

# SSW11

Aug 26-28

Speech Synthesis  
Workshop

2021  
Hungary

Gárdony, Vital Hotel Nautis\*\*\*\*



## Speech Synthesis Workshop (SSW)

At an international conference on speech processing, a speech scientist once held up a tube of toothpaste (whose brand was "Signal") and, squeezing it in front of the audience, coined the phrase "This is speech synthesis; speech recognition is the art of pushing the toothpaste back into the tube."

One could turn this very simplistic view the other way round: users are generally much more tolerant of speech recognition errors than they are willing to listen to unnatural speech. There is magic in a speech recognizer that transcribes continuous radio speech into text with a word accuracy as low as 50%; in contrast, even a perfectly intelligible speech synthesizer is only moderately tolerated by users if it delivers nothing more than "robot voices". Delivering both intelligibility and naturalness has been the holy grail of speech synthesis research for the past 30 years. More recently, expressivity has been added as a major objective of speech synthesis.

Add to this the engineering costs (computational cost, memory cost, design cost for making another synthetic voice or another language) which have to be taken into account, and you'll start to have an idea of the challenges underlying text-to-speech synthesis.

Major challenges call for major meetings: the Speech Synthesis Workshops (SSWs), which are held every three years under the auspices of ISCA's SynSIG. In 2019 it was decided to have an SSW every two years, since the technology is advancing faster these days. SSWs provide a unique occasion for people in the speech synthesis area to meet each other. They contribute to establishing a feeling that we are all participating in a joint effort towards intelligible, natural, and expressive synthetic speech.

## Call for Papers

The workshop program will consist of a single track with invited talks, oral and poster presentations. Prospective authors are invited to submit original, full-length, 4-6 page papers, including figures and references. Papers can be submitted via the EasyChair website in April-May, 2021.

Please follow the INTERSPEECH 2021 guidelines and templates when preparing your paper. Papers have to follow these guidelines except that we allow up to 6 pages, including figures and references.

Papers will be published in the ISCA online archive.

## Call for Demos

We are planning to have a demo session to showcase new developments in speech synthesis. If you have some demonstrations of your work that does not really fit in a regular oral or poster presentation, please let us know.

## Keynotes

Thomas Drugman *Amazon, Germany*

*Expressive Neural TTS*

István Winkler *Research Centre for Natural Sciences, Hungary*

*Early Development of Infantile Communication by Sound*

Lior Wolf *Facebook AI Research and Tel Aviv University, Israel*

*Deep Audio Conversion Technologies and Their Applications in Speech, Singing, and Music*

## Workshop Topics

Papers in all areas of speech synthesis technology are encouraged to be submitted, including but not limited to:

- Grapheme-to-phoneme conversion for synthesis
- Text processing for speech synthesis (text normalization, syntactic and semantic analysis, intent detection)
- Segmental-level and/or concatenative synthesis
- Signal processing/statistical model for synthesis
- Speech synthesis paradigms and methods; articulatory synthesis, articulation-to-speech synthesis, parametric synthesis etc.
- Prosody modeling, transfer and generation
- Expression, emotion and personality generation
- Voice conversion and modification, morphing (parallel and non-parallel)
- Concept-to-speech conversion speech synthesis in dialog systems
- Avatars and talking faces
- Cross-lingual and multilingual aspects for synthesis (e.g. automatic language switching)
- Applications of synthesis technologies to communication disorders
- TTS for embedded devices and computational issues
- Tools and data for speech synthesis
- Quality assessment/evaluation metrics in synthesis
- End-to-end text-to-speech synthesis
- Direct speech waveform modelling and generation
- Neural vocoding for speech synthesis
- Speech synthesis using non-ideal data ("found", user-contributed, etc.)
- Natural language generation for speech synthesis
- Special topic: Speech uniqueness and deep learning (generating diverse and natural speech)

## Venue

**VITAL HOTEL NAUTIS \*\*\*\*** wellness and conference hotel in Gárdony, the capital of Lake Velence directly on the lakeshore, next to the port and the beach.

## Organizing Committee

Géza Németh *Budapest University of Technology and Economics,*

*Department of Telecommunications and Media Informatics (BME TMIT), Chairman*

Junichi Yamagishi *National Institute of Informatics Japan, University of Edinburgh, UK*

Sébastien Le Maguer *ADAPT Centre/TC.D, Ireland*

Esther Klabbbers *Readspeaker, Netherlands*

Mátyás Bartalis *BME TMIT, Hungary*

Tamás Gábor Csapó *BME TMIT, Hungary*

Bálint Gyires-Tóth *BME TMIT, Hungary*

Gábor Olasz *BME TMIT, Hungary*

Csaba Zainkó *BME TMIT, Hungary*

## Important Dates

April-May, 2021 Paper submission

May, 2021 Registration opens

June, 2021 Notification of acceptance

26-28 August, 2021 Workshop