

Appendix 2.2

Report of International Conference Presentation

Name: (Presenter)	Aye Mya Hlaing
Affiliation:	University of Computer Studies, Yangon, Myanmar
Project Title:	Spoof Detection for Automatic Speaker Verification
Name of International Conference: (Link to website)	7th International Conference on Natural Language and Speech Processing, ICNLSP 2024 https://www.icnlsp.org/2024welcome/
Title of Research Paper:	Generative Adversarial Network based Neural Vocoder for Myanmar End-to-End Speech Synthesis
Name of all Co-authors (if any)	Prof. Dr. Win Pa Pa
<p>Comments or feedback received at the conference: (e.g. Questions or comments received by your presentation)</p> <p>The paper contributes as the first effort to Myanmar speech synthesis using neural vocoders. The authors provide a sufficient bibliographical review that contextualizes their work. Additionally, the paper includes details on the training and evaluation processes, ensuring reproducibility and transparency. The results are robust and supported by statistical analyses, including Mean Opinion Score tests, which strengthen the validity of their findings.</p> <p>Questions and comments:</p> <ol style="list-style-type: none">1. Are phoneme sequences used in training based on the spelling rules or orthographic rules?2. The total time of the waveforms used in the corpus should be presented though it is presented with total utterances.3. The quality of the presented models that achieve good results despite the low resource setting (with regard to both language and computing resources) and can therefore serve as an example for other low resource languages.	
<p>Contribution to the project: (e.g. Summary of your session or other sessions related with your presentation)</p> <p>Building the spoof datasets for Myanmar language by applying two neural vocoders and TTS model proposed and developed in the paper</p> <p>This conference covers the scope of –</p> <p>Speech Recognition</p> <p>Speech Translation</p>	

Speech Synthesis

Speaker verification and identification

Paralinguistics of pathological speech and language

Speech technology for disordered speech/hearing

Natural Language Processing

Photos

The screenshot shows a Zoom meeting interface. At the top, there's a toolbar with icons for Audio, Video, Participants (22), Chat, Share, Pause, Annotate, Remote control, Show meeting, and More. Below the toolbar, the title 'Myanmar End-to-End Speech Synthesis' is displayed in blue. The main content area shows a diagram of the speech synthesis process: a 'Phoneme Sequence' is input into 'Tacotron2', which outputs a 'Mel-spectrogram'. This is then processed by a 'Vocoder' block containing 'Parallel WaveGAN' and 'HiFi-GAN', resulting in a 'Speech Waveform'. Below the diagram, there are four bullet points: Tacotron2 (Shen et al., 2018) model was trained for phoneme to mel-spectrogram generation; Tacotron2 is a recurrent sequence-to-sequence feature prediction network with attention that maps phoneme embeddings to mel-spectrograms; Parallel WaveGAN (Yamamoto et al., 2020) and HiFi GAN (Kong et al., 2020) are separately trained on Myanmar speech dataset; The generated mel-spectrogram is then used to synthesize speech waveform. At the bottom, there are four video thumbnails of participants: Aye Mya Hlaing, Nicolas Ballier, aref, and Teodora Vukovic.

2024 ICNLSP

Myanmar End-to-End Speech Synthesis

Phoneme Sequence → Tacotron2 → Mel-spectrogram → Vocoder (Parallel WaveGAN, HiFi-GAN) → Speech Waveform

- Tacotron2 (Shen et al., 2018) model was trained for phoneme to mel-spectrogram generation
- Tacotron2 is a recurrent sequence-to-sequence feature prediction network with attention that maps phoneme embeddings to mel-spectrograms
- Parallel WaveGAN (Yamamoto et al., 2020) and HiFi GAN (Kong et al., 2020) are separately trained on Myanmar speech dataset.
- The generated mel-spectrogram is then used to synthesize speech waveform

Participants: Aye Mya Hlaing, Nicolas Ballier, aref, Teodora Vukovic

[Required Documents]

A) Presentation Materials (e.g., PPT slides)

B) Final Program of the conference

Reporter:

Aye Mya Hlaing

Date:

22/10/2024