# Unified Simplified Grapheme Acoustic Modeling for Medieval Latin LVCSR







**THINKTech** 

Lili Szabó, Péter Mihajlik, András Balog, Tibor Fegyó

lili@speechtex.com



#### Motivation

• Digitizing medieval charters when optical character recognition in not sufficient

### Challenges

- Latin is not spoken natively
- 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
- The pronunciation mainly depends on
- the **era** of the read text
- the **georaphical region** where the text originates from
- the **native language** of the speaker

#### Text data

- In-domain (Monasterium): medieval charters (HU)
- -480k/35k token/type
- Background (Latin Library): historical texts
- 1.3M/115k token/type

**Spelling variants** 

iam

iudex

gratia

### Speech data • CZ: 76 hours

- HU:
- -G2P model: 567 hours
- -GRA and USG models: 112 hours
- PL: 31 hours
- RO: 35 hours

### Test data

- Independent medieval charters read by historians
- Region of test text origin: CZ, HU, PL
- Native language of test speakers: CZ, HU, PL, SK

### Language model

- Kneser-Ney smoothing
- Interpolating the two corpora
- SRILM [2]

jam

judex

gracia

### **Acoustic model**

- 6-hidden-layer DNN
- 2000 neurons per layer
- p-norm activation function
- 7000-11000 senones (softmax size)
- Kaldi toolkit [1]

### • 3-gram language model

System diagram

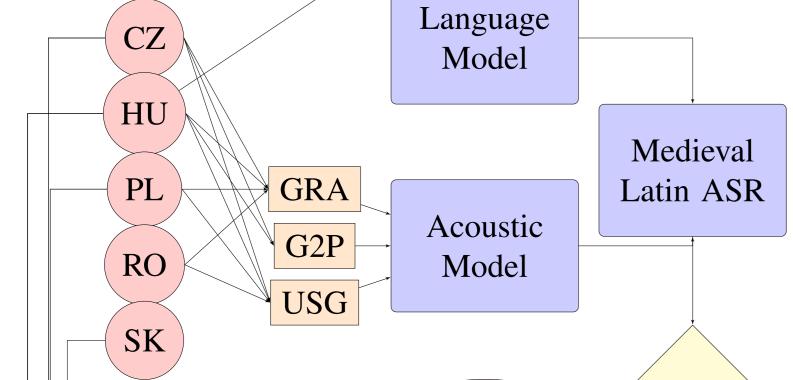
## Perplexity measures on

Table 1: Perplexity/OOV rate (%)

Corpus	CZ	HU	PL	All
Monasterium	551/11.8	82/0.9	3130/18.3	479/10.5
Latin Library	3266/7.8	3549/1.6	2305/5.5	2992/9.7
Interpolated	924/3.9	82/0.9	2288/5.5	672/3.5

### **Dimensions of data**

- Region of training text: Kingdom of Hungary (HU), mixed
- Region of test text origin: Kingdom of Bohemia (CZ), Kingdom of Hungary (HU), Kingdom of Poland (PL)
- Speech data: Czech (CZ), Hungarian (HU), Polish (PL), Romanian (RO)
- Native language of test speakers: CZ, HU, PL, Slovak (SK)
- Model type: GRA, G2P, USG



Training text

**GRA**: baseline grapheme model **G2P**: grapheme-to-phoneme model **USG**: Unified Simplified Grapheme model

Speaker

Test text

Figure 1: Medieval Latin Speech Recognizer

Evaluate

### **Baseline Grapheme Model**

- All graphemes are trained
- Only those grapheme models are retained that are part of the Latin alphabet, e.g.
- -keeping model of r
- throwing away model of ř

Table 2: Word Error Rate (WER[%]) results for monolingual grapheme-based acoustic models of Czech, Hungarian, Polish and Romanian (CZ, HU, PL, RO).

	S				
AM Language	CZ	HU	PL	SK	$\sum$
CZ	53.6	73.8	62.9	45.7	59.0
HU			47.1		l
PL	65.0	67.6	46.4	51.1	57.5
RO	53.6	69.1	44.7	43.8	52.8

### Knowledge-based grapheme-to-phoneme (G2P) mapping

Figure 2: Latin digraph context-insensitive rewrite rules and context-sensitive rewrite rules. V: vowel, VP: palatal vowel, ^VP: everything but a palatal vowel, C: consonant, \*: zero or any, ^: beginning of word,  $\lceil stx \rceil$ : not s, t or x.

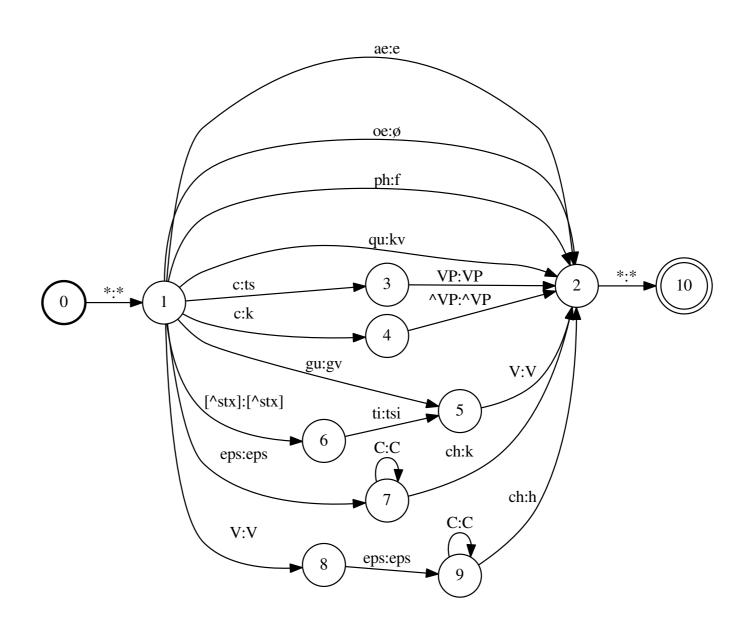


Table 3: WER[%] for Czech-Latin sourcetarget G2P model. Acoustic model training set: 76 hours.

	Latin Test Text					
Speaker	CZ	HU	PL	$\sum$		
CZ			49.1			
HU			58.7			
PL	53.3	18.2	53.2	41.6		
SK	30.3	30.0	44.0	34.8		
$\sum_{i}$	43.9	28.9	50.8	41.2		

Table 4: WER[%] for Hungarian-Latin source-target G2P model. Acoustic model training set: 567 hours.

	Latin Test Text						
Speaker	CZ	HU	PL	$\sum$			
CZ			28.0				
HU			20.2				
PL	28.9	15.4	41.3	28.5			
SK	20.4	9.1	22.9	17.5			
$\overline{\Sigma}$	22.6	12.5	28.1	21.1			

### Unified Simplified Grapheme (USG) Model

• Utilizing many available language resources in the hopes that statistical variations help generalizing over different pronunciations

Table 5: Simplification examples for the unified model.

Language	CZ	HU	PL	RC
Orthographic form	řekl	őz	miś	apa
USG transcription	rekl	ΟZ	mis	apa

Table 6: WER[%] for all the three-language

USG models.

	S				
AM Language	CZ	HU	PL	SK	$\sum$
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

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

Latin Test Text Speaker CZ HU PL \(\sum\_{\text{\tiny{\text{\tiny{\text{\tiny{\tiny{\titilex{\text{\text{\text{\text{\text{\text{\text{\text{\tiny{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\tiny{\text{\ti}}}}\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\texi{\text{\texi{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\ti}}\\ \tittt{\text{\text{\text{\text{\text{\text{\text{\texi}\text{\text{\text{\texi}}}\tint{\text{\text{\text{\text{\texi}\text{\text{\texit{\texi{\texi{\texi{\texi{\texi{\texi{\texi{\texit{\titil\titit{\texi}\til\titt{\texit{\texi{\texi{\texi{\texi{\texi{ 20.4 11.8 30.7 21.0 HU 21.1 14.6 25.7 20.5 23.0 **10.0** 33.0 22.0 SK 14.5 12.7 24.8 17.3 19.9 12.2 29.0 **20.4** 

### **Conclusions**

- Knowledge-based G2P modeling is good, but time consuming and restricted
- Four-language USG modeling is the best
- It is able to generalize over different speaker test sets

### References

- [1] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The kaldi speech recognition toolkit. In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. IEEE Signal Processing Society (2011)
- [2] Stolcke, A.: Srilm an extensible language modeling toolkit. In: In Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP). pp. 901–904 (2002)