

Unified Simplified Grapheme Acoustic Modeling for Medieval Latin LVCSR

Lili Szabó Péter Mihajlik, András Balog, Tibor Fegyó

Your logo here.

Introduction

What is the problem with Latin speech recognition?

- Latin is not a spoken language
- There is no available speech database, and it is resource-heavy to create one
- Many variants/dialects exists, and we can only make guesses about the pronunciation

Baseline Grapheme Model

Source-target grapheme-to-phoneme (G2P) mapping

Unified Simplified Grapheme Model

Conclusions

Four-language USG is the best.