Learning hidden gradient features: from muscular activation to featural representation

The motivation for gradience in phonology has received renewed support in recent work (e.g. Lionnet 2016; Rosen 2016; Smolensky & Goldrick 2016; Zimmermann to appear). The question addressed here is: where does gradience in phonology come from? Using a neural network model, this study argues that phonological gradience emerges from the mapping between continuous activations of tongue muscles and discrete phonological representations. The neural network model demonstrates that coarticulatory effects as critical sources of phonological phenomena are learned as hidden gradient features in the mapping.

Following cross-linguistic surveys (Bateman 2007; Kochetov 2011), high vowels tend to trigger coronal palatalization, but non-front high vowels palatalize coronals only when front high ones also do in the same language. In addition, the non-high back vowel /o/ never palatalizes coronals, while the non-high front vowel /e/ triggers coronal palatalization in some languages, as shown in Table 1. These typological patterns cannot be explained by simply assuming coronal palatalization as spreading of [+high] or [-back] feature from a trigger vowel to a target coronal.

Table 1. Triggers for coronal palatalization

	[-back]	[+back]
[+high, -low]		
[-high, -low]		
[-high, +low]		

	<Languages>
/i/	Apalai, Basque, Fongbe, Japanese, Korean, Luvale, Mandarin, Nishnaabemwin
/i,u/	Maori, Sentani
/i,e/	Amharic, Hausa, Polish, Tswana
/i,e,u/	Coatzospan Mixtec, Tohono O'odham

I propose that coronal palatalization is a coarticulatory effect on the basis of coordination of tongue muscles. Through articulatory simulations using a 3D tongue model in Artisynth (Lloyd et al. 2010), Jang (2018, 2019) has shown that different strengths of /i/ vs. /u/ in triggering palatalization of apical coronals come from activations of distinct tongue muscles that cause different degrees of lowering of the tongue tip in the vocalic articulation. In this study, I model mappings between muscular activations and phonological features using a neural network. The model shows that the typological asymmetry of vowels as triggers of coronal palatalization, not only /i/ vs. /u/ but also /e/ vs. /o/, emerges from the mapping by learning hidden and gradient vocalic features for coronals that result from coarticulation from the following vowel.

In this study, I model the mapping between activations of tongue muscles and phonological features using a feed-forward multi-layer neural network in Python's scikit-learn. In the network, the input is a matrix including temporal changes in activation values of tongue muscles and the output is a matrix of phonological features with values -1, 0, or 1 (e.g. [-high], unspecified, [+high]). A single hidden layer is a regressor using the rectified linear unit function, $f(x) = \max(0,x)$. In the training (1000 iteration, .001 alpha), muscular activations of isolated coronal consonants /d, dʒ/, isolated vowels /i, e, æ, u, o, a/, and CV sequences /di, du, da, dʒi, dʒu, dʒa/ map into featural representations. In the CV training data, tongue muscle activations for /d, dʒ/ were different depending on vocalic contexts to fully achieve articulatory targets for coronals that have the same featural representations. In the test data, the muscular activations of isolated /d/ overlap with the muscular activations of isolated vowels. The aim of tests is to check if the mapping model can learn the coarticulatory effects of the vocalic contexts on the featural representations for the same coronals that have the same muscular activations.

As a result of this mapping procedure, coronals acquire gradient activations of features depending on the following vowels. Activation values of the coronal feature [distributed] for /d/ show trimodal distribution as shown in Figure 1: /d/ before /i, e/ > /æ, u, o/ > /a/. Another coronal feature [anterior] shows the same grouping with the reversed distribution of activation values.

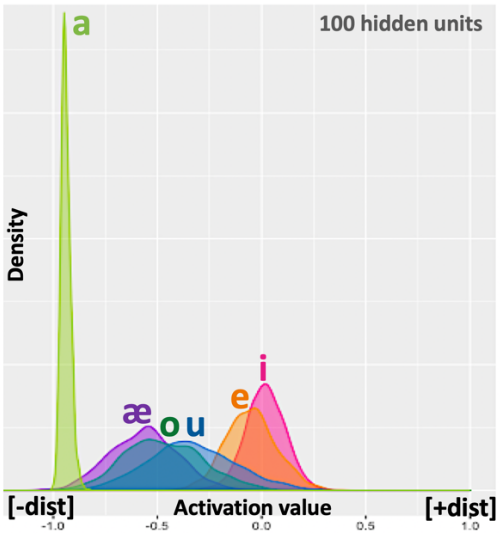

Figure 1. Distribution of [dist] activations for /d/ depending on vocalic contexts

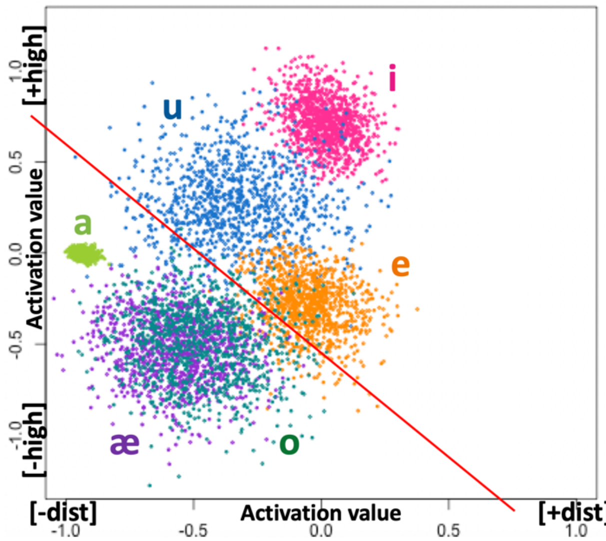

Figure 2. Distribution of [dist] and [high] activations for /d/ depending on vocalic contexts

The cluster of /æ, u, o/ requires further attention because /u/ palatalizes coronals in some languages but not /o/ and /æ/, as shown in Table 1. The appropriate typological asymmetries are obtained when activation values of the feature [high] for coronals are considered together with [distributed], as shown in Figure 2. In the phonetics-phonology mapping model, coronals acquire the feature [high] with gradient activation values as a coarticulatory effect of the vocalic contexts. Since the isolated coronals do not have the vocalic feature [high], the result shows that a hidden gradient feature emerges from coarticulatory effects that are learned based on muscular activations. The red line in Figure 2 is a suggested threshold for triggering coronal palatalization.

This study solves the puzzle of understanding the respective strengths of different vowels in triggering coronal palatalization using the neural network modeling and learning of the mapping between the low-level articulatory synergies on the basis of activations of tongue muscles and the high-level cognitive representation of speech sounds. The learning results demonstrate how hidden and gradient featural representations are obtained from coarticulation in CV contexts. This study and Lionnet's work (2016) share the idea that phonological representations must reflect coarticulatory effects, but the current model integrates coarticulatory effects into phonology without setting a separate representational tier for them. I propose that such gradient featural representations are in output of phonological computation (Faust & Smolensky 2017, Zimmermann 2018, Walker 2019).

Boersma et al. (2013) argue that a single framework that accounts for both phonological and phonetic phenomena can be based on neural networks. This study illustrates how neural network modeling combined with grammatical formalisms allows us to assess the extent to which phonological representations can be learned from physical substances (Pater 2018). This approach opens the possibility for a new understanding not only of phonological patterns that are problematic in the traditional perspectives but also of the relationship, specifically the interconnection between phonology and phonetics.