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]

The EFL University[1], University of Illinois at Urbana-Champaign[2], University of Florida[3], Indian Institute of Technology Guwahati[4]

INTERSPEECH 2017, Stockholm
August 20-24, 2017

Outline

Tonal coarticulation

- Tonal coarticulation is universally found to be greater in extent in the carryover direction compared to the anticipatory direction leading to assimilatory processes [Brunelle, 2009, Zhang, 2007, Xu, 1994, Gandour et al., 1992].
- 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 [Tilsen, 2007].
- Specific tone sequences, however, may result in bidirectional effects, where anticipatory effects tend to get realized through dissimilatory processes and carryover effects may result in assimilation [Xu, 1997, Shih, 1988].
- Asymmetrical patterns between anticipatory and carryover effects in terms of the extent of the spread of coarticulation are not attested for Mizo[Sarmah et al., 2015].

How do vowel harmony systems come about?

- Vowel harmony: long-distance assimilatory phonological process
- Acoustic variation due to V-to-V coarticulation, once phonologized - a conditioning factor for the development of vowel harmony [Przezdziecki, 2000, Ohala, 1994].
- Resultant reduction in phonetic distinctiveness between vowels compensated perceptually [Beddor et al., 2002]
- Vowel harmony develops from failure to perceptually compensate for reduction in acoustic distinctiveness due to coarticulation [Ohala, 1994].

Anticipatory coarticulation

- Anticipatory coarticulation understood in the context of articulatory planning
- Anticipatory coarticulation, at a cognitive level, results from interference of articulatory planning in the production of segments from neighboring segments [Martin and Bunnell, 1981, Recasens et al., 1995]

Carryover coarticulation

- Carryover coarticulation understood in the context of mechanico-intertial constraints
- Carryover coarticulation - inertial properties of articulators involved in prior segments interfere with the articulatory targets of current segments.
- More recently, 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 [Recasens, 1984, Tilsen, 2007, Tilsen, 2013].

Assymetry in the nature and extent of coarticulation

- Asymmetry in the magnitude and extent of anticipatory and carryover coarticulation looked into from several perspectives:
- Language specific [Manuel, 1990, Manuel, 1999],
- Based on syllable structure and the place of articulation of the intervening consonant [Modarresi et al., 2004]
- Vowel harmony systems develop as a consequence of lack of perceptual compensation - needed to maintain contrast between neighboring vowels
- That is to say assimilation of vowel features [Nguyen and Fagyal, 2008, Cole, 2009, Benus and Gafos, 2007].

Coarticulation: It's a good thing!

- Non-linear relationship between articulatory and acoustic parameters
- Articulatory sloppiness
- Large changes in the articulatory parameter don't always result in large changes in the acoustics; follows a sigmoid function
- Articulatory-gestural overlap between contiguous segments may aid vowel identification and recovery

$$V_1^{V_2} C^{V_1} V_2$$

Research questions

- Anticipatory coarticulation may have led to harmonic processes in Assamese - right-to-left direction
- What kind of predictions about non-harmonic sequences? Carryover or anticipatory?
- What role does contrast have to play in the magnitude and extent of coarticulation?

Data - Mizo

- Three female native Mizo speakers were recorded.
- 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. Tone being not specified in Mizo orthography, word meanings were also presented to the speakers.
- Manual segmentation of words was followed by f_0 extraction at 5% increments of duration in the vowel with a Praat script.
- f_0 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.

Classifiers - Mizo

- The neural net implementation from the nnet R package, which models networks with one hidden layer, was used [Ripley et al., 2015].
- We fixed a layer size of 20, and a decay value of .001 as the training parameters for the models.
- 10-fold cross validation was employed to ensure consistency and decrease bias within the trained models.
- Two sets of models were trained to predict $\sigma 1$ and $\sigma 2$ tones.
- Set 1 consisted of bilateral models trained on f_0 values of both $\sigma 1$ and $\sigma 2$.
- Set 2 had unilateral models trained to predict the tone of one syllable from the f_0 values of the other.
- Models trained to optimize the Cohen's kappa coefficient

Data - Assamese

- Annotated, studio recorded, read speech corpus
- 13 native speakers of Assamese, 5 male + 8 female
- 921 non-harmonic VCV sequences extracted
- FormantPro [Xu, 2015] used to calculate F_1, F_2, F_3, and μF_2_3 (mean of F_2 and F_3) at 20 time intervals for V_1, V_2 in each sequence.
- 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

Classifiers - Assamese

- Neural net implementation from the RSNNS package [Bergmeir and Benítez, 2012] for R [R Core Team, 2015], with two hidden layers
- SVM implementation with a radial kernel, from package Kernlab [Karatzoglou et al., 2004]
- Convenience training/optimisation wrappers provided in caret R package [Max Kuhn et al., 2016] used to tune hyper-parameters with 10-fold cross-validation

Classifiers - Assamese

- 5 pairs of classifiers (for V_1 and V_2 prediction) per algorithm
- **#1, #2** models trained to predict the identity of one vowel from **all** the acoustic features of the other or of both
- **#3** models trained to predict V_1 identity from **V_2-initial acoustics**, and V_2 from **V_1-final acoustics** (i.e. formant values of either vowel at times close to the POA)
- **#4** models trained on **V_1-initial** and **V_2-final** formants
- **#5** models trained on the middle portions of V_1 and V_2

Mizo Tone Predictions - Bilateral Models

- Prediction accuracies from the bilateral models (Set 1) showed that σ2-tone predictions were consistently better than σ1-tone predictions.
- Since the bilateral models were trained on contextual f_0 values from both σ1 and σ2, these models were not used to measure directionality of tonal coarticulation

Table: Performance measures of ANN models

	Bilateral Models	
	σ1-Tone (diag. acc/kappa)	σ2-Tone (diag. acc/kappa)
Set 1, 100%	0.88 / 0.84	0.95 / 0.93
Set 1, 66%	0.91 / 0.88	0.96 / 0.95
Set 1, 33%	0.88 / 0.84	0.93 / 0.91

Mizo Tone Predictions - Unilateral Models

- Prediction accuracies from the unilateral models (Set 2) show that tone of $\sigma 1$ is consistently better predicted than the tone of $\sigma 2$.
- Assimilatory effects of tonal coarticulation in Mizo are greater in the carryover direction, which is consistent with the universal claim and also claim specific to Mizo that carryover effects in tonal coarticulation are always assimilatory [Sarmah et al., 2015].

Table: Performance measures of ANN models

	Unilateral Models	
	$\sigma 1$-Tone (diag. acc/kappa)	$\sigma 2$-Tone (diag. acc/kappa)
Set 2, 100%	0.65 / 0.52	0.56 / 0.41
Set 2, 66%	0.60 / 0.46	0.53 / 0.37
Set 2, 33%	0.61 / 0.49	0.55 / 0.39

Assamese V_1/V_2 Predictions - Normalised per speaker - ANN

- V_2 predictions stronger, which implies greater anticipatory coarticulation

Table: Performance measures of ANN models

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**

Assamese V_1/V_2 Predictions - Normalised per speaker - SVM

- V_2 predictions stronger, which implies greater anticipatory coarticulation

Table: 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**

- Earlier vowel specific z-score normalizations led to lower accuracies and contrary results in the directionality effect

Other Languages

- Bengali: V_2 predictions stronger in both harmonic and non-harmonic contexts, which implies greater coarticulation in the anticipatory direction
- Telugu: Stronger V_1 predictions imply greater coarticulation in the carryover direction [Dutta et al., 2016]
- Hindi: V_1 and V_2 models evenly matched, only marginal differences in accuracies

Conclusions

- Coarticulation in tones and non-harmonic sequences in vowel harmony languages offer a unique test context for understanding the nature of interaction between coarticulation and contrast
- ANN classifiers exhibit sensitivity to acoustic variation due to coarticulation, in tones and VCV sequences

References I

Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (2002).
Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates.
Journal of Phonetics, 30(4):591 – 627.

Benus, S. and Gafos, A. I. (2007).
Articulatory characteristics of hungarian 'transparent' vowels.
Journal of Phonetics, 35(3):271 – 300.

Bergmeir, C. and Benítez, J. M. (2012).
Neural networks in R using the stuttgart neural network simulator: RSNNS.
Journal of Statistical Software, 46(7):1–26.

Brunelle, M. (2009).
Northern and Southern Vietnamese tone coarticulation: A comparative case study.
Journal of Southeast Asian Linguistics, 1:49–62.

Cole, J. (2009).
Emergent feature structures: harmony systems in exemplar models of phonology.
Language Sciences, 31(2–3):144 – 160.
Data and Theory: Papers in Phonology in Celebration of Charles W. Kisseberth.

References II

Dutta, I., Irfan, S., and Harsha, K. (2016).
Using ANNs for vowel identification from v-to-v coarticulation in non-harmonic vcv sequences.

Gandour, J., Potisuk, S., Dechongkit, S., and Ponglorpisit, S. (1992).
Tonal coarticulation in Thai disyllabic utterances: a preliminary study.
Linguistics of the Tibeto-Burman Area, 15(1):93–110.

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004).
kernlab – an S4 package for kernel methods in R.
Journal of Statistical Software, 11(9):1–20.

Manuel, S. (1990).
The role of contrast in limiting vowel-to-vowel coarticulation indifferent languages.
Journal of the Acoustical Society of America, 88:1286–1298.

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

References III

Martin, J. G. and Bunnell, H. T. (1981).
Perception of anticipatory coarticulation effects.
The Journal of the Acoustical Society of America, 69(2):559–567.

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

Modarresi, G., Sussman, H., Lindblom, B., and Burlingame, E. (2004).
An acoustic analysis of the bidirectionality of coarticulation in {VCV} utterances.
Journal of Phonetics, 32(3):291 – 312.

Nguyen, N. and Fagyal, Z. (2008).
Acoustic aspects of vowel harmony in french.
Journal of Phonetics, 36(1):1 – 27.

Ohala, J. J. (1994).
Towards a universal, phonetically-based, theory of vowel harmony.
In *Third International Conference on Spoken Language Processing*.

References IV

Przezdziecki, M. (2000).
Vowel harmony and vowel-to-vowel coarticulation in three dialects of yoruba.
Working Papers of the Cornell Phonetics Laboratory, 13:105–124.

R Core Team (2015).
R: A Language and Environment for Statistical Computing.
R Foundation for Statistical Computing, Vienna, Austria.

Recasens, D. (1984).
Vowel-to-vowel coarticulation in catalan vcv sequences.
The Journal of the Acoustical Society of America, 76(6):1624–1635.

Recasens, D., Fontdevila, J., and Pallarès, M. D. (1995).
Velarization degree and coarticulatory resistance for /i/ in catalan and german.
Journal of Phonetics, 23(1–2):37 – 52.

Ripley, B., Venables, W., and Ripley, M. B. (2015).
Package 'nnet'.

References V

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

Shih, C. (1988).
Tone and intonation in Mandarin.
Working Papers of the Cornell Phonetics Laboratory, 3:83–109.

Tilsen, S. (2007).
Vowel-to-vowel coarticulation and dissimilation in phonemic-response priming.
UC Berkeley Phonology Lab Annual Report.

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

Xu, Y. (1994).
Production and perception of coarticulated tones.
The Journal of the Acoustical Society of America, 95(4):2240–2253.

References VI

Xu, Y. (1997).
Contextual tonal variations in Mandarin.
Journal of phonetics, 25(1):61–83.

Xu, Y. (2007-2015).
Formantpro.praat.

Zhang, J. (2007).
A directional asymmetry in Chinese tone sandhi systems.
Journal of East Asian Linguistics, 16(4):259–302.