

Age-Related Vowel Variation and Automatic Classification in Taravo Corsican

Hi-Yon Yoo¹ Lesia Collard-Buresi¹

¹*Université Paris Cité*

While in adulthood, anatomical stabilisation and refined articulatory control lead to a relative stability of vowel targets, although effects related to sex, speaking style, and individual identity remain robust (Hillenbrand et al., 1995), vowels undergo systematic modifications with ageing such as a centralisation of the vowel space, a reduction of overall vowel space area, and an increase in within-speaker variability (Benjamin, 1981; Torre & Barlow, 2009). These effects, however, are not uniformly distributed across vowels. High vowels appear particularly sensitive to age-related effects, especially in their dynamic properties, including formant trajectories and temporal stability (Albuquerque et al., 2014; Reubold, Harrington, & Kleber, 2010). The purpose of this study is to test age classification using machine learning methods for under-resourced languages (Besacier et al., 2014) to understand if acoustic measurements of vowels alone can enable age classification, and reveal the existence of specific patterns in age-related vowel production in Taravo Corsican, an under-resourced Italo-romance variety spoken in the south of Corsica we examined the possibility of performing machine learning classification on spontaneous speech data.

The classification was tested on an excerpt of the *O'Sà* corpus (unpublished) which is constituted of 50 hours of Taravo Corsican spontaneous speech data. The audios used for this study was selected based on the presence of external noise and the clarity of the audio. The selected speaker allowed us to form two groups: a first group of “young” speakers aged from 30 to 52 years old, and a second group of “old” speakers aged from 68 to 78 years old. Both groups were constituted of one woman and two men, all natives of the dialect with both parents being also native of this dialect to avoid possible contact with other dialects of the island. The methodology used for the analysis consisted of three steps, each corresponding to a specific part of the analysis. The first step was the optimization of the audio files. For this we used WebMaus (Kisler et al., 2017) for segmentation and alignment and used the Italian settings. With the Webmaus alignment we then worked on the TextGrids for further optimization using the TextGridTools (Buschmeier and Włodarczak, 2013) package in python. Through a combined method using, Parselmouth (Jadoul et al. 2018) we extracted acoustic measurements such segment duration, formant values and MFCCs values, before realizing the classification using supervised classification using Random Forest (Breiman 2001), SMOTE (synthetic minority over-sampling) (Chawla et al 2002) to manage imbalance, and Leave-One-Speaker-Out cross validation to control for speaker identity.

This analysis showed that overall classification (43.3% accuracy) was limited by high speaker-specific variation and the small data sample. However, specific vowel models revealed that accuracy differed following the vowel (/i/ reached a 55.6% accuracy for 44,6% for /u/ and 32,3% for /a/ a 32.3%. However, despite its relatively higher accuracy, /i/ showed no difference for F1/F2 at midpoint while /a/ and /u/ presented clear differences. However dynamic formant trajectories (figures 1, 2 and 3) revealed vowel specific age patterns, with young speakers displaying a more dynamic movement. But despite the acoustic differences, classification remained limited ($p=0.002$) indicating that in the case of Taravo Corsican speaker specific features dominates age effect. Improvement in the methodology are required in order to understand how to make classification possible for under-resourced languages and understand the methodological requirements, which in this case requires a higher number of speakers to bypass speaker specific influence.

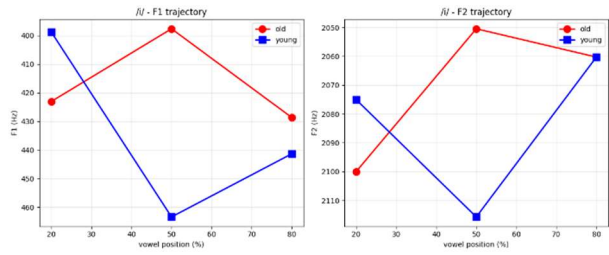


Fig. 1: /i/ dynamic formants trajectory

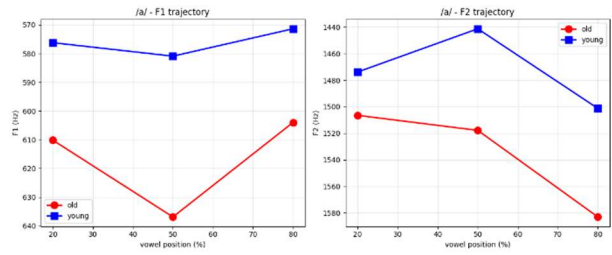


Fig.2: /a/ dynamic formant trajectory

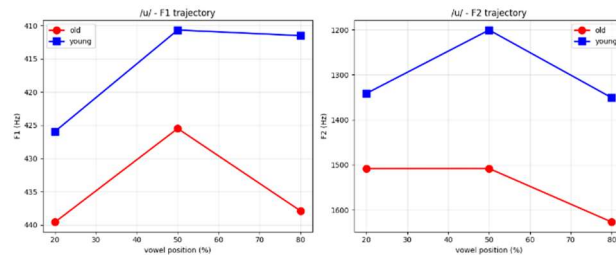


Fig 3: /u/ dynamic formant trajectory

References

- [1] Albuquerque, L., Cunha, C., Silva, C., & Teixeira, A. (2014). Age-related changes in European Portuguese vowel acoustics. *Speech Communication*, 65, 20–35.
- [2] Albuquerque, L., Oliveira, C., Teixeira, A., & Figueiredo, D. (2021). Eppur si muove: Formant dynamics is relevant for the study of speech aging effects. In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021)*
- [3] Benjamin, B. J. (1981). Frequency variability in the aged voice. *Journal of Gerontology*, 36(6), 722–726.
- [4] Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, 56, 85-100.
- [5] Breiman, L. (2001) Random Forests. *Machine Learning*, 45, 5-32.
- [6] Buschmeier, H. and Włodarczak, M. (2013). TextGridTools: A TextGrid Processing and Analysis Toolkit for Python. In P. Wagner (ed.) *Tagungsband der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013)*. Dresden: TUDpress, pp. 152-157
- [7] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- [8] Georgiou, G. P. (2023). Comparison of the prediction accuracy of machine learning algorithms in crosslinguistic vowel classification. *Scientific Reports*, 13, 15594.
- [9] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3099-3111.
- [10] Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1-15
- [11] Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326-347.
- [12] Ladefoged, P., & Johnson, K. (2014). *A Course in Phonetics* (7th ed.). Cengage Learning.
- [13] Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2), 175-184.
- [14] Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F0 and vowel formants in spontaneous speech. *Proceedings of Interspeech 2010*, 1974–1977
- [15] Torre, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, 42(5), 324–333.
- [16] Ulrich Reubold, Jonathan Harrington, Felicitas Kleber, Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers, *Speech Communication*, Volume 52, Issues 7–8, 2010, Pages 638-651