The development of syllable timing by children with cochlear implants and their peers with normal hearing.

Gibson[1], M., Bunta[2], F., Johnson[1,3], C. & Huárriz[1], M.

1. Universidad de Navarra, 2. University of Houston, 3. UCSB

12th International Seminar on Speech Production (ISSP), 2020

Introduction

- Gibson et al. (2018) examined the Spanish productions of syllable-level timing parameters (**VOT, C1 duration, lateral duration, vowel duration**) in **C** and **C1C2** onsets by bilingual Spanish- and English-speaking children with cochlear implants (**CI** group) and their peers with normal hearing (**NH** group), while Gibson et al. (under revision) reports results for the same children's English productions.
- Specifically, they examined: VOT of word initial stops in C and C1C2 (where the stop is C1), lateral duration in C and C1C2 (where /l/ is C2), and vowel duration following C and C1C2 onsets.
- In line with adult monolingual speakers for both languages, the authors found a compression effect for the lateral whereby the lateral shortens in C1C2 onsets, though stability (measured in terms of their coefficient of variation) was higher for the CI group.
- Against prediction, these authors found a lengthening effect for VOT in both languages (though it was only marginally hypothesized for English), in addition to a lengthening effect for vowels following C1C2 onsets.
- The present studies uses Random Forest analyses in a supervised learning environment to tease apart effects based on language and group differences, and Pearson Product-Moment Correlation to address the possible correlations between durational cues and speech rate and age.

Research questions

- **Do the bilingual children produce qualitatively different timing patterns for each language, or apply the same timing schemes to two different languages (or alternately construe a hybrid timing configuration for both languages)**
- **Is greater temporal stability for the lateral by the CI group a function of age?**
- **Is speech rate responsible for the longer VOT for stops in C1C2 clusters, and longer vowels following C1C2 onsets?**

Data for models

Data collection

- Acoustic signals were obtained at a sampling frequency of 44,100 Hz and 16 bits with a Marantz PMD 661 MKII Professional Field.

Subjects

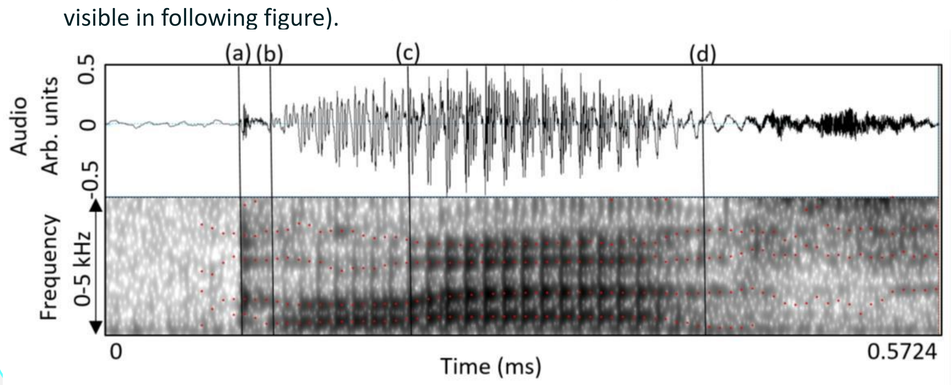
- 22 English- and Spanish speaking bilingual children. 11 (CI), 11 (NH), mean age 5 years and 2 months for both groups (full description of both groups in Gibson et al., 2018).
- NH group completed a hearing screening using pure tones at 0.5, 1, 2, and 4 kHz at 25 dB HL, bilaterally.
- Cochlear implants of the CI group were tested prior to testing.
- Children were matched for chronological age (NH group) vs hearing age (CI group).

Corpus

- A corpus of 80 target words containing word initial singleton stops and stops in clusters was elicited using a picture naming paradigm.

Variables

- VOT (b-a), lateral duration (c-b), vowel duration (d-c), interconsonantal interval (not visible in following figure).

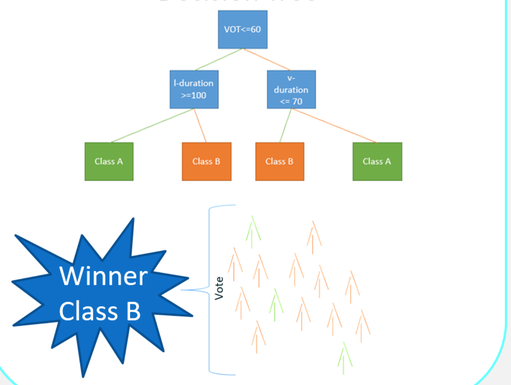


Random Forest analysis

- Random Forest classifiers are a machine learning model which takes the consensus of a number of decision trees to determine the probability of a single instance of data belonging to a particular class.
- For our models, we used scikitlearn's Random Forest Classifer class to implement the Random Forests.
- Random Forest Classifiers consisted of 100 decision trees, each of which had a maximum depth of 2 (only two boolean conditions max were allowed to classify any piece of data).
- Each decision tree uses a set of Boolean conditions on features (such as l-duration <= 45) to classify data into one of a number (in this case 2) classes.

Model schematic

Decision Tree



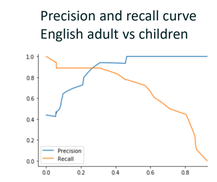
Model input

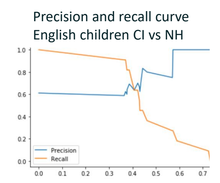
- For Spanish, acoustic data from Gibson et al. (2016, 2019) were used to train the RF models.
- For English, random data were generated for **lateral** and **vowel** duration based on results of Marin & Pouplier (2010), and **VOT** based on Klatt (1975).

Random Forest results

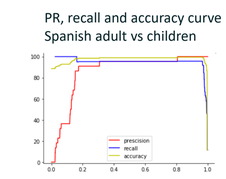
- For both Spanish and English productions, the models were able to classify adults versus children with very high accuracy (Spanish accuracy higher than Spanish because we also use IPI as classifier).
- Important classifiers in distinguishing adult versus child speech patterns were VOT and vowel duration for both languages.
- Models did not however distinguish productions from the CI and NH groups better than chance.
- Models were able to distinguish children based on language, but again did not perform better than chance in distinguishing between the two groups.

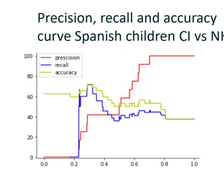
English results





Spanish results

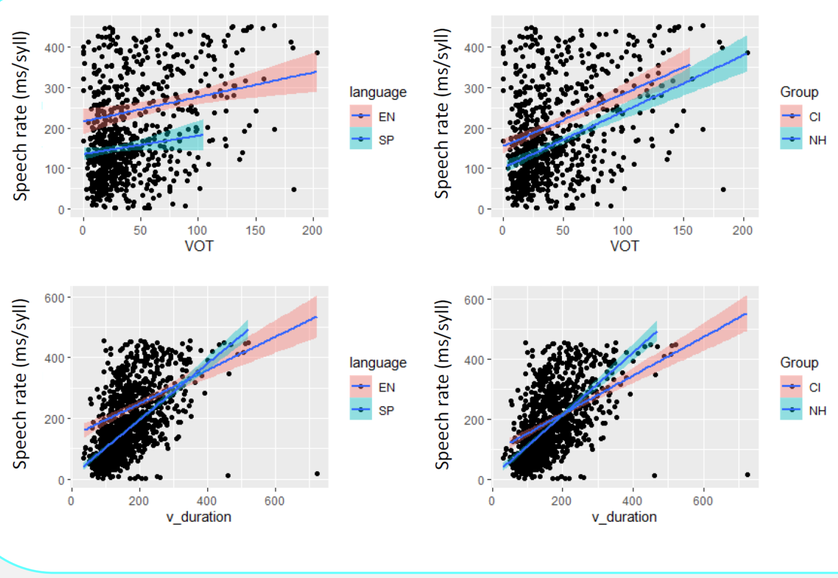




Feature importance

Experiment	Model evaluators for binary classifiers				Relative importance			
	precision at best f-score	recall at best f-score	best f-score	accuracy at best f-score	VOT	/l/ dur	V dur	ICI
English								
NH vs CI (CV)	0.6	0.87	0.71	0.65	0.26	0.15	0.59	0.0
NH vs CI (ClV)	0.68	1.0	0.81	0.74	0.30	0.47	0.23	0.0
Adult vs Youth (CV)	0.79	0.7	0.74	0.88	0.15	0.27	0.58	0.0
Adult vsYouth (ClV)	0.95	0.88	0.91	0.98	0.13	0.44	0.44	0.0
Spanish								
NH vs CI (CV)	0.57	0.98	0.72	0.61	0.18	0.13	0.69	0.0
NH vs CI (ClV)	0.8	0.88	0.84	0.81	0.21	0.37	0.42	0.0
Adult vsYouth (CV)	1.0	0.95	0.97	0.98	0.04	0.25	0.71	0.0
Adult vs Youth (ClV)	0.92	0.65	0.76	0.97	0.10	0.08	0.78	0.06
English/Spanish								
EN vs SP (CV)	0.64	0.98	0.77	0.67	0.61	0.34	0.06	0.0
EN vs SP (ClV)	0.97	0.91	0.94	0.95	0.03	0.41	0.19	0.36

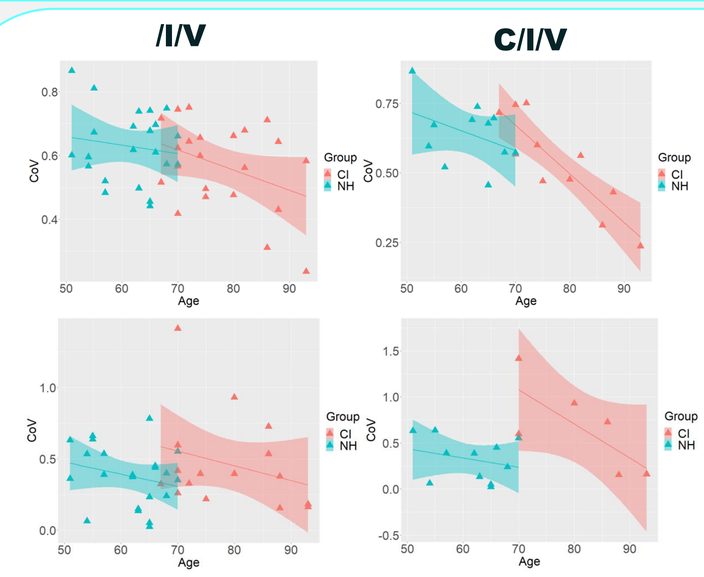
Speech rate and duration (VOT and vowel duration)



For pooled data for both languages and groups there was a positive correlation between speech rate and VOT [r = 0.38, n = 1215, p < 0.001] such that VOT increases in relation to increases in speech rate (i.e., as speech rate becomes slower. Though differences between groups and language were prevalent.

For vowel duration, correlations are generally stronger for both groups and languages. For pooled data across both languages and groups, there is a strong correlation between speech rate and vowel duration [r = 0.55, n = 1215, p < 0.001] which holds across groups and language.

Age and lateral stability



For Spanish word initial lateral tokens ([r = -0.37, n = 44, p = 0.01]) and laterals in clusters ([r = -0.71, n = 22, p < 0.001]), there is a correlation between the temporal stability of the lateral and age such that as age increases the coefficient of variation decreases (i.e., stability increases

For English, there is no correlation between the temporal stability of the lateral and age in word initial lateral tokens ([r = -0.017, n = 44, p = 0.91]) or laterals in clusters([r = 0.11, n = 44, p = 0.64]) .

Conclusions

- Results of the Random Forest models suggest that the children produce qualitatively different timing configurations for each language.
- True voice/short lag VOT paradigm for the Spanish voice contrast in word initial stops, versus a short lag/long lag VOT paradigm to express the voicing contrast in English word initial stops.
- Results of the Pearson Product-Moment Correlation models show a relation between speech rate and syllable complexity. Children slow down their speech rate before cluster production.
- Lateral stability is correlated to age, which we use as an umbrela term to reflect differences in fine motor skills and phonological awareness (we could not disentangle these so we assumed an umbrella term 'age'), but only for Spanish (probably due to the fact that Spanish has one clear lateral and English has two, which is determined by syllable affiliation).

References

Caldwell, A., & Nittrouer, S. (2013). Speech perception in noise by children with cochlear implants. *Journal of Speech, Language and Hearing Research*, 56, 13–30.

Gibson, M., Bunta, F., Goodin-Mayeda, L. & Hernández, A. (2018). The acquisition of syllable-level timing contrasts by English- and Spanish-speaking bilingual children with normal hearing and English- and Spanish-speaking bilingual children with cochlear implants. *Journal of Phonetics 7* (1), 98-112.

Gibson, M., Sotiroupoulou, S., Tobin, S., & Gafos, A. (under review). Temporal aspects of word initial single consonants and consonants in clusters in Spanish.

Gibson, M., Sotiroupoulou, S., Tobin, S., & Gafos, A. (2017). On some temporal properties of Spanish consonant-liquid and consonant-rhotic clusters. In M. Belz, S. Fuchs, S. Jannedy, C. Mooshammer, O. Rasskazova, & M. Zygis (Eds.), *Tagungsband der 13. agung Phonetik und Phonologie im deutschsprachigen Raum* (pp. 224–228). Berlin: Leibniz-Centre General Linguistics (ZAS) & Humboldt-Universität zu.

Gibson, M., Fernández Planas, A. M., Gafos, A., & Ramirez, E. (2015). Consonant duration and VOT as a function of syllable complexity and voicing in a sub-set of Spanish clusters. *Proceedings of the international speech communication association (Interspeech),* 1690–1694.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word initial consonant clusters. *Journal of Speech, Language and Hearing Research*, 18, 686–706.

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, 14, 380–407.