Nature of contrast and coarticulation: Evidence from Mizo tones and Assamese vowel harmony

Indranil Dutta[1], *Irfan S*[2], *Pamir Gogoi*[3], *Priyankoo Sarmah*[4]

[1]Dept. of Computational Linguistics, The EFL University, Hyderabad
[2] University of Illinois at Urbana-Champaign
[3] University of Florida, Gainesville
[4] Indian Institute of Technology Guwahati

indranil@efluniversity.ac.in

Abstract

Tonal coarticulation is universally found to be greater in extent in the carryover direction compared to the anticipatory direction ([1], [2], [3], [4], [5]) leading to assimilatory processes. In general, carryover coarticulation has been understood to be due to intertio-mechanical forces, and, anticipatory effects are seen to be a consequence of parallel activation of articulatory plans ([6]). In this paper, we report on results from a set of Artificial Neural Networks (ANN) trained to predict adjacent tones in disyllabic sequences. Our results confirm the universal pattern of greater carryover effects in Mizo leading to tonal assimilation. In addition, we report on results from single-layered ANN models and Support Vector Machines (SVM) that predict the identity of V_2 from V_1 (anticipatory) consistently better than V_1 from V_2 (carryover) in Assamese non-harmonic #…V_1CV_2…# sequences. The directionality in the performance of the V_1 and V_2 models, help us conclude that the directionality effect of coarticulation in Assamese non-harmonic sequences is greater in the anticipatory direction, which is the same direction as in the harmonic sequences. We argue that coarticulatory propensity exhibits a great deal of sensitivity to the nature of contrast in a language.

Index Terms: coarticulation, vowel harmony, Mizo tones, Assamese vowel harmony

1. Introduction

While there may be several sources that are responsible for acoustic variation in speech, one of the main sources of variation is coarticulatory overlap in gestures [7]. Coarticulatory propensity, namely, the extent and nature of coarticulation, has been shown to be a consequence of several factors such as differences in vocal tract shape, size, shape, and density of the segmental inventory [8, 9], and by extension the density of the tonal inventory as well. Coarticulatory processes may seem to be inimical to the nature of contrast and maintenance of contrast, coarticulation, and the resultant acoustic variation is known to aid perception and offer perceptual salience where contrast may indeed be endangered due to contextual variation. In this paper, we have two major goals; one, we examine the relationship between coarticulation and contrast in two separate phonological processes, namely, tonal coarticulation in Mizo, and vowel harmony in Assamese. Second, we use automatic methods of classification of tones and vowels, in Mizo and Assamese, respectively, and on the basis of the classification accuracies argue that coarticulatory propensity exhibits sensitivity to the contrastive nature of the tones and vowel identities. Effectively, a phonological process that leads to neutralization of a particular contrast in question, be it tonal or segmental may lead coarticulatory patterns that seek to maintain contrast in phonological contexts where that particular phonological process does not apply. Universally, the greater extent of tonal coarticulation in the carryover direction suggests that the intertio-mechanical forces (see also residual planning in [10]) tend to be greater than the parallel activation of articulatory planning for contiguous or upcoming tones. An analogous phenomenon would be the extent of coarticulation in non-harmonic sequences in vowel harmony languages. While, greater carryover tonal coarticulation may lead to assimilatory processes such as tone sandhi, in the anticipatory direction $\sigma 1$ tones may exhibit dissimilation to the $\sigma 2$ tones, since in this direction tonal assimilation may lead to neutralization of contrasts. The rest of the paper is organized as follows: In section 2, in addition to motivating the need to examine the interaction between coarticulatory propensity and contrast, we also provide brief phonological descriptions of tones in Mizo and vowel harmony in Assamese. In section 3, we describe the speech materials and the classification methods that have been employed in our study. In section 4, we present the results from our classification of tones and vowels in Mizo and Assamese, respectively. In section 5, we present and motivate our conclusions.

2. Coarticulatory propensity and contrast

In this section, first, we describe the phonological processes that involve tones in Mizo and vowel harmony in Assamese. Following that, we examine how these processes interact with notions of contrast and coarticulation. While tonal coarticulation often leads to tone sandhi, i.e., tonal assimilation, V-to-V coarticulation can result in vowel harmony patterns. The crucial distinction, however, between tonal assimilation and vowel harmony, is that tone sandhi may also lead to tonal dissimilation in specific contexts, while vowel harmony patterns can be seen to affect specific vowels and may lead to opacity of application of vowel harmony rules in other contexts. In our analysis, we unify tonal dissimilation and opacity in vowel harmony, as processes that come about from a typical interaction between coarticulation and contrast preservation.

2.1. Lexical tones in Mizo

Mizo is a Tibeto-Burman language spoken most extensively in the state of Mizoram in the northeastern part of India., and also parts of Myanmar and Bangladesh. One of the first descriptive studies on Mizo claims the existence of four lexical tones ([11], [12], [13], [14]). However, these studies are not consistent with the description of the tones as found in later acoustic studies. [11]'s analysis observes that Mizo has H(igh), F(alling) and R(ising) tones and that the fourth tone is a mid level tone which can also be realized as a mid to low tone. On the other hand, [15], suggests that Mizo exhibits four tones; H, L, F and R.

[14], however, states that Mizo has H, L and F and an unmarked tone which is a mid or a low tone. A later study by [16] conducted with a Visi-Pitch meter, finds that there are two level tones High (H) and Low (L) and two contour tones Falling (F) and Rising (R). The L mentioned by [16] is a low level tone in the underlying form, but can surface as a mid or a low level tone. [16] also separates the L tones of derived words with a glottal stop in the coda from the L tone of non-derived words, as they consistently surface as an extra low tone. A recent acoustic analysis of lexical tones of Mizo [17], found the results to be consistent (to some extent) with the findings of [16]. The H and the F tone started at around 250 Hz, the H tone maintained the level throughout its realization and the F tone showed a falling contour. The L tone, although called a level tone, had a falling contour but started much lower than the H tone at about 220 Hz and fell to 205 Hz. The R tone started around 220 Hz, and had an initial dip for about 40% of its realization and then rose to 220 Hz. However, the extra low tone did not find mention in [17].

2.2. Vowel harmony in Assamese

Assamese is a modern Indo-Aryan language spoken largely in state of Assam in India. According to [18], Assamese has an inventory of eight vowels —/i, u, ʊ, e, o, ɛ, ɔ, ɑ/. Vowel harmony proceeds strictly in the leftward, i.e. anticipatory, direction and is triggered by suffixal /i/ and /u/ which spread the feature [+ATR] to preceding vowels /ɛ/, /ɔ/, and /ʊ/ changing them to /e/, /o/, and /u/, respectively. The harmony process is non-neutralizing, and it is only the ATR feature of the triggering vowel that gets shared with the stem vowels. [18] cites the data in 1 as examples of this pattern.

Table 1: *Assamese Vowel Harmony*

noun	suffix	adjective	Gloss
tɛz	-i	tezi	'strong'
bɔl	-i	boli	'strong'
zʊr	-i	zuri	'strong'

2.3. Coarticulatory directionality and contrast

Universally, tonal coarticulation is found to be greater in extent in the carryover direction compared to the anticipatory direction ([1], [2], [3], [4], [5]). In general, carryover coarticulation is understood to be due to intertio-mechanical forces, and, anticipatory effects are seen as a consequence of parallel activation of articulatory plans ([6]). Within our context, greater carryover coarticulation in Mizo leads to tonal assimilation such that σ 2 tones assimilate to σ 1 tones, and greater anticipatory coarticulation in non-harmonic Assamese vowel sequences are interpretable as assimilatory due to the non-neutralizing nature of the Assamese vowel harmony processes. Here, we offer a unified account of these processes by arguing that phonological processes that lead to neutralization of contrast give rise to coarticulatory directionality patterns that seek to maintain contrast (see [19] for vowel harmony processes in Telugu which lead to neutralization). On the other hand, if the outcome of the phonological process is non-neutralizing and does not endanger contrast then coarticulatory directionality can continue to proceed in the same direction as the phonological process, even in contexts where the phonological process does not apply. In Mizo, underlying σ 1 rising tones become low tones in the context of σ 2 high onset tones as an outcome of a dissimilatory process. We argue that this dissimilation is essential for maintaining a contrast between rising and falling/high tones. Coarticulatory assimilation in Mizo, however, doesn't endanger contrast, and hence can follow universal patterns where greater coarticulatory propensity is found in the carryover direction. In a contrastive study of Bengali (non-neutralizing harmony) and Telugu (neutralizing harmony), [19] show that, while Bengali non-harmonic sequences exhibit greater anticipatory coarticulation, which follows the same direction as the harmonic sequences, it is the opposite in Telugu non-harmonic sequences, where the harmonic process is anticipatory and neutralizing in nature. We, therefore, invoke maintenance of contrast, as crucial to determining coarticulatory propensity and directionality in Mizo and Assamese. These patterns are motivated by the observation that non-neutralizing phonological processes that come about due to coarticulation, be they vowel harmony or tone sandhi, still offer perceptual salience as a by-product of articulatory-gestural overlap. Neutralizing phonological processes, however, as an outcome of coarticulation, need to be countered, as can be seen in greater degree of carryover coarticulation in Telugu non-harmonic sequences [19].

3. Materials and methods

3.1. Materials

3.1.1. Mizo tone production

Three female native Mizo speakers were recorded. The material consisted of disyllabic ($\sigma 1 \sigma 2$) compound words of all possible combinations for the four tones; Falling, High, Low and Rising. For each combination, two words were taken which made it a list of 15*2=30 disyllabic compound words. Recordings were conducted in a soundproof room with Computer Speech Laboratory (CSL).The data was presented to the speakers in Roman script. Since tone is not specified in Mizo orthography, the word meanings were also presented to the speakers. Words were manually segmented and f0 was extracted at 5% increments of duration in the vowel with a Praat script. For voiced codas, f0 was marked from the onset of the vowel to the offset of the coda and for voiceless codas, till the offset of the vowel. f0 values were converted from Hz to Mel with the following equation:

$$mel = 1127 \log_e(1 + \frac{f}{700}) \quad (1)$$

Z-score normalized Mel values were analyzed to eliminate speaker specific effects. The data was split 80-20 into train and test sets, respectively. The distribution of word tones in the train/test sets was matched.

3.1.2. Assamese speech production

The corpus consisted of annotated read speech studio recorded by 13 native speakers of Assamese, 8 male and 5 female subjects. 1071 non-harmonic #…V_1CV_2…# sequences were isolated within the corpus, where non-harmonic refers to sequences involving the following pairs of vowels in any order: i-ɑ, i-o, i-ɔ, i-e, i-ɛ, ɛ-u, ɛ-ʊ, e-u, e-ʊ, ɑ-e, ɑ-ɛ, ɑ-o, ɑ-ɔ, ɑ-u, and ɑ-ʊ. No selection was made with respect to the intervening consonants in isolating these sequences. The classifier models were trained to predict vowel identity from formant measures from these isolated V_1CV_2 sequences. The formants were extracted using the FormantPro Praat script ([20]).

3.2. Method

3.2.1. *Tone prediction*

The neural net implementation from the nnet R package [21], which models networks with one hidden layer was used . We fixed a layer size of 20, and a decay value of .001 as the training parameters for the models. All other parameters were default for the implementation. Two sets of models were trained to predict σ1 and σ2 tones. Set 1 consisted of bilateral models trained on f0 values of both σ1 and σ2. Set 2 had unilateral models trained to predict the tone of one syllable from the f0 values of the other. Each set had three pairs of models: the first trained on f0 measures across the entire σ1 and σ2 duration; the second and third, on 66% and 33% of the duration (σ1-final and σ2-initial), respectively. Convenience training functions provided by the Caret R package [22] were used to find optimal model parameters, and 10-fold cross-validation was used to find the best fits to the data. Cohen's kappa coefficient was used as a performance metric to evaluate models during training. Both diagonal percentage accuracy and the kappa measure are reported for the test sets.

3.2.2. *Vowel identity prediction*

The models were trained on F1, F2, F3 and μF2-3 (mean of F2 and F3) values measured at 10 intervals each across half the duration of V_1 and V_2; last 50% of V_1 duration and first 50% of V_2 duration. The formant measurements and V_1 and V_2 class labels became the training feature set for the models. Five pairs of models of each kind (ANN and SVM) were trained to predict the V_1 and V_2 class labels from varying combinations of the formant features. The first pair (#1) trained on all formant measures from both vowels, the second pair (#2) trained on all formant measures from one vowel to predict the identity of the other, and subsequent pairs (#3,#4, and #5) were trained on limited sections of the feature set, taking formant measures only from the vowel initial, medial, and final portions. The SVM implementation is from the kernlab R package [23], and the ANNs utilized the multi-layer perceptron (MLP) implementation from the RSNNS package [24]. This implementation allowed us to train ANN models with three hidden layers, which performed better than the single hidden-layered models trained for the Mizo data. Each layer was set to a size of 80 cells, all other parameters were kept at the defaults. The train function provided by the Caret R package was used to find the best fitting models during training, as described in section 3.2.1.

4. Results

4.0.1. *Mizo - Tone Prediction*

Table 2: *Prediction accuracies of Bilateral Models*

	Accuracy (% Acc./Kappa)	
	σ1-Tone	σ2-Tone
#1 100%	0.88 / 0.84	**0.95 / 0.93**
#2 66%	0.91 / 0.88	**0.96 / 0.95**
#3 33%	0.88 / 0.84	**0.93 / 0.91**

Table 3: *Prediction accuracies of Unilateral Models*

	Accuracy (% Acc./Kappa)	
	σ1-Tone	σ2-Tone
#1 100%	**0.65 / 0.52**	0.56 / 0.41
#2 66%	**0.60 / 0.46**	0.53 / 0.37
#3 33%	**0.61 / 0.49**	0.55 / 0.39

Prediction accuracies from the bilateral models showed that σ2-tone predictions were consistently better than σ1-tone predictions (Table 2). Prediction accuracies from the unilateral models show that tone of σ1 is consistently better predicted than the tone of σ2 (Table 3). Since the bilateral models were trained on contextual f_0 values from both σ1 and σ2, we refrain from making an inference on the overall directionality of tonal coarticulatory effects in Mizo from the classifier accuracies in 2. Instead, based on the pattern of results in Table 3, we determine that assimilatory effects of tonal coarticulation in Mizo are greater in the carryover direction, which is consistent with the universal claim and also the claim specific to Mizo that carryover effects in tonal coarticulation are always assimilatory [5].

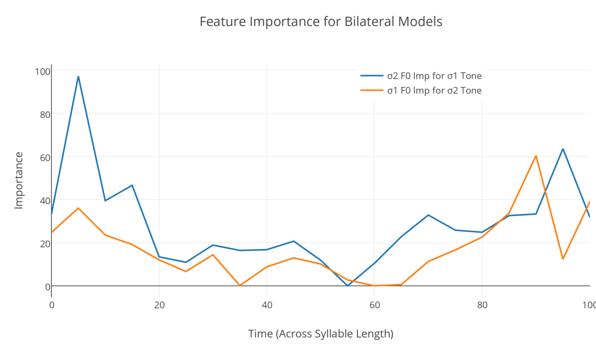

Figure 1: *Variable f_0 Importance for tone prediction across syllable duration - Bilateral Models*

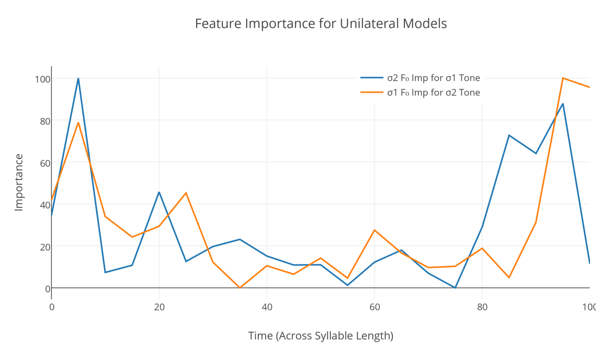

Figure 2: *Variable f_0 Importance for tone prediction across syllable duration - Unilateral Models*

Figures 1 and 2 show plots of the importance of f_0 values at different time-points across syllable durations to the ANN classifier predictions. f_0 information from the initial part of σ2 is most important in predicting σ1, while f_0 information from the final part of σ1 is most important in predicting σ2. We expect coarticulatory effects on syllable tones to be stronger at syllable final and initial portions, and the ANN classifier models are indeed more sensitive to those features. This gives us confidence in our interpretation of directionality patterns of tonal coarticulation based on the performance of these classifier models. Variable importance was calculated using an implemented function 'varImp' in [22].

4.0.2. Assamese - Vowel Prediction

Table 4: *Performance measures of ANN models - V_1 vs. V_2 Prediction*

ANN Models	Accuracy (% Acc./Kappa)	
	V1	V2
#1 All features	68% / 0.58	**69% / 0.60**
#2 All V_1 / All V_2	61% / 0.48	**63% / 0.52**
#3 V_1 final / V_2 initial	60% / 0.48	**62% / 0.51**
#4 V_1 initial / V_2 final	59% / 0.46	**60% / 0.49**
#5 V_1 medial / V_2 medial	63% / 0.52	**64% / 0.54**

Table 5: *Performance measures of SVM models - V_1 vs. V_2 Prediction*

SVM Models	Accuracy (% Acc./Kappa)	
	V1	V2
#1 All features	63% / 0.51	**67% / 0.57**
#2 All V_1 / All V_2	55% / 0.40	**60% / 0.47**
#3 V_1 final / V_2 initial	53% / 0.37	**60% / 0.47**
#4 V_1 initial / V_2 final	50% / 0.33	**56% / 0.42**
#5 V_1 medial / V_2 medial	55% / 0.39	**58% / 0.45**

Performance measures (percentage diagonal accuracy and Cohen's Kappa) of the ANN models are presented in Table 4 and the SVM models in Table 5. Both sets of models perform consistently better at V_2 prediction, even when the features are constrained. While the difference in the performance of the ANN V_1 and V_2 models is relatively marginal, the SVM models show clear and consistent directionality in their performance. The performance of the ANN models is superior to the SVM models, and it is notable that this difference is greater in the constrained models (#3,#4, and #5). Based on the directionality in the performance of the V_1 and V_2 models, we conclude that the directionality effect of coarticulation in Assamese non-harmonic sequences is greater in the anticipatory direction, which is the same direction as in the harmonic sequences. [25] report similar prediction accuracies in the anticipatory direction for American English.

5. Conclusions

In this paper, we have argued that tonal dissimilation and non-harmonic sequences in vowel harmony languages offer a unique test context for understanding the nature of interaction between coarticulation and contrast. We have shown that automatic classification of tones and vowels are one, affected by coarticulatory propensity, and two, by the nature of contrast in these languages. Artificial Neural Networks exhibit sensitivity towards modeling acoustic variation as a result of coarticulation, both in tones, and in non-harmonic sequences in vowel harmony languages. While, in tonal languages, dissimilatory processes function as a way of maintaining and preserving tonal contrasts, in spite of a greater propensity for tonal coarticulation in the carryover direction, in vowel harmony languages, especially ATR or featural harmony systems, coarticulation in non-harmonic sequences could yet proceed in the same direction as the harmony process because the harmonic process does not endanger contrast. The nature of the outcome of phonological processes; neutralizing or non-neutralizing, therefore, are central in understanding the nature of coarticulatory propensity and directionality. We have shown that in Mizo tonal sequences, in general, tonal coarticulation is greater in the carryover direction, which is also a universal pattern. While this reinforces our understanding that laryngeal adjustments in the carryover direction are preferred, dissimilatory processes may proceed in the opposite direction in contexts where contrast may get endangered. The uniformity in the directionality of coarticulatory propensity in Assamese non-harmonic and harmonic sequences, we see as a consequence of the non-neutralizing nature of the vowel harmony process. An additional goal of our study has also been to show that artificial neural networks offer a novel way of approaching and understanding acoustic variation as a result of coarticulation. We have also shown that artificial neural networks and automatic classification methods, such as SVMs, are extremely sensitive to acoustic variation as a consequence of articulatory-gestural overlap and adjustments. At the same time, ANNs also provide interpretable models of coarticulation and articulatory-acoustic mapping, even in the absence of articulatory data. We have argued that the interaction between coarticulatory propensity and contrastive patterns is crucially determined by the outcome towards preservation of contrast. Coarticulatory processes that are non-neutralizing can offer perceptual salience and therefore may continue to have the same directionality as the phonological process in question, while those that endanger contrast, are countered through other processes, which may include tonal dissimilation and change in the directionality of coarticulation in non-harmonic sequences as in Telugu ([19]).

6. References

[1] M. Brunelle, "Northern and Southern Vietnamese tone coarticulation: A comparative case study," *Journal of Southeast Asian Linguistics*, vol. 1, pp. 49–62, 2009.

[2] J. Zhang, "A directional asymmetry in Chinese tone sandhi systems," *Journal of East Asian Linguistics*, vol. 16, no. 4, pp. 259–302, 2007.

[3] Y. Xu, "Production and perception of coarticulated tones," *The Journal of the Acoustical Society of America*, vol. 95, no. 4, pp. 2240–2253, 1994.

[4] J. Gandour, S. Potisuk, S. Dechongkit, and S. Ponglorpisit, "Tonal coarticulation in Thai disyllabic utterances: a preliminary study," *Linguistics of the Tibeto-Burman Area*, vol. 15, no. 1, pp. 93–110, 1992.

[5] P. Sarmah, L. Dihingia, and W. Lalhminghlui, "Contextual variation of tones in Mizo," in *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015*, pp. 983–986, 2015.

[6] S. Tilsen, "Vowel-to-vowel coarticulation and dissimilation in phonemic-response priming," *UC Berkeley Phonology Lab Annual Report*, 2007.

[7] S. E. Öhman, "Coarticulation in vcv utterances: Spectrographic measurements," *The Journal of the Acoustical Society of America*, vol. 39, no. 1, pp. 151–168, 1966.

[8] S. Manuel, "The role of contrast in limiting vowel-to-vowel coarticulation indifferent languages," *Journal of the Acoustical Society of America*, vol. 88, pp. 1286–1298, 1990.

[9] S. Manuel, "Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts," in *Coarticulation: Theory, data and techniques* (W. Hardcastle and N. Hewlett, eds.), Cambridge University Press, 1999.

[10] S. Tilsen, "Inhibitory mechanisms in speech planning maintain and maximize contrast," in *Origins of Sound Change: Approaches to Phonologization* (A. C. L. Yu, ed.), Oxford University Press, 2013.

[11] W. Bright, "Singing in Lushai," *Indian Linguistics*, vol. 17, pp. 24–28, 1957.

[12] R. Burling, "Lushai phonemics," *Indian Linguistics*, vol. 17, pp. 148–155, 1957.

[13] F. Lehman, "On certain aspects of Mizo (Lushai) grammar," in *Paper for the 8th International Conference on Sino-Tibetan Languages and Linguistics, Berkeley*, 1975.

[14] L. Chhangte, "A preliminary grammar of Mizo language (unpublished ma dissertatbn)," *Arlington: University of Texas*, 1986.

[15] A. Weidert, *Componential Analysis of Lushai Phonology*. John Benjamins Publishing Company, jan 1975.

[16] L. Fanai, *Some aspects of the lexical phonology of Mizo and English: An autosegmental approach*. PhD thesis, Ph. D. dissertation, CIEFL, Hyderabad, India, 1992.

[17] P. Sarmah and C. R. Wiltshire, "A preliminary acoustic study of Mizo vowels and tones," *J. Acoust. Soc. Ind*, vol. 37, no. 3, pp. 121–129, 2010.

[18] S. Mahanta, *Directionality and locality in vowel harmony: With special reference to vowel harmony in Assamese*. Netherlands Graduate School of Linguistics, 2008.

[19] I. Dutta, S. Irfan, and K. Harsha, "Using ANNs for vowel identification from v-to-v coarticulation in non-harmonic vcv sequences," Presented at the 15th Conference on Laboratory Phonology, Ithaca, NY, 2016.

[20] Y. Xu, "Formantpro.praat," 2007-2015.

[21] B. Ripley, W. Venables, and M. B. Ripley, "Package 'nnet'," 2015.

[22] C. J. W. Max Kuhn, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt., *caret: Classification and Regression Training*, 2016. R package version 6.0-72.

[23] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab – an S4 package for kernel methods in R," *Journal of Statistical Software*, vol. 11, no. 9, pp. 1–20, 2004.

[24] C. Bergmeir and J. M. Benítez, "Neural networks in R using the stuttgart neural network simulator: RSNNS," *Journal of Statistical Software*, vol. 46, no. 7, pp. 1–26, 2012.

[25] J. Cole, G. Linebaugh, C. Munson, and B. McMurray, "Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach," *Journal of Phonetics*, vol. 38, no. 2, pp. 167 – 184, 2010.