ANNs for vowel identification from V-to-V coarticulation in non-harmonic VCV sequences

Indranil Dutta, Irfan S. & K.R.S. Harsha
Department of Computational Linguistics
The English and Foreign Languages University
indranil@efluniversity.ac.in

Introduction

- Primary factor influencing acoustic variation in speech is the variable nature of overlap between articulatory gestures, more generally known as coarticulation [6]
- A special case of coarticulation is vowel-to-vowel (henceforth, V-to-V) coarticulation
- Phonologized V-to-V patterns are a conditioning factor for development of vowel harmony patterns [8, 5]
- Reduction in phonetic distinctiveness between vowels is known to be compensated perceptually [1]
- Vowel harmony develops from failure to perceptually compensate for reduction in acoustic distinctiveness due to coarticulation [5]
- Telugu vowel harmony patterns advanced from the anticipatory direction leading to neutralization of vowel contrasts between /iCu/->[uCu] sequences
- Vowel harmony patterns in Telugu result in a contrast neutralization

What we find

- Extent of carryover coarticulation is still greater in Telugu non-harmonic contexts
- Trained on non-harmonic #…V_1CV_2…# sequences, a single-layered Artificial Neural Network (ANN) model predicts the identity of V_1 from V_2 (carryover) consistently better than the V_2 from V_1 (anticipatory)

Carryover and anticipatory coarticulation in vowel harmony languages

- Anticipatory coarticulation is related to articulatory planning
- Carryover coarticulation is said to result from mechanico-intertial constraints
- At a cognitive level, anticipatory coarticulation results from the interference of articulatory planning of neighboring segments in the production of current segments[4, 10]
- Carryover coarticulation has been understood in terms of mechanico-inertial constraints, where inertial properties of articulators involved in prior segments interfere with the articulatory targets of current segments
- Carryover coarticulation has also been shown to have a cognitive basis, in that residual articulatory planning may get integrated into the production of V_2 from V_1 [12, 13].
- Nature, extent, and magnitude of coarticulation in vowel harmony languages not well understood, and not operationalized for determining weights for acoustic features for ASR and other automatic classification tasks

Vowel harmony and contrast neutralization

- Telugu vowel harmony
- Trigger is the suffixal vowel, [u] $\Leftarrow$ in the anticipatory direction
- Target $\rightarrow$ stem vowels till unaffected stem initial vowel
- The low vowel, [a], is opaque to the harmony process
- Unlike, Turkish, Mongolian, Bengali and other featural harmony languages, Telugu vowel harmony thought of as long distance vowel copying - target vowels are copies of the trigger, [u], [3, 11, 15]
- Consequence of vowel harmony in Telugu - neutralization of contrast between [i] and [u], in the stem

Materials and Methods

- First three formants are well accepted as primary indicators of the spectral identity of vowels.
- Formant features have also been reported to be adequate for training ANNs at vowel classification tasks [17].
- Training a set of single-layered ANN models to predict vowel identity from acoustic features in isolated VCV sequences extracted from phonetically balanced speech recordings of 1 Telugu speaker, and testing the predictions

Data

- Consisted of a 1000 sentence read-speech corpus from the IIIT-H Indic Speech Database [7].
- #…V_1CV_2…# sequences were extracted, where V_1 and V_2 were any two distinct choices from the following: [a, i, u]
- We found 1782 instances of these sequences in Telugu
- FormantPro [16] was used to calculate F1, F2, F3, and μF2_3 (mean of F2 and F3) at 20 time intervals for V_1 and V_2 in each sequence
- Formant measures for the 10 intervals across the second half of V_1 and the first half of V_2 retained for training purposes

Classifier

- Neural net implementation - nnet R [9] package - models simple neural networks with one hidden layer [14]
- Initial cross-validated training passes set a layer size of 20 and a decay parameter of 0.001
- On each of the 2 datasets, 2 models were trained, learning to predict V_1 and V_2 identity, respectively.
- Each model trained to predict V_1 or V_2 from features across all 10 time intervals of the other vowel, and the 5th and 6th interval of the vowel being predicted.
- To ensure consistency and decrease bias within the trained models, the models were trained on 10 bootstrap resamples from the datasets.
- Training set to optimize the Kappa measure of classification accuracy
- Accuracy and Kappa measures of each model, returned after the bootstrap training iterations, taken as the indicators of performance of the models

Results

- Prediction accuracies of the models
- Feature importance for the models, calculated using a method based on [2] as implemented in the varImp.nnet method in the caret R package.
- Inferences about the directionality of coarticulatory effects

Prediction accuracies

V1		V2	
92.48%	**0.868**	84.66%	0.762

Table 1: Telugu V1/V2 Predictions in Non-Harmonic context

- V_1 prediction accuracy from V_2 features is better than V_2 prediction accuracy from V_1 features

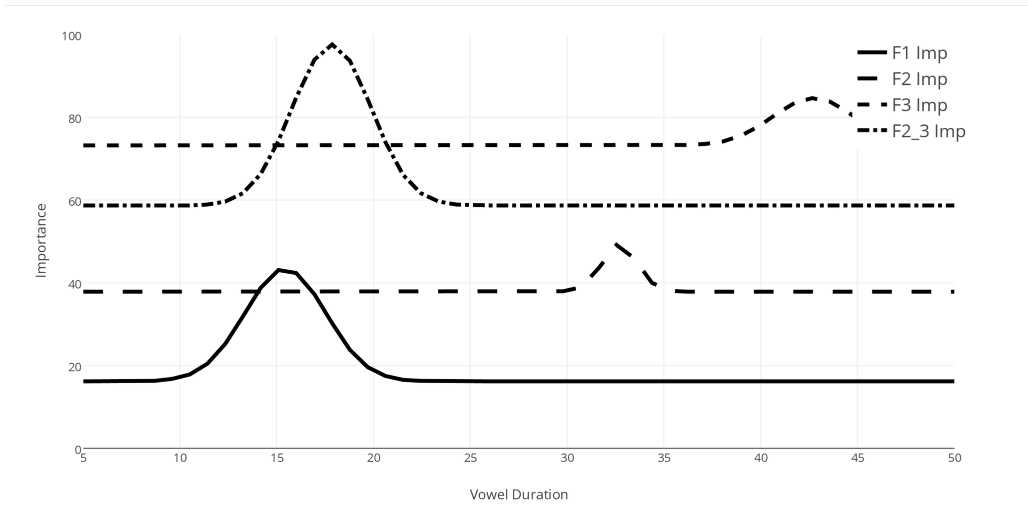

Figure 1: V_2 F_n Importance for V_1 Prediction - Non-Harmonic Telugu Contexts

- F_3 is the overall most important feature across V_2 duration (Figure 1)
- V_2-initial position, however, we see a peak in F_1 importance, which otherwise is the least important feature for V_1 prediction (Figure 1)
- Indicative of F_1 susceptibility to coarticulatory perturbation in this position (Figure 1)

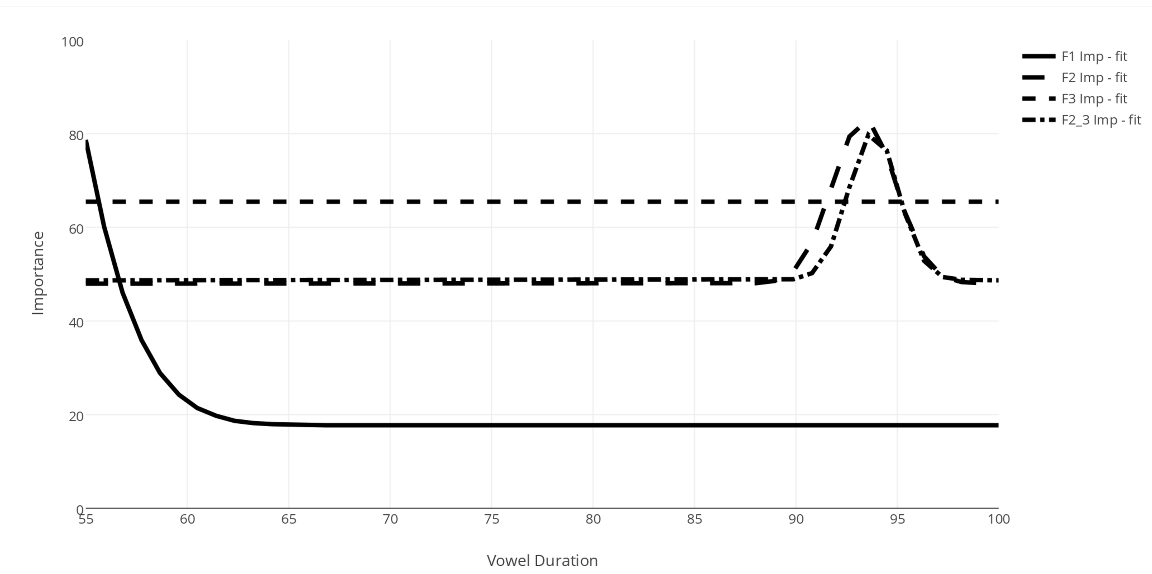

Figure 2: V_1 F_n Importance for V_2 Prediction - Non-Harmonic Telugu Contexts

- F_3 is again overall, the most important feature (Figure 2)
- F_1 is also the least important feature, but in the V_1-medial position, its importance is at a peak, making it one of the most important features at this position (Figure 2)

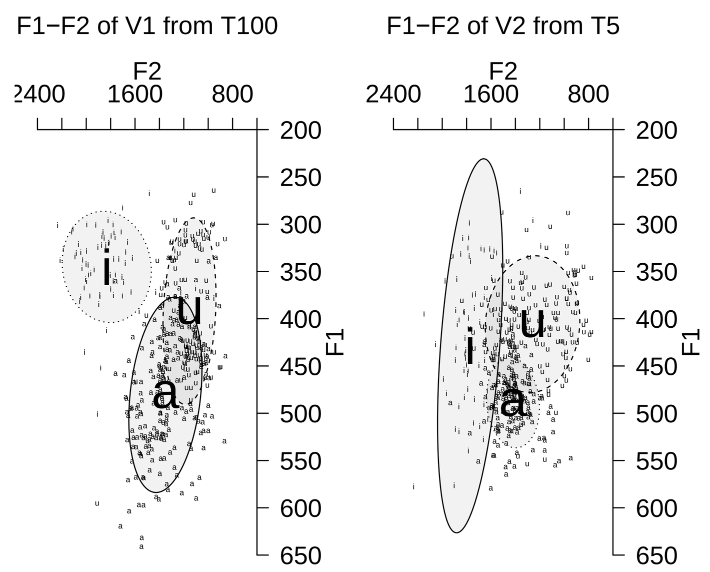

Figure 3: F1-F2 plots for V_1 (left panel) and V_2 (right panel), at time points T100 and T5 respectively, for Telugu

- V_1 and V_2 in the F_1-F_2 space at time points most proximal to the intervening consonant (Figure 3)
- V_2 vowels distributed closer to each other, implying greater perturbation due to carryover coarticulatory effects in the V_2 position
- [i] is most affected, with F_1 at its mean position lowering between 400-450Hz (Figure 3, right panel) vs. around 350Hz in the V_1 position (Figure 3, left panel)
- In Telugu, [i-C-u] turn into [u-C-u] sequences, **but in non-harmonic cases, [i] in the V_1 position appears to be resistant to coarticulatory perturbation**

Conclusions

- Directionality effect of coarticulation in Telugu non-harmonic sequences is found to be greater in the carryover direction
- Unlike feature spreading harmonies, in Telugu vowel harmony, the suffix vowel replaces all the stem vowels except the initial
- Leads to complete assimilation and neutralization from the anticipatory direction
- Anticipatory coarticulation, therefore, does not offer perceptual benefits in non-harmonic sequences, instead, carryover coarticulation increases perceptually relevant contrasts

References

[1] Patrice Speeter Beddor, James D. Harnsberger, and Stephanie Lindemann. Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4):591 – 627, 2002.

[2] Muriel Gevrey, Ioannis Dimopoulos, and Sovan Lek. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling*, 160(3):249–264, 2003.

[3] Dworak Kissock. Telugu vowel assimilation: Harmony, umlaut, or neither? In *Proceedings of the Seventeenth Manchester Phonology Meeting*, 2009.

[4] James G. Martin and H. Timothy Bunnell. Perception of anticipatory coarticulation effects. *The Journal of the Acoustical Society of America*, 69(2):559–567, 1981.

[5] John J Ohala. Towards a universal, phonetically-based, theory of vowel harmony. In *Third International Conference on Spoken Language Processing*, pages 491–494, 1994.

[6] Sven EG Öhman. Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(1):151–168, 1966.

[7] Kishore Prahallad, Naresh Kumar Elluru, Venkatesh Keri, Rajendran S, and Alan W. Black. The IIIT-H indic speech databases. In *INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association, Portland, Oregon, USA, September 9-13, 2012*, pages 2546–2549, 2012.

[8] Marek Przezdziecki. Vowel harmony and vowel-to-vowel coarticulation in three dialects of yoruba. *Working Papers of the Cornell Phonetics Laboratory*, 13:105–124, 2000.

[9] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[10] Daniel Recasens, Jordi Fontdevila, and Maria Dolors Pallarès. Velarization degree and coarticulatory resistance for /i/ in catalan and german. *Journal of Phonetics*, 23(1–2):37 – 52, 1995.

[11] P Sailaja. Syllable structure of telugu. 4:743–746, 1999.

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

[13] Sam Tilsen. Inhibitory mechanisms in speech planning maintain and maximize contrast. In Alan C. L. Yu, editor, *Origins of Sound Change: Approaches to Phonologization*. Oxford University Press, 2013.

[14] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.

[15] Robert W Wilkinson. Tense/lax vowel harmony in telugu: the influence of derived contrast on rule application. *Linguistic Inquiry*, pages 251–270, 1974.

[16] Yi Xu. Formantpro.praat, 2007-2015.

[17] SA Zahorian and Amir Jalali Jagharghi. Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*, 94(4):1966–1982, 1993.