Normalization in End-to-End TTS for Low-Resource Morphologically Complex Languages

Aleksandr Jan Smoliakov[1]

Supervisor: Gerda Ana Melnik-Leroy

[1]Vilnius University, Faculty of Mathematics and Informatics

2025–10–08

Table of Contents

Problem statement

- **Text-to-Speech (TTS) systems:** Converting written text into human-like speech
- **End-to-End (E2E) neural models:** State-of-the-art approach achieving high quality
- **Challenge:** E2E models struggle with:
  - Non-Standard Words (NSWs): Numbers, dates, abbreviations
  - Low-resource languages with limited training data
  - Morphologically complex languages like Lithuanian
- **Example:** 21 (dvidesimt vieneri) metai vs. 21 (dvidesimt vienas) saldainis

Research Questions:

1. What impact does Lithuanian text normalization have on E2E TTS quality?
2. What is the trade-off between normalization complexity and TTS performance?

Experimental design & Evaluation

Experimental design:

- **Dataset:** Lithuanian speech–text corpora with NSWs and annotations: Common Voice, Liepa 2, audiobooks (?)
- **Normalization levels:** Baseline (none), Rule-based (several levels), Neural (if data permits)
- **TTS models:** Popular E2E architectures (Tacotron 2, FastSpeech 2, VITS)

Evaluation:

Objective:

- Mel cepstral distortion (MCD)
- Duration modeling accuracy

Subjective:

- Mean Opinion Score (MOS)
- NSW intelligibility tests

Progress summary

Completed:

- ✓ Built a homelab
- ✓ Literature review
- ✓ Problem formulation
- ✓ Initial dataset collection: Common Voice, Liepa 2
- ✓ First TTS model trained: Tacotron 2

In Progress:

- Improving baseline TTS model
- Lithuanian normalization rules
- Experimental setup

Planned:

- × Experiments with all normalization levels and TTS models
- × Experimental evaluation
- × Results analysis

Example of generated output

Paskui Tomas issikeite keleta marmuriniu rutuliuku i tris raudonus bilietelius ir dar kelis niekniekius i pora melynu

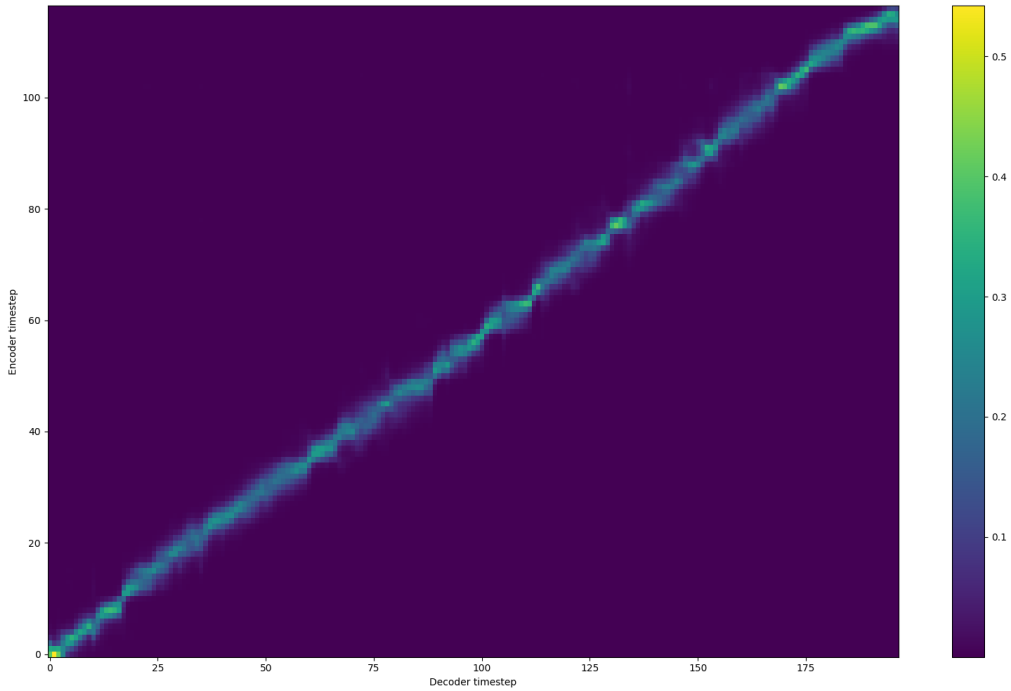

Figure: Sample's input-output alignment from Tacotron 2 model

Audio Sample: Click to play

Thank You!

Thank you for your attention!