Towards articulatory comparison between French and German from MRI data

Antoine Serrurier*, Christiane Neuschaefer-Rube
Clinic for Phoniatrics, Pedaudiology & Communication Disorders, University Hospital and Medical Faculty of the RWTH Aachen University, Germany
*aserrurier@ukaachen.de

Introduction

Although well documented phonetically, quantitative comparisons between languages in the articulatory domain remain arduous due to the large inter-speaker variability. In addition to their language, speakers differ indeed by their morphologies and their idiosyncratic articulatory strategies. These two components can be handled by normalising the speakers to remove the variations due to the morphology on the one hand and by considering several speakers to reveal the common language-related features despite the different strategies on the other hand. This study attempts to compare the articulatory spaces of two languages, namely French (FR) and German (DE), based on these two principles. The analyses are based on anatomical Magnetic Resonance Imaging (MRI) data of the vocal tract area and make use of articulatory modelling.

Material & methods

Static midsagittal MRI data have been recorded for 11 French and 5 German native speakers sustaining artificially 62 phonemes representative of the articulatory repertoire of their respective language. The articulators surrounding the vocal tract have been manually segmented and aligned on a common anatomical reference system based on the cranium, as described in [4] for the FR dataset. Due to the relatively large and balanced corpus, the mean articulation of a speaker is assumed to be free of any idiosyncratic articulatory strategy and to bear the morphology characteristics [4]. The normalisation between speaker is therefore achieved by subtracting from each articulation the mean articulation of the speaker and by replacing it by the overall mean articulation computed over the 16 speakers, following a procedure described in [3]. The two datasets have been compared in terms of variance. The Standard Deviation (STD) of each contour point over the 11×62 observations and 5×62 observations for the FR and DE datasets has been calculated.

Following a method proposed by [2] in the context of speech *vs.* feeding and also implemented for bilingual studies [1], the performance of the cross-reconstructions of each dataset by articulatory models built on the other dataset has been evaluated. For that purpose, individual organ-based articulatory models based on guided Principal Component Analysis with 14 components [4] have been built. Each speaker articulations have been reconstructed by each articulatory model and the Root-Mean-Square reconstruction errors calculated. The accuracy of the reconstructions of the FR data by the FR models represents the baseline and the gap with the accuracy of the reconstructions of the same FR data by the DE models represents the deficit of the DE models to reconstruct the FR data, and vice versa. Models with lower deficit could be considered as more general, suggesting a larger articulatory space.

Finally, the articulations of all speakers have been projected into the same articulatory space for comparing the control parameters. For that purpose, a single cross-language articulatory model has been built on the 62+62 articulations calculated by averaging each dataset over the speakers. The articulations of each speaker have been reconstructed by this model, leading to control parameters corresponding to the same articulatory space. The control parameters have subsequently been compared between the speakers of each dataset.

Results

Figure 1 represents the differences between the overall STD between the DE and FR data for the X and Y dimensions. More variability is observed for the jaw and tongue in the FR data while DE data exhibit more variability for the other articulators.

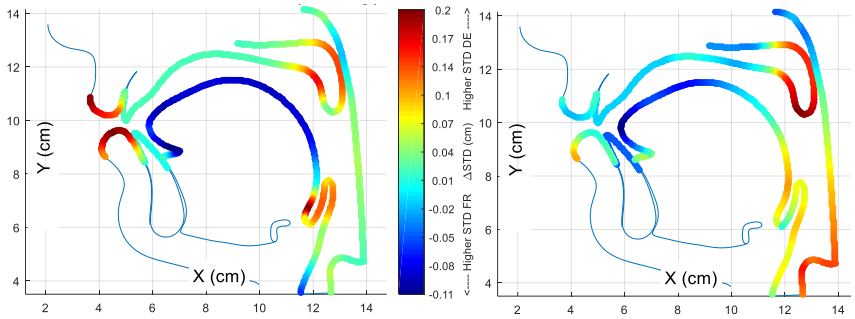

Figure 1: Average midsagittal contours of the vocal tract; colour denotes the difference between the STD of the DE data and the FR data for the X (left) and Y (right) dimensions; dark blue (resp. red) represents higher STD for the FR (resp. DE) data.

Deficit of the FR and DE models have been calculated from the pairwise reconstructions for all speakers. While overall results show similar performance of the FR and DE models, further detailed analyses revealed a very slight lower deficit for the DE models, suggesting that DE articulations might encompass more FR articulations than FR articulations might encompass DE articulations.

Finally, a relative homogeneity was observed for the control parameters of the speakers projected in the same articulatory space. The most salient difference was observed for the main control parameter of the velum, for which DE data require a larger range of variation, in line with the observations on the data variabilities (Figure 1).

Discussion & perspectives

This study proposes a framework for the quantitative articulatory comparison between FR and DE. Slightly higher variability of the DE data and slightly better performance of the DE models over the FR models was observed, suggesting that DE articulations might encompass more FR articulations than FR articulations might encompass DE articulations. Further analyses will be conducted to verify these preliminary results. Although the variability ascribable to the morphology has been removed from the data, the relatively low number of speakers, even though still high for such demanding manual processing, makes the distinction between common language-related features and idiosyncratic strategy-related features still challenging. Informal tests show indeed a sensitivity of the results to the numbers of speakers included in each dataset. Further analyses may therefore include more speakers. In addition, modalities recording articulations for a dynamic speech task, such as provided by real-time MRI, may open the possibility to compare directly the articulations between languages, and not only the articulatory spaces as done in this study.

References

[1] Badin, P., Sawallis, T.R., Crépel, S. and Lamalle, L. 2014. Comparison of articulatory strategies for a bilingual speaker: Preliminary data and models. *ISSP10*, 17–20.

[2] Serrurier, A., Badin, P., Barney, A., Boë, L.-J. and Savariaux, C. 2012. The tongue in speech and feeding: Comparative articulatory modelling. *J Phon*. 40(6), , 745–763.

[3] Serrurier, A., Badin, P., Boë, L.-J., Lamalle, L. and Neuschaefer-Rube, C. 2017. Inter-speaker variability: speaker normalisation and quantitative estimation of articulatory invariants in speech production for French. *Interspeech*, 2272–2276.

[4] Serrurier, A., Badin, P., Lamalle, L. and Neuschaefer-Rube, C. 2019. Characterization of inter-speaker articulatory variability: a two-level multi-speaker modelling approach based on MRI data. *J Acoust Soc Am* 145(4), 2149–2170.