

Unified Simplified Grapheme Acoustic Modeling for Medieval Latin LVCSR

Lili Szabó, Péter Mihajlik, András Balog, Tibor Fegyó

SpeechTex logo here.

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

Text data

Speech data

Test data

Native language of test speakers: CZ, HU, PL, SK

Region of read text: CZ, HU, PL

Baseline Grapheme Model

Languages: Czech (CZ), Hungarian (HU), Polish (PL), Romanian (RO)

- All graphemes are trained
- Only those grapheme models are retained that are part of the Latin alphabet

Table 1: Word Error Rate (WER[%]) results for monolingual grapheme-based acoustic models of Czech, Hungarian, Polish and Romanian (CZ, HU, PL, RO).

AM Language	Speaker				\sum
	CZ	HU	PL	SK	
CZ	53.6	73.8	62.9	45.7	59.0
HU	33.7	28.6	47.1	29.1	34.6
PL	65.0	67.6	46.4	51.1	57.5
RO	53.6	69.1	44.7	43.8	52.8

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

Table 2: WER[%] for Czech-Latin source-target G2P model. Acoustic model training set: 76 hours.

Speaker	Latin Test Text			
	CZ	HU	PL	\sum
CZ	43.8	28.2	49.1	40.4
HU	48.7	40.0	58.7	49.1
PL	53.3	18.2	53.2	41.6
SK	30.3	30.0	44.0	34.8
\sum	43.9	28.9	50.8	41.2

Table 3: WER[%] for Hungarian-Latin source-target G2P model. Acoustic model training set: 567 hours.

Speaker	Latin Test Text			
	CZ	HU	PL	\sum
CZ	19.4	6.4	28.0	17.9
HU	25.0	25.4	20.2	23.5
PL	28.9	15.4	41.3	28.5
SK	20.4	9.1	22.9	17.5
\sum	22.6	12.5	28.1	21.1

Unified Simplified Grapheme Model

Table 4: WER[%] for all the three-language USG models.

AM Language	Speaker				\sum
	CZ	HU	PL	SK	
CZ+HU+PL	28.2	28.2	27.7	22.4	26.6
CZ+HU+RO	23.3	21.4	23.9	19.2	21.9
CZ+PL+RO	24.6	33.1	25.6	19.8	25.8
HU+PL+RO	24.8	21.5	25.7	20.7	23.2

Table 5: WER[%] for USG model of Czech, Hungarian, Polish and Romanian (CZ+HU+PL+RO).

Speaker	Latin Test Text			
	CZ	HU	PL	\sum
CZ	20.4	11.8	30.7	21.0
HU	21.1	14.6	25.7	20.5
PL	23.0	10.0	33.0	22.0
SK	14.5	12.7	24.8	17.3
\sum	19.9	12.2	29.0	20.4

Conclusions

Four-language USG is the best.