



LATVIJAS
UNIVERSITĀTE

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

Viesturs Jūlijs Lasmanis (1. gads)

Darba vadītājs: Dr. sc. comp. Normunds Grūzītis

Darba izstrādes vieta



LU MII AiLab – zinātniskais asistents

- Mūsdienu latviešu valodas pareizrūnas standartizācija (ParRuna), Izp-2024/1-0613
- Biežākās kļūdas latviešu valodā: korpusā balstīta kļūdu analīze un teksta labošana (Norma), Izp-2023/1-0481



LU EZTF – vecākais eksperts

- Valodu tehnoloģiju iniciatīva (VTI):
- Asistents datorikas nodaļas BSP kursam “Valodu tehnoloģiju pamati (DatZB022)”



LATVIJAS
UNIVERSITĀTE

Darba mērķi

Uzlabot runas sintēzes precizitāti, pievienojot tekstam fonētisko informāciju.

Spēt papildināt tekstu ar:

- Patskaņu lietojumu (*platais/šaurais e, o/ō/uo*);
- Intonācijām (*zāle/zâle/zàles*);
- Uzsvāru notāciju.

Integrēt fonētisko informāciju balss sintēzes procesā, lai lietotājs varētu rediģēt.

Literatūras izpēte

- Kādas ir latviešu valodas fonētiskās īpašības, kuras ir problemātiskas runas sintēzes modeļiem?
- Kādi ir eksistējošie rīki priekš fonētiskās transkripcijas un balss sintēzes?
- Kas ir veikts šajā jomā citām mazresursu valodām?
- Kādi dati un rīki ir pieejami latviešu valodai?

Literatūras izpēte

Latviešu valodas fonētika un latviešu valodas mašīnlasāmais fonētiskais alfabēts:

Burts	Fonētiskās transkripcijas sistēmas				Piemēri				
	LFA ²	IPA ³	LVMFA (iekšējais) ⁴	LVMFA (fonLATE) ⁵	Vārds ortogr.	LFA	IPA	LVMFA (iekšējais)	LVMFA (fonLATE)
Stieptā zilbes intonācija (~)					<i>zāle</i>	[zāl ^ē]	[za::l ^ē]	zā=lex	zā=lex
~	:	=	=	<i>skola</i>	[buŋg ^ā s]	[skuo:lā]	sku_o=lax	skuo=lax	
				<i>bungas</i>	[skuōl ^ā]	[buŋ:gās]	buN=gaxs	buN=gaxs	
Krītošā zilbes intonācija (`)					<i>lēni</i>	[l ^ē n ⁱ]	[læ:nī]	lĒnix	lĒnix
`	(nenorād a)	(nenorāda)	(nenorāda)	<i>maize</i>	[māiz ^ē]	[maiz ^ē]	ma_izex	maizex	
				<i>salna</i>	[saln ^ā]	[salnā]	salnax	salnax	
Lauztā zilbes intonācija (^)					<i>zāle</i>	[zāl ^ē]	[za:ʔl ^ē]	zāqlax	zāqlax
^	_ʔ	_q	_q	<i>ieļa</i>	[iēl ^ā]	[iēʔlā]	i_eqlax	ieqlax	
				<i>darbs</i>	[darps]	[darʔps]	darqps	darqps	

Pētījumā aplūkotie datu korpusi



fonLATE

LATE fonētiski marķēts runas korpus

2012–2024, 4 stundas (48 000 tekstvienību)

Izstrādātāji: LU MII

BalsuTalka

Balsutalka.lv runas korpus (Common Voice 17.0)

2023–2024, 277 stundas (1,3 milj. tekstvienību)

Izstrādātāji: LU MII, LU LFMI, LATA



Dataset for Latvian Phonetic Analysis

✍ Authors	Trumpa, Edmunds ; Ozola, Anete and Jansone, Laura Paula
🔗 Item identifier	http://hdl.handle.net/20.500.12574/122
📅 Date issued	2024-12-19
📁 Type	audio, corpus
📏 Size	855 utterances
🗣 Language(s)	Latvian

Igauņu valodai veikto pētījumu analīze

Estonian TTS with Non-autoregressive Transformers

453

Method	Mari	Vesta	Meelis	Average
Grapheme	3.68 ± 0.2	3.83 ± 0.17	3.51 ± 0.24	3.67 ± 0.12
Vabamorf	3.99 ± 0.2	4.11 ± 0.17	3.52 ± 0.23	3.87 ± 0.12
Phoneme	2.66 ± 0.19	2.92 ± 0.22	2.54 ± 0.23	2.71 ± 0.12
Grapheme, multi-speaker	3.96 ± 0.18	3.84 ± 0.16	3.98 ± 0.18	3.93 ± 0.1
Vabamorf, multi-speaker	4.04 ± 0.2	3.98 ± 0.16	4.2 ± 0.18	4.07 ± 0.11
Phoneme, multi-speaker	2.9 ± 0.21	2.93 ± 0.22	2.63 ± 0.23	2.82 ± 0.13

Table 3: Mean opinion scores with 95% confidence intervals on out-of-domain data.

ÜKSIKSÕNADE SÜNTEES

Sisesta sõna KOOS MÄRKIDEGA:

- < kolmandavärtelise silbi täishääliku ees (k<oeri, kub<ism)
- ? ebaregulaarne rõhk rõhulise silbi vokaali ees (rak?etiga, kr<eekl?anna)
-] palataliseeritud kaashääliku (l, n, s, t, d) järel (p<a]k, k<õ]lama)
- _ liitsõnapiir (rõdu_<uks)

Kui tunnete EKI sõnastike märke paremini kui Vabamorfi märke, siis kasutage ` ` +

Sisesta sõna

[5] Rätsep, L., Lellep, R. and Fishel, M., (2022). Estonian Text-to-Speech Synthesis with Non-autoregressive Transformers. *Baltic Journal of Modern Computing*, 10(3).

[6] Kiissel, I., Piits, L., Sähkai, H., Hein, I., Ermus, L. and Mihkla, M., (2025, March). Estonian isolated-word text-to-speech synthesiser. *In Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025)* (pp. 302-306).

Eksperimentālā daļa

Fonētiskā pierakstā esoša teksta sintezēšanai nepieciešama datu kopa ar:

- kvalitatīviem audio ierakstiem,
- tiem atbilstošas fonētiskās transkripcijas.

Galvenais avots ar šāda veida datiem in fonLate

fonLate

- Trūkst intonāciju marķējuma
- Satur tikai 4 stundas ar marķētiem datiem

fonLATE

LATE fonētiski marķēts runas korpuss

2012–2024, 4 stundas (48 000 tekstvienību)

Izstrādātāji: LU MII

1	0.0000000	0.0322433	k	1	0.0000000	0.1264931	kad
2	0.0322433	0.0669669	a	2	0.1264931	0.5258144	Rīgā
3	0.0669669	0.1264931	d	3	0.5258144	1.2686512	plosījās
4	0.1264931	0.1996606	r				
5	0.1996606	0.3311142	ii				
6	0.3311142	0.3831996	g				
7	0.3831996	0.5258144	aa				
8	0.5258144	0.6076628	p				
9	0.6076628	0.6746298	l				
10	0.6746298	0.7887216	uo				
11	0.7887216	0.9102541	s				
12	0.9102541	1.0888326	ii				
13	1.0888326	1.1111549	j				
14	1.1111549	1.2004441	aa				
15	1.2004441	1.2686512	s				

Datu papildināšanas iespējas

- Datu kopā, kuru paredzēts pielietot fonētiskajai balss sitēzei, trūkst informācijas par intonācijām.
- Fonētiskā marķēšana ir valodniekiem laikietilpīgs process, intonāciju marķēšana būtu būtiski paildzinājusi šo procesu.
- Intonāciju modelis ļautu vokāļa skaņas audio segmentam paredzēt un automātiski piešķirt intonācijas marķējumu.

Intonāciju izšķiršanas modeļa apmācība

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	Espeak	download	download
XLSR-53	CommonVoice	26	Espeak	download	download
XLSR-53	CommonVoice	21	Phonetisaurus	download	download
XLSR-53	CommonVoice, BABEL	21, 19	Phonetisaurus	download	download

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 šķērsvalidācija

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

Intonāciju izšķiršanas modelis

Pašlaik uz LaVI intonāciju datiem apmācītais modelis:

- Spēj atšķirt 2-toņu sistēmu: nestiepto un stiepto

Runātājs evaluācijas datu kopā	AT1972		DT2009		LJ2001		OR2001		VB1936	
Kļūdu matrica	533	26	520	39	536	23	519	40	506	53
	43	249	27	265	34	258	38	254	44	248
F1-score	0.878		0.889		0.901		0.867		0.836	

F1 = 0.874 ± 0.031

- 3-toņu sistēmu uz esošajiem datiem neizdevās sekmīgi apmācīt

Runātājs evaluācijas datu kopā	AT1972			DT2009			LJ2001			OR2001			VB1936		
Kļūdu matrica	272	0	23	273	0	22	271	0	24	274	0	21	128	156	11
	256	0	8	248	0	16	243	0	21	247	0	17	100	158	6
	45	0	247	26	0	266	28	0	264	52	0	240	83	12	197
Svērtais F1-score	0.515			0.531			0.526			0.508			0.580		

F1 = 0.532 ± 0.035

Fonētiskās transkripcijas modelis

Kādēļ papildus transkripcijas modeli:

- Salīdzināt teksta sintēzes modeļa apmācībā vai ir iespējams uzlabot precitāti ar automātiski papildinātiem datiem
- Iegūt fonētisko izrunu konkrētiem vārdiem no runāta teksta

Automātiskas fonētiskās transkripcijas modeļa arhitektūra

Pašlaik izstrādes procesā:

- Pielietojot fonLate korpusa datus
- wav2vec 2.0 priekšapmācīta modeļa fine-tuning
- Pagaidām ir izveidota arhitektūra, lai varētu apmācīt modeli uz esošajiem datiem, kā arī padot CommonVoice audio failus marķēršanai.

Fonētiskās transkripcijas modelis

3 atšķirīgi modeļi:

- Intonāciju modelis
- Fonētiskās transkripcijas modelis
- Whisper latviešu valodas modelis vārdu līmenim

Fonētiskās transkripcijas modeļa izvade

Labā gadījumā iegūst izrunāto vārdu fonētiskā un ortogrāfiskā pierakstā ar laika intervālu.

1	0.92 1.42	ta_iqznī~ba_iq	taishnībai
2	1.60 1.86	da_uqt_s	daudz
3	1.98 2.68	preteni_e~kux	pretinieku
4	2.74 3.08	labamq	labam
5	3.16 3.88	t_sil~vĒ~kamq	cilvēkam
6	3.96 4.22	da_u~d_z	daudz
7	4.38 5.30	i_eqna_i~ni_e~kux	ienaidnieku

1	0.98 1.96	svisealāqb	viss
2	2.00 2.02	ax	ir
3	2.24 3.02	sna_u~šatstū~	labi

```
{  
  "epoch": 28.571428571428573,  
  "eval_cer": 0.13826281592936862,  
  "eval_loss": 0.6204245686531067,  
  "eval_runtime": 9.0854,  
  "eval_samples_per_second": 61.527,  
  "eval_steps_per_second": 7.705,  
  "step": 4000  
}
```

Pašreizējās fonētiskā modeļa problēmas

- Trokšņa noskīršana
- Uzsvāra noteikšana
- Pusgaru un dubultotu līdzskaņu atpazīšana
- Piekārtošana Whisper vārdiem

Turpmākie mērķi

- Publicēt rakstu(s) par intonāciju un fonētiskās transkripcijas modeļiem
- Apmācīt jaunu intonāciju modeli uz jauniem datiem vai arī kopējiem CommonVoice datiem.
- Optimizēt modeļa apmācības hiperparametrus un patestēt citus wav2vec 2.0 priekšapmācītos modeļus
- Transkribēt CommonVoice latviešu valodas datus fonētiskajā notācijā

Izmanotā literatūra

- [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., Reinsone, S., Dejus, R., Klavinska, A., Gruzitis, N. (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] Rätsep, L., Lellep, R. and Fishel, M. (2022). Estonian Text-to-Speech Synthesis with Non-autoregressive Transformers. *Baltic Journal of Modern Computing*, 10(3).
- [6] Kiissel, I., Piits, L., Sahkai, H., Hein, I., Ermus, L. and Mihkla, M. (2025, March). Estonian isolated-word text-to-speech synthesiser. *In Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025)* (pp. 302-306).
- [7] Baeveski, 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.
- [8] Dargis, R., Auziņa, I. (2022). Ilvars - Latvian Male VITS Text-to-Speech Model (vers. 2022), *CLARIN-LV digital library at IMCS, University of Latvia*, <http://hdl.handle.net/20.500.12574/71>.

Paldies par uzmanību!