SSW10

The 10th ISCA Speech Synthesis Workshop

The Austrian museum of folk life and folk art, Laudongasse 15-19, A-1080 Vienna

SSW10 Sponsors

We would like to thank our workshop sponsors for their support.

Acoustics Research Institute

Amazon Alexa

Austrian Academy of Sciences

The International Speech Communication Association (ISCA)
Friday, 20th September

08:00  Registration Opens
09:15  Welcome
09:30 – 10:30  Keynote 1: Deep learning for speech synthesis
               Aäron van den Oord, Google DeepMind, London, UK
10:30 – 10:50  Coffee Break
10:50 – 12:10  Oral Session 1: Neural vocoder
               o  Neural Harmonic-plus-Noise Waveform Model with Trainable Maximum
                 Voice Frequency for Text-to-Speech Synthesis
                 Xin Wang and Junichi Yamagishi
               o  A Comparison of Recent Neural Vocoders for Speech Signal Reconstruction
                 Prachi Govalkar, Johannes Fischer, Frank Zalkow and Christian Dittmar
               o  Deep neural network based real-time speech vocoder with periodic and
                 aperiodic inputs
                 Keiichiro Oura, Kazuhiro Nakamura, Kei Hashimoto, Yoshihiko Nankaku and
                 Keiichi Tokuda
               o  Generative Adversarial Network based Speaker Adaptation for High Fidelity
                 WaveNet Vocoder
                 Qiao Tian and Xucheng Wan

12:10 – 14:00  Lunch Break
14:00 – 15:00  Oral Session 2: Adaptation
               o  Neural Text-to-Speech Adaptation from Low Quality Public Recordings
                 Qiong Hu, Erik Marchi, David Winarsky, Yannis Stylianou, Devang Naik and
                 Sachin Kajarekar
               o  Neural VTLN for Speaker Adaptation in TTS
                 Bastian Schnell and Philip N. Garner
               o  Problem-Agnostic Speech Embeddings for Multi-Speaker Text-to-Speech
                 with SampleRNN
                 David Álvarez de la Torre, Santiago Pascual de la Puente and Antonio Bonafonte
                 Cávez

15:00 – 15:15  SynSIG meeting
15:15 – 15:30  Coffee Break
15:30 – 17:00  Poster Session 1: Voice conversion and multi-speaker TTS
               o  Multi-Speaker Modeling for DNN-based Speech Synthesis Incorporating
                 Generative Adversarial Networks
                 Hiroki Kanagawa and Yusuke Iijima
               o  Speaker Adaptation of Acoustic Model using a Few Utterances in DNN-
                 based Speech Synthesis Systems
                 Ivan Himawan, Sandesh Aryal, Iris Ouyang, Shukhan Ng and Pierre Lanchantin
               o  DNN-based Speaker Embedding Using Subjective Inter-speaker Similarity
                 for Multi-speaker Modeling in Speech Synthesis
                 Yuki Saito, Shinnosuke Takamichi and Hiroshi Saruwatari
               o  Generalization of Spectrum Differential based Direct Waveform
                 Modification for Voice Conversion
                 Wen-Chin Huang, Yi-Chiao Wu, Kazuhiro Kobayashi, Yu-Huai Peng, Hsin-Te
                 Hwang, Patrick Lumban Tobing, Yu Tsao, Hsin-Min Wang and Tomoki Toda
               o  Statistical Voice Conversion with Quasi-periodic WaveNet Vocoder
                 Yi-Chiao Wu, Patrick Lumban Tobing, Tomoki Hayashi, Kazuhiro Kobayashi
                 and Tomoki Toda
               o  Voice Conversion without Explicit Separation of Source and Filter
                 Components Based on Non-negative Matrix Factorization
Hitoshi Suda, Daisuke Saito and Nobuaki Minematsu

- **Voice conversion based on full-covariance mixture density networks for time-variant linear transformations**
  Gaku Kotani and Daisuke Saito

- **Unsupervised Learning of a Disentangled Speech Representation for Voice Conversion**
  Tobias Gburrek, Thomas Glarner, Janek Ebbers, Reinhold Haeb-Umbach and Petra Wagner

- **Novel Inception-GAN for Whispered-to-Normal Speech Conversion**
  Maitreya Patel, Mihir Parmar, Savan Doshi, Nirmesh Shah and Hemant Patil

- **Implementation of DNN-based real-time voice conversion and its improvements by audio data augmentation and mask-shaped device**
  Riku Arakawa, Shinnosuke Takamichi and Hiroshi Saruwatari

**17:30**

Welcome Reception: Gartenpalais Schönborn, Austrian museum of folk life and folk art
Saturday, 21st September

09:30 – 10:30  **Keynote 2: Synthesizing animal vocalizations and modelling animal speech**
Tecumseh Fitch, *University of Vienna, Austria*
Bart de Boer, *Vrije Universiteit Brussel, Belgium*

10:30 – 10:50  **Coffee Break**

10:50 – 12:10  **Oral Session 3: Evaluation and performance**
- Evaluating Long-form Text-to-Speech: Comparing the Ratings of Sentences and Paragraphs
  Rob Clark, Hanna Silen, Tom Kenter and Ralph Leith
- **Speech Synthesis Evaluation — State-of-the-Art Assessment and Suggestion for a Novel Research Program**
  Petra Wagner, Jonas Beskow, Simon Betz, Jens Edlund, Joakim Gustafson, Gustav Eje Henter, Sébastien Le Maguer, Zofia Malisz, Éva Székely, Christina Tännander and Jana Voše
- **Rakugo speech synthesis using segment-to-segment neural transduction and style tokens — toward speech synthesis for entertaining audiences**
  Shuhei Kato, Yusuke Yasuda, Xin Wang, Erica Cooper, Shinji Takaki and Junichi Yamagishi
- **Voice Puppetry: Exploring Dramatic Performance to Develop Speech Synthesis**
  Matthew Aylett, David Braude, Christopher Pidcock and Blaise Potard

12:10 – 14:00  **Lunch Break**

14:00 – 15:00  **Oral Session 4: Speech science**
- **Measuring the contribution to cognitive load of each predicted vocoder speech parameter in DNN-based speech synthesis**
  Avashna Govender, Cassia Valentini-Botinhao and Simon King
- **Statistical parametric synthesis of budgerigar songs**
  Lorenz Gutscher, Michael Pucher, Carina Lozo, Marisa Hoeschele and Daniel Mann
- **GlottDNN-based spectral tilt analysis of tense voice emotional styles for the expressive 3D numerical synthesis of vowel [a]**
  Marc Freixes, Marc Arnela, Francesc Alías and Joan Claudi Socoró

15:00 – 15:15  **Best paper award of past SSW workshops**

15:15 – 15:30  **Coffee Break**

15:30 – 17:00  **Poster Session 2: Applications**
- **Preliminary guidelines for the efficient management of OOV words for spoken text**
  Christina Tännander and Jens Edlund
- **Loss Function Considering Temporal Sequence for Feed-Forward Neural Network-Fundamental Frequency Case**
  Noriyuki Matsunaga, Yamato Ohtani and Tatsuya Hirahara
- **Sparse Approximation of Gram Matrices for GMMN-based Speech Synthesis**
  Tomoki Koriyama, Shinnosuke Takamichi and Takao Kobayashi
- **Speaker Anonymization Using X-vector and Neural Waveform Models**
  Fuming Fang, Xin Wang, Junichi Yamagishi, Isao Echizen, Massimiliano Todisco, Nicholas Evans and Jean-Francois Bonastre
- **V2S attack: building DNN-based voice conversion from automatic speaker verification**
  Taiki Nakamura, Yuki Saito, Shinnosuke Takamichi, Yusuke Ijima and Hiroshi Saruwatari
- **Evaluation of Block-Wise Parameter Generation for Statistical Parametric...**
Speech Synthesis
Nobuyuki Nishizawa, Tomohiro Obara and Gen Hattori

- **Impacts of input linguistic feature representation on Japanese end-to-end speech synthesis**
  Takato Fujimoto, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku and Keiichi Tokuda

- **Low computational cost speech synthesis based on deep neural networks using hidden semi-Markov model structures**
  Motoki Shimada, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku and Keiichi Tokuda

- **Neural iTTS: Toward Synthesizing Speech in Real-time with End-to-end Neural Text-to-Speech Framework**
  Tomoya Yanagita, Sakriani Sakti and Satoshi Nakamura

18:00 **Dinner: Buschenschank Fuhrgassl-Huber, Stadl, Buschenschank Fuhrgassl-Huber, Neustift am Walde 68, A-1190 Wien**
Sunday, 22nd September

09:30 – 10:30  Keynote 3: Natural Language Generation: Creating Text  
Claire Gardent, CNRS/LORIA, Nancy, France

10:30 – 10:50  Coffee Break

10:50 – 12:10  Oral Session 5: Language varieties
- Enhancing Myanmar Speech Synthesis with Linguistic Information and LSTM-RNN  
  Aye Mya Hlaing, Win Pa Pa and Ye Kyaw Thu
- Building Multilingual End-to-End Speech Synthesisers for Indian Languages  
  Anusha Prakash, Anju Leela Thomas, Umesh S and Hema A Murthy
- Diphthong interpolation, phone mapping, and prosody transfer for speech synthesis of similar dialect pairs  
  Michael Pucher, Carina Lozo, Philip Vergeiner and Dominik Wallner
- Subset Selection, Adaptation, Gemination and Prosody Prediction for Amharic Text-to-Speech Synthesis  
  Elshadai Tesfaye Biru, Yishak Tofik Mohammed, David Tofu, Erica Cooper and Julia Hirschberg

12:10 – 14:00  Lunch Break

14:00 – 15:00  Oral Session 6: Sequence to sequence model
- Initial investigation of encoder-decoder end-to-end TTS using marginalization of monotonic hard alignments  
  Yusuke Yasuda, Xin Wang and Junichi Yamagishi
- Where do the improvements come from in sequence-to-sequence neural TTS?  
  Oliver Watts, Gustav Eje Henter, Jason Fong and Cassia Valentini-Botinhao
- A Comparison of Letters and Phones as Input to Sequence-to-Sequence Models for Speech Synthesis  
  Jason Fong, Jason Taylor, Korin Richmond and Simon King

15:00 – 15:15  Best paper award of past SSW workshops

15:15 – 15:30  Coffee Break

15:30 – 17:00  Poster Session 3: Prosody
- Generative Modeling of F0 Contours Leveraged by Phrase Structure and Its Application to Statistical Focus Control  
  Yuma Shirahata, Daisuke Saito and Nobuaki Minematsu
- Subword tokenization based on DNN-based acoustic model for end-to-end prosody generation  
  Masashi Aso, Shinnosuke Takamichi, Norihiro Takamune and Hiroshi Saruwatari
- Using generative modelling to produce varied intonation for speech synthesis  
  Zack Hodari, Oliver Watts and Simon King
- How to train your fillers: uh and um in spontaneous speech synthesis  
  Éva Székely, Gustav Eje Henter, Jonas Beskow and Joakim Gustafson
- An Investigation of Features for Fundamental Frequency Pattern Prediction in Electrolaryngeal Speech Enhancement  
  Mohammad Eshghi, Kou Tanaka, Kazuhiro Kobayashi, Hirokazu Kameoka and Tomoki Toda
- PROMIS: a statistical-parametric speech synthesis system with prominence control via a prominence network  
  Zofia Malisz, Harald Berthelsen, Jonas Beskow and Joakim Gustafson
Deep Mixture-of-Experts Models for Synthetic Prosodic-Contour Generation
Raul Fernandez

Prosody Prediction from Syntactic, Lexical, and Word Embedding Features
Rose Sloan, Syed Sarfaraz Akhtar, Bryan Li, Ritvik Shrivastava, Agustin Gravano and Julia Hirschberg

Sequence to Sequence Neural Speech Synthesis with Prosody Modification Capabilities
Slava Shechtman and Alex Sorin

17:00 – 17:30  Closing Ceremony

INFORMATION

Dinner on Saturday

- The busses for the dinner on Saturday, 21st September to Buschenschank Fuhrgassl-Huber, Neustift am Walde 68, A-1190 Wien leave at 17:30 at Piaristengasse, which is 5 minutes walk from the venue.

- To return back from Buschenschank Fuhrgassl-Huber you need to take the public transport Bus 35A from station Neustift am Walde to the U6 or a taxi. If you do not have a ticket for public transport you need to buy it beforehand, you cannot buy it in the bus.
Pictures and Videos

- During the event pictures and videos are taken.

WLAN

- WLAN: VKM nutzedeimuseum, Password: Frauenwahlrecht
- WLAN: FunkFeuer.at

Lunch options around the venue

- Bierheuriger zum Gangl, Alser Str. 4, 1090 Wien, 01 4091994
- Universitätsbräuhaus, Alser Str. 4, 1090 Wien, 01 4091815
- Gasthaus zur böhmischen Kuchl, Schlößlgasse 18, 1080 Wien, 01 4025731
- La Tavolozza, Florianigasse 37, 1080 Wien, 01 4063757
- Via Toledo Enopizzeria, Laudongasse 13, 1080 Wien, 0660 7551672
- Tunnel Vienna Live, Florianigasse 39, 1080 Wien, 01 9904400

Plan of the venue