarasipura Narasanna, thesis “*Graph convolutional network based friendship recommender system for a social network*”. Company supervisor Erwan Lemonnier, GoFrendly AB.
- 2020 **B.Sc. student supervision.**
Jorge García Pueyo, project “*Replicating MelNet: A generative model for audio in the frequency domain*” for the course DD2465 Advanced, Individual Course in Computer Science (6.0 credits).
- 2019– **Ph.D. student supervision.**
Ghazaleh Esfandiari Baiat. Principal supervisor Prof. Jens Edlund.
- 2019–2021 **Ph.D. student supervision.**
Taras-Svitozar Kucherenko, thesis topic: Gesture in human-robot communication. Principal supervisor Prof. Hedvig Kjellström.
- 2019 **M.Sc. student supervision.**
Helen From, thesis “*Predicting the ground effect in drone landing with online learning*”. Jointly supervised with Prof. Christian Wolf at INSA Lyon, France.
- 2019 **M.Sc. student supervision.**
Ulme Wennberg, thesis “*An evaluation of BERT for a span-based approach for jointly predicting entities, coreference clusters and relations between entities*”. Jointly supervised with Prof. Hannaneh Hajishirzi at University of Washington, Seattle.
- 2018 **M.Sc. student supervision.**
Per Näslund, thesis “*Artificial neural networks in Swedish speech synthesis*”.
- 2016 **M.Sc. student supervision.**
Mironas Bitinis, thesis “*Performance of deep neural networks employing predicted durations and stacked bottleneck features for speech synthesis*”. Jointly supervised with Prof. Simon King, University of Edinburgh.
- 2015–2016 **Visiting Ph.D. student supervision.**
Takenori Yoshimura, visiting Ph.D. student from Nagoya Institute of Technology (NITech), working on automatic prediction of subjective naturalness ratings of synthetic speech, published at Interspeech 2016. Jointly supervised with Dr. Oliver Watts, University of Edinburgh.
- 2010–2011 **M.Sc. student supervision.**
Anders Blomqvist, thesis “*Modelling pulse timing patterns with a GMM/HMM framework*”. Jointly supervised with Prof. Arne Leijon, KTH.

Examination

- 2020 **Faculty opponent, Aalto University, Helsinki, Finland.**
Faculty opponent for Ph.D. candidate Lauri Juvela, thesis “*Neural waveform generation for source-filter vocoding in speech synthesis*”. Custos Prof. Paavo Alku, Department of Signal Processing and Acoustics, Aalto University.

Grants awarded

- 2021 **Wallenberg Launchpad, Knut and Alice Wallenberg Foundation.**
Joint principal investigator with Dr. Simon Alexanderson of grant application “*MoGlow WALP*”. 2,900,000 SEK awarded (corresponding to 285,175 EUR) over 14 months.
- 2019 **WASP Starting Grant in AI/MLX, Wallenberg AI, Autonomous Systems and Software Program, Knut and Alice Wallenberg Foundation.**
Fully funded assistant professorship with a start-up package of 18,000,000 SEK (corresponding to 1,725,289 EUR) over five years.

2017 **KAKENHI Wakate B (Grant-in-Aid for Young Scientists)**, *Japan Society for the Promotion of Science*.

Principal investigator of grant application “*Harnessing Latent Variation in DNN-Based Speech Synthesis*”. 3,250,000 JPY awarded (corresponding to 27,486 EUR) over fiscal years 2017–2018.

Selected talks and tutorials

(Conference, meeting, and workshop presentations excluded.)

2021 **AI and Animation**, *WASP4ALL 2021 – Building excellence for the era of AI*, Online event, Sweden.

Invited talk.

2021 **Move over, MSE! – New probabilistic models of motion**, *Department of Computer Science, University of California, Davis, Davis, USA*.

Computer Science Colloquium Seminar. Presented online due to the COVID-19 pandemic.

2021 **Move over, MSE! – New probabilistic models of motion**, *The Robotics Institute, Carnegie Mellon University, Pittsburgh, USA*.

Robotics Institute Seminar. Presented online due to the COVID-19 pandemic.

2021 **Move over, MSE! – New probabilistic models of motion**, *Department of Computing Science, Umeå University, Umeå, Sweden*.

Invited talk. Presented online due to the COVID-19 pandemic.

2021 **Move over, MSE! – New probabilistic models of motion**, *Yamagishi lab, Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo, Japan*.

Invited talk. Presented online due to the COVID-19 pandemic.

2020 **New probabilistic models of motion**, *Research group on Computer Graphics and Visualisation, The University of Edinburgh, Edinburgh, UK*.

Invited talk with Dr. Simon Alexanderson. Presented online due to the COVID-19 pandemic.

2018 **Cyborg speech and other controllable synthesisers**, *Centre for Speech Technology Research, The University of Edinburgh, Edinburgh, UK*.

Invited talk.

2018 **Wagging speech by the tail: The case for robust data generation**, *ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden*.

Invited ACCESS seminar.

2018 **Perceptual debugging of speech synthesis**, *Tokuda and Nankaku Laboratory, Nagoya Institute of Technology, Nagoya, Japan*.

Invited talk.

2017 **Perceptual debugging of speech synthesis**, *Department of Speech, Music, and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden*.

Invited talk.

2016 **Wagging speech by the tail: The case for robust speech synthesis**, *Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo, Japan*.

Invited talk.

2016 **Wagging speech by the tail: The case for robust speech synthesis**, *Department of Speech, Music, and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden*.

Invited talk.

- 2016 **Wagging speech by the tail: The case for robust speech synthesis**, *Department of Architecture, Design, and Media Technology, Aalborg University, Aalborg, Denmark*.
Invited talk.
- 2015 **Wagging speech by the tail: The case for robust speech synthesis**, *Department of Signal Processing and Acoustics, Aalto University, Helsinki, Finland*.
Invited talk.
- 2015 **Measuring the perceptual effects of speech synthesis modelling assumptions**, *Department of Speech Sciences, University of Helsinki, Finland*.
Invited talk.
- 2015 **Make yourself heard! Man and machine, speaking in noise**, *The Royal Institution, London, UK*.
Public engagement talk on speech intelligibility pre-enhancement.
- 2014 **Manipulating natural speech**, *INSPIRE Summer School – Use of speech synthesis for psychoacoustic and phonetics research, Edinburgh, UK*.
Hands-on tutorial.

Organisation

- 2021 **GENEA Workshop 2021**, *23rd ACM International Conference on Multimodal Interaction*, Montréal, Canada.
Official workshop of ICMI 2021 on generation and evaluation of non-verbal behaviour for embodied agents. Accepted papers published in ICMI adjunct proceedings. Co-organised with Taras Kucherenko at KTH, Zerrin Yumak at Utrecht University, Pieter Wolfert at Ghent University, Youngwoo Yoon at ETRI & KAIST, and Patrik Jonell at KTH. Hybrid event due to the COVID-19 pandemic.
- 2021– **ISCA SynSIG**, *International Speech Communication Association*.
ISCA Special Interest Group on speech synthesis. SynSIG officer and secretary.
- 2021– **WARA Media and Language**, *Wallenberg AI, Autonomous Systems and Software Program (WASP) Research Arena*, Sweden.
Organisation for increased collaboration between media industry and academia in Sweden. Steering group member and core organiser of the gaming focus area.
- 2021– **Synthesis reading group**, *Division of Speech, Music and Hearing, KTH Royal Institute of Technology*, Stockholm, Sweden.
Weekly reading group in machine learning for data generation. Organiser.
- 2020 **GENEA Workshop 2020**, *20th ACM International Conference on Intelligent Virtual Agents*, Glasgow, UK.
Official satellite workshop of IVA 2020 on generation and evaluation of non-verbal behaviour for embodied agents. Co-organised with Taras Kucherenko at KTH, Pieter Wolfert at Ghent University, Youngwoo Yoon at ETRI & KAIST, and Patrik Jonell at KTH. Held online due to the COVID-19 pandemic.
- 2020 **GENEA Challenge 2020**.
Large-scale challenge and evaluation of co-speech gesture-generation methods with 5 participating teams. Co-organised with Taras Kucherenko at KTH, Pieter Wolfert at Ghent University, Youngwoo Yoon at ETRI & KAIST, and Patrik Jonell at KTH.
- 2018– **TMH seminars**, *Division of Speech, Music and Hearing, KTH Royal Institute of Technology*, Stockholm, Sweden.
Responsible for seminars and invited talks.

- 2016–2018 **Sing!**, *Digital Content and Media Sciences Research Division, National Institute of Informatics, Japan.*
Weekly speech synthesis paper-reading group. Founder and organiser.
- 2015–2016 **Speak!**, *The Centre for Speech Synthesis Research, The University of Edinburgh, UK.*
Weekly speech synthesis paper-reading group. Organiser.
- 2014 **INSPIRE summer school**, Edinburgh, UK.
Co-organised with Dr. Catherine Mayo and Prof. Simon King.
- 2011 **LISTA project meeting**, Stockholm, Sweden.
Co-organised with Petko N. Petkov and Prof. Arne Leijon.
- 2010 **SIP lunch seminars**, *Sound and Image Processing Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden.*
Biweekly, catered scientific presentations by research group members. Organiser.
- 2008 **ACORNS scientific advisory committee and project meeting**, Norrtälje, Sweden.
Co-organised with Christos Koniaris and Prof. W. Bastiaan Kleijn.

Databases and corpora

- 2021 **GENEA Challenge 2020 data release.**
With Taras Kucherenko, Patrik Jonell, Youngwoo Yoon, and Pieter Wolfert.
Improved version of the Trinity Speech-Gesture Dataset. Available under the same license as the original data at trinityspeechgesture.scss.tcd.ie. Additional material from the challenge available via svito-zar.github.io/GENEAchallenge2020 under open licenses.
- 2014 **Repeated Harvard Sentence Prompts (REHASP) corpus version 0.5.**
With Thomas Merritt, Matt Shannon, Catherine Mayo, and Simon King.
Available under a Creative Commons license at dx.doi.org/10.7488/ds/39.

Peer review

- 2021 **Reviewer**, *Young Female Researchers in Speech Workshop.*
2 submissions reviewed.
- 2020–2021 **Reviewer**, *SIGGRAPH.*
3 submissions reviewed.
- 2020 **Reviewer**, *SIGGRAPH Asia.*
1 submission reviewed.
- 2010–2020 **Reviewer**, *Interspeech.*
31 submissions total reviewed.
- 2015–2019 **Reviewer**, *ICASSP.*
6 submissions total reviewed.
- 2019 **Session chair**, *Interspeech.*
1 poster session chaired.
- 2019 **Reviewer**, *Speech Synthesis Workshop.*
3 submissions reviewed.
- 2014–2016 **Reviewer**, *IEEE/ACM Transactions on Audio, Speech, and Language Processing.*
3 submissions total reviewed.
- 2015 **Reviewer**, *SPIRE workshop.*
3 submissions reviewed.

- 2014 **Reviewer**, *IEEE Signal Processing Magazine*.
1 submission reviewed.
- 2013 **Reviewer**, *IEEE Journal of Selected Topics in Signal Processing*.
1 submission reviewed.
- 2010 **Reviewer**, *EUSIPCO*.
3 submissions reviewed.

Language skills

Swedish	Native	<i>Excellent grades.</i>
English	Near native	<i>Excellent grades. Extensive practice in English-speaking countries.</i>
German	Fundamental	<i>Excellent grades.</i>

Computer skills

OS	macOS, Linux, Windows.	
Languages	Matlab, Python, Pytorch, bash, C/C++.	
Development	Development and maintenance of MUSHRA test software in Matlab , used for conducting and analysing listening tests at the University of Edinburgh.	
	Long-term development and maintenance of an HMM toolbox in Matlab , used in teaching and research at Communications Theory at KTH.	
	Implemented and extended several machine learning models in C++ .	
Support	Sound and Image Processing lab designated computer support person 2009–2011.	

Awards and recognition (selected)

- 2021 **ACM IVA Extended Abstract Honourable Mention**, 21st *ACM International Conference on Intelligent Virtual Agents*, Kyoto, Japan.
Honourable mention given to the best extended abstract (awarded to 2 out of 16 accepted extended abstracts) for the paper “*Speech2Properties2Gestures: Gesture-property prediction as a tool for generating representational gestures from speech*” co-authored with Taras Kucherenko, Rajmund Nagy, Patrik Jonell, Prof. Michael Neff, and Prof. Hedvig Kjellström. Awarded online due to the COVID-19 pandemic.
- 2020 **ACM ICMI Best Paper Award**, 22nd *ACM International Conference on Multimodal Interaction*, Utrecht, The Netherlands.
Best paper award (awarded to 2 out of 159 long paper submissions) for the paper “*Gesticulator: A framework for semantically-aware speech-driven gesture generation*” co-authored with Taras Kucherenko, Patrik Jonell, Sanne van Waveren, Dr. Simon Alexanderson, Assoc. Prof. Iolanda Leite, and Prof. Hedvig Kjellström. Awarded online due to the COVID-19 pandemic.
- 2020 **ACM IVA Best Paper Award**, 20th *ACM International Conference on Intelligent Virtual Agents*, Glasgow, UK.
Best paper award (awarded to 1 out of 137 full paper submissions) for the paper “*Let’s face it: Probabilistic multi-modal interlocutor-aware generation of facial gestures in dyadic settings*” co-authored with Patrik Jonell, Taras Kucherenko, and Prof. Jonas Beskow. Awarded online due to the COVID-19 pandemic.

2020 **Honourable Mention for the Günter Enderle Award**, 41st Annual Conference of the European Association for Computer Graphics (*EUROGRAPHICS*) 2020, Norrköping, Sweden.

Top 4 out of 141 submissions and best paper award nomination for the paper “*Style-controllable speech-driven gesture synthesis using normalising flows*” co-authored with Dr. Simon Alexanderson, Taras Kucherenko, and Prof. Jonas Beskow. Awarded online due to the COVID-19 pandemic.

2020 Supervised recipient of the **SAIS Best AI Master’s Thesis Award 2020**, Swedish AI Society, Gothenburg, Sweden.

Awarded to M.Sc. student Ulme Wennberg for the thesis “*An evaluation of BERT for a span-based approach for jointly predicting entities, coreference clusters and relations between entities*”, which I supervised jointly supervised with Prof. Hannaneh Hajishirzi at University of Washington, Seattle. Awarded online due to the COVID-19 pandemic.

2019 **Interspeech 2019 Best Show and Tell Paper**, Interspeech, Graz, Austria.

Best demo out of 37, awarded for the paper “*Off the cuff: Exploring extemporaneous speech delivery with TTS*” co-authored with Dr. Éva Székely, Prof. Jonas Beskow, and Prof. Joakim Gustafson.

References

References provided upon request.

Publications

Publications are listed in reverse chronological order per section. Asterisks signify equal contribution.

Journal publications

- [1] G. Valle-Pérez, **G. E. Henter**, J. Beskow, A. Holzapfel, P.-Y. Oudeyer, & S. Alexanderson, “Transflower: Probabilistic autoregressive dance generation with multimodal attention,” *ACM Transactions on Graphics*, vol. 40, no. 6, 1:1–1:13, 2021. Impact factor 5.414 in 2020. Presented at SIGGRAPH Asia 2021.
- [2] P. Jonell,* B. Moëll,* K. Håkansson,* **G. E. Henter**, T. Kucherenko, O. Mikheeva, G. Hagman, J. Holleman, M. Kivipelto, H. Kjellström, J. Gustafson, & J. Beskow, “Multimodal capture of patient behaviour for improved detection of early dementia: Clinical feasibility and preliminary results,” *Frontiers in Computer Science*, vol. 3, article no. 642633, pp. 1–22, 2021.
- [3] T. Kucherenko, D. Hasegawa, N. Kaneko, **G. E. Henter**, & H. Kjellström, “Moving fast and slow: Analysis of representations and post-processing in speech-driven automatic gesture generation,” *International Journal of Human–Computer Interaction*, vol. 37, no. 14, 1300–1316, 2021. Impact factor 3.353 in 2020.
- [4] **G. E. Henter**,* S. Alexanderson,* & J. Beskow, “MoGlow: Probabilistic and controllable motion synthesis using normalising flows,” *ACM Transactions on Graphics*, vol. 36, no. 6, 236:1–236:14, 2020. Impact factor 5.414 in 2020. Presented at SIGGRAPH Asia 2020.
- [5] S. Alexanderson, **G. E. Henter**, T. Kucherenko, & J. Beskow, “Style-controllable speech-driven gesture synthesis using normalising flows,” *Computer Graphics Forum*, vol. 39, no. 2, pp. 487–496, 2020. Impact factor 2.078 in 2020. **Honourable mention and best paper award nominee at EUROGRAPHICS 2020 (in the top 4 of 141 submissions).**
- [6] J. Lorenzo-Trueba, **G. E. Henter**, S. Takaki, J. Yamagishi, Y. Morino, & Y. Ochiai, “Investigating different representations for modeling and controlling multiple emotions in DNN-based speech synthesis,” *Speech Communication*, vol. 99, pp. 135–143, 2018. Impact factor 1.661 in 2018.
- [7] **G. E. Henter** & W. B. Kleijn, “Minimum entropy rate simplification of stochastic processes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 12, pp. 2487–2500, 2016. Impact factor 8.329 in 2016.
- [8] A. Leijon, **G. E. Henter**, & M. Dahlquist, “Bayesian analysis of phoneme confusion matrices,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 469–482, 2015. Impact factor 1.225 in 2015.
- [9] P. N. Petkov, **G. E. Henter**, & W. B. Kleijn, “Maximizing phoneme recognition accuracy for enhanced speech intelligibility in noise,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 1035–1045, 2013. Impact factor 2.625 in 2013.
- [10] **G. E. Henter** & W. B. Kleijn, “Picking up the pieces: Causal states in noisy data, and how to recover them,” *Pattern Recognition Letters*, vol. 34, no. 5, pp. 587–594, 2013. Impact factor 1.062 in 2013.

Conference publications (peer reviewed)

- [11] G. T. Döhler Beck, U. Wennberg, **G. E. Henter**, & Z. Malisz, “Wavebender GAN: Controllable speech synthesis for speech sciences,” extended abstract accepted for publication at the *1st International Conference on Tone and Intonation (TAI)*, 2021.
- [12] S. Wang, S. Alexanderson, J. Gustafson, J. Beskow, **G. E. Henter**, & É. Székely, “Integrated speech and gesture synthesis,” *Proc. ICMI*, 2021, pp. 117–185. CORE rank B.
- [13] P. Jonell, Y. Yoon, P. Wolfert, T. Kucherenko, & **G. E. Henter**, “HEMVIP: Human evaluation of multiple videos in parallel,” *Proc. ICMI*, 2021, pp. 707–711 (short paper). CORE rank B.
- [14] M. Sorkhei, **G. E. Henter**, & H. Kjellström, “Full-Glow: Fully conditional Glow for more realistic image generation,” accepted for publication at the *4^{3rd}* DAGM German Conference on Pattern Recognition (GCPR), 2021.

- [15] T. Kucherenko, R. Nagy, P. Jonell, M. Neff, H. Kjellström & **G. E. Henter**, “Speech2Properties-2Gestures: Gesture-property prediction as a tool for generating representational gestures from speech,” *Proc. IVA*, 2021, pp. 145–147 (extended abstract). CORE rank B. **Extended Abstract Honourable Mention (given to 2 of 8 accepted extended abstracts)**.
- [16] U. Wennberg & **G. E. Henter**, “The case for translation-invariant self-attention in transformer-based language models,” *Proc. ACL-IJCNLP Vol. 2: Short Papers*, 2021, pp. 130–140. CORE rank A*. **13.6% acceptance rate for short papers**.
- [17] T. Kucherenko,* P. Jonell,* Y. Yoon,* P. Wolfert, & **G. E. Henter**, “A large, crowdsourced evaluation of gesture generation systems on common data: The GENEA Challenge 2020,” *Proc. IUI*, 2021, pp. 11–21. CORE rank A.
- [18] T. Kucherenko, P. Jonell, S. van Waveren, **G. E. Henter**, S. Alexanderson, I. Leite, & H. Kjellström, “Gesticulator: A framework for semantically-aware speech-driven gesture generation,” *Proc. ICMI*, 2020, pp. 242–250. CORE rank B. **Best Paper Award (given to 2 of 159 long paper submissions)**.
- [19] P. Jonell, T. Kucherenko, **G. E. Henter**, & J. Beskow, “Let’s face it: Probabilistic multi-modal interlocutor-aware generation of facial gestures in dyadic settings,” *Proc. IVA*, 2020, article no. 31, pp. 1–8. CORE rank B. **Best Paper Award (given to 1 of 137 full paper submissions)**.
- [20] S. Alexanderson, É. Székely, **G. E. Henter**, T. Kucherenko, & J. Beskow, “Generating coherent spontaneous speech and gesture from text,” *Proc. IVA*, 2020, article no. 1, pp. 1–3. CORE rank B.
- [21] A. Ghosh, A. Honoré, D. Liu, **G. E. Henter**, & S. Chatterjee, “Robust classification using hidden Markov models and mixtures of normalizing flows,” *Proc. MLSP*, 2020, pp. 1–6.
- [22] S. Alexanderson & **G. E. Henter**, “Robust model training and generalisation with Studentising flows,” *Proc. INNF+ (2nd Workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models)*, 2020, article no. 25, pp. 1–9.
- [23] É. Székely, **G. E. Henter**, J. Beskow, & J. Gustafson, “Breathing and speech planning in spontaneous speech synthesis,” *Proc. ICASSP*, 2020, pp. 7649–7653. CORE rank B.
- [24] É. Székely, **G. E. Henter**, J. Beskow, & J. Gustafson, “How to train your fillers: uh and um in spontaneous speech synthesis,” *Proc. SSW*, 2019, pp. 245–250.
- [25] O. Watts, **G. E. Henter**, J. Fong, & C. Valentini-Botinhao, “Where do the improvements come from in sequence-to-sequence neural TTS?,” *Proc. SSW*, 2019, pp. 217–222.
- [26] P. Wagner, J. Beskow, S. Betz, J. Edlund, J. Gustafson, **G. E. Henter**, S. Le Maguer, Z. Malisz, É. Székely, C. Tännander, & J. Voße, “Speech synthesis evaluation – State-of-the-art assessment and suggestion for a novel research program,” *Proc. SSW*, 2019, pp. 105–110.
- [27] É. Székely, **G. E. Henter**, J. Beskow, & J. Gustafson, “Spontaneous conversational speech synthesis from found data,” *Proc. Interspeech*, 2019, pp. 4435–4439. CORE rank A.
- [28] É. Székely, **G. E. Henter**, J. Beskow, & J. Gustafson, “Off the cuff: Exploring extemporaneous speech delivery with TTS,” *Proc. Interspeech*, 2019, pp. 3687–3688. CORE rank A. **Best Show & Tell award (given to 1 of 37 demos at the conference)**.
- [29] Z. Malisz, **G. E. Henter**, C. Valentini-Botinhao, O. Watts, J. Beskow, & J. Gustafson, “Modern speech synthesis for phonetic sciences: a discussion and an evaluation,” *Proc. ICPHS*, 2019, pp. 487–491.
- [30] T. Kucherenko, D. Hasegawa, **G. E. Henter**, N. Kaneko, & H. Kjellström, “Analyzing input and output representations for speech-driven gesture generation,” *Proc. IVA*, 2019, pp. 97–104. CORE rank B.
- [31] É. Székely, **G. E. Henter**, & J. Gustafson, “Casting to corpus: Segmenting and selecting spontaneous dialogue for TTS with a CNN-LSTM speaker-dependent breath detector,” *Proc. ICASSP*, 2019, pp. 6925–6929. CORE rank B.
- [32] T. Kucherenko, D. Hasegawa, N. Kaneko, **G. E. Henter**, & H. Kjellström, “On the importance of representations for speech-driven gesture generation,” *Proc. AAMAS*, 2019, pp. 2072–2074 (extended abstract). CORE rank A*.

- [33] **G. E. Henter**, J. Lorenzo-Trueba, X. Wang, M. Kondo, & J. Yamagishi, “Cyborg speech: Deep multilingual speech synthesis for generating segmental foreign accent with natural prosody,” *Proc. ICASSP*, 2018, pp. 4799–4803. CORE rank B.
- [34] J. Lorenzo-Trueba, **G. E. Henter**, S. Takaki, J. Yamagishi, Y. Morino, & Y. Ochiai, “Investigating different representations for modeling multiple emotions in DNN-based speech synthesis,” *Proc. ASMMC (3rd International Workshop on Affective Social Multimedia Computing)*, 2017.
- [35] **G. E. Henter**, J. Lorenzo-Trueba, X. Wang, & J. Yamagishi, “Principles for learning controllable TTS from annotated and latent variation,” *Proc. Interspeech*, 2017, pp. 3956–3960. CORE rank A.
- [36] J. Lorenzo-Trueba, C. Valentini-Botinhao, **G. E. Henter**, & J. Yamagishi, “Misperceptions of the emotional content of natural and vocoded speech in a car,” *Proc. Interspeech*, 2017, pp. 606–610. CORE rank A.
- [37] H.-T. Luong, S. Takaki, **G. E. Henter**, & J. Yamagishi, “Adapting and controlling DNN-based speech synthesis using input codes,” *Proc. ICASSP*, 2017, pp. 4905–4909. CORE rank B.
- [38] S. Ronanki, O. Watts, S. King, & **G. E. Henter**, “Median-based generation of synthetic speech durations using a non-parametric approach,” *Proc. SLT*, 2016.
- [39] T. Yoshimura, **G. E. Henter**, O. Watts, M. Wester, J. Yamagishi, & K. Tokuda, “A hierarchical predictor of synthetic speech naturalness using neural networks,” *Proc. Interspeech*, 2016, pp. 342–346. CORE rank A.
- [40] S. Ronanki, **G. E. Henter**, Z. Wu, & S. King, “A template-based approach for speech synthesis intonation generation using LSTMs,” *Proc. Interspeech*, 2016, pp. 2463–2467. CORE rank A.
- [41] M. Wester, O. Watts, & **G. E. Henter**, “Evaluating comprehension of natural and synthetic conversational speech,” *Proc. Speech Prosody*, 2016, pp. 736–740.
- [42] **G. E. Henter**, S. Ronanki, O. Watts, M. Wester, Z. Wu, & S. King, “Robust TTS duration modelling using DNNs,” *Proc. ICASSP*, 2016, pp. 5155–5159. CORE rank B.
- [43] O. Watts, **G. E. Henter**, T. Merritt, Z. Wu, & S. King, “From HMMs to DNNs: where do the improvements come from?” *Proc. ICASSP*, 2016, pp. 5130–5134. CORE rank B.
- [44] R. Dall, S. Brognaux, K. Richmond, C. Valentini-Botinhao, **G. E. Henter**, J. Hirschberg, J. Yamagishi, & S. King, “Testing the consistency assumption: Pronunciation variant forced alignment in read and spontaneous speech synthesis,” *Proc. ICASSP*, 2016, pp. 5155–5159. CORE rank B.
- [45] M. Wester, C. Valentini-Botinhao, & **G. E. Henter**, “Are we using enough listeners? No! An empirically-supported critique of Interspeech 2014 TTS evaluations,” *Proc. Interspeech*, 2015, pp. 3476–3480. CORE rank A.
- [46] **G. E. Henter**, T. Merritt, M. Shannon, C. Mayo, & S. King, “Measuring the perceptual effects of modelling assumptions in speech synthesis using stimuli constructed from repeated natural speech,” *Proc. Interspeech*, 2014, pp. 1504–1508. CORE rank A.
- [47] M. P. Aylett, R. Dall, A. Ghoshal, **G. E. Henter**, & T. Merritt, “A flexible front-end for HTS,” *Proc. Interspeech*, 2014, pp. 1283–1287. CORE rank A.
- [48] P. N. Petkov, W. B. Kleijn, & **G. E. Henter**, “Enhancing subjective speech intelligibility using a statistical model of speech,” *Proc. Interspeech*, 2012, pp. 166–169. CORE rank A.
- [49] P. N. Petkov, W. B. Kleijn, & **G. E. Henter**, “Speech intelligibility enhancement using a statistical model of clean speech,” *Proc. The Listening Talker Workshop*, 2012, p. 77.
- [50] **G. E. Henter**, M. R. Frean, & W. B. Kleijn, “Gaussian process dynamical models for nonparametric speech representation and synthesis,” *Proc. ICASSP*, 2012, pp. 4505–4508. CORE rank B.
- [51] **G. E. Henter** & W. B. Kleijn, “Intermediate-state HMMs to capture continuously-changing signal features,” *Proc. Interspeech*, 2011, pp. 1817–1820. CORE rank A.
- [52] **G. E. Henter** & W. B. Kleijn, “Simplified probability models for generative tasks: a rate-distortion approach,” *Proc. EUSIPCO*, 2010, pp. 1159–1163. CORE rank B.

Conference publications (not peer reviewed)

- [53] T. Kucherenko, P. Jonell, Y. Yoon, P. Wolfert, Z. Yumak, & **G. E. Henter**, “GENEA Workshop 2021: The 2nd workshop on generation and evaluation of non-verbal behaviour for embodied agents,” *Proc. ICMI*, 2021, pp. 872–873 (extended abstract). CORE rank B.
- [54] T. Kucherenko,* P. Jonell,* Y. Yoon,* P. Wolfert, & **G. E. Henter**, “The GENE Challenge 2020: Benchmarking gesture-generation systems on common data,” *Proc. GENE Workshop*, 2020.
- [55] K. Håkansson, J. Beskow, H. Kjellström, J. Gustafsson, A. Bonnard, M. Rydén, S. Stormoen, G. Hagman, U. Akenine, K. Morales Pérez, **G. E. Henter**, M. Sundell, & M. Kivipelto, “Robot-assisted detection of subclinical dementia: Progress report and preliminary findings,” *Alzheimer’s & Dementia*, vol. 16, no. S6, article no. e043311, 2020 (extended abstract). Presented at Alzheimer’s Association International Conference (AAIC) 2020.
- [56] Z. Malisz, **G. E. Henter**, C. Valentini-Botinhao, O. Watts, J. Beskow, & J. Gustafson, “Modern speech synthesis and its implications for speech sciences,” *Proc. UK Speech*, 2019, pp. 11–12. Selected for oral presentation.
- [57] É. Székely, **G. E. Henter**, J. Beskow, & J. Gustafson, “Spontaneous conversational TTS from found data,” *Proc. UK Speech*, 2019, p. 31.
- [58] O. Watts, **G. E. Henter**, J. Fong, & C. Valentini-Botinhao, “Sequence-to-sequence neural TTS: an assessment of the contribution of various ingredients,” *Proc. UK Speech*, 2019, p. 64.
- [59] Z. Malisz, **G. E. Henter**, C. Valentini-Botinhao, O. Watts, J. Beskow, & J. Gustafson, “The speech synthesis phoneticians need is both realistic and controllable,” *Proc. Fonetik*, 2019, pp. 103–107.
- [60] É. Székely, **G. E. Henter**, J. Beskow, & J. Gustafson, “Spontaneous conversational speech synthesis: The making of a podcast voice – breathing, uhs & ums and some ponderings about appropriateness,” presentation at *Fonetik*, 2019.
- [61] **G. E. Henter**, J. Lorenzo-Trueba, X. Wang, M. Kondo, & J. Yamagishi, “Generating segment-level foreign-accented synthetic speech with natural speech prosody,” *IPSSJ SIG Technical Reports*, vol. 2018-SLP-120, no. 8, pp. 1–3, 2018.
- [62] J. Lorenzo-Trueba, **G. E. Henter**, S. Takaki, J. Yamagishi, “Analyzing the impact of including listener perception annotations in RNN-based emotional speech synthesis,” *IPSSJ SIG Technical Reports*, vol. 2017-SLP-119, no. 8, pp. 1–2, 2017.
- [63] **G. E. Henter**, S. Ronanki, O. Watts, & S. King, “Non-parametric duration modelling for speech synthesis with a joint model of acoustics and duration,” *IEICE Technical Report*, vol. 116, no. 414, pp. 11–16, 2017.
- [64] **G. E. Henter**, S. Ronanki, O. Watts, M. Wester, Z. Wu, & S. King, “Robust text-to-speech duration modelling with a deep neural network,” *Proc. ASA/ASJ*, vol. 140, no. 4, p. 2961, 2016.
- [65] **G. E. Henter**, S. Ronanki, O. Watts, M. Wester, Z. Wu, & S. King, “Robust text-to-speech duration modelling using DNNs,” *Proc. UK Speech*, 2016, p. 44.
- [66] S. Ronanki, **G. E. Henter**, Z. Wu, & S. King, “A template-based approach for intonation generation using LSTMs,” *Proc. UK Speech*, 2016, p. 22.
- [67] **G. E. Henter**, T. Merritt, M. Shannon, C. Mayo, & S. King, “Measuring the perceptual effects of speech synthesis modelling assumptions,” *Proc. UK Speech*, 2014, p. 37.

Books and book chapters

- [68] **G. E. Henter**, S. King, T. Merritt, & G. Degottex, “Analysing shortcomings of statistical parametric speech synthesis” in *Natural Speech Technology* (chapter for book proposal accepted by Cambridge University Press), Eds. S. Renals, M. J. F. Gales, T. Hain, P. C. Woodland, S. King, & S. Cunningham, 2016, 34 pages.
Complete draft available at arxiv.org/abs/1807.10941 (July 2018).

- [69] A. Leijon, & **G. E. Henter**, *Pattern Recognition: Fundamental Theory and Exercise Problems*, 2012. Course book at the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. Most recently updated 2015. 283 pages.
Book webpage: people.kth.se/~ghe/pattern_recognition

Theses

- [70] **G. E. Henter**, “Probabilistic sequence models with speech and language applications,” Ph.D. dissertation, School of Electrical Engineering (EES), KTH Royal Institute of Technology, Stockholm, Sweden, Dec. 2013.
- [71] **G. E. Henter**, “Mathematical techniques for applied demand estimation and profit maximization,” M.Sc. thesis, School of Engineering Sciences (SCI), KTH Royal Institute of Technology, Stockholm, Sweden, May 2007.

Preprints

- [72] S. Mehta, É. Székely, J. Beskow, & **G. E. Henter**, “Neural HMMs are all you need (for high-quality attention-free TTS),” Sept. 2021, 5 pages.
Available at arxiv.org/abs/2108.13320.
- [73] T. Kucherenko, R. Nagy, M. Neff, H. Kjellström & **G. E. Henter**, “Multimodal analysis of the predictability of hand-gesture properties,” Aug. 2021, 10 pages.
Available at arxiv.org/abs/2108.05762.
- [74] A. Ghosh, A. Honoré, D. Liu, **G. E. Henter**, & S. Chatterjee, “Normalizing flow based hidden Markov models for classification of speech phones with explainability,” July 2021, 12 pages.
Available at arxiv.org/abs/2107.00730.
- [75] S. S. Sarfjoo, X. Wang, **G. E. Henter**, J. Lorenzo-Trueba, S. Takaki, & J. Yamagishi, “Transformation of low-quality device-recorded speech to high-quality speech using improved SEGAN model,” Nov. 2019, 5 pages.
Available at arxiv.org/abs/1911.03952.
- [76] **G. E. Henter**, J. Lorenzo-Trueba, X. Wang, & J. Yamagishi, “Deep encoder-decoder models for unsupervised learning of controllable speech synthesis,” July 2018, 17 pages.
Available at arxiv.org/abs/1807.11470 (last updated Sept. 2018).
- [77] S. Phan, **G. E. Henter**, Y. Miyao, & S. Satoh, “Consensus-based sequence training for video captioning,” Dec. 2017, 11 pages.
Available at arxiv.org/abs/1712.09532.
- [78] **G. E. Henter**, A. Leijon, & W. B. Kleijn, “Kernel density estimation-based Markov models with hidden state,” Nov. 2016, 14 pages.
Available at arxiv.org/abs/1807.11320 (last updated July 2018).

Appendix

Excerpts from course evaluations

Gustav Eje Henter
KTH Royal Institute of Technology
Lindstedtsvägen 24
SE-100 44 Stockholm, SWEDEN
☎ +46 70 526 0185
✉ ghe@kth.se
🏠 people.kth.se/~ghe

These quotes are excerpted from the course evaluations of the course EN2200/EN2202 in years 2008, 2009, 2010, and 2012, for which I had access to course evaluations. The quotes are reproduced “as is”, without editing for spelling or grammar. Comments on the course project are included since the project primarily was designed by me.

How good were the exercise assistants at explaining and inspiring to learn?

- I learnt some important points from Gustav in the couple of days before the exam days. Big thanks to him! (2010)
- If a student asked any questions Gustaf was highly motivated on answering the question as if the theme was really interesting. And that motivation from the exercises assistant may gets contagious to the students. Which is really nice! Definitely a bonus point! (2010)
- A very good teacher, he describes things logically and easy to understand, I liked him (2012)
[NB: I was the only exercise assistant this year.]

How useful was the project in helping to learn the course?

- A very good way to learn and to remember when and how to use different algorithms. It made the course extra funny and encouraged me to attend to lectures. I like that the teachers allowed us to have our own project instead of following those described in the book. Keep it! (2012)
- Perfect project, perfect way of learning (2012)

What was the best thing with this course?

- Gustav Henter. Very supportive and giving constructive feedback. (2008)
- Projektet (2008)
- The hands-on-experience in the project. (2009)
- project!!
- Project – fun and perfectly prepared (2009)
- project work teacher and project/exercise contact persons have been always very nict, compentend and helpful! (2010)
- The project was very much fun to do. (2010)
- The project is very interesting. (2010)

- The project: to make a real application (2010)
- Making the recognition system working. (2010)
- Idea itself is very interesting, concept of the course is very good (lectures and project are correlated, project itself is very fun). (2010)
- The project assignment. (2012)
- The practical project. We learned practical knowledge that can be used right a way to build real applications (not topical for the majority of the courses in KTH) (2012)
- The project is certainly the best thing of the course, it is very concrete and matched perfectly with the course (2012)
- of course: the project (2012)
- Project and actual doing something “real”, great view how engineering can actually be later on and specially in these days with smartphones and Co. (2012)
- The project. (2012)
- the practical application (2012)
- The project. (2012)
- course project (2012)
- the project! Partly because we didn't have to implement the whole thing ourselves – that would have made it into another boring programming exercise and not leave much time to understand the theory of the HMM classifier, and partly because of the many inspiring applications demonstrated during the presentations. (2012)