Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Unsupervised modeling of vowel harmony using WaveGAN

Sneha Ray Barman [1], *Shakuntala Mahanta* [1], *Neeraj Kumar Sharma* [1]

[1]Centre for Linguistic Science & Technology, IIT Guwahati
{sneha.barman, smahanta, neerajs}@iitg.ac.in

Abstract

Neural network models of phonological learnability are said to learn the phonotactics of a language better than traditional models of learnability[1]. Our paper explores whether the Featural InfoWaveGAN architecture (fiwGAN [2]; inspired by WaveGAN [3] and InfoGAN [4]) can capture regressive vowel harmony patterns when trained unsupervised on raw acoustic data without any supply of prosodic cues. We train the model with Assamese speech data recorded by 15 native speakers. Assamese is one of the few Indian languages that exhibit phonologically regressive and word-bound vowel harmony. [+high, +ATR] vowels [i, u] trigger right-to-left harmony of [-ATR] vowels [ɛ, ɔ, ʊ] resulting in [e], [o], and [u], respectively. We analyze the outputs generated by the fiwGAN model and observe that it learns the regressive directionality of harmony. It produces innovative items by stringing together vowels and consonants from the training dataset. It showcases its capability of learning the phonotactics of Assamese and iterative harmony patterns over a longer domain without any relevant prosodic information in the output. We assume the model treats the outputs as abstract prosodic units without external prosodic cues triggering vowel harmony.
Index Terms: generative adversarial networks, phonological learnability, prosody, vowel harmony, Assamese

1. Introduction

Among other aspects of grammar, phonology is also a crucial aspect of modeling human acquisition. Understanding the model's inherent bias and its ability to make human-like generalizations is important. Additionally, sound systems are not organized arbitrarily but contain structural generalizations and interdependence[5]. Therefore, learning a sound system requires learning to acoustically recognize segments (phonetics), mapping them to an inventory characterized by distinctive features, and learning distributional constraints on segment sequences (phonology). To address how a grammar derives surface phonetic outputs from phonological inputs, researchers devised learning theories, commonly executed as machine-implemented models that mirror the capacities of human children. The Generative approach or the rule-based approach [6] derives outputs from input through derivation. Optimality Theory [7] and other related proposals like Harmonic Grammar and Maximum Entropy Grammar[8, 9, 10, 11] model phonological grammar as an input-output pairing; they choose the most optimal output based on an input.

In his in-depth assessment of previous generative linguistics research, [1] argued the significance of implementing neural network approaches to generative linguistics. Recently, [12] and [13] reported that a neural network autoencoder learns phoneme-like representations without explicit labeling. [13]'s autoencoder neural network model is trained on pre-segmented acoustic data. It inputs the acoustic data and outputs corresponding values to phonological features. Given the autoencoder architecture, the model does not generate innovative data but attempts to match the outputs as closely as possible to the input.

Generative Adversarial Network (GAN) is claimed to model learning more naturally as phonetic and phonological processes are computed as the mapping from latent space to generative data [14, 15]. [3] apply the GAN architecture based on the DCGAN model for audio data [16] to learn language features from continuous speech signals. The WaveGAN model for phonetic and phonological learning [15] successfully captures the linguistic representations as a one-to-one mapping from latent space to generated data, similar to how humans construct underlying phonological representations by listening to a speech stream in a language. However, a crucial aspect of human acquisition is lexical learning- the ability to store unique information attached to each meaning-bearing unit or phonemes and to generate new lexical items while also conforming to the phonotactics of their language. [2] modified the WaveGAN architecture by adding a lexical learning component- Q-network [4], which enables the Generator to produce linguistically categorical meaningful sounds. Featural InfoWaveGan (fiwGAN) has the advantage of proposing a new latent space structure to learn featural representations of phonetic and phonological learning. The Generator network learns how to generate acoustic data that encodes unique lexical information and outputs innovative acoustic data such that each lexical item is associated with a unique code. [5] examined the robustness of the fiwGAN model by testing English and French nasality. The results meet [15]'s claim that the GANs can learn phonological representation where single latent variables correspond to identifiable phonological features. The ability of fiwGAN to distinguish between the English non-contrastive and French contrastive nasality makes the model a good phonological learner, given it is tasted in a more controlled setting in which the same feature is compared across languages.

Vowel harmony is a complex phonological process requiring the learner to grasp crucial aspects like directionality, domains, features, iterativity, locality, and opacity [17]. Additionally, prosodic factors like morphological units such as root and stem significantly influence vocalic alternation [18]. The existing models examining the learnability of the vowel harmony process [10, 19] relied on text data and supervised training. Considering that human language acquisition is best modeled unsupervised from raw speech [20], we take an unsupervised approach, treating raw unannotated acoustic waveforms as the principal input for learning. In Assamese, the [+ATR] (advanced tongue root) vowels trigger harmony of [-ATR] vowels. We examine whether the fiwGAN architecture learns long-distance iterative harmony and regressive directionality without explicit prosodic cues. We pose two questions- i. Can the GAN model learn ATR harmony patterns from real-life data? ii. If yes, what does it take as a prosodic cue? We train the

 10.21437/SpeechProsody.2024-41

model with Assamese speech data containing harmonic and non-harmonic words and also examine whether lexical learning emerges from training the model on vowel harmony.

2. Assamese

Assamese is an Indo-Aryan language spoken in the state of Assam by 15.3 million speakers [21]. The variety examined for this research is the colloquial Assamese spoken in the eastern region of Assam, also referred to as Upper Assam. Assamese has 8 surface vowels [i, e, ɛ, ɑ, ɔ, o, ʊ, u] and 20 consonants [p, p^h, b, b^h, t, t^h, d, d^h, k, k^h, g, g^h, m, n, ŋ s, z, x, h, ɹ, j, w, l] (see Table 1).

Table 1: *Vowel inventory of Assamese [22]*

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

2.1. Assamese vowel harmony

Vowel harmony is a phonological process where adjacent vowels in a word change their vocalic property to match the neighboring vowels. Assamese is one of the few Indian languages that exhibits rich vowel harmony. [22] has extensively studied, and more recently, [23] gave us new perspectives on vowel harmony in Assamese.

The surface vowels' occurrence is restricted in the following ways:

- The two high vowels /i/ and /u/ are pronounced with an advanced tongue root [ATR], as are the mid vowels /e/ and /o/.
- The mid vowels / ɛ/ and / ɔ/ are slightly lower than /e/ and /o/ and are not realized with an advanced tongue root, i.e., they are specified as [-ATR].
- /e/ and /o/ cannot occur word-finally or initially; they appear only under circumstances of vowel harmony.

Assamese exhibits right-to-left regressive ATR harmony where the high vowels /i/ and /u/ trigger [+ATR] harmony in [-ATR] high and mid vowels in the language, i.e., /ɛ/, /ɔ/ and /ʊ/, except the [–ATR] low vowel /ɑ/. Exceptional patterns emerge when a root morpheme containing the vowel /ɑ/, in suffixes like /-ijɑ/ and /-uwɑ/, is added. An otherwise opaque vowel [ɑ] becomes either /e/ or /o/. Therefore, mid [-ATR] vowels /ɛ,ɔ/ become [e] and [o] if they are following [+high,+ATR] vowels /i/ and /u/; [+high,-ATR] vowel /ʊ/ becomes [+ATR] /u/; /ɑ/ does not change when it is not adjacent to the triggering suffixes (see Table 2).

Table 2: *Illustrated examples of vowel harmony in Assamese*

Assamese	Gloss	-suffix	Harmonised	Gloss
/pɛt/	'belly'	-u	[petu]	'pot-bellied'
/bɛpɑr/	'trade'	-i	[bɛpɑri]	'trader'
/zʊnɑk/	'firefly-M'	-i	[zʊnɑki]	'firefly-F'
/pɑgɔl/	'mad-M'	-i	[pɑgoli]	'mad-M'

From the above discussion and the examples of Table 2, we can argue that the stem vowels alternate under the influence of suffixes in Assamese. The [+ATR] acts as a dominant feature that spreads from both the root and the suffix vowels in the leftward direction and also affects vowels on the left side of the triggering vowel. The examples show that harmony applies to all the [-ATR] vowels preceding a [+ATR] vowel exhibiting iterative long-distance harmony. /ɑ/ appears to be an opaque vowel that blocks harmony from its targeting vowels to its left (zʊnɑki 'firefly' *zunɑki).

3. Model architecture

The fiwGAN neural network model comprises three primary components: a Generator, a Q-network, and a Discriminator (see Figure 1). Unlike traditional GAN models, where the Generator typically takes uniformly distributed latent variables as input, fiwGAN's Generator operates within a latent space composed of binary codes (ϕ) and latent variables. Each binary variable (ϕ_n) corresponds to a distinct feature [2]. Trained to maximize the Discriminator's error rate while minimizing the Q-network's error, the Generator produces a vector containing approximately 16384 data points for a 1 second of audio data (sampled at 16 kHz). These generated outputs, along with real acoustic data, are then assessed by the Discriminator, which computes the Wasserstein distance [24] between the generated and real data.

In parallel, the Q-network, which shares the same input data as the Discriminator, operates differently. Its final layer comprises nodes corresponding to categorical variables (ϕ). Unlike the Discriminator, the Q-network's loss function drives updates not only to its weights but also to the Generator's. This mechanism enables the Generator to associate each lexical item with a unique latent code, allowing the Q-network to retrieve lexical information solely from the acoustic signals, resulting in lexical learning. While the Generator produces raw acoustic data similar to but not an exact replication of the real input data, it effectively captures essential characteristics of the acoustic signals. [1]

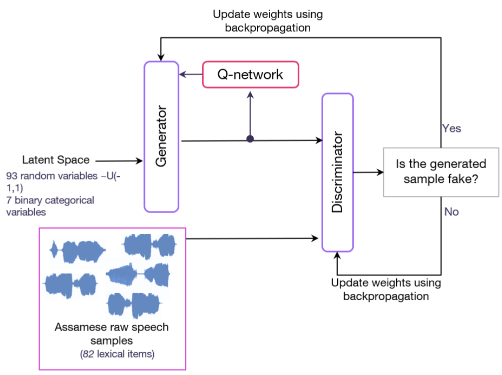

Figure 1: *A diagram of fiwGAN*

[1]We used the Pytorch implementation of fiwGAN made available by [2] at `https://github.com/gbegus/ciwganFiwGAN-pytorch.git`.

4. Materials

4.1. Participants and Data

For this experiment, 15 native Assamese speakers (7 males, 8 females) between the age group 19-35 were consulted. The data was recorded with a Tascam DR-100 MKII recorder in the soundproof booth at the Phonetics and Phonology lab of IIT Guwahati. The dataset consists of 82 words in total. Each target word was in a carrier sentence written in Assamese script ('moi X buli kolu' in Assamese; 'I say X' in English) (see Table 3). The participants were asked to utter each sentence at least four times. The collected speech data was then manually sliced using PRAAT [25] to separate the target words for training the model. This resulted in 5000 tokens; some words were repeated more than four times.

Table 3: *Example dataset for training fiwGAN*

Assamese	suffix	Harmonised
ɛlah	-uwɑ	elehuwɑ
ɑlɑx	-uwɑ	ɑloxuɑ
dilɛ	-i	dilei
nokorilɛ	-u	nokorileu

4.2. Methodology

Among 5000 tokens, 4789 (3169 harmonic and 1620 non-harmonic) items are used for training based on their quality. The fiwGAN model takes 1-second-long unannotated single-channel audio files sampled at 16 kHz as raw waveforms. The latent space of the model is designed with 93 uniformly distributed random variables (z∼u(-1,1)) and 7 binary categorical variables to accommodate 82 unique lexical items [2].The Generator and Discriminator were trained with Adam optimizer while the Q-network was trained with RMSProp algorithm at a .0001 learning rate with a batch size 64. The Generator data was loaded for analysis following 960 training epochs (∼ 44000 steps).

The generated items are manually tested in Praat [25]. Examining the noise quality and intelligibility, 64 out of 100 generated items were analyzed for our experiment after manual listening and segmentation in Praat [25]. We collect the F1 and F2 values at 10-time points of vowels in both training data and generated outputs. We quantify the presence of ATR harmony by examining the target vowel's mean F1 and F2 values in the vicinity of the trigger vowels. The formant values are fed to a linear regression mixed-effects model (lmer) [26] and then a linear regression model (lm) [27] in R [28] to assess the direction of vowel harmony.

5. Results

In the previous research in English allophonic distribution [29] and English and French nasality [5], the Generator learned to produce human speech-like sounds after 649 epochs of training, but in our case, the human-like speech was generated after 720 epochs. Intelligible sounds were generated after 800 epochs. After obtaining the outputs, the audios are manually annotated and segmented in Praat [25]. Analyzing the spectrograms, we observe that out of 64 audio outputs, 29 are novel items, and 35 are identical to the training data. About 20 items are innovative and grammatical, while 9 are ungrammatical or illicit. Out of 35 identical items, 25 have locally or non-locally licit vowel patterns, while 10 are totally ungrammatical.

After 960 epochs, the model generates lexical items identical to the training dataset ([prohori],[ɛkɛbɑrɛ],[pɔlɔx]), innovative outputs([dekhisi],[korobe],[korisuwɑ],[debeku] etc.) as well as some illicit forms ([nɔkorilu], [pɔdobi] etc.). We analyze the innovative outputs and observe that the model is stringing together vowels and consonants from the training data, for example, [dekhisi] is assumed to be influenced by the word [dekhisu] in the input, while [korobe] may stem from [korobi]. Since our focus of the study is vowel harmony, the vowel sequences are studied carefully in the results. The novel items [debeku], [dekhisi], and [korisuwɑ] follow the harmony pattern where [+high,+ATR] vowels [i,u] trigger harmony of the underlying [-ATR] vowels turning them into [+ATR]. In our above examples, we observed that [e]/[o] occurs when followed by [i]/[u] locally [korisuwɑ/dekhisi] as well as over a longer domain [debeku] (see Figure 2). This suggests the model's capability to learn the phonotactics of Assamese along with iterative long-distance harmony. We further probed that vowels in the illicit items follow a specific pattern where the trigger vowels impact the immediately preceding vowel, suggesting non-iterative harmony([pɔdobi] instead of [podobi]) (see Figure 3). This observation indicates that the model learns iterative harmony is myopic [30]. Moreover, [korisuwɑ] is close to a real-world compound word [kori suwɑ'do see'] and a word the model has not encountered in the input, implying the emergence of lexical learning from the training. We then statistically examine the generated outputs to examine the directionality of the vowel sequences.

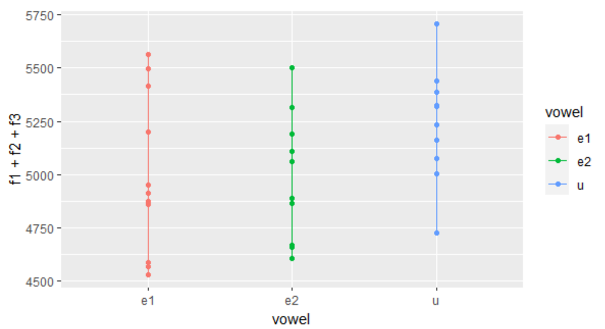

Figure 2: *First three formants of vowels in [debeku] (innovative item)*

5.1. Statistical analysis

We examined the generated outputs statistically to quantify the presence of directionality in the vowel sequences. The formant values of vowels in the training data were analyzed. It was observed that the first formant frequency (F1) of the underlying [-ATR] vowels is much lower in the vicinity of the [+ATR] vowels (Figure 4), triggering harmony. We hypothesize that if V2 in the V1CV2 setting explains V1 better than V1 explains V2, the dataset follows regressive vowel harmony. If not, the directionality is assumed to be left-to-right. A linear mixed-effects regression analysis was carried out on the training dataset [26] (using R [28]). This helped examining the relationship between

[2]In fiwGAN, n features (ϕ) are used to categorize 2^n lexical classes. For 82 lexical items, we need 82 lexical classes. 2^6=64, 2^7=128. So, we use 2^7 binary variables in the latent space.

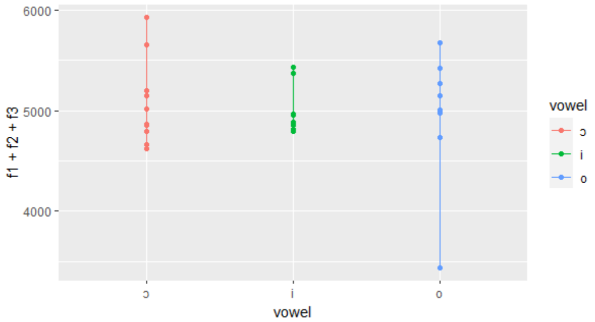

Figure 3: *First three formants of vowels in [pɔdobi](illicit item)*

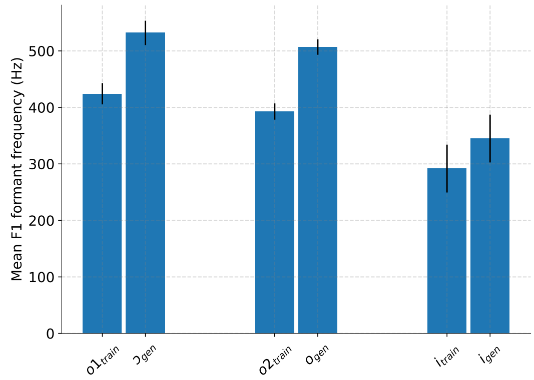

Figure 4: *F1 comparison of [podobi] (training data; shown in bars) and [pɔdobi] (generated data; shown in hatched bars). Here, o1 and o2 denote the first and second vowel, and i denotes the third vowel, in the input training data "podobi".*

the dependent variable F1V1 and predictor variables V1 and V2 [3]. V1 and V2 are the fixed effects, while 'word' is the random effect in the model [4]. The likelihood ratio test compares the full model to a null model. The AIC score revealed a statistically significant difference between the two models [5]. The chi-square test indicates a significant difference in model fit between the null and full models (χ^2=33.062, df=13, p<0.001), reflecting a significant effect of V2 on the F1 value of V1 (Table 4). The dataset is then categorized into two subsets: harmonic and non-harmonic. Both the subsets are analyzed using the lmer model. The fixed-effects results from the full model for harmony[6] showed a distinction between V1, V2, and the intercept representing F1 of V1. The estimate of the intercept is about 552.129 Hz with a p-value<2e-16. A comparison of the full and null model using the likelihood ratio test showed that V2 affects V1 (χ^2= 27.829, df = 7, p = 0.0002361), which reveals that the contrast is statistically significant [7](Table 4). Assuming the results from statistical models for training data as a baseline, the generated items are annotated and then fit to linear regression models in R to examine the relationship between the dependent variable F1V1 and predictor variables V1 and F1V2 (see Table 5). The model was statistically significant (F(6, 15) = 9.504, p = 0.0002087), indicating that at least one predictor variable had a non-zero effect. The Adjusted R-squared value of 0.7084 suggests that the model explained approximately 70.84% of the variability in F1V1. The coefficients for individual predictor variables were significant at various levels. Our regression model suggests that V1 and, to some extent, F1V2 are associated with variations in F1V1. Further analysis shows that the coefficients of V2[T.i] are significantly higher than other V2 variables (Estimate=-279.11, t-value=-3.376,p= 0.00817) (Table 5). This supports the idea that [i], a [+high,+ATR] vowel in Assamese, acts as a triggering vowel in the machine-generated outputs. The overall analysis suggests that the model learns the right-to-left harmony pattern, which is evident in the influence of V2 on V1 for both first and second formant frequencies.

6. Discussion and Conclusion

Our research explores long-distance computing of vowel patterns, emphasizing the learning of a series of vowel changes to

Table 4: *Results from 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: *Results from 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

the underlying forms required by learners to grasp vowel harmony. Given limited data, fiwGAN is an effective phonotactic learner, as demonstrated by its novel outputs. It learns a complex system like iterative long-distance harmony. The error items show non-iterative local harmony, implying that unbounded harmony may be myopic in Assamese [30]. The dominance of V2[i] as a trigger vowel indicates that the model can also learn the trigger feature. The statistical analysis further suggests that the model learns regressive directionality. Moreover, lexical learning emerges after the training. It puts proper sequences together without any other relevant prosodic information in the input, so we propose the system treats its outputs as abstract prosodic units. This further supports that the GAN model can identify word boundaries and learn the underlying representations in an unsupervised setting just like humans do from continuous speech streams or articulatory movements. Despite not having other widely recognized prosodic cues, we posit it may construe word-based iterative harmony as a prosodic unit. We'll analyze the Generator network's intermediate layers to grasp the features influencing vowel harmony learning. Our current results lack output for the opaque vowel /ɑ/, leaving uncertainty about the model's ability to recognize opacity. We aim to streamline training epochs for improved quality and less noise. We trained the model on a relatively simpler dataset of individual lexical items, which leaves room for deeper investigation with trans-word utterances in the future.

[3] V1full.model=lmer(F1V1∼V1+V2+(1|word), data=formants, REML=FALSE)

[4] V1 and V2 are the vowels while F1V1 is the mean F1 value of V1

[5] AIC(full)=1588.8, AIC(null)=1595.9

[6] V1vhfull=lmer(F1V1∼V1+V2+(1|word),data=har, REML=FALSE)

[7] AIC(full)=467.04, AIC(null)=480.87

7. References

[1] J. Pater, "Generative linguistics and neural networks at 60: Foundation, friction, and fusion," *Language*, vol. 95, no. 1, pp. e41–e74, 2019.

[2] G. Beguš, "Ciwgan and fiwgan: Encoding information in acoustic data to model lexical learning with generative adversarial networks," *Neural Networks*, vol. 139, pp. 305–325, 2021.

[3] C. Donahue, J. McAuley, and M. Puckette, "Adversarial audio synthesis," *arXiv preprint arXiv:1802.04208*, 2018.

[4] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in neural information processing systems*, vol. 29, 2016.

[5] J. Chen and M. Elsner, "Exploring how generative adversarial networks learn phonological representations," *arXiv preprint arXiv:2305.12501*, 2023.

[6] N. Chomsky and M. Halle, *Sound Patterns of English*. New York, Evanston and London: Harper and Row, 1968.

[7] A. Prince and P. Smolensky, "Optimality theory: Constraint interaction in generative grammar," *Optimality Theory in phonology: A reader*, pp. 1–71, 2004.

[8] G. Legendre, Y. Miyata, and P. Smolensky, *Can connectionism contribute to syntax?: Harmonic Grammar, with an application*. University of Colorado, Boulder, Department of Computer Science, 1990.

[9] S. Goldwater, M. Johnson, J. Spenader, A. Eriksson, and Ö. Dahl, "Learning ot constraint rankings using a maximum entropy model," in *Proceedings of the Stockholm workshop on variation within Optimality Theory*, vol. 111, 2003, p. 120.

[10] B. Hayes and C. Wilson, "A maximum entropy model of phonotactics and phonotactic learning," *Linguistic inquiry*, vol. 39, no. 3, pp. 379–440, 2008.

[11] J. Pater, "Weighted constraints in generative linguistics," *Cognitive science*, vol. 33, no. 6, pp. 999–1035, 2009.

[12] O. Räsänen, T. Nagamine, and N. Mesgarani, "Analyzing distributional learning of phonemic categories in unsupervised deep neural networks," in *CogSci... Annual Conference of the Cognitive Science Society. Cognitive Science Society (US). Conference*, vol. 2016. NIH Public Access, 2016, p. 1757.

[13] C. Shain and M. Elsner, "Measuring the perceptual availability of phonological features during language acquisition using unsupervised binary stochastic autoencoders," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 69–85.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[15] G. Beguš, "Generative adversarial phonology: Modeling unsupervised phonetic and phonological learning with neural networks," *Frontiers in artificial intelligence*, vol. 3, p. 44, 2020.

[16] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[17] D. Archangeli and D. Pulleyblank, *Harmony*, ser. Cambridge Handbooks in Language and Linguistics. Cambridge University Press, 2007, p. 353–378.

[18] M. Krämer, "On the role of prosodic and morphological categories in vowel harmony and disharmony: a correspondence approach," in *Proceedings of ConSOLE*, vol. 7, 1999, pp. 183–199.

[19] C. Mayer and M. Nelson, "Phonotactic learning with neural language models," *Proceedings of the Society for Computation in Linguistics*, vol. 3, no. 1, pp. 149–159, 2020.

[20] E. Dupoux, "Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner," *Cognition*, vol. 173, pp. 43–59, 2018.

[21] Government General of India, "Census of India," 2011.

[22] S. Mahanta, "Directionality and locality in vowel harmony," Ph.D. dissertation, Utrecht University, 2007.

[23] D. B. Archangeli and J. Yip, "Assamese vowels and vowel harmony," *Journal of South Asian Languages and Linguistics*, vol. 6, no. 2, pp. 151–183, 2020.

[24] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks," in *Proc. Intl. Conf. Machine Learning (ICML)*, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 2017, pp. 214–223.

[25] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 5.1.13)," 2009. [Online]. Available: http://www.praat.org

[26] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[27] A. Zeileis and T. Hothorn, "Diagnostic checking in regression relationships," *R News*, vol. 2, no. 3, pp. 7–10, 2002. [Online]. Available: https://CRAN.R-project.org/doc/Rnews/

[28] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[29] G. Beguš, "Identity-based patterns in deep convolutional networks: Generative adversarial phonology and reduplication," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 1180–1196, 2021.

[30] C. Wilson, "Unbounded spreading is myopic," in *Current Perspectives on Phonology workshop, Phonology Fest*, 2006.