COMPARISON OF THE FRENCH AND GERMAN ARTICULATORY SPACES

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

Abstract: Comparing languages in terms of articulations remains an arduous task due to the inherent different phonetic repertoires and the large inter-speaker variability. This study proposes a model-based approach to compare the articulatory spaces of French (FR) and German (DE) and to explore the discrepancy between the range of articulations of these languages. The approach consists in building articulatory models for the two languages and to reconstruct representative articulations of these languages by the two models. The accuracy of the reconstructions of the FR articulations by the FR model represents the baseline and the gap with the accuracy of the reconstructions of the same FR articulations by the DE model represents the deficit of the DE model to reconstruct the FR articulations, and vice versa. Static midsagittal Magnetic Resonance Images of 11 FR and 5 DE speaker sustaining articulations representative of their respective phonetic repertoire have been considered and the articulator contours manually segmented and aligned. After normalising the data over the speakers, individual articulator-based linear articulatory models have been derived and pairwise cross-reconstructions of each speaker data by each model performed. Comparative analyses of the performance of the cross-reconstructions tend to show a similar articulatory space for the two languages, suggesting that the articulatory degrees of freedom of FR speakers are enough to produce DE articulations and vice versa. However, the large inter-speaker variability highlighted during analysis suggests that the discrepancy between the speakers' individual strategies might be larger than the discrepancy between the languages' articulatory spaces.

1 Introduction

Learning a second language (L2) may involve the difficult task to form new vocal tract articulations not present in the native language (L1). This task may be all the more difficult when the new articulations are distant from any known L1 articulation [1]. The articulatory discrepancy between the two languages may therefore be an indicator of the challenge for a speaker to learn a L2 language. Measuring this discrepancy constitutes the focus of the current study.

This question lies in the more general framework of comparing different articulatory datasets. A very large variety of approaches are observed in the literature to deal with this issue, reflecting the various motivations driving these studies. Articulatory comparisons can indeed be motivated for instance by comparing languages [2–8], comparing dialects of a same language [9], exploring L2 learning strategies [10] or exploring individual strategies in the context of bilingualism or multilingualism [11, 12]. Within this variety, most studies focus on specific phonetical characteristics, *e.g.* the realisation of coronal consonants [2] or the lip shape [8]. The data are also obtained through very different techniques, *e.g.* X-rays [3], Electromagnetic Articulography (EMA) [4–7,9], Magnetic Resonance Imaging (MRI) [11,12], Electropalatography [6] or photographs [8], leading to different levels of analysis. A few studies attempt to compare more globally the range of articulations. Li *et al.* [7] attempt to hierarchically cluster the phonemes of Mandarin Chinese and English on the basis of the Mahalanobis distances of EMA data. Based on midsagittal contours obtained from MRI, Badin *et al.* [11] compare the analogous articulatory components derived from two articulatory models, one build on a set of French articulations and the other on a set of English articulations, for a bilingual speaker. Finally, based on EMA data, Serrurier *et al.* [13] compare the speech and feeding articulatory spaces by reconstructing one dataset by an articulatory model built on the other dataset for a same speaker and by comparing the performance of the cross-reconstructions. Note that this approach was also attempted by Badin *et al.* [11] but the results are not explicitly reported in their study. The current study was directly inspired by this approach and intends more specifically to compare the articulatory spaces of French (FR) and German (DE) by means of modelling.

Attempts have already been made to compare FR and DE articulations. Delattre [3] compares for instance midsagittal contours obtained from X-rays for the vowel /i/, Hoole *et al.* [6] explore the coarticulation strategies for the consonants /s/ and /ʃ/ from EMA data, Bombien *et al.* [14] explore the effect of voicing on the oral articulations, Zimmerer *et al.* [15] analyse from acoustic data the production of DE /h/ by FR and DE speakers and Gendrot *et al.* [4] compare the realisation of FR and DE /ʁ/ from EMA data. To our knowledge, no study addresses the issue of articulatory space. The question in the current study is therefore to determine whether the articulatory repertoire of one of the two languages encompasses the other one and which articulatory dimensions might be missing in one language in comparison to the other.

Although specific studies mentioned above involve the same speaker performing two different tasks [11–13], comparing articulatory realisations of languages involves in the general case different speakers. The comparison becomes then arduous due to the large inter-speaker variability already reported in the literature (*e.g.* [16]). Indeed, in addition to the speech task, speakers differ also by their morphology, *i.e.* the position and shape of the articulators irrespective of the speech task, and their articulatory strategy, *i.e.* the displacement and deformation of the articulators to achieve the articulatory targets. The inter-speaker variability ascribable to the morphology can be addressed by normalising the articulatory data on the speakers, for which various *ad hoc* methods have been proposed in the literature [9,17]. In this study, we rely on an original method aiming at completely removing this variability from the data [18]. The issue of the inter-speaker variability related to the idiosyncratic articulatory strategies of the speakers can be addressed by considering a large set of speakers in order to disentangle the common and the speaker-specific articulatory features. This study constitutes a preliminary attempt towards this objective.

The general objective of the study consists thus in comparing the articulatory spaces of FR and DE by means of articulatory modelling. In addition to describe the data, articulatory models have the power to make predictions according to the data on which they have been built. This study aims at taking advantage of this property by attempting to predict FR data from an articulatory model based on DE data and vice versa. The analysis of the performance of these cross-reconstructions aims at providing an indication of the discrepancy of the two articulatory spaces. To describe this procedure, the manuscript is organised as follows: the section 2 describes the data and the methods, the section 3 presents the results and the section 4 comments the results and provides a general discussion.

2 Material and methods

2.1 Data and corpora

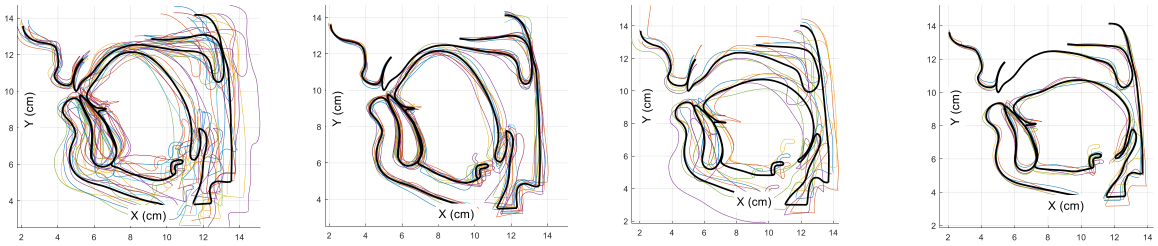
The data consist in two datasets of static midsagittal MRI images of the vocal tract. The FR dataset has been recorded on 11 French speakers (6 males, 5 females) sustaining artificially 62 articulations: 10 oral and 2 nasal vowels [i e ɛ a y ø œ u o ɔ ã ɔ̃], and each of the 10 consonants [p t k f s ʃ m n ʁ l] in the 5 symmetric vowel contexts [i e ɛ a u]. The DE dataset has been recorded on 5 German male speakers sustaining artificially also 62 articulations: 10 vowels [a: e: ɛ: i: o: u: y: ø:], the consonants [p t k f s ʃ m n ŋ l] in the 5 symmetric vowel contexts [i: ɛ: y: a: u:] and the consonants [ç x] in the 2 respective vowel contexts [i: ɛ:] and [a: u:]. Note that the two datasets have the same number or articulations, which was not initially targeted but is a result of a similar design. The set of articulations of each dataset is considered as balanced and representative of the articulatory repertoire of each language, making them suitable to analyse the articulatory spaces.

The contours of all articulators surrounding the vocal tract on each image have been manually segmented and the resulting articulations aligned per speaker on the contour of the hard palate and between speakers on corresponding landmarks taken on the palate bone. Further details on the FR data and the processing procedure can be found in [16]. The data consist in the end in two datasets of contour coordinates of respective size 11×62×1037×2 and 5×62×1037×2 for the FR and DE dataset corresponding to the 11 FR and 5 DE speakers, the 62 FR and DE articulations, the 1037 contour points and the 2 x-y dimensions. Note that the articulation contours will be sometimes referred to as articulations in this manuscript for simplicity reasons.

2.2 Methods

2.2.1 Speaker normalisation

As mentioned in the introduction, in addition of using different languages, the speakers differ by their morphologies and their idiosyncratic articulatory strategies. Normalising the speakers' articulations removes the variations ascribable to the morphology in the data. Various methods have been proposed to normalise speakers according to their morphologies [9,17]. The method used in this study follows a procedure initially proposed by Serrurier *et al.* [18] supposed to remove exactly all variability related to morphology variations. Due to the large and balanced corpus supposedly sampling the articulatory space of a speaker, the mean articulation of each speaker can be considered as free from the articulatory strategy and represent her/his morphology. In addition, each articulation of a speaker can be considered as the deformation of the mean articulation towards the target articulation, implying that the mean articulation is present in each articulation. The normalisation consists in replacing in all articulations the corresponding mean articulation by the overall mean articulation calculated over the overall datasets. Practically, it consists in subtracting for each articulation the speaker mean articulation and adding the overall mean articulation. By this method, all speakers have at the end the same mean articulation, *i.e.* the same morphology. Note however that individual strategies, including strategies possibly deriving from morphology constraints, remain present in the data. It can also be seen as removing for each articulation the marginal morphology of each speaker between the overall mean articulation and the speaker mean articulation. An illustration of the effect of normalisation on the data can be seen in Figure 1. Further details on the procedure can be found in [18] and on the mean articulation in [16]. All the further processing described in this article are performed on the normalised data.



FR speakers – phoneme /i/ DE speakers – phoneme /a/

Figure 1 – Superposition of the articulation contours for the FR speakers for /i/ (left and middle left) and for the DE speakers for /a/ (middle right and right) before (left and middle right) and after (middle left and right) applying the normalisation.

2.2.2 Variability analysis

The variability of the two datasets has been measured in terms of Standard Deviation (STD), *i.e.* in terms of variations around the mean value. The STD has been preferred over the variance in order to obtain values in interpretable units. Given the matrix format of the data, the STD can be calculated in overall as well as per speaker and per contour point, as presented in the section 3.

2.2.3 Modelling analysis

As explained in the introduction, this study aims at tackling the problem of comparing the articulatory spaces of two datasets by means of modelling. Two modelling approaches are considered.

In the first approach, the underlying idea is to perform cross-reconstructions of the data by the model corresponding to the other dataset and to evaluate whether the articulatory model built on one dataset could outperform the articulatory model built on the other dataset. For this purpose, individual articulatory models have been built for the 16 speakers of the study. Following a principle detailed in [16], the models are articulator-based and are obtained by so-called *guided* Principal Component Analysis (PCA), *i.e.* an iterative PCA where the deformations corresponding to the components aim at being realistic in terms of biomechanics. It results in linear models made of 14 articulatory components. In such articulatory models, the data articulations can be represented as a linear combination of eigenvectors, the articulatory components, weighted by control parameters, also referred to as predictors. Further details can be found in [16].

The accuracy of the articulatory models is measured in terms of Root-Mean-Square (RMS) errors of the reconstructions, *i.e.* the RMS error between the reconstructed articulations and the data articulations, expressed in cm. As for the STD, the matrix form of the reconstructed data allows the calculation of the RMS in overall as well as per speaker, per phoneme, per articulator or per contour point. The individual models present similar performance, with an RMS error of 0.1±0.02 cm for the FR dataset and 0.1±0.01 cm for the DE dataset. They demonstrate therefore a similar power of prediction on the data on which they are built.

Pairwise reconstructions of the 16 speakers' data by the 16 articulatory models have been performed. For that purpose, the predictors corresponding to the data of a specific speaker to be reconstructed by a specific articulatory model have been iteratively estimated by inversing the model, *i.e.* by multiplying the considered articulation data with the inversed of the considered eigenvector matrix. The articulations have then been reconstructed by multiplying these predictors and the eigenvector matrix. The reconstructions have been evaluated in terms of RMS error in overall and per speaker.

The articulatory models built on the data of one language are supposed to be optimal to reconstruct data of the same language. The reconstructions of the FR dataset by the FR models constitute therefore the performance baseline for the reconstruction of the FR dataset, and inversely for the DE dataset. The reconstructions of the same FR dataset by the DE models are supposed to present higher reconstruction errors. The gap between the RMS error of the reconstructions of the FR dataset by the DE models and the RMS error of the reconstructions of the same FR dataset by the FR models, referred to as *ΔRMS* in this study, can therefore be considered as the deficit of the DE models to reconstruct the FR dataset, and vice versa. The ΔRMS of the DE models and of the FR models have been calculated from the pairwise reconstructions and are reported per contour point and phoneme together. Furthermore, an illustration of the reconstructions corresponding to the highest ΔRMS of the DE and FR models is provided.

The second modelling approach consists in projecting all speakers in the same articulatory space, *i.e.* in decomposing the FR and DE articulations according to similar articulatory components, and to compare the range of use made by the FR and DE speakers of these components. For that purpose, a single cross-language articulatory model of 14 components following the same principles as described earlier [16] has been built on the 124 articulations obtained by pooling the 62 FR articulations averaged over the 11 speakers and the 62 DE articulations averaged over the 5 speakers. This model is therefore supposed to cover both FR and DE articulations. The predictors corresponding to the data of a specific speaker to be reconstructed by this cross-language model have then been estimated as described earlier for the pairwise reconstructions, leading to two sets of respectively 11×62×14 and 11×62×14 predictors for the FR and DE datasets. The range of the two sets has then been compared per articulatory component to determine whether one dataset would use some articulatory components at a different degree than the other dataset.

3 Results

3.1 Variability analysis

The DE dataset shows a higher overall variability, with an overall STD of 0.29 cm against 0.26 cm for the FR dataset. Figure 2 (left) plots the STD per speaker. It shows a higher inter-speaker variability for the 11 FR speakers than for the 5 DE speakers. A finer analysis per dimension and per contour point is displayed in Figure 2 (middle and right). It shows in general a higher variability for the DE dataset, except notably for the tongue.

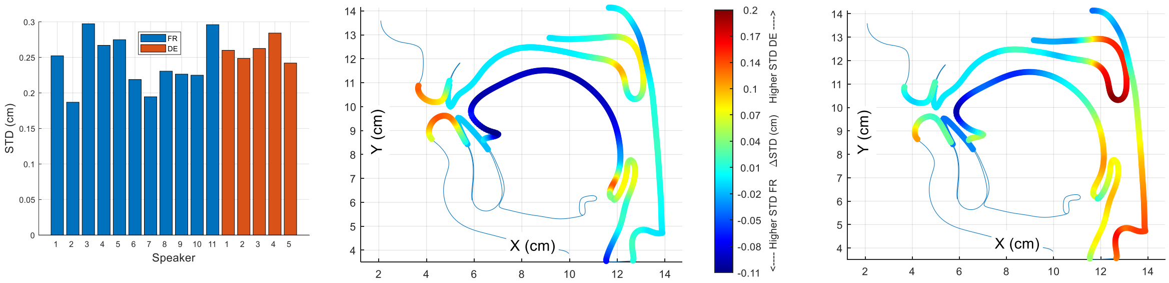


Figure 2 – Left: STD of the data per speaker. Middle and right: Overall mean articulation where colour codes the difference between the STD of the contour points of the DE dataset and the STD of the contour points of the FR dataset for the X (middle) and Y (right) dimensions; higher STD for the DE dataset (*resp*. FR dataset) is coded in red (*resp*. blue).

3.2 Modelling analysis

For the first modelling approach, the cross-reconstructions show similar overall performance for the FR and DE datasets: the RMS errors for both datasets reconstructed by the models built on themselves are 0.15 cm while the RMS errors reconstructed by the models built on the other dataset are 0.16 cm, leading to similar ΔRMS of 0.01 cm for both DE and FR models. It means that the models built on the other dataset are very slightly suboptimal to reconstruct one dataset, but this very slight level of suboptimality is comparable between the FR and DE models. In other words, the power of prediction of the FR models to reconstruct the DE articulations appears similar to the power of prediction of the DE models to reconstruct the FR articulations. These overall performances hide however a large variability between speakers. Figure 3 illustrates the RMS errors of the cross reconstructions per speaker and per model. It can be seen notably high errors for the model built on DE speaker number 4 reconstructing the FR articulations and conversely high errors for the FR models reconstructing the articulations of the same DE speaker number 4. Note also for instance the incompatibility between FR speaker 7 and DE speaker 4.

The ΔRMS of the DE and FR models per contour point and articulation are presented in Figure 4. The high values observed for the contour points near the glottis emphasize for this region the difficulty for the models built on one dataset to reconstruct the data of the other dataset. More interestingly, it can be seen higher values in the ΔRMS of the DE models for contour points corresponding to the tip of the tongue and in the ΔRMS of the FR models for contour points corresponding to the tip of the velum. It means that DE models tend to miss more than FR models the reconstruction of the tip of the tongue in FR articulations and that FR models tend to miss more than DE models the reconstruction of the tip of the velum in DE articulations. This suggests that some articulatory degrees of freedom regarding the tip of the tongue (*resp*. velum) might be missing in the DE (*resp*. FR) articulations in comparison to the FR (*resp*. DE) articulations. Figure 5 illustrates the data and reconstructions corresponding to the extreme phonemes in that case identified on Figure 4.

The second modelling analysis did not reveal any difference in the use of the same articulatory components between the FR and DE speakers: two-sample t-tests between the 14 predictors of the FR and DE datasets did not show statistically significant differences between the two sets, suggesting that FR and DE speakers use the same articulatory components in a similar range.

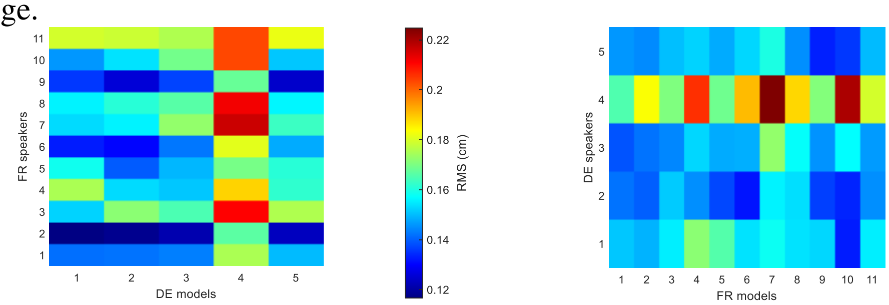


Figure 3 - Images of the RMS errors of the reconstructions of the 11 FR speaker articulations by the 5 DE models (left) and of the 5 DE speaker articulations by the 11 FR models (right).

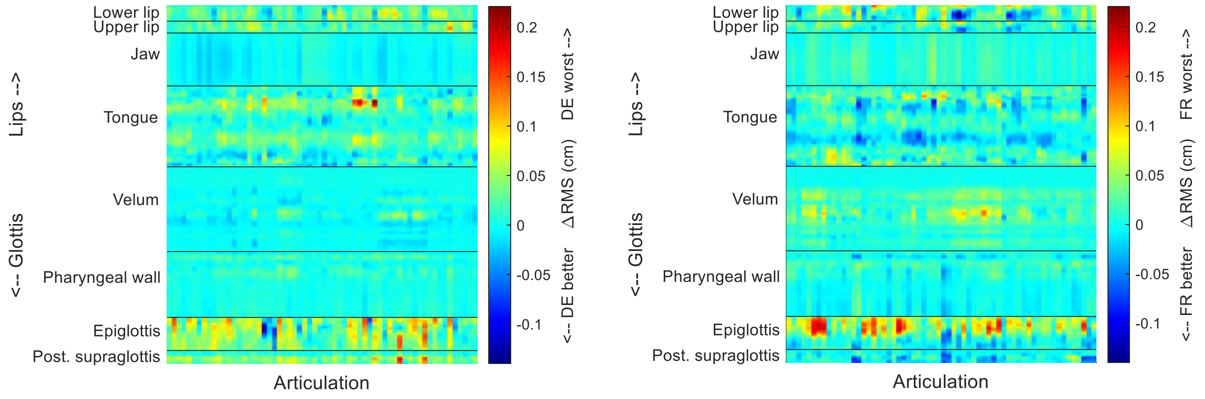


Figure 4 - Images of the ΔRMS of the DE models reconstructing the FR data (left) and of the FR models reconstructing the DE data (right) calculated per contour point (y dimension) and articulation (x dimension).

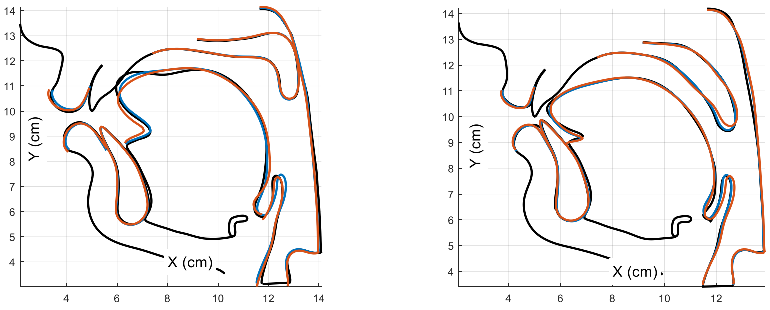


Figure 5 - Data contours for the phonemes [l^u] (black left) and [n^a] (black right) respectively averaged over the 11 FR and the 5 DE speakers superposed with the corresponding reconstructions obtained from the models built on the same dataset (blue) and from the models built on the other dataset (red).

4 Discussion & conclusion

This study applies the methodology initially proposed by Serrurier *et al.* [13] to compare the articulatory spaces represented by two datasets to the case of FR and DE. According to the preliminary results presented here, a slight deficit of the DE models to reconstruct the tip of the tongue of FR articulations and of the FR models to reconstruct the tip of the velum of DE articulations is observed. This corresponds interestingly to the larger variability observed for the tongue in the FR data and for the velum in the DE data. Nonetheless, despite this slight deficit, it seems that FR and DE present rather similar articulatory spaces. This would have for consequences that the articulatory degrees of freedom of native FR speakers are enough to produce DE articulations and vice versa for the DE native speakers. FR and DE L2 learning should therefore not be problematic for DE and FR native speakers regarding the formation of new articulations. This general observation hides however a large inter-speaker variability, making difficult to uncover language-specific features independent from the speaker, in general agreement with observations from Dart [2] on FR and English coronal consonants. Although the inter-speaker morphology was supposedly removed from the data, the remaining inter-speaker variability related to the speakers' idiosyncratic articulatory strategies may be larger than the differences between the languages' articulatory spaces, limiting the possibilities to characterise this discrepancy. A larger set of speakers may help to solve this issue.

This study is based on static articulations artificially sustained by the speakers. As emphasized by Delattre [3], static articulations may not be representative of the corresponding articulations in dynamic speech, limiting their analysis in terms of phonetics. However, such data have already proved to sample the articulatory space of the speaker and to be representative of her/his articulatory capacities [19]. This make them valid for modelling approaches and for extracting the articulatory degrees of freedom as done in the current study.

As emphasized by Serrurier *et al.* [13], the design of the articulatory models is crucial to compare two datasets by means of cross-reconstructions: adding more articulatory components in a model increases the number of degrees of freedom taken into account in the model and should lead to better reconstructions, boosting the observed performance of the considered model in comparison to others. For this reason, the choice of guided PCA seems appropriate, as only biomechanically plausible components are retained in the models. The chosen approach to design similarly the 16 models of the study might however lead to miss some specific degrees of freedom for some speakers. Individual assessment should be performed to solve this issue. Note also that the difference in the performance of the models to perform the cross-reconstructions approaches sometimes the accuracy of the models. Finer and more speaker-specific analyses might be therefore necessary in the future.

Despite these limitations, this study presents a promising methodology to compare the articulatory spaces corresponding to two different languages. Further research should involve more speakers in order to draw conclusions regarding the languages despite the variable speakers' idiosyncratic strategies.

Acknowledgments

The authors are very grateful to P. Badin for providing the FR data and to L. Lamalle and S. Romanzetti for participating at recording the FR and DE data. The DE corpus has been designed with the help of Bernd Kröger and Cornelia Busch. The initial tracing was performed by J.-A. Valdés Vargas, G. Ananthakrishnan, Mohammad Eslami, Arne Hülsdünker and Frederic Stepp. This research project is supported by the START-Program of the Faculty of Medicine, RWTH Aachen. The recording of the FR data has been partially funded by the French ANR (grant ANR-08-EMER-001-02 'ARTIS').

References

[1] Ellis, R.: *The Study of Second Language Acquisition*. Oxford University Press, 1994.

[2] Dart, S. N.: *Comparing French and English coronal consonant articulation*. In *Journal of Phonetics*, 26(1), 71–94, 1998.

[3] Delattre, P.: *Comparing the Vocalic Features of English, German, Spanish, and French*. In *IRAL - International Review of Applied Linguistics in Language Teaching*, 2(1), 71–98, 1964.

[4] Gendrot, C., Kuhnert, B., and Demolin, D.: *Articulatory and acoustic realization of French and German /R/*. In *The Journal of the Acoustical Society of America*, 140(4), 3221–3221, 2016.

[5] Geng, C.: *A cross-linguistic study on the phonetics of dorsal obstruents*. PhD Thesis, Humboldt-Universität zu Berlin, 2008.

[6] Hoole, P., Nguyen-Trong, N., and Hardcastle, W.: *A Comparative Investigation of Coarticulation in Fricatives: Electropalatographic, Electromagnetic, and Acoustic Data*. In *Language and Speech*, 36(2–3), 235–260, 1993.

[7] Li, S. and Wang, L.: *Cross Linguistic Comparison of Mandarin and English EMA Articulatory Data*. In," in *Proc. of Interspeech*, 2012, 903–906.

[8] Zerling, J.-P.: *Frontal lip shape for French and English vowels*. In *Journal of Phonetics*, 20(1), 3–14, 1992.

[9] Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., and Baayen, R. H.: *Investigating dialectal differences using articulography*. In *Journal of Phonetics*, 59, 122–143, 2016.

[10] Wilson, I. L.: *Articulatory settings of French and English monolingual and bilingual speakers*. PhD Thesis, The University of British Columbia, 2006.

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

[12] Badin, P., Tabain, M., and Lamalle, L.: *Comparative study of coarticulation in a multilingual speaker: Preliminary results from MRI data*. In Proc. of ICPhS, 2019, 3453–3457.

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

[14] Bombien, L. and Hoole, P.: *Articulatory overlap as a function of voicing in French and German consonant clusters*. In *The Journal of the Acoustical Society of America*, 134(1), 539–550, 2013.

[15] Zimmerer, F. and Trouvain, J.: *Productions of /h/ in German: French vs. German speakers*. In *Proc. of Interspeech*, 2015, 1922–1926.

[16] Serrurier, A., Badin, P., Lamalle, L., and Neuschaefer-Rube, C.: *Characterization of inter-speaker articulatory variability: a two-level multi-speaker modelling approach based on MRI data*. In *The Journal of the Acoustical Society of America*, 145(4), 2149–2170, 2019.

[17] Geng, C. and Mooshammer, C.: *How to stretch and shrink vowel systems: Results from a vowel normalization procedure*. In *The Journal of the Acoustical Society of America*, 125(5), 3278–3288, 2009.

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

[19] Beautemps, D., Badin, P., and Bailly, G.: *Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling*. In *The Journal of the Acoustical Society of America*, 109(5), 2165–2180, 2001.