

Predicting the Age of Emergence of Consonants

Luis M. T. Jesus¹, Jihen Trabelsi²
(lmtj@ua.pt, jihen.trabelsi@ua.pt)

¹School of Health Sciences (ESSUA), Intelligent Systems Associate Laboratory (LASI) and Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro, Aveiro, Portugal

²School of Health Sciences (ESSUA), University of Aveiro, Aveiro, Portugal



16th International Conference on Computational Processing of Portuguese (PROPOR 2024) | Santiago de Compostela Spain | 12th to 15th March 2024



This work was supported by National Funds through the FCT - Foundation for Science and Technology, in the context of the project UIDB/00127/2020

Introduction

Models of phonological acquisition must account for **ambient language effects** and **articulatory complexity** of speech (Bybee & Easterday 2022).

There are, however, very limited assets that allow researchers of under-resourced languages to analyse the effect of predictors such as **ambient frequency** **AF**, **functional load** **FL** and **articulatory complexity** **AC** on the **age of emergence (AoE)** of phonemes (Lindblom et al. 2011; Stokes & Surendran 2005).

This poster describes the development of a new open access resource, the **Preditores da Aquisição de Consoantes (PAC) database** (available in open access from acsa.web.ua.pt), which allows the exploration of these issues for **European Portuguese (EP)** and comparing the results with a typologically unrelated language, **Tunisian Arabic (TA)**.

Novel **AF**, **FL**, and **AC** values, calculated for **EP** and **TA** consonant inventories, are presented.



Method

PAC database: Orthographic transcription, frequency, rank, parts of speech and phonemic transcriptions of the 1000 most frequently used lemmas in Portuguese and Arabic.

List of lemmas compiled from a corpus of 20 (Portuguese)/ 30 (Arabic) million words from Portuguese (Davies & Bay, 2008) and Arabic (Buckwalter & Parkinson, 2011) frequency dictionaries, were used to compile the corpus. Phonemic transcriptions were produced for each word.

Consonant inventory and AoE based on Portuguese and Arabic languages' literature

19 EP phonemes /p, b, t, d, k, g, m, n, ɲ, r, f, v, s, z, ʃ, ʒ, ɥ, l, ʎ/ (Jesus et al., 2015)

26 TA phonemes /b, t, tʰ, d, dʰ, k, q, ʔ, m, n, r, f, θ, ð, ðʰ, s, sʰ, z, ʃ, x, ɣ, ħ, ʕ, h, ɟʃ, l/ (Thelwall & Sa'Adeddin, 1990; Tice, 2021)

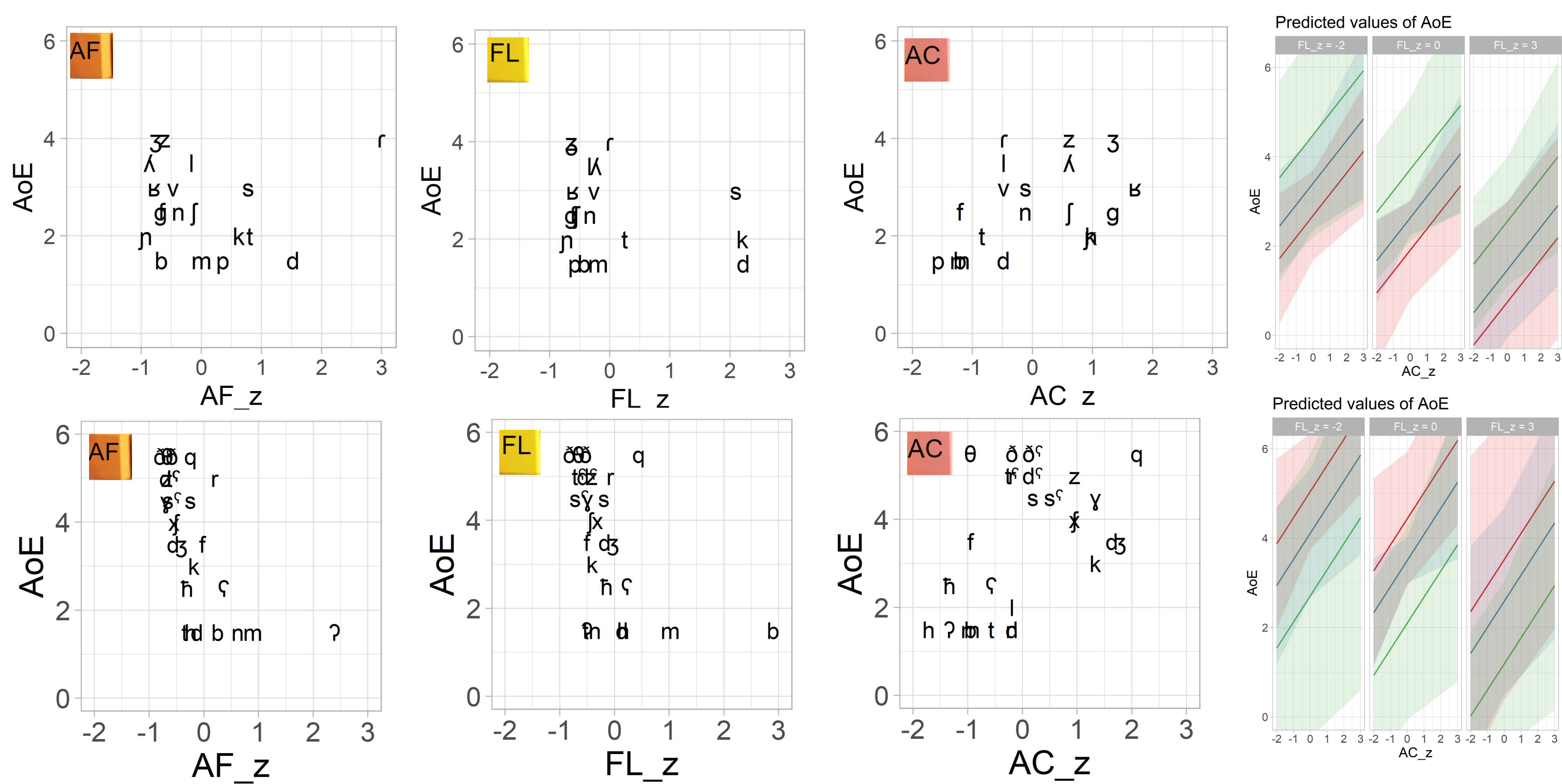
AF derived from the frequency and phonemic transcription of the 1000 lemmas in the **PAC** database.

FL calculated with the Phonological Corpus Tools 1.5.1

AC calculated (Stokes & Surendran, 2005; Bybee & Easterday, 2022) with an articulatory cost defined by Lindblom et al. (2011).

Multiple linear regression models developed in R 4.3.1, with **AoE** as outcome variable, and **AC**, **AF**, and **FL** as predictors, for the two languages.

Results



AoE in years for the EP (top) and TA (bottom) consonant inventory, as a function of z-scored AF, FL, and AC. Also shown, marginal effects computed for the z-scored model predictors (AF_z, FL_z and AC_z) at three different levels: -2, 0 and 3.



Model accounted for **20%** of the variance in the **AoE**: Adjusted $R^2 = 0.197$. The **AC** predictor had the largest effect on the **AoE**: For each increase in **AC** by one standard deviation (holding all variables constant), the **AoE** increased significantly by **6 months** (slope = 0.48; $SE = 0.18$; $p = 0.030$).

Results also suggest that the more complex and frequent a phoneme is, the later it is acquired in EP.



Model accounted for **37%** of the variance in the **AoE**: Adjusted $R^2 = 0.365$. The **AC** predictor had the largest effect on the **AoE**: For each increase in **AC** by one standard deviation, the **AoE** increased significantly by **7 months** (slope = 0.58; $SE = 0.26$; $p = 0.038$).

Results also suggest that the more frequent a phoneme is and the more meaningful a contrast is, the earlier in Tunisian children's lives it is acquired.

Conclusions

AF, **FL**, **AC** and **AoE** values discussed in this poster can provide guidance in establishing intervention strategies to facilitate the acquisition of consonants for effective communication in Portugal and Tunisia.

The multiple regression models presented in this poster can be used by researchers, educators, and clinicians to estimate a typical range for the **AoE** of consonants.

Current results showed that **AC** was the only significant predictor.

We are currently expanding the size of the **PAC** database to more than 2000 lemmas.

References

- Buckwalter & Parkinson (2011). A Frequency Dictionary of Arabic Core Vocabulary for Learners. Routledge.
- Bybee & Easterday (2022). Primal consonants and the evolution of consonant inventories. *Language Dynamics and Change*, 13(1), 1-33.
- Davies & Bay (2008). *Frequency Dictionary of Portuguese*. Routledge.
- Jesus et al. (2015). Is the Portuguese version of the passage 'The North Wind and the Sun' phonetically balanced? *Journal of the International Phonetic Association*, 45(01), 1-11.
- Lindblom et al. (2011). Sound systems are shaped by their users. In Clements & Ridouane (Eds.), pp. 65-98. John Benjamins.
- Stokes & Surendran (2005). Articulatory Complexity, Ambient Frequency, and Functional Load as Predictors... *Journal of Speech, Language, and Hearing Research*, 48(3), 577-591.
- Thelwall & Sa'Adeddin (1990). Arabic. *Journal of the International Phonetic Association*, 20(2), 37-39.
- Tice (2021). *Language and Performance in Post-revolution Tunisia*. Ph.D. Thesis. The Ohio State University.