Using Random Forests to compare production patterns by bilingual English-Spanish speaking children with cochlear implants and their peers with normal hearing

Mark Gibson[1], Ferenc Bunta[2], Charles Johnson[1,3] , Miriam Huárriz[1]

[1]Universidad de Navarra, [2]University of Houston, [3]Brigham Young University

Four random forest models in a supervised learning environment were constructed and used to compare (1) Spanish production data from Spanish- and English-speaking bilingual children with normal hearing (NH group) and Spanish- and English-speaking bilingual children with hearing loss who use cochlear implants (CI group), and (2) the children's productions with the speech of adult native Spanish speakers. The twenty two participants were simultaneous bilinguals from birth. Participants were matched on the basis of chronological age for the NH group (M=5;1) and hearing age for the CI group (M=5;1). The models were fitted for four temporal parameters related to different aspects of syllable structure: (1) lateral duration, (2) vowel duration, (3) voice onset time (VOT), and (4) interplateau intervals (IPI), or the period between two gestural plateaus for a given consonant stop+liquid cluster.

Random forests were preferred over neural networks because they generally require fewer data points to train, and because they have a natural measure for parameter importance based on their use in discriminating between categories. For the current models we used Scikit-learn's ensemble learning package [1] in Python to construct four distinct 100 tree random forest classifiers fitted with articulatory data (electromagnetic articulography, EMA) related to intergestural and oral-laryngeal timing patterns from native Spanish and speaking adults. The trees in each random forest were allowed to have a maximum depth of two and a maximum of six leaf nodes. Two of our classifiers compare adult speech to the CI and NH groups' productions. The other two compare the speech of children with cochlear implants to those without. For each of these two groups, one classifier compares /l/ duration (in /l/ and stop+l contexts), vowel duration and VOT and the other compares interplateau intervals.

Results of the first model show that lateral and vowel duration serve as important parameters when classifying complex onsets, while VOT and IPI are less so (as shown in Table 1), though these latter variables may offer viable cues related to the place categories that can occupy different syllable positions. Additionally, we found that the two classifiers which compare child and adult speech could be thresholded to predict the nature of the speaker with perfect or near-perfect precision for recall and accuracy (see Figure 1). However, the same classifiers performed little better than chance when trying to predict the difference between children with and without cochlear implants (see Figure 2). We interpret the success of these models in the first two instances and their failure in the second two cases to confirm our hypothesis that children with and without cochlear implants in both languages are nearly indistinguishable with respect to the temporal parameters tested.

Results will be discussed in relation to existing work regarding the acquisition of syllable level timing patterns by bilingual children and children with hearing loss who use cochlear implants.

(1) Feature performance values for Spanish timing variables

	NH/CI discrimination				Adult/child discrimination			
	/l/ dur	V dur	VOT	IPI	/l/ dur	V dur	VOT	IPI
C	0.26865	0.54756	0.18377		0.42909	0.54383	0.02707	
CC	0.42860	0.3057	0.11415	0.15153	0.23739	0.70217	0.01682	0.04361

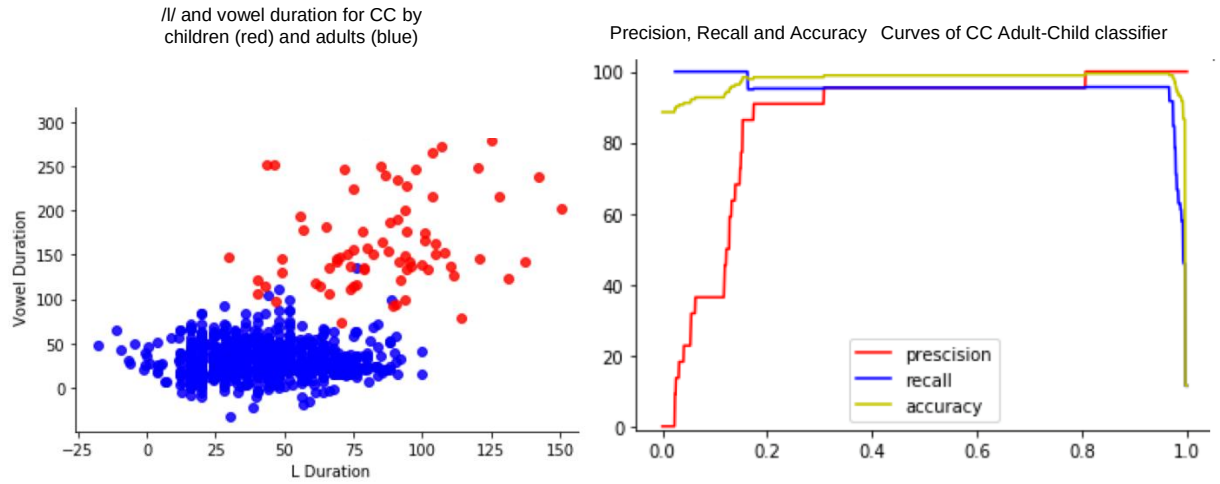

Figure 1. *(left) scatterplot comparing adult (blue dots) and children (red dots) /l/ duration and vowel duration productions in CC clusters. (Right) precision, recall and accuracy curves in discriminating between children and adults based on /l/ duration and vowel duration. The model was able to discriminate children from adults based almost entirely on /l/ duration and vowel duration (though as shown in Table 1, VOT and IPI were also on occasion used).*

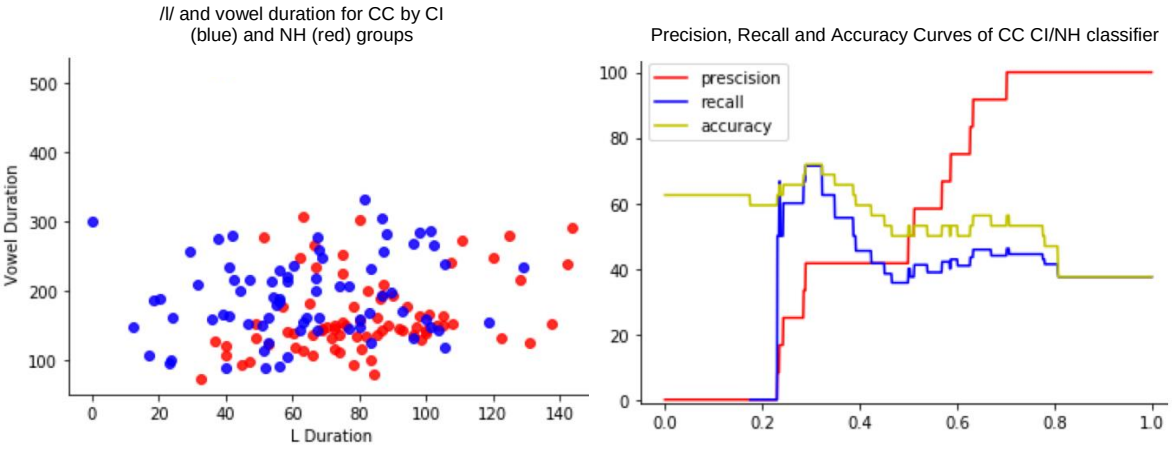

Figure 2. *(left) scatterplot comparing CI (blue dots) and NH (red dots) /l/ duration and vowel duration productions in CC clusters. (Right) precision, recall and accuracy curves in discriminating between CI and NH based on /l/ duration and vowel duration. The model performed no better than chance in discriminating the groups*

[1] Pedregosa et al., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825-2830.