ARIA
APPLIED RESEARCH IN ACTION

Speech Generation for Indigenous Language Education

Improving Speech Generation for Low Resource Languages via Spectrogram Denoising

David Guzmán

Annie En-Shiun Lee and Gerald Penn
ACADEMIC SUPERVISORS

Aidan Pine
INDUSTRY SUPERVISOR

Metric	FastSpeech 2	FastSpeech 2 + denoiser
Mel-ceptral distortion	5.231	**5.229**
F0 correlation	0.190	**0.201**
PESQ	1.285	**1.297**

Table 1: Mel-ceptral distortion (smaller is better), F0 correlation (bigger is better) and PESQ (bigger is better)

PROJECT SUMMARY

Recent progress in neural Text-to-Speech (TTS) has resulted in the development of models capable of generating synthetic speech of exceptional quality. However, these models demand substantial volumes of training data, often making these methods inaccessible to low-resource languages, thus leaving them behind. This poses a significant concern, especially in the context of language revitalization, where many endangered languages could greatly benefit from TTS systems, but lack the necessary volume of data. To bridge this gap, we propose an intermediate spectrogram denoising stage that can improve the quality of the speech synthesis output without the need of more training data. This method is especially beneficial for low-resource languages, where obtaining new training data is particularly challenging.

REFERENCES

A. Pine, D. Wells, N. Brinklow, P. Littell, and K. Richmond. Requirements and motivations of low-resource speech synthesis for language revitalization. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 7346–7359, Dublin, Ireland, May 2022. Association for Computational Linguistics.

National Research Council Canada | Conseil national de recherches Canada

Computer Science
UNIVERSITY OF TORONTO
Master of Science in Applied Computing