



LATVIJAS  
UNIVERSITĀTE

# Latviešu valodas runas sintēzes sistēmu attīstīšana, integrējot lingvistiskās un neironu modeļu pieejas

Viesturs Jūlijs Lasmanis (2<sup>nd</sup> year)

Research supervisor: Dr. sc. comp. Normunds Grūzītis

# About me



## UL IMCS AiLab – scientific assistant

- Mūsdienu latviešu valodas pareizrūnas standartizācija (ParRuna) Izp-2024/1-0613  
[eng. **Standardization of Modern Latvian language pronunciation**]
- Biežākās kļūdas latviešu valodā: korpusā balstīta kļūdu analīze un teksta labošana (Norma) Izp-2023/1-0481  
[eng. **Most common errors in Latvian: corpus-based error analysis and text correction**]



## UL Faculty of Science and Technology – teaching assistant

- Teaching assistant for the Bachelor 3<sup>rd</sup> year course “Fundamentals of Natural Language Processing (DatZB022)”



# Research aims

Improve Latvian language speech recognition and text-to-speech (TTS) synthesis technologies through using phonetic information.

Ability to augment text with information regarding:

- Ambiguous vowel pronunciation (*wide/narrow e, o/ō/uo*);
- Intonations (*zāle/zâle/zàles*);
- Syllable stress notation.

Integrate the ability to provide phonetic information into TTS synthesis pipeline and allow human-in-the-loop corrections.

# Aims for phonetic transcription model

Augment existing speech corpora with automatically generated phonetic transcriptions:

- Conduct data augmentation for TTS synthesis models using audio, phonetic transcription pairs.
- Extract phonetic information from existing speech corpora.

# Intonation recognition model






## wav2vec 2.0

wav2vec 2.0 learns speech representations on unlabeled data as described in [wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations \(Baevski et al., 2020\)](#).

Pretrained Model	Finetune Dataset	# Languages	Phonemizer	Model	Dictionary
LV-60	CommonVoice	26	<a href="#">Espeak</a>	<a href="#">download</a>	<a href="#">download</a>
XLSR-53	CommonVoice	26	<a href="#">Espeak</a>	<a href="#">download</a>	<a href="#">download</a>
XLSR-53	CommonVoice	21	<a href="#">Phonetisaurus</a>	<a href="#">download</a>	<a href="#">download</a>
XLSR-53	CommonVoice, BABEL	21, 19	<a href="#">Phonetisaurus</a>	<a href="#">download</a>	<a href="#">download</a>

We release 2 models that are finetuned on data from 2 different phonemizers. Although the phonemes are all [IPA](#) symbols, there are still subtle differences between the phonemized transcriptions from the 2 phonemizers. Thus, it's better to use the corresponding model, if your data is phonemized by either phonemizer above.

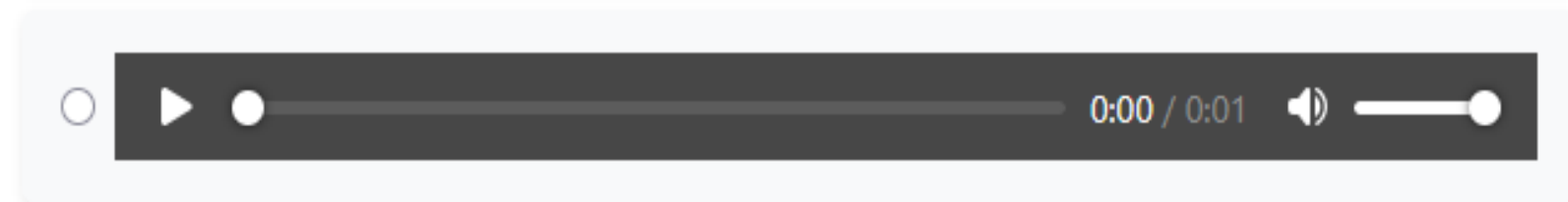
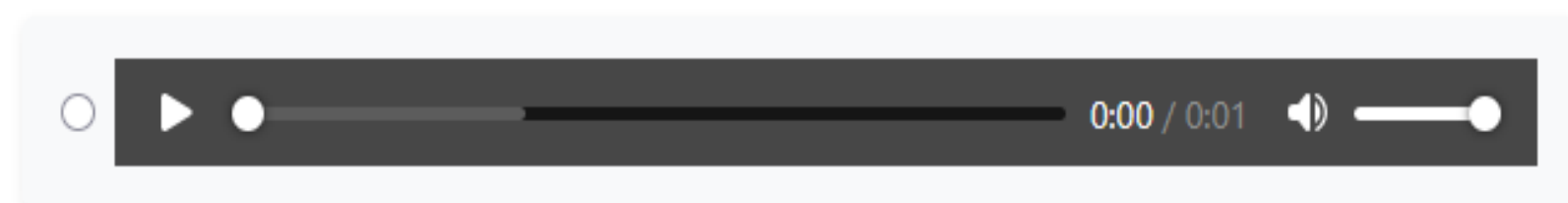
## Leave-One-Subject-Out cross-validation

-  AT1972 (anotēts, pārbaudīts)
-  DT2009 (anotēts, pārbaudīts)
-  LJ2001 (anotēts, pārbaudīts)
-  OR2001 (anotēts, pārbaudīts)
-  VB1936 (anotēts, pārbaudīts)

# Intonation recognition model

Additional evaluation was done on a separate audio file dataset.

Kalns ir ļoti *augsts*.

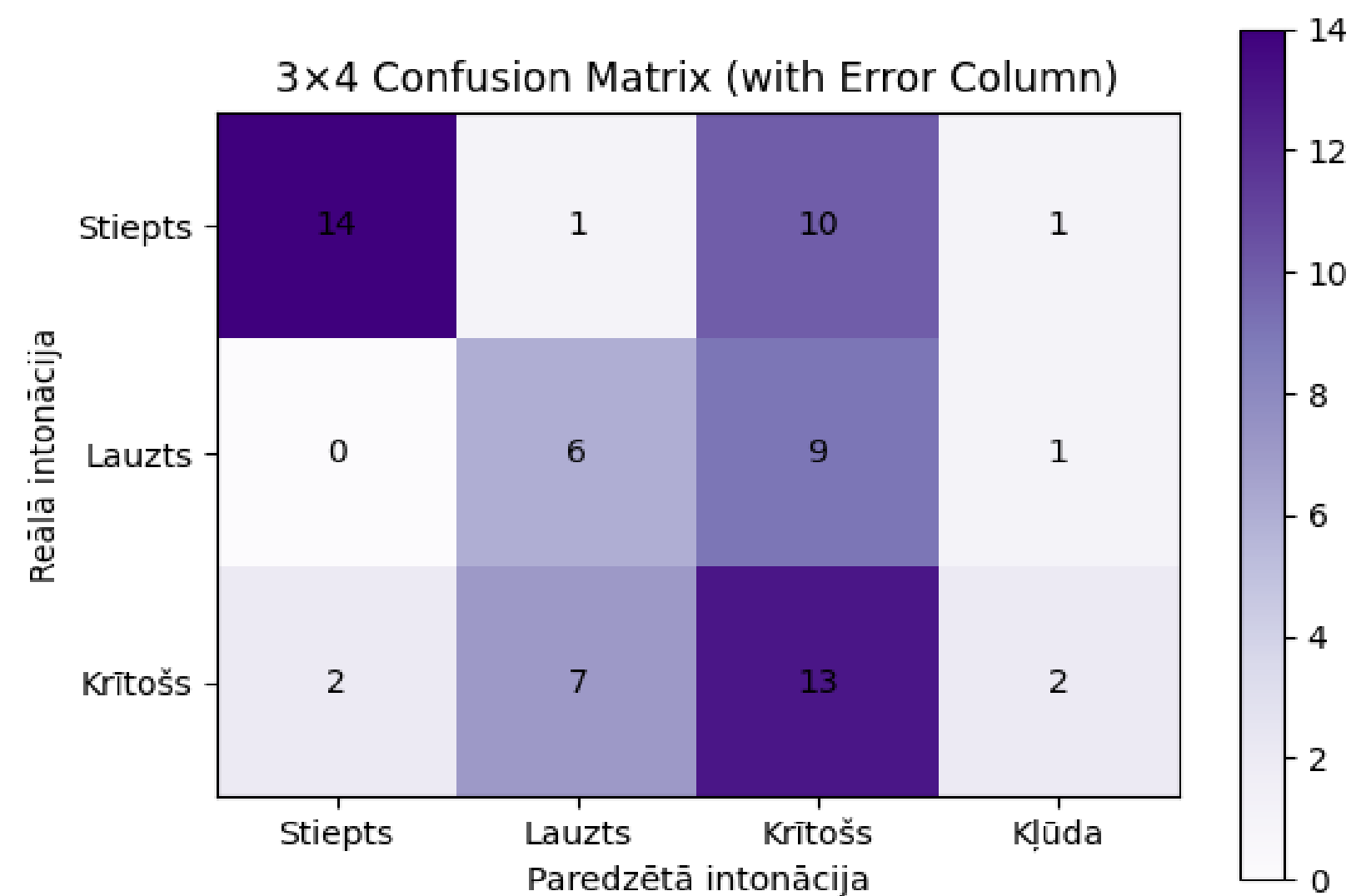


Abi varianti atbilstoši

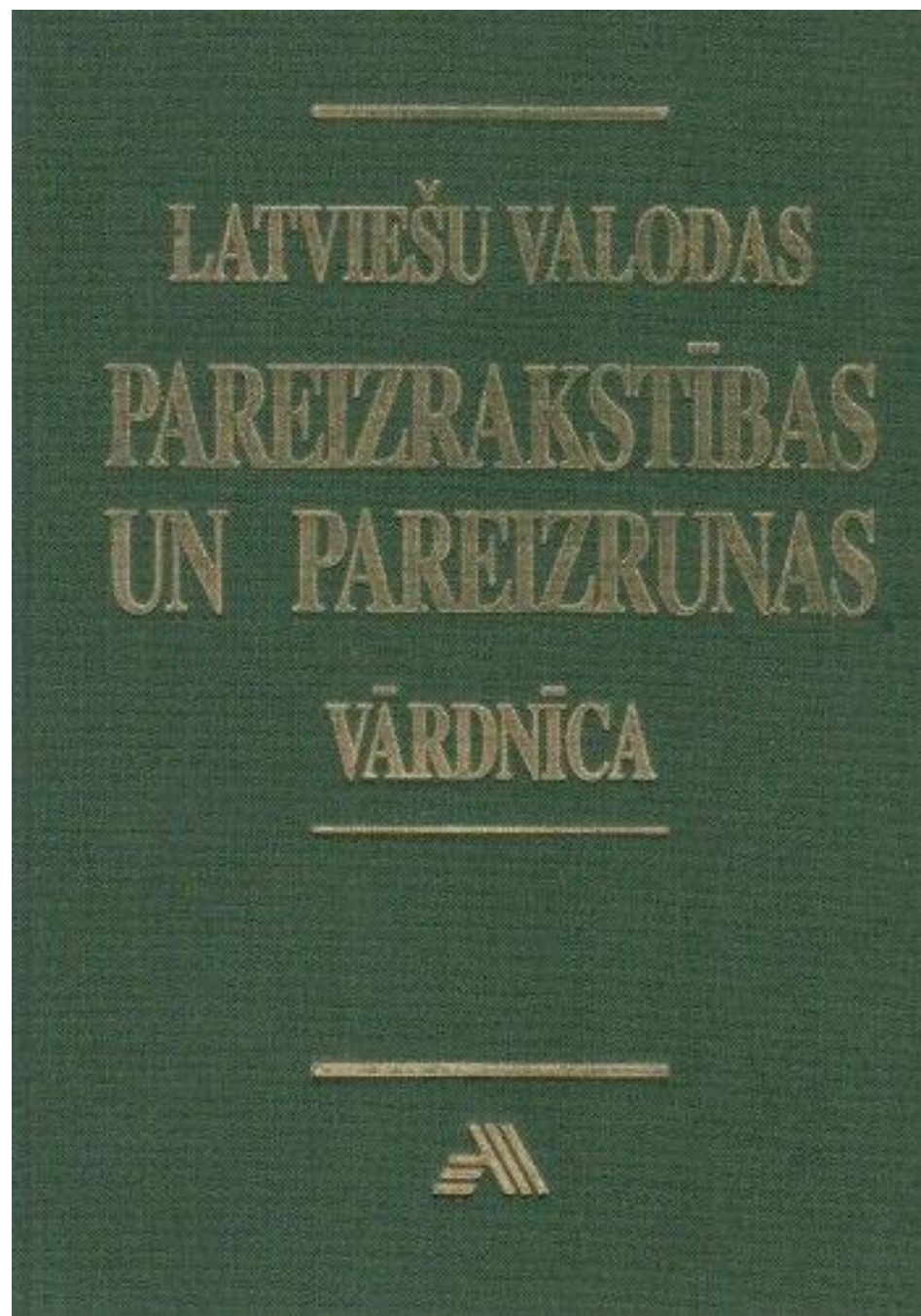
Cits variants

- augsts\_q.wav
- augsts\_q\_t.wav
- augsts\_q\_v.wav
- auksts=.wav
- auksts=t.wav
- auksts=v.wav

3x4 Confusion Matrix (with Error Column)



# Latvian spelling and pronunciation dictionary



**dvest**, dvešu, dves, dveš, dvešam, dvesāt, *pag.* dvesu, dvesām, dvesāt; *pav.* dves, dvesiet; dvesdams [dvēzdāms], dvestu [dvēstu]

**dvie/lis** [dviēlis], -la

**dvinga** [dviņga]

**dvingains** [dviņgaîns]

**dvīņi** [dviņi], *usk.* dvī/nis, -ņa; *s.* -ne, -ņu

**dvīņkristāls** [dviņkristāls]

**dvīņledājs** [dviņlēdājs]

**dvīņogas** [dviņūôgas], *usk.* dvīņoga

**dvīņsēkle/nis** [dviņsēklenis], -ņa

**dvīņser/de** [dviņsefde], -žu

à â ã à

ì î ï ì

è ê ë è

ı ū

ù û ũ ù

ò ô õ ò

ę ę ę ę ę

î ï

î ï

: - \* •

î ï

ı ū

ı ū

ı ū

ı ū

ı ū

# Latvian spelling and pronunciation dictionary

Main problem: No clear quality metrics for evaluation

- Comparison with MLVV

koks<sup>1</sup>

koks [kùoks] lietvārds; vīriešu dzimte

6117 such entries as of  
Autumn 2025 version of MLVV

- Comparison with rules-based transcriptions

# Latvian spelling and pronunciation dictionary

Linguist corrected OCR data could then be used further in research to extract specific intonation data from speech corpora.

	A	B	C	D	E	F
1	Lemma	Biežums	OCR transkripcija	Likumbāzētā transkripcija	Sakriti	
54	iespēja	108589	i_espēja	i_espējax	1	
55	likums	107829	likums	lik:ums	0	
56	punkts	104902	puņkc	puNkc	0	
57	iespējams	103938	i_espējams	i_espējams	1	
58	savukārt	102454	savukārt	savukārt	1	
59	pilsēta	101522	pilsĒ=ta	pilsĒtax	1	
60	noteikt	101507	nu_ote_ikt	nu_ote_ikt	1	
61	gadījums	101440	gadī=jums	gadījums	1	
73	rezultāts	93813	rezul=tā=c	rEzultāc	0	

# Automatic Phonetic Recognition model

## Evaluating 2 fine-tuned Wav2Vec2.0 models on phonetic recognition task.

### Universal Automatic Phonetic Transcription into the International Phonetic Alphabet

*Chihiro Taguchi<sup>1</sup>, Yusuke Sakai<sup>2</sup>, Parisa Haghani<sup>3</sup>, David Chiang<sup>1</sup>*

<sup>1</sup>University of Notre Dame, United States

<sup>2</sup>Nara Institute of Science and Technology, Japan

<sup>3</sup>Google, United States

ctaguchi@nd.edu, sakai.yusuke.sr9@is.naist.jp, parisah@google.com, dchiang@nd.edu

### BalsuTalka.lv – Boosting the Common Voice Corpus for Low-Resource Languages

**Roberts Dargis<sup>\*</sup>, Artūrs Znotiņš<sup>\*</sup>, Ilze Auziņa<sup>\*</sup>, Baiba Saulīte<sup>\*</sup>,  
Sanita Reinsone<sup>†</sup>, Raivis Dejus<sup>‡</sup>, Antra Kļavinska<sup>¶</sup>, Normunds Grūzītis<sup>§</sup>**

<sup>\*</sup>Institute of Mathematics and Computer Science, University of Latvia, Raina bulvaris 29, Riga, Latvia

<sup>†</sup>Institute of Literature, Folklore and Art of the University of Latvia, Mukusalas iela 3, Riga, Latvia

<sup>‡</sup>Latvian Open Technology Association, Raina bulvaris 19-9, Riga, Latvia

<sup>¶</sup>Rezekne Academy of Technologies, Atbrivosanas aleja 115, Rezekne, Latvia

<sup>§</sup>Faculty of Computing, University of Latvia, Raina bulvaris 19, Riga, Latvia

Corresponding authors: {roberts.dargis, arturs.znotins, ilze.auzina, baiba.valkovska}@lumii.lv

ctaguchi/wav2vec2-large-xlsr-japlmthufielta-ipa1000-ns

Automatic Speech Recognition Transformers PyTorch wav2vec2

AiLab-IMCS-UL/wav2vec-xls-r-300m-lv-cv17 like 0 Follow AI Lab at IMCS, Univers... 22

Automatic Speech Recognition Transformers PyTorch wav2vec2 License: apache-2.0

# Automatic Phonetic Recognition model

The fine-tuned Latvian language ASR model outperforms the MultiIPA model.

Pretrained model	Training iters	Testing CER	Testing PER
ctaguchi wav2vec2-large-xlsr-53	3500	0.1370	0.1307
AiLab-IMCS-UL wav2vec2-xls-r-300m	3500	0.0854	0.0826
	5000	0.0835	0.0803

```
=====
MOST FREQUENT PHONEME SUBSTITUTIONS (PER ANALYSIS)
Direction: REFERENCE → PREDICTED
=====
[ā] → [a] | 20 occurrences
[i^] → [j] | 18 occurrences
[ix] → [i] | 15 occurrences
[ux] → [u] | 13 occurrences
[e] → [E] | 13 occurrences
[j] → [i^] | 13 occurrences
[z] → [s] | 13 occurrences
[s] → [z] | 11 occurrences
[a] → [ā] | 10 occurrences
[ē] → [Ē] | 10 occurrences
```

# Current aims

- Completing and submitting an article regarding the Automatic Phonetic Recognition models.
- Work on additional articles regarding intonation recognition and pronunciation dictionary OCR.
- Begin literary review and practical testing of TTS synthesis models using phonetic transcription information.

# Sources

- [1] Auziņa, I.; Rābante-Buša, G. and Dargis, R., (2024). LATE Phonetically Annotated Speech Corpus V1 (fonLATE), *CLARIN-LV digital library at IMCS, University of Latvia*, <http://hdl.handle.net/20.500.12574/115>
- [2] Dargis, R.; Znotins, A.; Auzina, I.; Saulite, B.; S. Reinsone, R. Dejus, A. Klavinska, N. Gruzitis, (2024). BalsuTalka.lv – Boosting the Common Voice Corpus for Low-Resource Languages
- [3] Trumpa, E.; Ozola, A. and Jansone, L. P., (2024). Dataset for Latvian Phonetic Analysis, *CLARIN-LV digital library at IMCS, University of Latvia*, <http://hdl.handle.net/20.500.12574/122>.
- [4] Auziņa, I., (2005). Latviešu valodas izrunas datormodelēšana. *Latvijas Universitāte*
- [5] Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33, 12449-12460.
- [6] Taguchi, C., Sakai, Y., Haghani, P., Chiang, D. (2023) Universal Automatic Phonetic Transcription into the International Phonetic Alphabet. *Proc. Interspeech 2023*, 2548-2552, doi: 10.21437/Interspeech.2023-2584
- [7] Roberts Dargis, Arturs Znotins, Ilze Auzina, Baiba Saulite, Sanita Reinsone, Raivis Dejus, Antra Klavinska, and Normunds Gruzitis. 2024. BalsuTalka.lv - Boosting the Common Voice Corpus for Low-Resource Languages. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2080–2085, Torino, Italia. ELRA and ICCL.
- [8] L. Ceplītis, A. Miķelsons, T. Porīte, and S. Raģe. (1995). Latviešu valodas pareizrakstības un pareizrunas vārdnīca. *Avots*.
- [9] Strankale, L., & Paikens, P. (2020). OCR Challenges for a Latvian Pronunciation Dictionary. In *Human Language Technologies–The Baltic Perspective* (pp. 199-206). IOS Press.

Thank you for your attention!