

## An AI revolution? Not without speech and language sciences. How much further can AI go with technology and sciences working together?

Organized by Zofia Malisz (KTH Royal Institute of Technology, Stockholm)

Machine learning methods such as deep neural networks have revolutionised speech and language technologies. It seems modern machines could learn anything just given enough data and powerful algorithms. For example, deep learning-based speech synthesis is now capable of imitating human speech so well that it is indistinguishable from natural speech in behavioural studies (Malisz et al. 2019a). In Natural Language Processing, language prediction models based on very large networks (e.g. GPT-3 by Open AI) continue “writing” a text - perfectly matching the purpose, style and logic of the input sequence - automatically, generating news, novels, poetry or computer code.

This rapid progress has consequences for both language sciences and technology - as new opportunities and goals for research at the intersection of these disciplines emerge. On the one hand, engineers respond to the needs of basic research and create new exploration and modeling methods for e.g.: linguistic and phonetic problems, together with linguists (Doehler Beck et al. 2021). On the other hand, machine learning continues its rapid evolution by involving the human, naive or expert, to optimize both the data it is using (active learning, Einfeldt et al. 2021) as well as its algorithms (human-in-the-loop approaches, Fallgren et al. 2019).

Additionally, fundamental questions remain: we still know very little about how much these systems are able to “understand” about speech and human communication or about the ways in which we can use powerful neural models to test and explain hypotheses in linguistics. Ideally, as scientists, we would like to be able to look into the algorithmic black box and interpret the AI.

This session's goal is to showcase work that explores the above problems and intersections. We invite contributions particularly in the areas of:

- Digital Linguistics, including Digital Phonetics, wherein fundamental linguistic problems and daily practice receive new methodological treatment using modern technologies (including but not limited to machine learning)
- Speech Technology wherein linguistic and, particularly, speech science knowledge is shown to contribute to the improvement and innovation in ASR, TTS, speech conversion or any other classic or new speech and language technology task
- Machine Learning approaches that involve humans or expertise in linguistics i.e.: in active learning, human-in-the-loop approaches or Interpretable AI.

Contributions can take form of original experimental work. We will also accept a small number of position papers on these topics.

## References:

Einfeldt, M., Sevastjanova, R., Zahner-Ritter, K., Kazak, E., & Braun, B. (2021). Reliable estimates of interpretable cue effects with Active Learning in psycholinguistic research. *Proceedings of Interspeech 2021, Brno*, (pp. 1743-1747).

Döhler Beck, G.T., Wennberg, U., Malisz, Z., Henter, G.E. (2022). Wavebender GAN: An architecture for phonetically meaningful speech manipulation. *Proceedings of ICASSP 2022, Singapore*. Fallgren, P., Malisz, Z., & Edlund, J. (2019). How to Annotate 100 Hours in 45 Minutes. In *Proceedings of Interspeech 2019, Graz*, (pp. 341-345)

Malisz, Z., Henter, G. E., Valentini-Botinhao, C., Watts, O., Beskow, J., & Gustafson, J. (2019a). Modern speech synthesis for phonetic sciences: A discussion and an evaluation. In *19th International Congress of Phonetic Sciences, Melbourne, Australia*.