Modeling Regressive Vowel Harmony from Continuous Speech Stream

Sneha Ray Barman, Shakuntala Mahanta

Indian Institute of Technology Guwahati, India

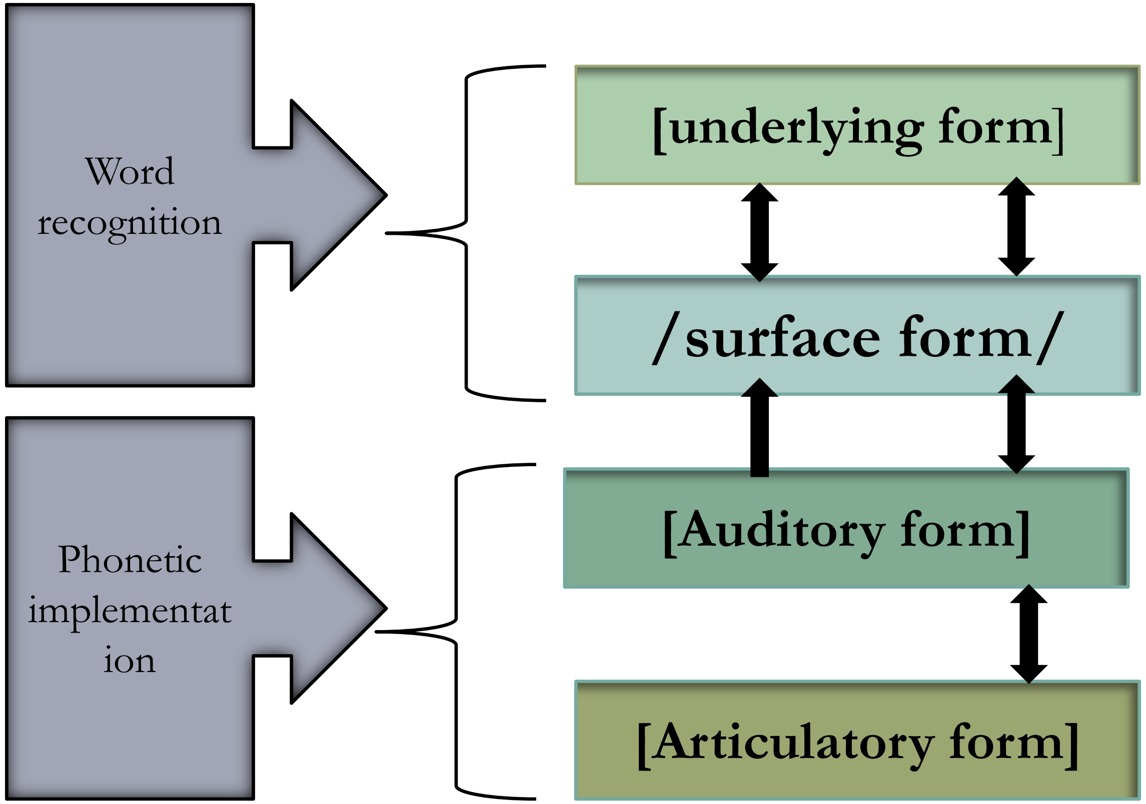
sneha.barman@iitg.ac.in, smahanta@iitg.ac.in

CLST IITG CENTRE FOR LINGUISTIC SCIENCE & TECHNOLOGY

SCiL SOCIETY FOR COMPUTATION IN LINGUISTICS

27th June 2024

Phonological learning



Phonological production

Prelexical perception

Boersma et al. (2012)

Existing models and more!

Approach	Author(s)	Common Factor
Optimality Theory	Prince and Smolensky (2004)	Curated **text** data; Supervised and/or semi-supervised training.
Harmonic Grammar	Legendre et al. (1990)	
Maximum Entropy Grammar	Goldwater & Johnson (2003); Hayes & Wilson (2008)	
Computational Models	Kirove & Cotterell (2018); Mayer & Nelson (2020); Prickett et al. (2022)	

Unsupervised Modeling of Vowel Harmony- **WHY?**

- Vowel harmony is widespread yet nowhere near universal.
- Involves learning crucial factors like features, domains, directionality, iterativity, and opacity (Archangeli & Pulleybank 2007).
- Regular patterns while also accommodating exceptions.
- Raw speech ~ The input received by a child (approx.)
- Unsegmented, unlabeled.
- Probably easy to infer phonology from readily available language data.

Assamese

- An Indo-Aryan language spoken across the state of Assam
- Spoken by 15 million people, according to 2011 Census of India
- 20 consonants and 8 surface vowels.

Phonemic inventory of Assamese

Consonants:	Bilabial		Alveolar		Palatal	Velar		Glottal
Stops	p	b	t	d		k	g	
	p^h	b^h	t^h	d^h		k^h	g^h	
Nasals		m		n			ŋ	
Fricatives			s	z		x		h
Approximants				ɹ	j		w	
Lateral approximant				l				

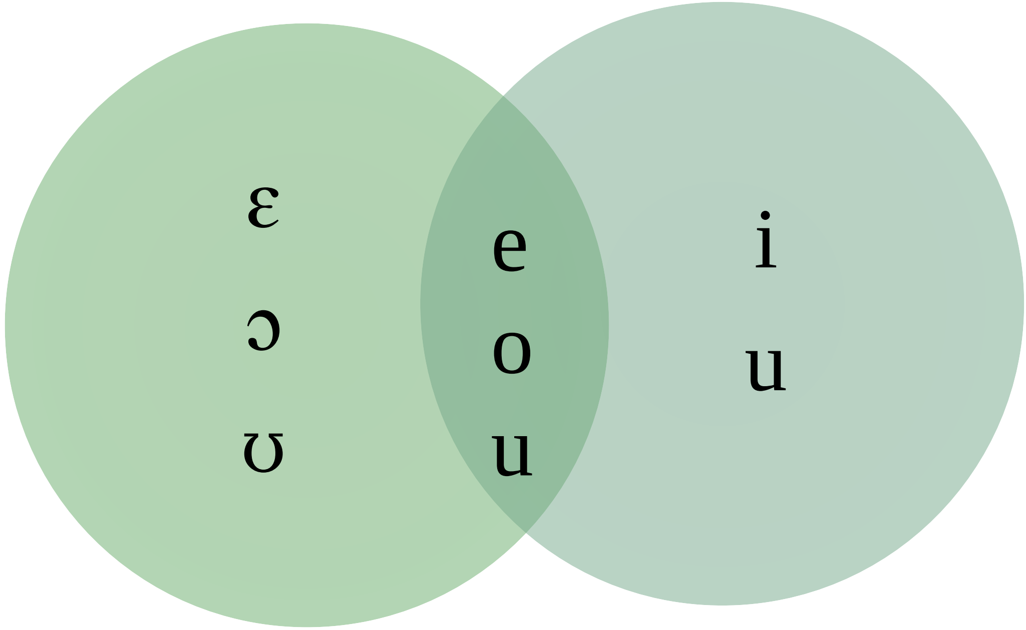
Fig 1. Consonants in Assamese (Mahanta 2007)

Vowels:	Front	Back	ATR
High	i	u	+ATR
		ʊ	-ATR
Mid	e	o	+ATR
	ɛ	ɔ	-ATR
Low		ɑ	-ATR

Fig 2. Vowels in Assamese (Mahanta 2007)

Assamese Vowel Harmony

- Feature
- Domain
- Directionality
- Iterativity
- Opacity



Features

- Target: The vowel that changes its vocalic properties
- Trigger: The vowel that induces the change

- **Example:**

- [-ATR] vowels become [+ATR] when followed by [+high, +ATR] vowels.

p **ɛ** t 'belly' → p ***e*** t -**u** 'pot-bellied'

p ɑ g **ɔ** l 'mad-M' → p ɑ g **o** l –**i** 'mad-F'

Vowels:	Front	Back	ATR
High	i	u	+ATR
		ʊ	-ATR
Mid	e	o	+ATR
	ɛ	ɔ	-ATR
Low		ɑ	-ATR

Fig 2. Vowels in Assamese (Mahanta 2007)

Directionality

- **Regressive:** Righ-to-left harmony
- Examples:

k **o** b **i** t̪ **ɑ** ‘poem’

p **ɑ** g **ɔ** l ‘mad-M’   p **ɑ** g **o** l –**i** ‘mad-F’

Domain

Non-local harmony: Trigger vowels target all the vowels.

- Example: p **o** d **o** b **i** ‘surname’

  b **o** s **o** r **i** ‘yearly’

Iterativity

- **Long-distance iterative harmony**
- Examples:

p **o** d **o** b **i** - 'surname'

b **o** s **o** r **i** - 'yearly'

Opacity

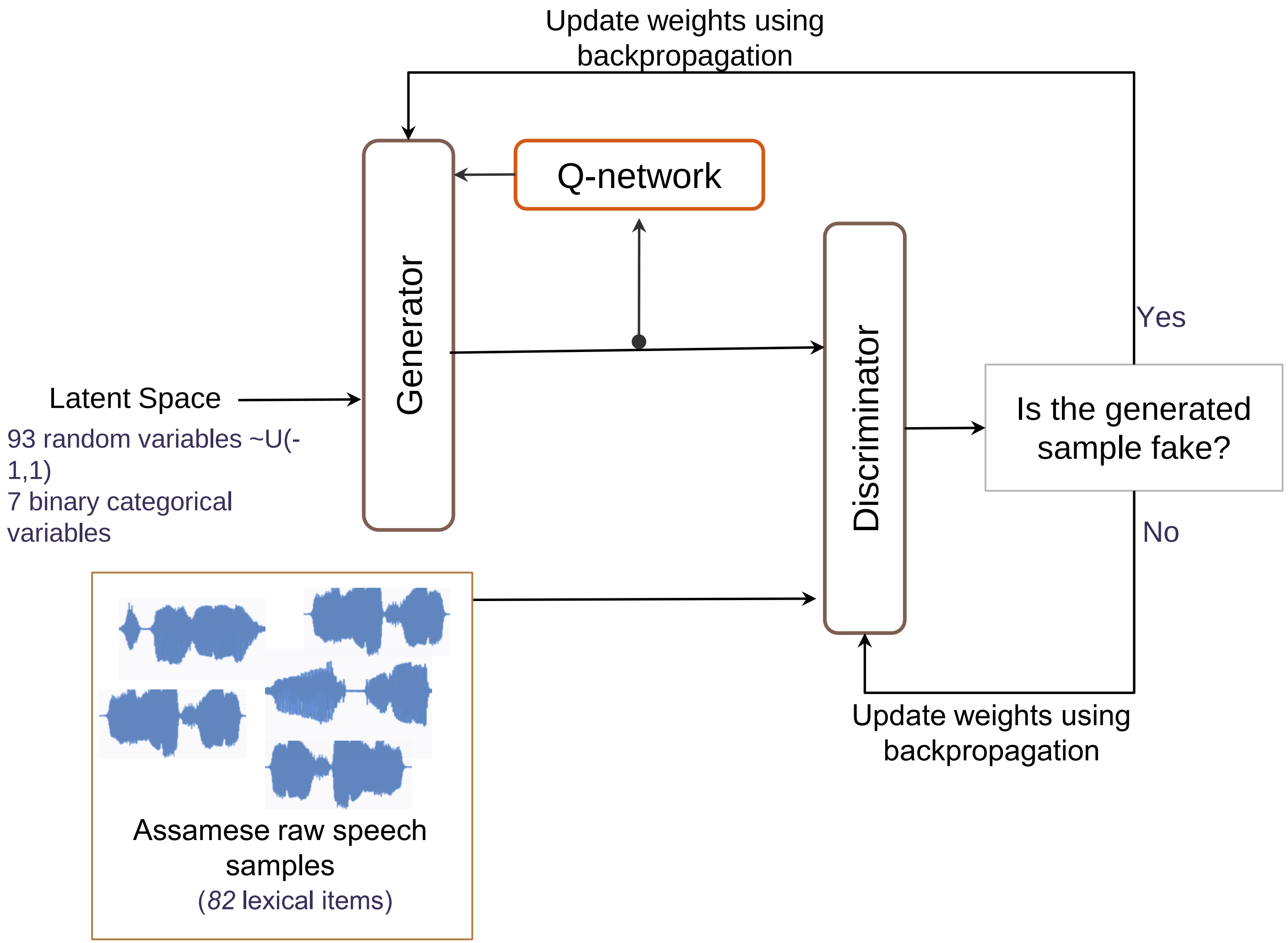
- **/ɑ/ [-high, -ATR] blocks harmony.**
- Examples:

z **ʊ** n **ɑ** k	'firefly-M'	*z **u** n **ɑ** k -**i**	z **ʊ** n **ɑ** k –**i** 'firefly-F'
b **ɛ** p **ɑ** r	'trade'	*b **e** p **ɑ** r -**i**	b **ɛ** p **ɑ** r –**i** 'trader'

Exception:

ɑ l **ɑ** x	'luxury'	**ɑ** l **o** x -**u** w **ɑ**	'pampered'	***ɑ** l **ɑ** x u w **ɑ**
m i s **ɑ**	'lie'	m i s **o** –l **-i j ɑ**	'liar'	*m i s **ɑ** -l **-i j ɑ**

Featural InfoWaveGAN for Vowel Harmony



FiwGAN architecture (Beguš 2021; Beguš and Zhou 2022) **trained on Assamese**

Experiment

- **Research questions:**

1. Can we model Assamese vowel harmony, especially iterative long-distance patterns, using fiwGAN?
2. How far can the model learn the discrete categories related to harmony?

- **Participants:**

1. 15 native Eastern Assamese speakers from the campus. 8 females and 7 males between 18-35 years. All of them were educated in vernacular medium.
2. Recorded at the Phonetics and Phonology lab at IIT Guwahati with a DR-100 MKII recorder.

Data

English	Assamese	Recorded Sentence
(will) Tell	kobo	মই ক'ব বুলি ক'লো
Something worth mentioning	kobologija	মই ক'বলগীয়া বুলি ক'লো
To tell (you)	koboloi	মই ক'বলৈ বুলি ক'লো
Tell (me)	koba	মই ক'বা বুলি ক'লো
Meanwhile	ɛnɛtɛ	মই এনেতে বুলি ক'লো

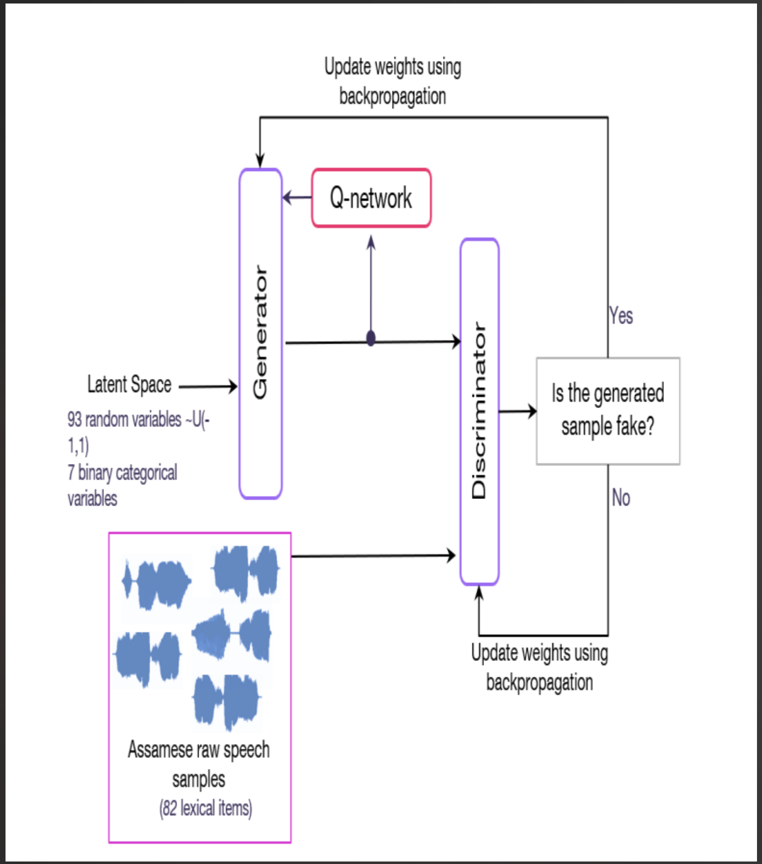
- 82 words harmonic and non-harmonic words in total.
- Each word was in a carrier sentence মই X বুলি ক'লো in Assamese; 'I say X' in English.
- Each sentence was repeated at least 4 times.
- 5000 tokens in total. 4789 tokens used for the training.

Data

English	Assamese	Recorded Sentence
(will) Tell	kobo	মই ক'ব বুলি ক'লো
Something worth mentioning	kobologija	মই ক'বলগীয়া বুলি ক'লো
To tell (you)	koboloi	মই ক'বলৈ বুলি ক'লো
Tell (me)	koba	মই ক'বা বুলি ক'লো
Meanwhile	ɛnɛtɛ	মই এনেতে বুলি ক'লো

Stem	Suffix	Surface	Category
dilɛ	-i	dilei	Harmonic
nokorilɛ	-u	nokorileu	Harmonic
gɔrɔm	-ɔ-t	gɔrɔmɔt	Non-harmonic
bɛpɑr	-i	bɛpɑri	Non-harmonic

Model Implementation



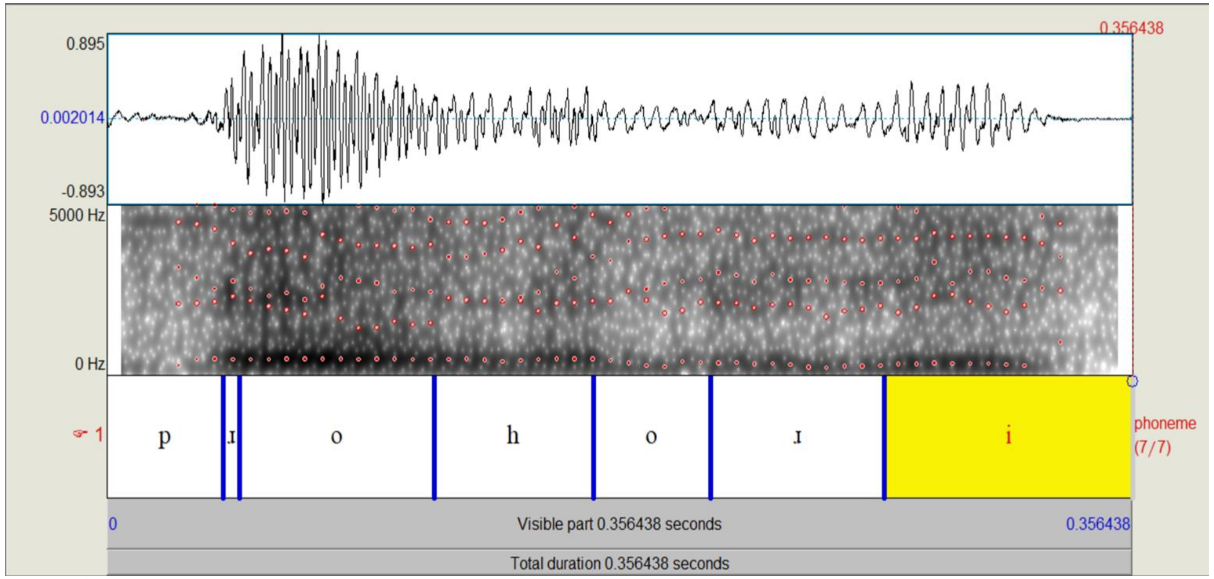
- Latent space has 93 uniformly distributed random variables (z).
- 7 binary latent codes (φ) accommodate 82 unique lexical items. $2^7 = 128$ lexical classes.
- Each word is represented as a one-hot vector [1,0,0,0,0,0,0; 0,1,0,0,0,0,0 etc.].
- Batch size = 64.
- Generator and Discriminator – Adam optimizer.
- Q-network- RMSProp Algortihm.
- The model was trained for 960 epochs.
- Each epoch generated 100 outputs.
- 64 out of 100 outputs were analyzed.

Method

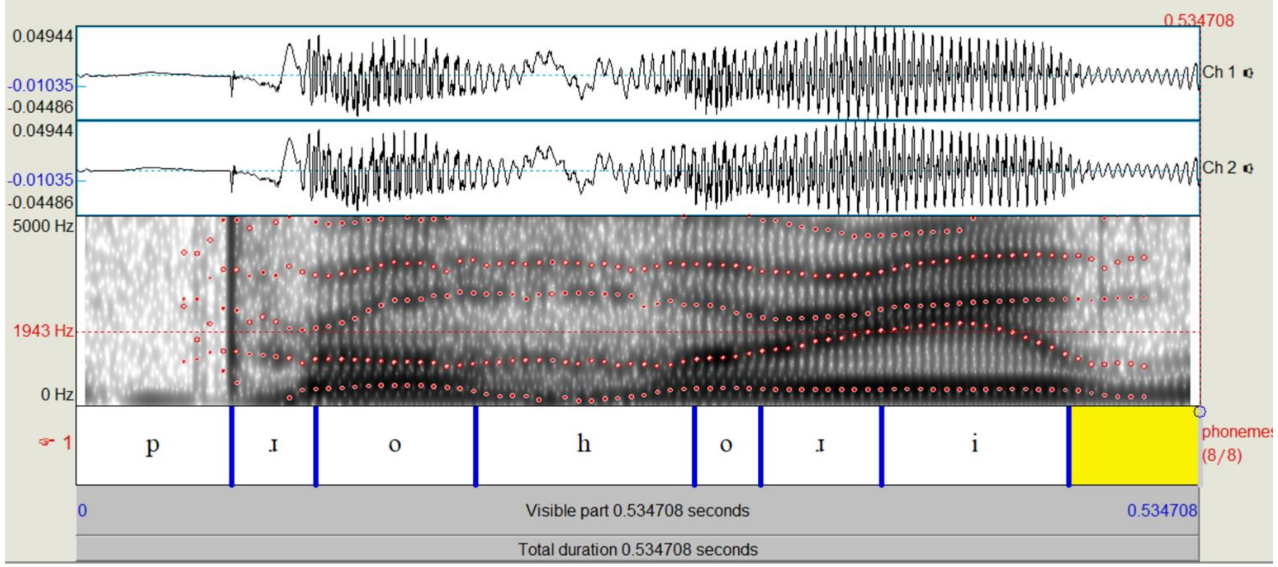
- The recorded data was sliced in PRAAT (Boersma & Weenick 2009).
- Recorded data sampling rate 48 kHz with 16-bit quantization.
- Downsampled the data at 16 kHz using the Sox program. Converted to single-channel .wav files.
- Training dataset contained 3169 harmonic and 1620 non-harmonic words.
- At least 60 data points for each lexical element.
- The PyTorch version of the model was used.
- The model ran on the CLST lab's GPU for 3 consecutive days.

Results

Results at 960 epochs (Identical)

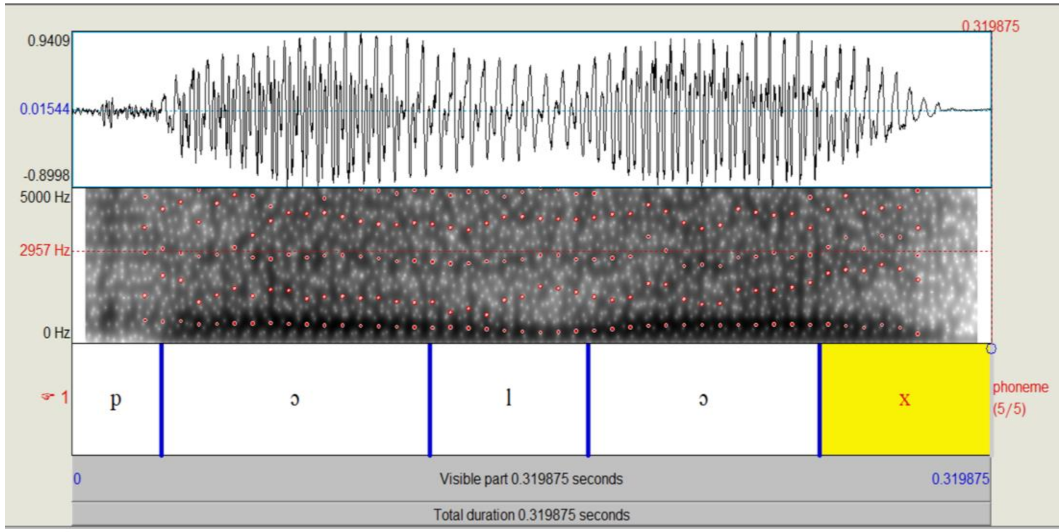


Generated item

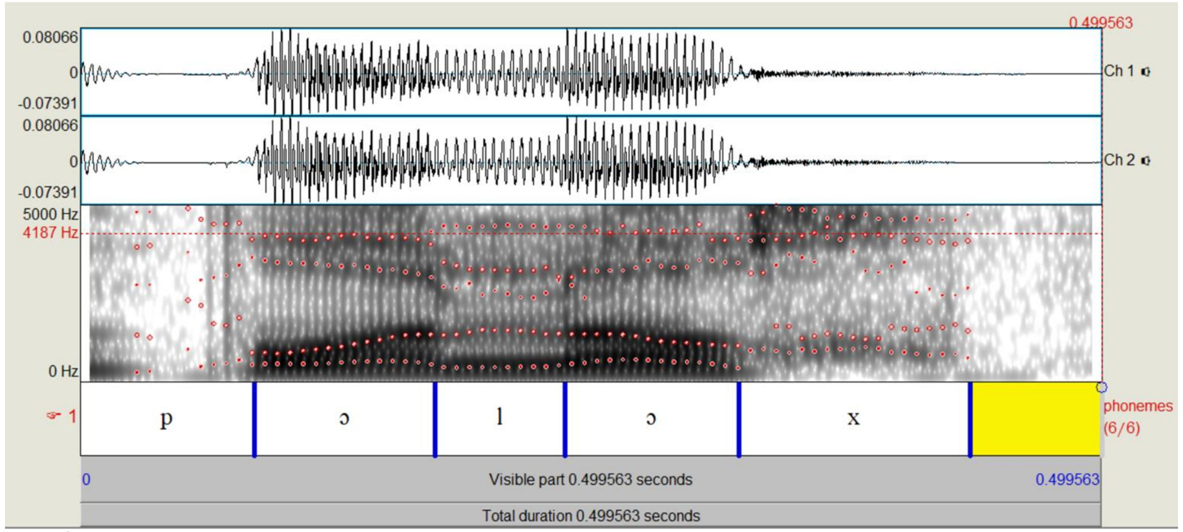


Training item

More identical outputs…

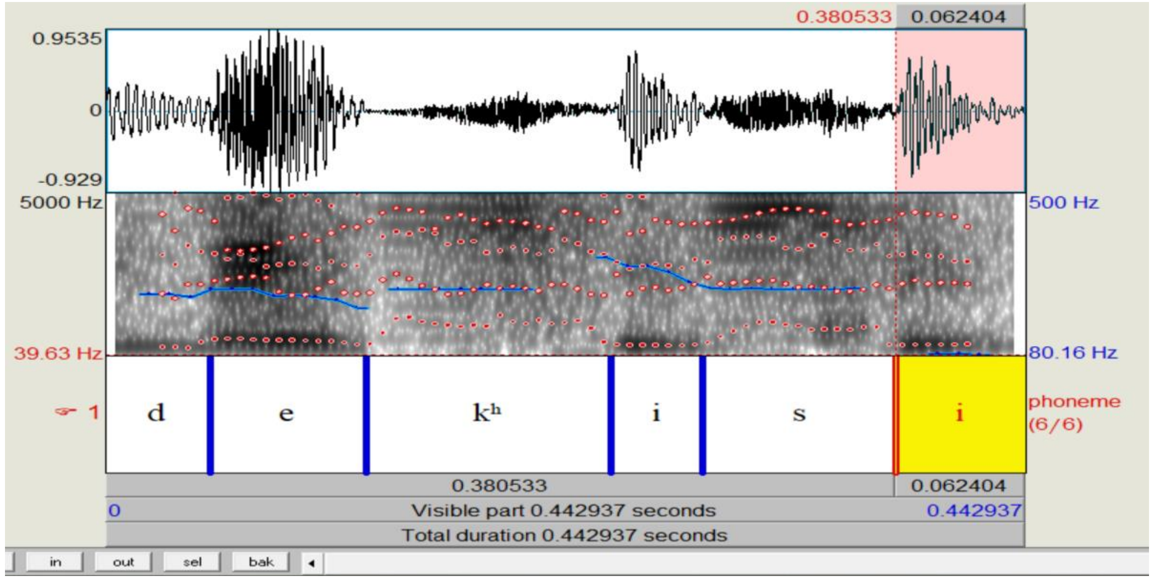


Generated item

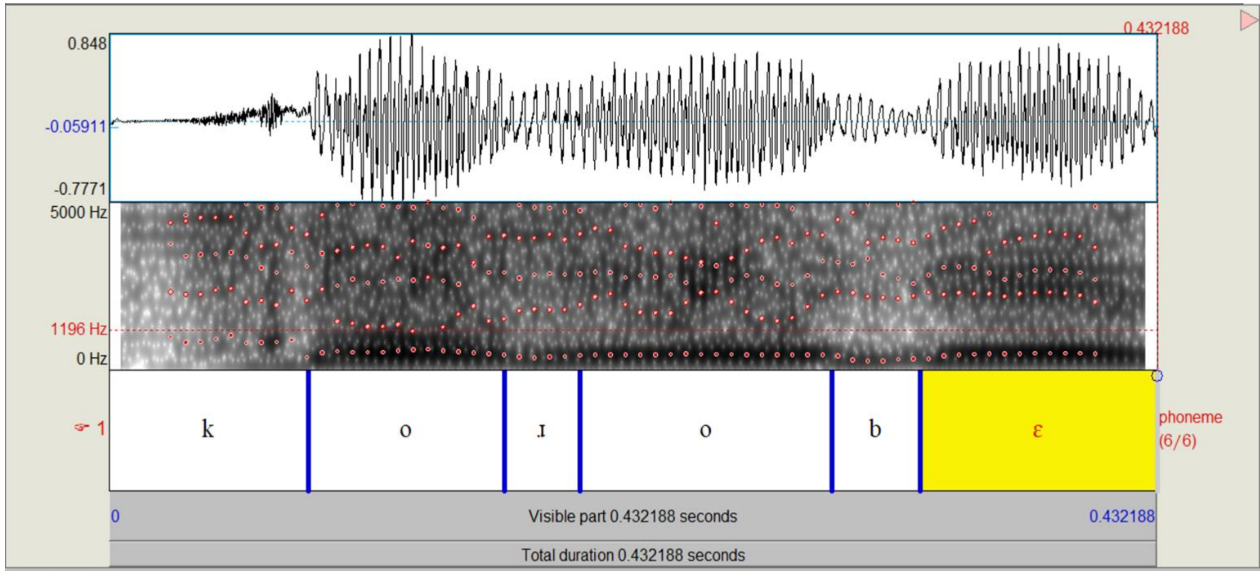


Training item

Results at 960 epochs (Innovative)

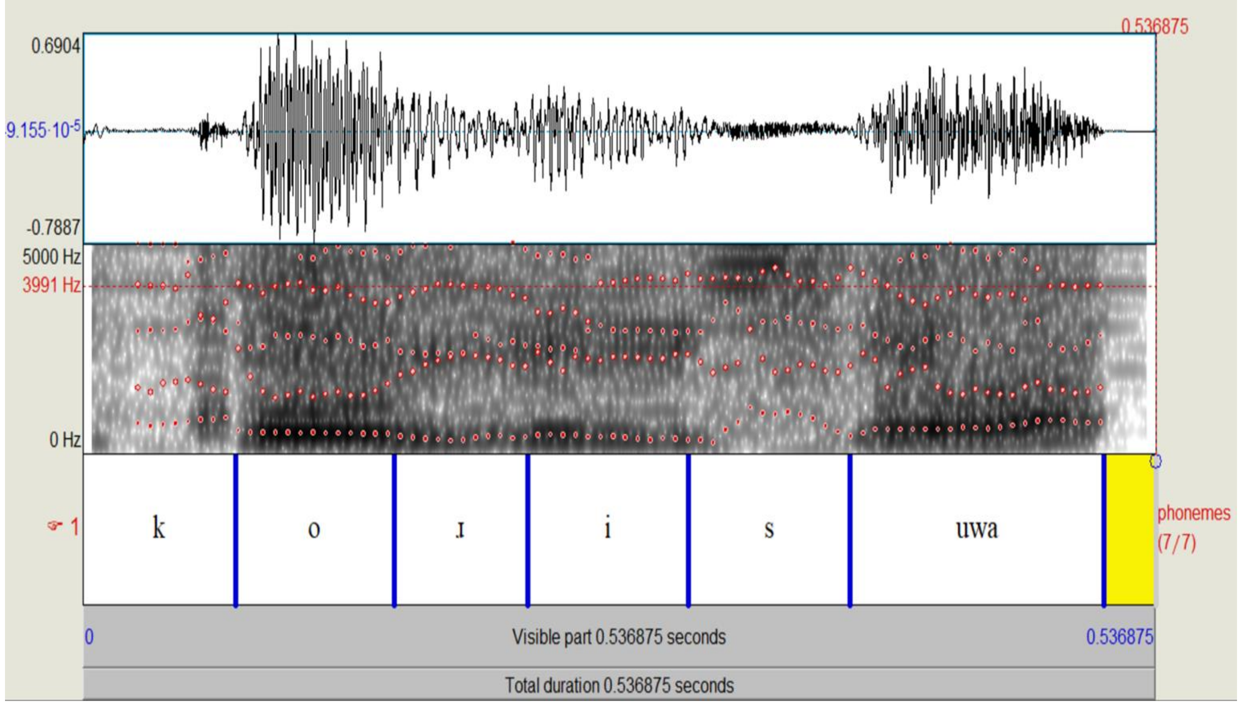


Generated item

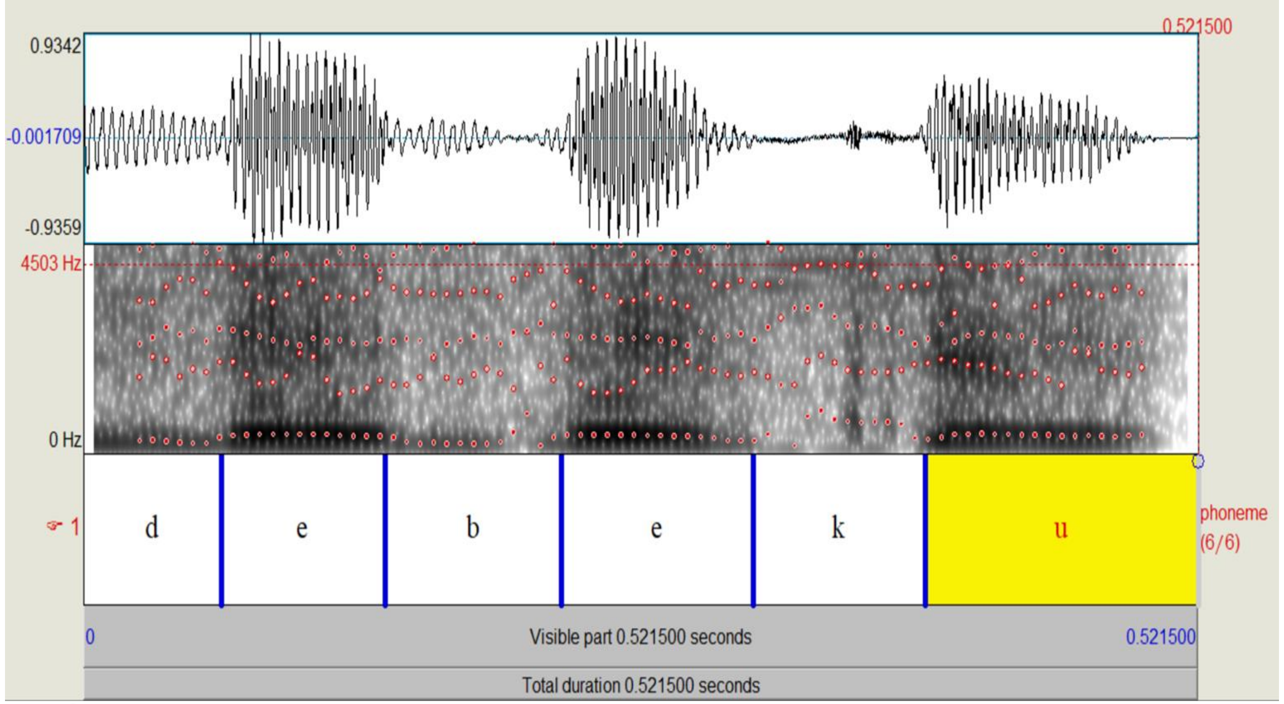


Generated item

More innovative outputs…

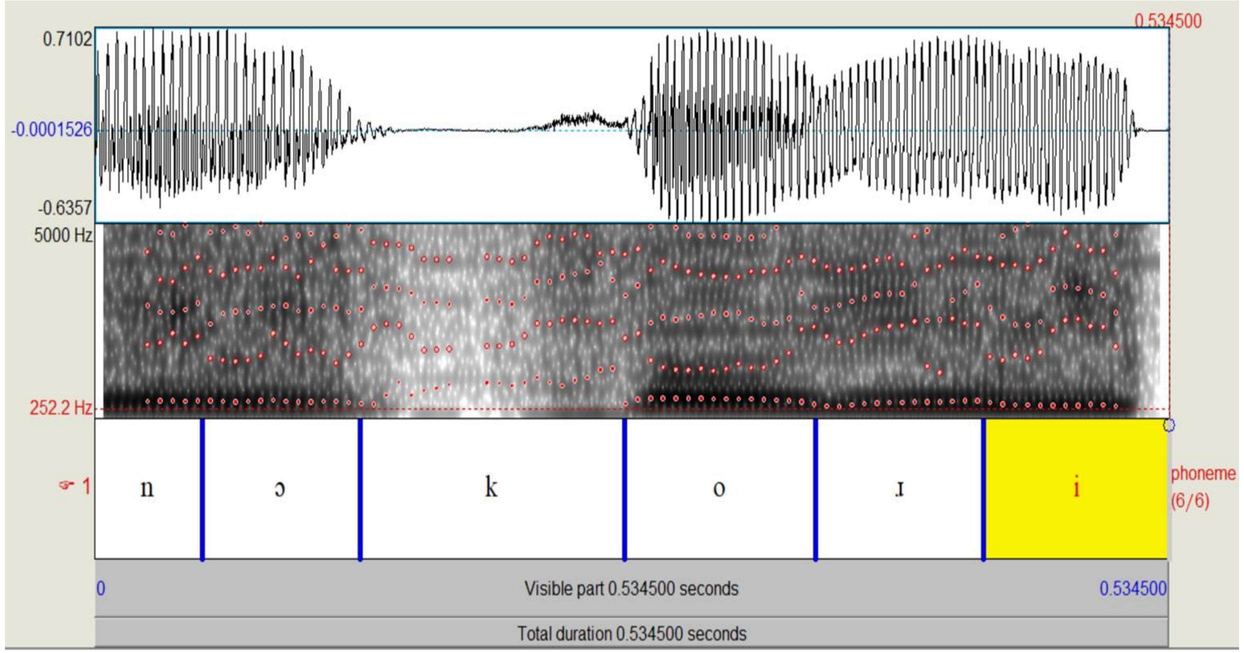


Generated item

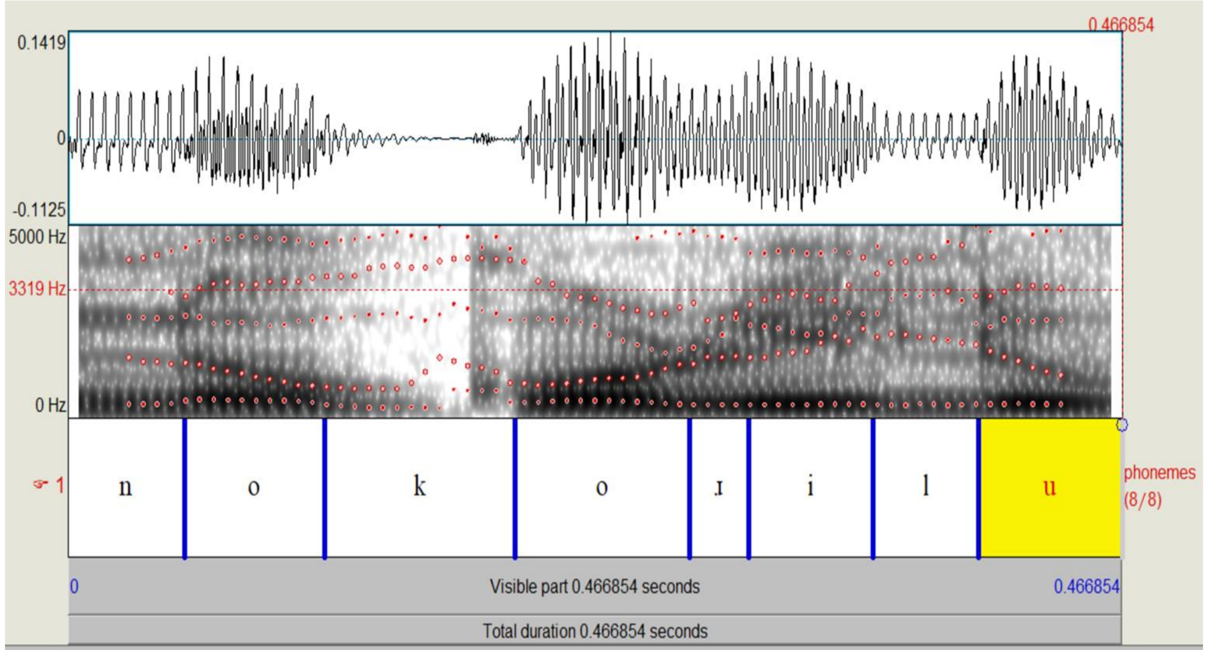


Generated item

Results at 960 epochs (Shortened)

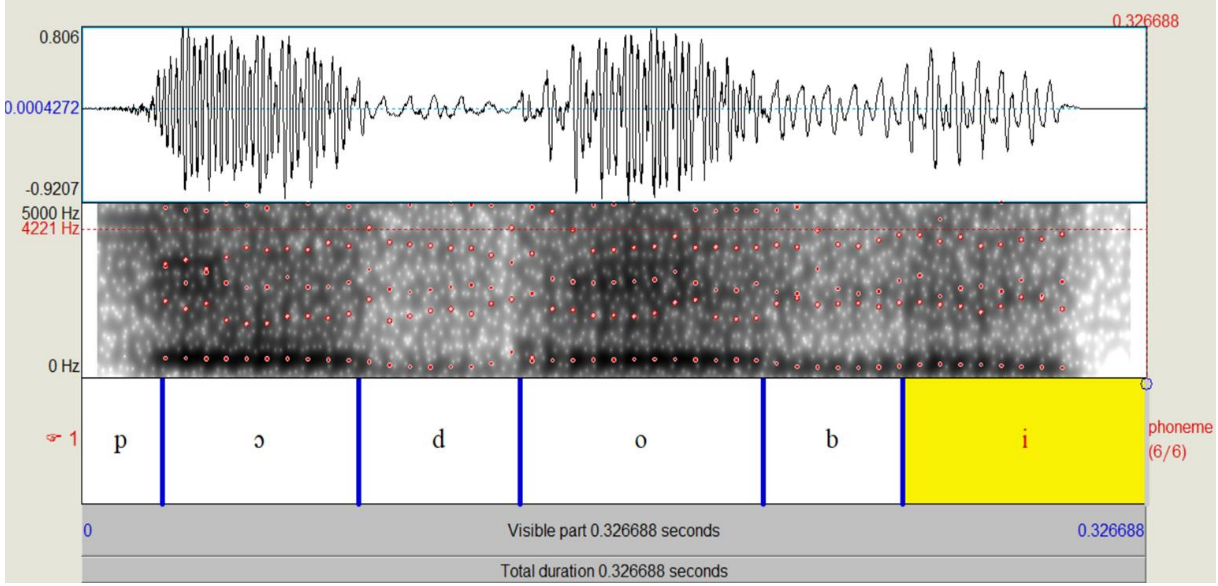


Generated item

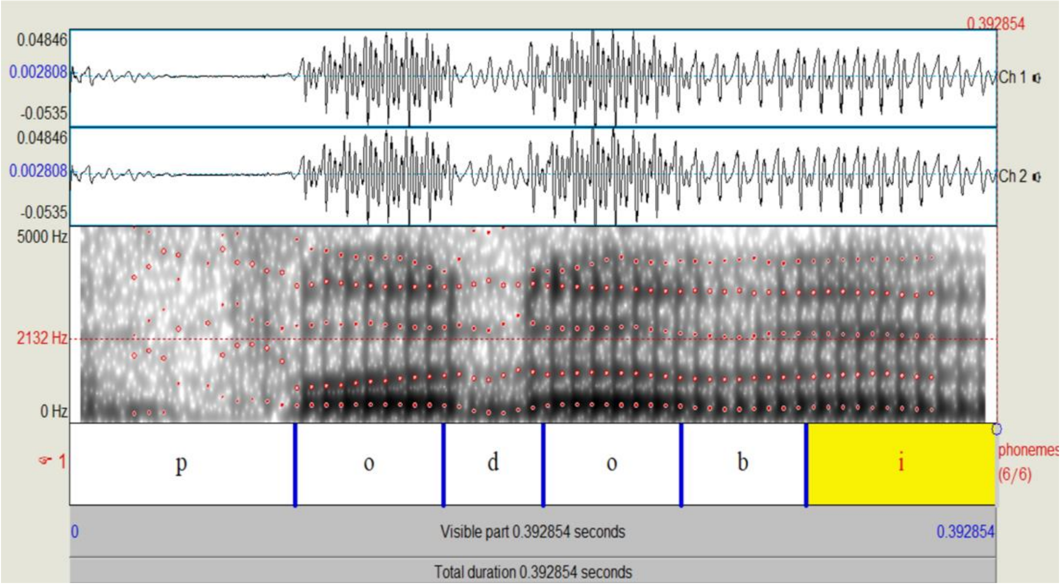


Training item

Ungrammatical outputs



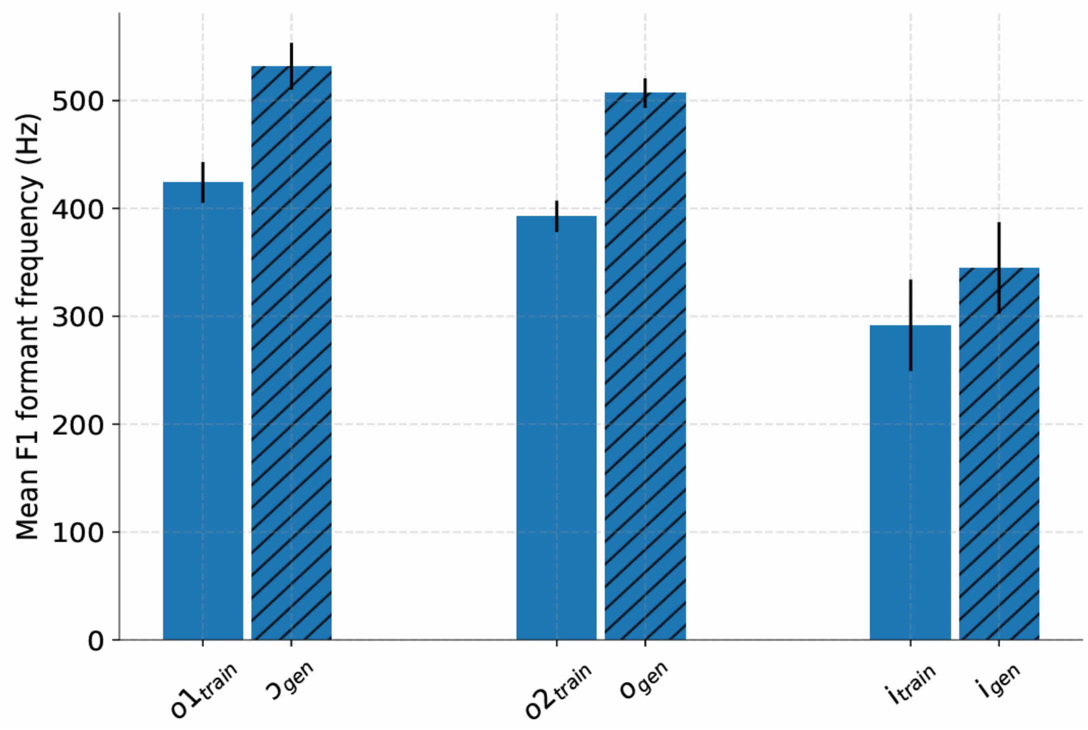
Generated item



Training item

Analysis

- The outputs were manually annotated in PRAAT; and collected F1, F2, F3 values at 10-time points.
- The mean first formants of the vowels in training and generated data to quantify the presence of ATR vowel harmony in PRAAT.
- Regression analysis in R (R Core Team 2021) to assess the presence of directionality.



F1 comparison of [podobi] (training data; shown in bars) and [pɔdobi] (generated data; shown in hatched bars). Here, o1 and o2 denote the first and second vowel and i denotes the third vowel, in the input training data “podobi”.

Table 4: *LMER model for the training dataset*

Data	Directionality	Fixed effects	DF	χ^2	p
Whole	right-to-left	F1V1~V1+**V2**	13	33.062	<0.001
	left-to-right	F1V2~V2+**V1**	10	6.5156	0.77
Only [+ATR]	right-to-left	F1V1~V1+**V2**	7	27.829	< 0.001
	left-to-right	F1V2~V2+**V1**	2	1.6522	0.43

Table 5: *Linear regression model for machine-generated items*

Data	Estimate	t-value	p-value
Whole	605.25	7.793	< .001
only V2[i] coefficient	-279.11	3.376	.01

Discussion

- Computation of long-distance iterative vowel harmony.
- Feature learning.
- Emergence of lexical learning.
- Ungrammatical outputs with local harmony.
- Lack of results with opaque vowel /ɑ/. Difficult to learn non-frequent/irregular items? (Marcus et al. 1995; McCurdy et al. 2020)

Future Direction

- Need more outputs to assess the learnability
- How does this learning take place? What are the cues? Latent space analysis.
- More epochs than previous experiments in English aspiration and French nasality.
- Can the model learn trans-word utterances?

Selected References

- Alan Prince and Paul Smolensky. 2004. Optimality 631 theory: Constraint interaction in generative grammar. Optimality Theory in phonology: A reader, pages 1–71.
- Bruce Hayes and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. Linguistic Inquiry, 39(3):379–440.
- Christo Kirov and Ryan Cotterell. 2018. Recurrent neural networks in linguistic theory: Revisiting pinker and prince (1988) and the past tense debate. Transactions of the Association for Computational Linguistics, 6:651–665.
- Connor Mayer and Max Nelson. 2020. Phonotactic learning with neural language models. Proceedings of the Society for Computation in Linguistics, 3(1):149–159.
- Diana Archangeli and Douglas Pulleyblank. 2007. Harmony, Cambridge Handbooks in Language and Linguistics, page 353–378. Cambridge University Press.

Selected References

- Gašper Beguš. 2021. Ciwgan and fiwgan: Encoding information in acoustic data to model lexical learning with generative adversarial networks. Neural Networks, 139:305–325.
- Gašper Beguš and Alan Zhou. 2022. Interpreting intermediate convolutional layers of generative cnns trained on waveforms. IEEE/ACM Transactions on 556 Audio, Speech, and Language Processing, 30:3214–3229.
- Géraldine Legendre, Yoshiro Miyata, and Paul Smolen603 sky. 1990. Can connectionism contribute to syntax?: Harmonic Grammar, with an application. University of Colorado, Boulder, Department of Computer Science.
- Paul Boersma and David Weenink. 2009. Praat: doing phonetics by computer (version 5.1.13).
- R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Selected References

- Shakuntala Mahanta. 2007. Directionality and locality 608 in vowel harmony. Ph.D. thesis, Utrecht University.
- Shakuntala Mahanta. 2008. Directionality and locality 610 in vowel harmony: With special reference to vowel harmony in Assamese. Netherlands Graduate School of Linguistics.
- Sharon Goldwater, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Proceedings of the Stockholm Workshop on Variation within Optimality Theory, ed. by Jennifer Spenader, Anders Eriksson, and O¨sten Dahl, 111–120. Stockholm: Stockholm University, Department of Linguistics

Dataset and codes available at:
https://github.com/sneha2599/FiwGAN-Assamese.git

Acknowledgement

- To the 15 data consultants.
- To my supervisor, Prof. Shakuntala Mahanta;
- To Gasper Begusˇ for his encouragement and help with the model.
- The three anonymous reviewers of SCiL for their valuable and detailed feedback.

THANK YOU!

Ungrammatical outputs