Machine learning shows that clusters representations have a temporal component

[1]Tom Lentz, [2]Marianne Pouplier, [3]Phil Hoole
[1]University of Amsterdam; [2,3]IPS Munich, Ludwig Maximilians Universität
[1]T.O.Lentz@uva.nl , [2]pouplier@phonetik.uni-muenchen.de, [3]hoole@phonetik.uni-muenchen.de

Languages differ not only in their inventory of consonant clusters [1,2], but also in the coarticulation, or timing, of such clusters [3]. Thus the same segmental sequence can be articulated quite differently across languages. There are different proposals on the underlying control structures that give rise to the observed articulatory differences [4], but as hypotheses about timing patterns exist in a high-dimensional space the best formalisation might still be undetected. The high dimensionality arises because coordination between consonants might be relative to onsets, offsets, and/or plateaus, yielding a myriad of logically possible coarticulation patterns. Without knowing which underlying relation best describes the temporal component of cluster representations, it is hard to know if temporal representations truly differ between languages and if speakers can adapt their timing to another language. However, deep learning techniques allow to separate patterns with minimal a priori assumptions about them. The present study explores Support Vector Machines (SVM), in which two categories are separated by an (n-1)-dimensional hyperplane fitted computationally through two sets of n-dimensional data points representing each category [5]. The separation can be non-linear (with the *kernel trick)* Although in principle not hidden, SVM fits do not offer insight into the way the categories are separated, only into the practical separability. However, if two datasets can be separated on given dimensions, it is theoretically possible to find a representation of the categories based on these dimensions.

We illustrate the usefulness of SVMs with data from a study comparing CCV onset clusters recorded from ten Georgian and eight German native speakers. The languages have been claimed to differ in their consonant cluster timing pattern, with Georgian having lower consonant overlap [6] than what has independently been reported for German [7]. In our study, each speaker shadowed cluster productions from two auditorily presented model speakers, one German, one Georgian. In the native condition, participants heard the model (and hence coarticulatory pattern) corresponding to their native language whereas in the non- native condition, participants imitated the model from the other language. All recorded clusters were phonotactically legal in both languages, but differed in their temporal overlap between the consonants depending on whether they were spoken by the German or Georgian model. Articulatory movement data (EMA) were recorded. For each consonant and vowel six articulatory landmarks were identified. Time points were registered relative to the velocity peak of the first consonant, yielding 17 other measurements per production for 602 cluster produced by German speakers (160 native productions, 442 non-native, i.e. imitations of Georgian) and 767 of Georgian speakers (230 native, (537 non-native). The native productions in the data were divided randomly in a training (95%) and a test (5%) set. An SVM was fitted to the training set, separating German and Georgian. It then classified the test set, as well as the non-native productions. The procedure was repeated 20000 times (1000 x 20-fold crossvalidation). Figure 1 shows the performance on the test set is good (d' = 3.8, 95% CI: 3.84-3.90). Figure 2 shows how German's non-native productions fooled the SVM: most were classified as Georgian (p= 0.54; for all reported differences $p<0.001$***). Yet Georgians' non-native productions were mostly *not* classified as Germans when imitating the German model (p = 0.26), meaning they did not reach the German coarticulation pattern.

Results confirm that Georgian and German differ substantially on some temporal component of CCV articulation. Some speakers (German participants) can imitate the timing pattern of another language, indicating that this part of the representation is not completely opaque to them. We will discuss possible reasons for the asymmetric result between languages.

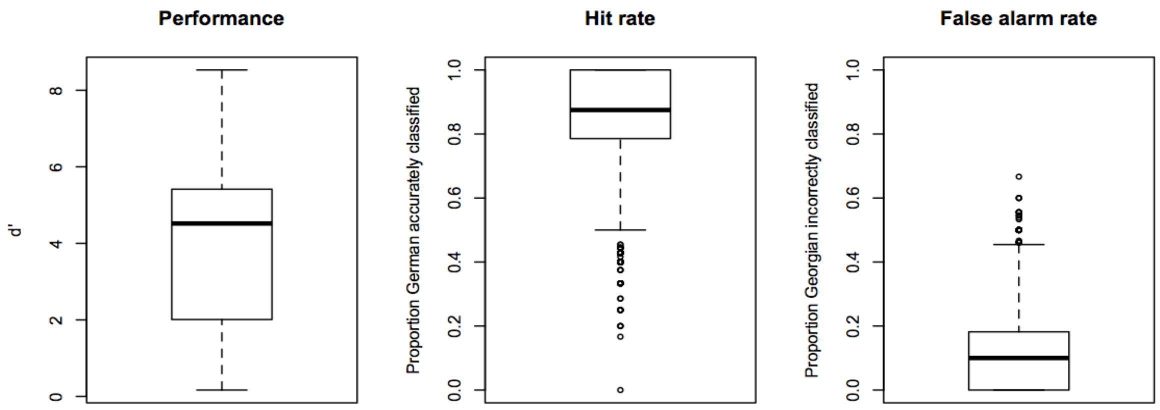

Figure 1. *Model performance on test set over 20000 runs; box indicates first to third quartile, whiskers extend to 1.5 interquartile range.*

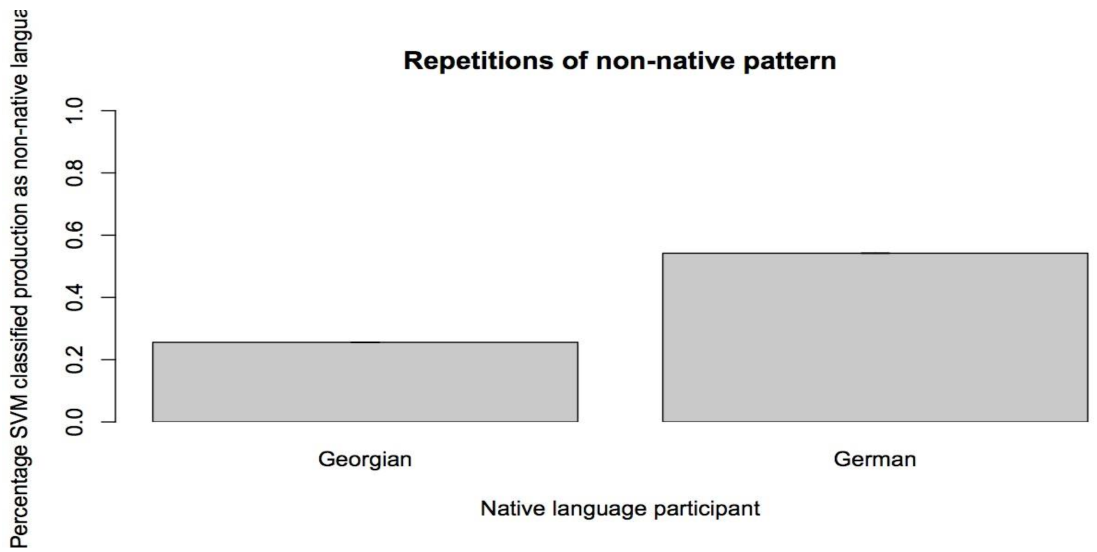

Figure 2. *Model's classification of speakers imitation of the other language as that language, per participant group, over 20000 models each fit to 95% of the native productions. Georgians repeating German are classified as German less than Germans repeating Georgian are classified as Georgian. Error bars (hardly visible) indicate s.d..*

[1] Trubetzkoy, N. S. 1939. Travaux du Cercle Linguistique de Prague 7: Grundzüge der Phonologie. Prague: Jednota Cěskoslovenských Matematiků a Fysiků.

[2] Chomsky, N. & Halle, M. 1965. Some controversial questions in phonological theory. Journal of Linguistics, 1, 97–138.

[3] Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, 20, 269—337.

[4] Browman, C. P. and Goldstein, L. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica*, 45, 140–155.

[5] Bennet, K. T. and Campbell, C. 2000. Support Vector Machines: Hype or hallelujah? *SIGKDD Explorations*, 2(2), 1–13.

[6] Chitoran, I., Goldstein, L., and Byrd, D. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. *Laboratory Phonology*, 7, 419–447.

[7] Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in word-initial clusters of German. Journal of Phonetics, 41(6), 546-561