Modelling Vowel Harmony with Neural Networks

Irfan S

May 2, 2018

1 Set-up

1.1 Vowel Harmony

Vowel harmony (henceforth VH) is a long-distance assimilatory process involving vowels within a certain morphological or prosodic domain. Long-distance here means that that the assimilatory effect affects vowels that are not produced in contiguity, and may have one or more intervening segments. Anderson [1980] presents a discussion of examples of VH attested in various languages with a view to arriving at a unified formal definition for the process. Acknowledging the variety with which VH is realised in different languages, Anderson finally concludes that the formal mechanism of VH is not distinct from that of other forms of metaphony, such as dissimilation or Umlaut rules, stating that "this means there is nothing special about vowel harmony". With specific reference to the phonetic motivatedness for VH, Anderson, while acknowledging that such motivations are apparent in the historical basis of VH, says that phonetic grounding cannot be considered a necessary condition for vowel harmony systems from a synchronic point of view.

1.2 Coarticulatory approaches to Vowel Harmony

A different approach is proposed by Ohala [1994] in the form of a universal and phonetically-based theory of VH. According to Ohala, VH is the "fossilised result of purely phonetic between-vowel assimilations", the result of sound change that occurs when two independent production events (such as a vowel and the perturbation caused in the production of that vowel by a neighbouring vowel, in the case of VH) are misperceived as one by the listener. If this happens repeatedly over time and with many speaker-listener pairs, it could cause the a new pronunciation norm to emerge that would be integrated into the language's grammar. In this way perturbatory phonetic effects, or coarticulatory effects, of a unit of speech on neighbouring units could be linked to phonological processes such as harmony. An inconclusive attempt to find experimental evidence in support of this was made by Busá and Ohala [1999]. Stronger supporting evidence was found by Przezdziecki [2005], who compared three different dialects of Yoruba, one attesting a VH process and two without. It was found that the effect and context of the high vowel harmony pattern in VH-attesting dialect was similar to the coarticulatory behaviour observed for the corresponding vowels in the other two dialects. Further, a decision-tree learner model applied to the data by Przezdziecki resulted in splits corresponding to the natural classes involved in the harmony process (±ATR), even when unconstrained by featural specifications. In the author's words, "pre-existing abstract features were not necessary for the learner to extract a phonological pattern from a coarticulatory pattern" (see Nevins, 2010 for a rejoinder asserting that VH and coarticulation are entirely distinct processes).

1.3 Temporal Domain of coarticulation

One of the questions raised against linking VH and coarticulation is based on the assumption that grounding VH in a phonetic process such as coarticulation would imply strict locality [Kimper, 2017], while VH rules can apply over long distances, across

syllable boundaries or even word boundaries in some cases. Grosvald [2009] shows that vowel-to-vowel coarticulatory effects can even be detected across at least 3-vowel distances in productions by speakers of American English, a non-VH language. Interestingly, Grosvald also found that these coarticulatory effects could also be perceived across similar spans, noting considerable inter-subject variability in both the production and perception tasks. Using nonce stimuli in a perception task setting, Kimper [2017] shows that vowel harmony is perceptually advantageous even for non-adjacent vowels.

1.4 Exemplar modelling approaches to Vowel Harmony

Cole [2009] introduces an alternative to the representational models of generative phonology in keeping with the evolutionary, phonetically grounded approach advocated by Blevins [2004] and Ohala [1994]. She proposes an exemplar-theoretic model, with no a priori feature representations imposed as inputs. Instead, she makes the claim that the feature structure required for a phonological account of harmony would be an emergent property of the exemplar model's encoding. Cole points to the Simple Recurrent Network model demonstrated by McClelland and Elman, 1986 (cited in Cole, 2009), as a suitable way to implement the exemplar model, which would essentially model transitional probabilities between input elements of speech.

Mailhot [2010] demonstrates LɪʙPʜoɴ[1], an implementation of an exemplar-based learning algorithm. The input *instances* to the model are acoustic features corresponding to word-level units. The model implements a perception (instance to label) and a production module (label to stored representation of instance). The reported results indicate that the model is able to generalise and produce harmonically correct morphological forms that it was not exposed to during training. Mailhot [2010] provides a proof-of-concept implementation of the modelling approach suggested by Cole [2009], even though this implementation directly produces a word-label and not phone transi-

---

[1]the Lazy Instance-based Phonologist

tional probabilities.

2 The proposed model

I propose that a Long Short Term Memory (LSTM) network of sufficient architecture, a modified kind of recurrent neural network, be used to implement an encoder-decoder model. The encoder would be trained to generate transition probabilities for a sequence of utterance level acoustic features. The decoder will be trained to regenerate those sequences from the probabilities output by the encoder. In tandem, the encoder-decoder would approximate a perception/production model of an individual speaker.

2.1 Data

Large corpora of natural speech recordings from multiple speakers of both harmony and non-harmony languages would be required. A large enough corpus would ideally be representative of the language community.

2.2 Experiments

Experiments would involve training the encoder/decoder model on instances of utterance level acoustic features. The variance in the input could be controlled to simulate various "settings" of the language community, which should result the model in learning differently. The outputs produced by the model could be tested to see if featural structures required for harmony (or other phonological processes) emerge, as predicted by Cole [2009].

References

Stephen R. Anderson. Problems and perspectives in the description of vowel harmony. *Issues in Vowel Harmony*, page 1, 1980. ISSN 0165-7763. doi: 10.1075/slcs.6.02and.

URL `http://dx.doi.org/10.1075/slcs.6.02and`.

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

M Busá and John Ohala. In search of the perceptual correlates of vowel harmony. In *Proceedings of the XIVth International Congress of Phonetic Sciences (San Francisco) Vol*, volume 1, pages 357–360, 1999.

Marek Przezdziecki. *Vowel harmony and coarticulation in three dialects of Yoruba: phonetics determining phonology*. PhD thesis, Cornell University, 2005.

A. Nevins. *Locality in Vowel Harmony*. EBSCO ebook academic collection. MIT Press, 2010. ISBN 9780262140973. URL `https://books.google.com/books?id=XIG8ZDzuAHcC`.

Wendell Kimper. Not crazy after all these years? perceptual grounding for long-distance vowel harmony. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1):nil, 2017. doi: 10.5334/labphon.47. URL `https://doi.org/10.5334/labphon.47`.

Michael Grosvald. Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, 37(2):173 − 188, 2009. ISSN 0095-4470. doi: https://doi.org/10.1016/j.wocn.2009.01.002. URL `http://www.sciencedirect.com/science/article/pii/S0095447009000035`.

Jennifer Cole. Emergent feature structures: harmony systems in exemplar models of phonology. *Language Sciences*, 31(2-3):144–160, 3 2009. ISSN 0388-0001. doi: 10.1016/j.langsci.2008.12.004.

Juliette Blevins. Evolutionary phonology. 2004. doi: 10.1017/cbo9780511486357. URL `http://dx.doi.org/10.1017/cbo9780511486357`.

Frédéric Mailhot. Instance-based acquisition of vowel harmony. In *Proceedings of the 11th Meeting of the ACL Special Interest Group on Computational Morphology and Phonology*, pages 1–8. Association for Computational Linguistics, 2010.