INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

Subjective Evaluation of Text-to-Speech Models: Comparing Absolute Category Rating and Ranking by Elimination Tests

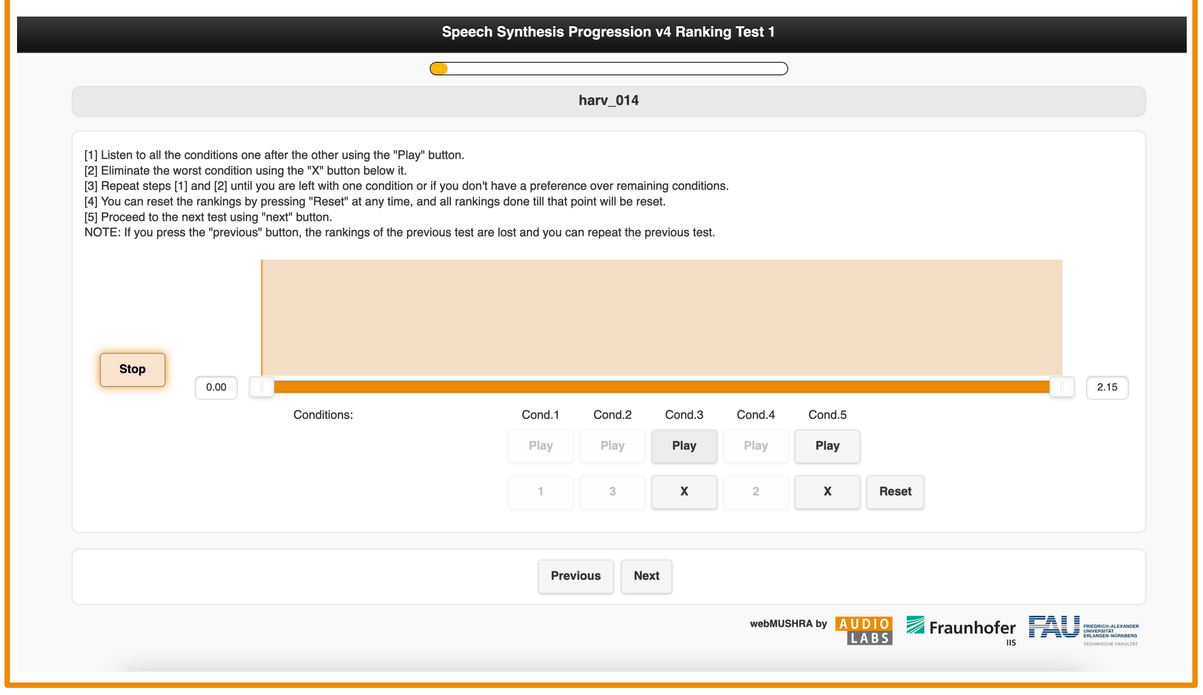
Kishor Kayyar Lakshminarayana, Christian Dittmar, Nicola Pia, Emanuël A.P. Habets

1. Introduction

- Absolute Category Rating (ACR) tests are popular for Text-to-Speech (TTS) model evaluation
- Difficult to evaluate subtle differences through ACR test
  - Conditions presented one after another
  - Conditions cannot be compared against each other
- Paired comparison test (ABX)
  - Can capture subtle differences
  - Exponential growth in number of tests with models
- Multi-Stimulus Hidden Reference and Anchor (MUSHRA) test
  - Not suitable for TTS , 3.5kHz anchor not good
  - Reference in different prosody than test samples
- Could a different test grade TTS models better than ACR, especially when dealing with subtle differences?

2. Ranking By Elimination (RBE) Test

- Eliminate conditions one by one from the worst to the best
- Comparable results to pairwise comparison in audio codec evaluation in less test time
- Resultant ranks analyzed through Plackett-Luce method – Worth score

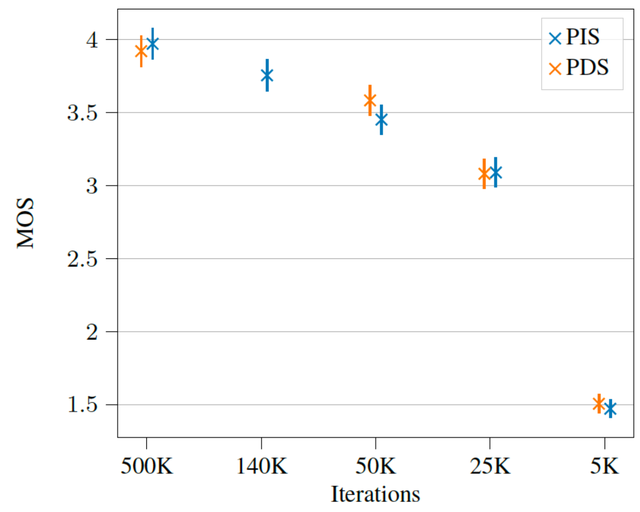


3. Controlled Test Conditions

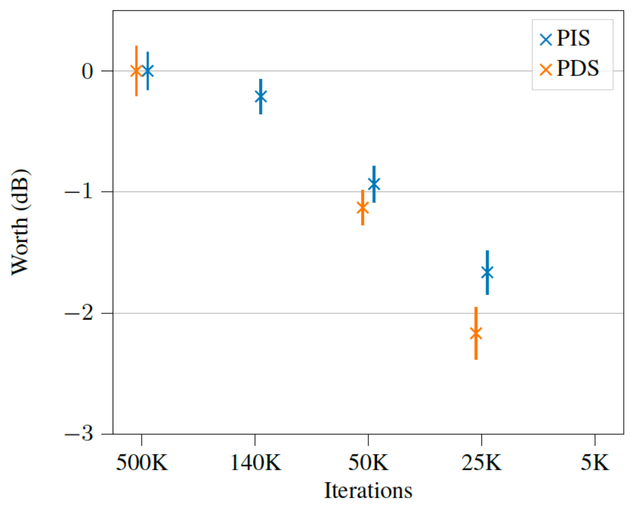
- Intermediate checkpoints of training a ForwardTacotron model with LJ Speech
- Test configuration 1 – Perceptually different – 4 conditions each significantly different from one another (PDS)
- Test configuration 2 – Perceptually similar – an additional condition perceptually close to the best condition (PIS)

Training Iterations	Training Loss	Mel-Cepstral Distortion	Mel-Spectrogram Distortion	F0-Root Mean Square Error
5K	1.382	17.84	32.20	238.87
25K	1.088	15.60	27.96	209.68
50K	0.960	14.40	26.16	198.01
140K	0.930	11.79	21.42	167.49
500K	0.725	-	-	-

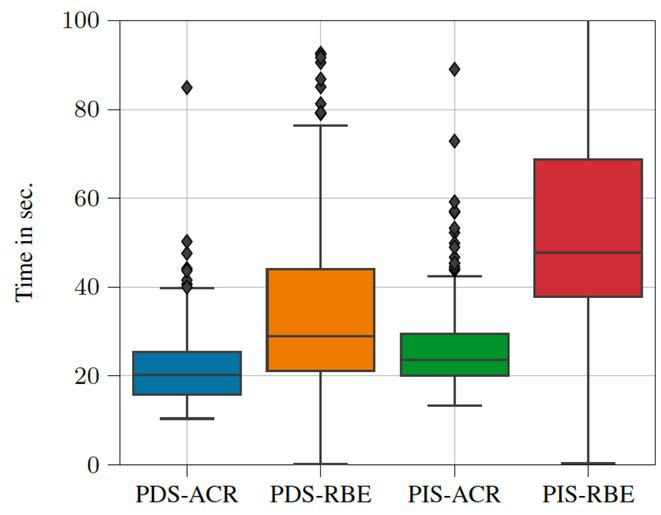
4. Performance Evaluation



ACR Test



RBE Test



Time taken to rate individual items

5. Conclusions

- Applied RBE test for evaluation of TTS models
- Test framework open sourced
- ACR – RBE results comparable

Friedrich-Alexander-Universität Erlangen-Nürnberg
Fraunhofer IIS