# Speak! 2016-2017

Speak! - speech synthesis meetings schedule for 2016/2017

<table>
<thead>
<tr>
<th>Date</th>
<th>Name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.09.16</td>
<td>Cassia</td>
<td>Semester 1 Planning Meeting</td>
</tr>
<tr>
<td>06.10.16</td>
<td>Felipe</td>
<td>A hybrid harmonics-and-bursts modelling approach to speech synthesis (pp. 224–229)</td>
</tr>
<tr>
<td>13.10.16</td>
<td>Gustav</td>
<td>Temporal modeling in neural network based statistical parametric speech synthesis (pp. 113–118)</td>
</tr>
<tr>
<td>20.10.16</td>
<td>Rasmus</td>
<td>Prediction of Emotions from Text using Sentiment Analysis for Expressive Speech Synthesis</td>
</tr>
<tr>
<td>03.11.16</td>
<td>Cassia</td>
<td>TTS quality prediction [paper1] and [paper2 - pp.139-145]</td>
</tr>
<tr>
<td>10.11.16</td>
<td>Srikanta</td>
<td>(1) End to end modelling using neural attention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) Waveform generation based on phase reconstruction from FFT spectral amplitudes</td>
</tr>
<tr>
<td></td>
<td>Junichi</td>
<td>[ongoing work presentation]</td>
</tr>
<tr>
<td>17.11.16</td>
<td>Simon</td>
<td>Multi-output RNN-LSTM for multiple speaker speech synthesis with alpha-interpolation model (p.119-124)</td>
</tr>
<tr>
<td>24.11.16</td>
<td>Catherine</td>
<td>The garden of forking paths and Beyond power calculations: Assessing Type S (sign) and Type M (magnitude)</td>
</tr>
<tr>
<td>01.12.16</td>
<td>[canceled]</td>
<td>Most people are away due to Tom Merritt’s visit</td>
</tr>
<tr>
<td>08.12.16</td>
<td>Oliver</td>
<td>Enhance the word vector with prosodic information for the recurrent neural network based TTS system</td>
</tr>
<tr>
<td>09.12.16</td>
<td></td>
<td>CSTR Christmas Lunch and ILCC Christmas Party</td>
</tr>
<tr>
<td>19.01.17</td>
<td>Sam</td>
<td>Semester 2 Planning Meeting</td>
</tr>
<tr>
<td>26.01.17</td>
<td>Korin</td>
<td>Articulatory-to-Acoustic Conversion with Cascaded Prediction of Spectral and Excitation Features Using Neural Networks</td>
</tr>
<tr>
<td>02.02.17</td>
<td>Gustav</td>
<td>Invertible conditional GANs for image editing (Bonus for interested readers: Inverting the generator of a generative adversarial network)</td>
</tr>
<tr>
<td>09.02.17</td>
<td>Cassia</td>
<td>DNN adaptation using speaker codes (An Investigation of DNN-Based Speech Synthesis Using Speaker Codes ) (extra: Speaker Representations for Speaker Adaptation in Multiple Speakers’ BLSTM-RNN-based Speech Synthesis )</td>
</tr>
<tr>
<td>Date</td>
<td>Name</td>
<td>Event</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>-----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>16.02.17</td>
<td>Oliver</td>
<td>Toshniwal et al. (2016) Jointly learning to align and convert graphemes to phonemes with neural attention models</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gonzalvo et al. (2014) Text-To-Speech with cross-lingual Neural Network-based grapheme-to-phoneme models</td>
</tr>
<tr>
<td>23.02.17</td>
<td>Zack</td>
<td>Feature extraction for emotion detection. Main focus on Florian Eyben et al. (2016) GeMAPS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. Chul Min Lee et al. (2004) Emotion Recognition based on Phoneme Classes</td>
</tr>
<tr>
<td>02.03.17</td>
<td>Gustav</td>
<td>SampleRNN: Unconditional waveform generator and an alternative to WaveNet. Links: Paper to read, reviews, official code, audio examples, discussion on tensorflow-wavenet GitHub, (very preliminary) application to TTS with output</td>
</tr>
<tr>
<td>09.03.17</td>
<td></td>
<td>ICASSP 2017 week and Interspeech 2017 submission approaching</td>
</tr>
<tr>
<td>17.03.17</td>
<td></td>
<td>Interspeech 2017 (final) deadline week - deadline 21.03</td>
</tr>
<tr>
<td>23.03.17</td>
<td>Sam</td>
<td>Draft Dissertation Defense - Suprasegmental representations for the modeling of fundamental frequency in statistical parametric speech synthesis</td>
</tr>
<tr>
<td>30.03.17</td>
<td>Sam</td>
<td>Planning Meeting</td>
</tr>
<tr>
<td>06.04.17</td>
<td>Gustav</td>
<td>Kaneko et al. (2017) Generative adversarial network-based postfilter for statistical parametric speech synthesis (ICASSP 17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Saito et al. (2017) Training algorithm to deceive anti-spoofing verification for DNN-based speech synthesis (ICASSP 17)</td>
</tr>
<tr>
<td>20.04.17</td>
<td>Sam</td>
<td>Montaño and Alias (2017) The role of prosody and voice quality in indirect storytelling speech: A cross-narrator perspective in four European languages</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Montaño and Alias (2016) The role of prosody and voice quality in indirect storytelling speech: Annotation methodology and expressive categories [earlier work - optional]</td>
</tr>
<tr>
<td>27.04.17</td>
<td>Felipe</td>
<td>Felipe’s new vocoder - “Direct Modelling of Magnitude and Phase Spectra for Statistical Parametric Speech Synthesis” [ongoing work presentation]</td>
</tr>
<tr>
<td>04.05.17</td>
<td>Simon</td>
<td>Wang et al (2017) Tacotron: a fully end-to-end text-to-speech synthesis model</td>
</tr>
<tr>
<td>11.05.17</td>
<td>Andrew</td>
<td>Language Modelling for Automatic Music Transcription [ongoing work presentation]</td>
</tr>
<tr>
<td>18.05.17</td>
<td>Rasmus</td>
<td>TTS at Jibo: lessons learned moving from research to industry [presentation]</td>
</tr>
<tr>
<td>25.05.17</td>
<td>Korin</td>
<td>Kakouros and Rasanen (2016) An unsupervised method for the automatic detection of sentence prominence in speech</td>
</tr>
<tr>
<td>Summer</td>
<td>Sam</td>
<td>Summer planning meeting</td>
</tr>
<tr>
<td>08.06.17</td>
<td>Srikant</td>
<td>Hybrid speech synthesis framework based on sequence-to-sequence learning [presentation]</td>
</tr>
</tbody>
</table>
### Suggestions for future presentations

- **tutorial suggestions**
  - statistical analysis of listening experiment results - review methodologies (suggested by Cássia)
  - any signal processing topic (Felipe and Cássia can organize presentation/tutorial)
    - MFCC vs Mel-cepstral coefficients vs MGCs vs Complex cepstrals
    - Alternative parametric representations of speech (LPCs, ...)

- **postponed sessions for future schedules**
  - Cassia’s Overview of SPPL2017 - Workshop on Speech Perception and Production across the Lifespan

- **paper suggestions**