Typological classification of European Portuguese fricatives: a cross-language forced alignment and pronunciation variants study

Anisia Popescu and **Lori Lamel** and **Ioana Vasilescu**
Laboratoire Interdisciplinaire des Sciences du Numérique (LISN)
Université Paris Saclay
anisia.popescu@universite-paris-saclay.fr

Abstract

Devoicing in European Portuguese fricatives is an extensive phenomenon, especially when compared to other languages. Small scale acoustic studies have shown that devoicing rates and voicing profiles of fricatives are more similar to those of Germanic languages, setting European Portuguese (EP) apart within the Romance language family. The present study tests whether voicing in EP fricatives diverges from its sister languages by using empirically motivated combinations of different languages (EP, Italian, German) acoustic phone models on large EP corpora, allowing an ASR system to choose the best fitting one when force aligning the data. Results confirm that voicing in EP fricatives is more similar to German, suggesting EP voicing patterns are shifting away from classic voicing systems known for Romance languages.

1 Introduction

Romance languages are generally known as "true voicing languages" - implementing the voice-voiceless contrast through the use of the [voice] feature (Lisker and Abramson, 1964), opposing prevoiced with unaspirated voiceless obstruents. "Aspirating languages", such as most of the Germanic languages, make use of the [spread glottis] feature (Jansen, 2004), contrasting unaspirated, phonetically voiceless obstruents with long-lag aspirated obstruents. There are however exceptions such as Dutch (van Alphen and Smits, 2004) or, more recently, European Portuguese (Pape and Jesus, 2011, 2015). More specifically, in the latter case, both small scale acoustic studies (Jesus and Shadle, 2003; Pape and Jesus, 2011) and large scale corpus-based studies (Wu et al., 2022; Hutin et al., 2022) have found higher rates of devoicing of phonemically voiced obstruents in EP than in other Romance languages. Furthermore evidence from voicing profiles shows that while Italian and Spanish voicing probability remains high (close to 1) throughout the obstruent, in EP and German there is a decrease in voicing probability starting with 30% of the obstruent. (Pape and Jesus, 2015; Shih and Möbius, 1999, 1998).
The present study tests these patterns on a much larger scale (100+ hours of speech) via forced alignment of the speech with the orthographic transcription. EP speech data is aligned using parallel multiple-language acoustic models and pronunciation variants for fricatives to answer the following theoretical research question:

- Does EP fricative voicing show consistency within the Romance languages family, or does it take a different path, more similar to languages that are both genetically and geographically different?

To answer this question we chose one Germanic - German - and one Romance - Italian - language. The choice of languages mirrors the set of languages tested in the original acoustic study (Pape and Jesus, 2015). Based on the chosen language set we can now fine-tune our experimental research question to:

- Is voicing in EP fricatives more similar to German than to Italian ?

2 Methods

To answer this experimental research question we analyzed an EP corpus consisting of 114 hours of mostly standard dialectal broadcast news speech from TV and radio shows. Multiple sources were used for acquiring the data: LDC, ELRA and international projects. The phone level segmentation was generated using a Portuguese acoustic model, estimated using language-specific annotated (manual transcription) training data and pronunciation dictionaries. The output is a sequence of phone segments with labels selected by aligning the reference transcriptions via a language specific dictionary.

To test whether voicing in EP fricatives is more similar to German than Italian, two additional sets of fricative phone models, one for German and one for Italian, were added in parallel to the original Portuguese one. The phone models for all other phonemes are kept in their original Portuguese form. For each language the acoustic models were all trained on roughly 100 hours of transcribed broadcast news data (Portuguese: 1.1 million word tokens, 46k word types; Italian: 1.8 million word tokens, 58.8k word types; German: 1.8 million word tokens, 90k word types). All three (EP, Italian and German) acoustic models are speaker-, context- and word-position-independent monophone models. Each phone model is a 3-state left-to-right continuous density hidden HMM with Gaussian mixtures with up to 32 Gaussians per state (silences are modeled by a single state with 256 Gaussians). Each acoustic model used the same acoustic parameterization (cepstral - PLP (Hermansky, 1990) and pitch (F0) features), similar to (Lamel et al., 2011). Figure 1 illustrates the speech modeling and alignment process. By using different combi-

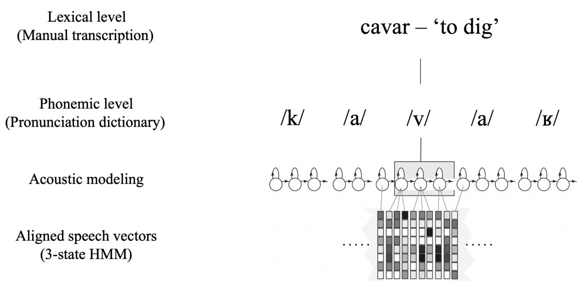

Figure 1: Illustration of the speech modeling and alignment process for the Portuguese word *cavar* ’to dig’

nations of language acoustic models on EP speech data we force the recognition system to choose the best fitting phone model (be it the original EP, or an Italian or German one) for each individual phonemically voiced fricative in the corpus (illustrated in Figure 2). The set of voiced fricatives in the corpus consisted of a total of 37.563 coronal /z/ and 36.354 labiodental /v/. The postalveolar /ʒ/ was left out of the study since it is not included in the Italian phoneme inventory and it appears only in loanwords in German.

Two different combinations of the three acoustic models were tested: (1) a three way choice of acoustic models between Portuguese, Italian or German, and (2) a two way choice of acoustic models between Italian or German (Portuguese fricative

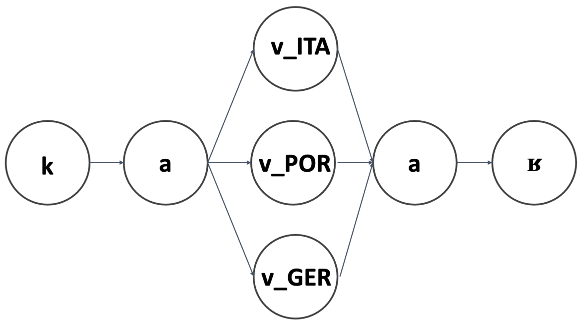

Figure 2: Combination of three acoustic models (Italian, EP and German) for the fricative /v/ in the Portuguese word *cavar* ’to dig’

phone models were no longer available to the ASR system). Each combination will be described in a different section. If voicing in EP fricatives is indeed more similar to German, as attested by previous acoustic studies, we would expect the system to choose the German fricative phone models to a higher degree than it does the Italian fricative phone models. If, however the opposite stands (i.e., EP voicing does not behave similarly to a Germanic language, but is still related to its sister language Italian), we would expect the system to prefer the Italian fricative phone models.

We ran a third experiment which involved using one acoustic model at a time (not in parallel) with the addition of pronunciation variants. The language specific (European Portuguese) dictionary was enriched with pronunciation variants for fricative voicing. For example the Portuguese word /vinho/ - ’wine’ had two possible pronunciations: the original [viɲo] and a devoiced variant [fiɲo]. The procedure is similar to that described for English and French in (Lamel and Adda, 1996; Adda-Decker and Lamel, 1999). The system then had to choose which phone model (phonetically voiced or phonetically voiceless) best fitted the phonemically voiced fricative in the data. This experiment will be described in section 5.

3 Experiment 1: Three-way choice of acoustic models - European Portuguese, Italian and German

In this first experiment, for each fricative /v,z / the system was presented with three phone models in parallel, the original EP phone model completed with the German and Italian ones. The system then had to choose which of the three phone models for the phonemically voiced /v,z/ best fitted the acoustic realization of the fricatives in European Portuguese. Figure 3 shows the percentages of selected phone models per language as a function of place of articulation (labiodental /v/, coronal /z/).

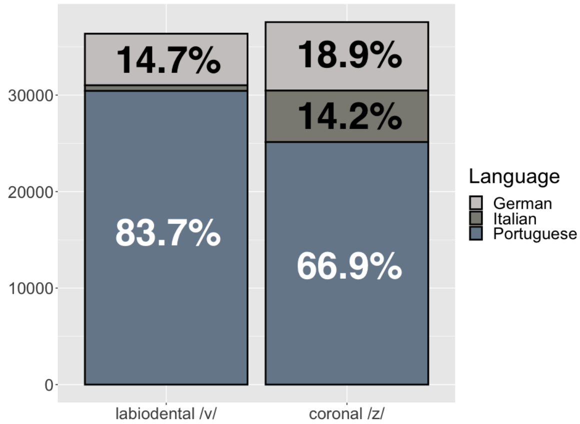

Figure 3: Percentages of phone occurrences aligned with either the original Portuguese, the Italian or the German acoustic model for the labiodental /v/ (left) and the coronal /z/ (right)

As expected, the original Portuguese phone model was markedly preferred (83.7% of cases for the labiodental /v/ and 66.9% of cases for the coronal /z/). For the rest of the cases the system preferred either the German or the Italian phone models. For both the labiodental /v/ and the coronal /z/ the German models were preferred. There is an effect of place of articulation with more mismatches (i.e., the original Portuguese model is less preferred) in the case of the more back fricative (coronal /z/).

4 Experiment 2: Two-way choice of acoustic models - Italian and German

In the second experiment the original EP phone models for fricatives was no longer an option, forcing the ASR system to choose between either an Italian or a German fricative phone model. Figure 4 shows the counts and percentages of phone occurrences aligned with the Italian or the German phone model as a function of place of articulation.

Results mirror those of experiment one, suggesting the German acoustic models seem to be preferred in 89.5% of cases for the labiodental /v/ and 61.7% of cases for the coronal /z/ over the Italian acoustic models. Similar to experiment 1 there is an effect of place of articulation with Italian acoustic models being chosen to a higher degree in the

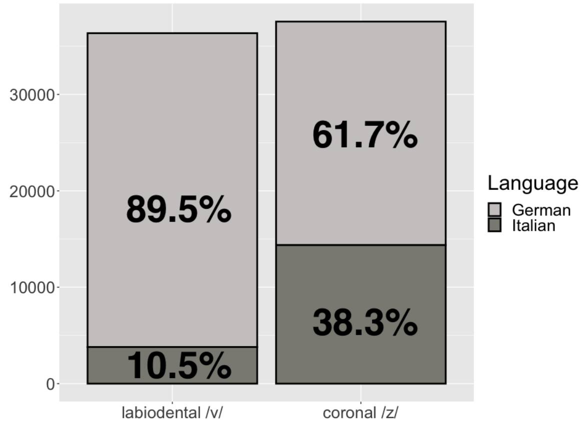

Figure 4: Percentages of phone occurrences aligned with either the Italian or the German acoustic model for the labiodental /v/ (left) and the coronal /z/ (right)

case of coronal /z/ as compared to the labiodental /v/.

5 Experiment 3: Italian and German with pronunciation variants

In this third experiment, the Portuguese language dictionary was enriched with pronunciation variants for fricative voicing (voiced fricatives /v,z/ could be produced either as phonetically voiced [v,z] or voiceless [f,s]) allowing the system to choose the best phone model (voiced or voiceless) for each phonemically voiced fricative in the data. Two separate alignments were run using either the Italian or the German [v,z - f,s] phone models. For example, when using the Italian acoustic model for fricatives on the data, if a Portuguese phonemically voiced fricative /v/ better matched the Italian phone model [v] the output would be [v]. If however the acoustic realization of the Portuguese [v] better matched the Italian [f] phone model, the output would be the Italian [f]. The same procedure was applied using the German fricative phone models. This experiment differs from the first two, in that it allows us to test the similarity/difference between EP and Italian/German from a different angle. Based on previous acoustic studies, we know that voicing profiles differ based on language: while Italian voicing probability remains high (close to 1) throughout the fricative, the EP and German voicing probability decreases starting with 30% of the fricative (Pape and Jesus, 2015). This suggests that both EP and German exhibit partial devoicing during the fricative, whilst Italian does not. If this is indeed the case we would expect to find higher percentages of voiceless variants when using the Italian acoustic model (i.e., Italian voiceless fricative models better match the partially devoiced phonetically voiced EP fricative).

Figure 5 shows the percentages of phonetically voiceless variants (greyer shades) identified when using the Italian and German voiced-voiceless acoustic models. White shades correspond to the phonetically voiced variants identified by the system. Results yet again confirm the higher degree

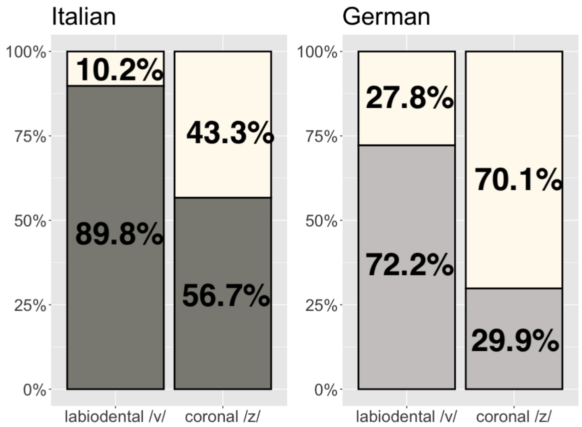

Figure 5: Percentages of phone occurrences aligned with either the voiceless (grey shades) or the voiced (white shade) for the labiodental /v/ (left columns) and coronal /z/ (right columns) per language (Italian on the left and German on the right)

of similarity between EP and German fricatives. As predicted when aligning the data with the Italian fricative phone models, the voiceless variants are preferred at higher rates than in the case of the German alignment: for the phonemically voiced Portuguese /v/ the Italian [f] models are preferred in 89.8% of cases compared to only 72.2% cases of German [f]. For the Portuguese phonemically voiced /z/ the Italian [s] models are preferred in 72.2% of cases compared to only 29.9% cases of German [s]. All the attested differences are statistically significant.

Limitations

The goal of the present study was to (in)validate typological classification results derived from small scale acoustic studies on large scale corpus data. The proposed methodology (i.e., using different combinations of trained acoustic phone models) does not permit a direct replication: while the acoustic studies relied on Praat's (Boersma and Weenink, 2019) autocorrelation (AC) pitch extraction algorithm, the present study relies on the acoustic models of the systems, which include multiple acoustic features. An acoustic analysis pinpointing the most relevant acoustic features is needed. A second limitation of the current study is the non-inclusion of several acoustic correlates of voicing in the analysis. It is known that adjacent segments, position in the word/syllable and stress have a significant effect on devoicing rates (Pape et al., 2003; Bybee and Easterday, 2019; Hutin et al., 2022). The phonological/phonetic context is all the more pertinent given the use of German as a comparison language, whose acoustic correlates for voicing are dependent on word position (word medial vs. initial vs. final) and stress (Jessen, 1998; Fuchs, 2005). Positional effects in the case of fricatives are more reduced in the case of EP since the only licensed consonants in coda position are /l/, /r/ and /ʃ/ and word initial /v,z/ are rare (in our data /v/: 1671 tokens and /z/: 122 tokens) and found mainly in loanwords. Stress on the other hand is relevant: EP, like German, and unlike Italian is said to be stress-timed (Cruz-Ferreira, 1999) or partially stressed and partially syllable-timed (Frota and Vigário, 2006). The corpora is not annotated for prosodic information which limits our analysis. Another well known correlate of voicing that has only indirectly been included in the analysis (via the trained acoustic models) is segmental duration. A more detailed analysis including phonological context and duration information is needed to better explain the patterns.

6 Conclusion

The present study tested whether voicing in European Portuguese fricatives is more similar to Italian, a closely related Romance language, or to German, a more distantly related language, on large scale corpora using ASR acoustic modeling and pronunciation variants. By allowing the system to choose a preferred model when making the alignment we replicated results from small scale acoustic studies that showed EP tends to diverge from other Romance languages when it comes to fricative voicing. The results also show that the effect seems to be modulated by place of articulation, with the more posterior fricative (the coronal /z/) behaving differently than the more anterior labiodental /v/.

Results support the use of speech technology methodologies to replicate and test phonological hypotheses on large amounts of data (Yuan and Liberman, 2011; Ryant et al., 2013).

Acknowledgements

This work was partially supported by the French National Research Agency (ANR) as part of the DIPVAR project (ANR-21-CE38-0019).

References

Martine Adda-Decker and Lori Lamel. 1999. Pronunciation variants across system configuration, language and speaking style. *Speech Communication*, 29(2-4):83–98.

Paul Boersma and David Weenink. 2019. Praat: doing phonetics by computer. *Computer program*.

Joan Bybee and Shelece Easterday. 2019. Consonant strenghtening: a crosslinguistic survey and articulatory proposal. *Linguistic Typology*, 2(23):263–302.

Madalena Cruz-Ferreira. 1999. *Portuguese (European)*. Cambridge University Press.

Sonia Frota and Marina Vigário. 2006. On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus*, 13(2):247–275.

Susanne Fuchs. 2005. *Articulatory correlates of the voicing contrast in alveolar obstruent production in German*. Phd thesis, ZAS, Berlin, Germany.

Hynek Hermansky. 1990. Perceptual linear prediction (plp) analysis for speech. *Journal of the Acoustical Society of America*, 87.

Mathilde Hutin, Martine Adda-Decker, Lori Lamel, and Ioana Vasilescu. 2022. When phonetics meets morphology: Intervocalic voicing within and across words in Romance languages. In *Proceedings of the 23rd Annual Conference of the International Speech Communication Association*, pages 3438–3442.

Wouter Jansen. 2004. *Laryngeal contrast and phonetic voicing: a laboratory phonology approach to English, Hungarian and Dutch*. Dissertation, University of Groningen.

Michael Jessen. 1998. *Phonetics and Phonology of Tense and Lax Obstruents in German*. John Benjamins Publishing Company.

Luis M. Jesus and Christine H. Shadle. 2003. Devoicing measures of european portuguese fricatives. In *Proceedings of the International Conference on Computational Processing of the Portuguese Language*, volume 2721.

Lori Lamel and Gilles Adda. 1996. On designing pronunciation lexicons for large vocabulary continuous speech recognition. In *Proc. 4th International Conference on Spoken Language Processing*, Philadelphia, USA.

Lori Lamel, Sandrine Courcinous, Julien Despres, Jean-Luc Gauvain, Yvan Josse, Kevin Kilgour, Florian Kraft, Viet-Bac Le, Hermann Ney, Markus Nußbaum-Thom, Ilya Oparin, Tim Schlippe, Ralf Schlüter, Tanja Schultz, Thiago Fraga da Silva, Sebastian Stüker, Martin Sundermeyer, Bianca Vieru, Ngoc Thang Vu, Alexander Waibel, and Cécile Woehrling. 2011. Speech recognition for machine translation in quaero. In *Proceedings of the 8th International Workshop on Spoken Language Translation: Evaluation Campaign*, pages 121–128, San Francisco, California.

Leigh Lisker and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, (20):384–422.

Daniel Pape and Luis M.T. Jesus. 2011. Devoicing of phonologically voiced obstruents: is European Portuguese different from other Romance languages. In *Proceedings of the 23rd International Conference of Phonetic Sciences*, pages 1566–1569.

Daniel Pape and Luis MT Jesus. 2015. Stop and fricative devoicing in European Portuguese, Italian and German. *Language and Speech*, 58(2):224–245.

Daniel Pape, Christine Mooshamer, Phil Hoole, and Susanne Fuchs. 2003. Devoicing of word-initial stops: a consequence of the following vowel? In *Proceedings of the 6th International Seminar on Speech Production*, pages 207–212, Sydney, Australia.

Neville Ryant, Jiahong Yuad, and Mark Liberman. 2013. Automating phonetic measurement: The case of voice onset time. In *In Proceedings of Meetings on Acoustics ICA2013*, Montreal, Canada.

Chilin Shih and Bernd Möbius. 1998. Contextual effects on voicing profiles of German and Mandarin consonants. In *Proceedings of the 14th International Congress of Phonetic Sciences*, pages 989–992.

Chilin Shih and Bernd Möbius. 1999. Contextual effects on consonantal voicing profiles: A cross-linguistic study. In *Proceedings of the 14th International Congress of Phonetic Sciences*, pages 989–992, San Francisco, US.

Petra M. van Alphen and Roel Smits. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics*, 32(4):455–491.

Yaru Wu, Mathilde Hutin, Ioana Vasilescu, Lori Lamel, and Martine Adda-Decker. 2022. Extracting linguistic knowledge from speech: A study of stop realization in 5 Romance languages. In *Proceedings of the International Conference on Language Resources and Evaluation*, pages 3257–3263.

Jiahong Yuan and Mark Liberman. 2011. /l/ variation in american english: A corpus approach. *Journal of Speech Sciences*, (2):35–46.