COMPARATIVE STUDY OF COARTICULATION IN A MULTILINGUAL SPEAKER: PRELIMINARY RESULTS FROM MRI DATA

Pierre Badin[1], Marija Tabain[2], Laurent Lamalle[3]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France
* Institute of Engineering Univ. Grenoble Alpes
[2] Department of Languages and Linguistics, La Trobe University, Melbourne, Australia
[3] Inserm US 17 - CNRS UMS 3552 - Univ. Grenoble Alpes & CHU de Grenoble, UMS IRMaGe, France
pierre.badin@gipsa-lab.grenoble-inp.fr, M.Tabain@latrobe.edu.au

ABSTRACT

This study explores the relative coarticulatory resistance of consonants across languages using static mid-sagittal MRI of one trilingual speaker sustaining consonants in symmetrical vowel contexts for Australian English (AuE), French (FR), and Croatian (CR). Semi-supervised segmentation provided articulators' contours that can be superposed and compared. A set of 7 vowels analogous for the 3 languages was determined and used to compute mean consonant contours and dispersion related to vowels. A focussed observation of /s/ and /ʃ/ showed that the tongue in CR /s/ is more resistant to coarticulation than in AuE or FR, whereas for /ʃ/ it is more resistant in FR than in AuE or CR. It was also found that, the stop and the fricative parts of the CR affricates /ʦ/ and /ʧ/ are more resistant than the single consonants /t/, /s/ and /ʃ/. This study will be extended to other consonants, and make use of articulatory modelling.

Keywords: Australian English, French, Croatian, co-articulation, MRI, articulatory measurement.

1. INTRODUCTION

Previous work on the sibilant fricatives /s ʃ/ has shown that these sounds are highly resistant to lingual coarticulation with adjacent vowels ([6], [23]). Fricative sounds block coarticulation on the high-low dimension, due to aerodynamic requirements for fricative production ([14]): in order to produce the acoustic turbulence needed for noise generation, the tongue must form a very narrow channel along the palate, thereby preventing lowering of the tongue body in anticipation of an adjacent low vowel. In the case of sibilant fricatives, this channel must direct airflow towards the lower teeth in order to amplify the spectral energy generated at the constriction ([24]), and this requirement places further constraints on the tongue configuration.

However, /s/ and /ʃ/ are articulated at different points along the palate, and also involve different portions of the tongue in their articulations. /s/ is typically articulated in the denti-alveolar region using the tongue tip (with exact location differing between languages), while /ʃ/ is typically articulated in the post-alveolar/pre-palatal region, using the tip-blade complex (*i.e.* a longer portion of the front part of the tongue). The constriction for /s/ tends to be very narrow, and for /ʃ/ a little wider ([3], [1]). As a consequence of these differences, competing predictions might be made regarding the relative amount of coarticulation possible for these two sibilant fricatives. On the one hand, the narrower constriction for /s/ may result in less coarticulation with adjacent vowels, since very precise lingual control would be required to maintain such a narrow channel ([18], [20], [16]). On the other hand, the greater involvement of the tongue body for /ʃ/, due to greater coupling with the tip-blade complex for /ʃ/ than with the tip alone for /s/, may result in less coarticulation for /ʃ/ than /s/. This latter prediction would be much in line with cross-linguistic research which has shown that laminal sounds are more resistant to coarticulation than are apical sounds ([8], [9]). Proctor et al. [22] observed greater variation in the production of [ʃ] than of [s], which was consistently produced with a more anterior constriction.

Note that until now, we have been considering high-low coarticulation with adjacent vowels. However, coarticulation may also occur on the front-back dimension, and in theory similar principles should apply: if the tongue is more highly constrained for one sibilant fricative than for the other sibilant fricative, coarticulation should be lesser with adjacent vowels on both the high-low dimension, and the front-back dimension.

In this study we explore the question of the relative coarticulatory resistance of /s/ *vs.* /ʃ/ using static mid-sagittal MRI of one female speaker recorded for three languages: Australian English, French, and Croatian. This approach is based on recent work by Badin *et al.* ([2]) and by Peters *et al.* [21], which use MRI and acoustic data respectively to examine coarticulatory patterns across different languages and dialects within the same speaker. The reasoning behind such an approach is that if speaker characteristics remain the same across dialects, language-specific coarticulatory patterns may emerge.

These languages have differing vowel inventories (see further below), and for the purposes of comparison in the current study, we restrict our study to vowels which are broadly similar across the three languages.

In this respect it should be noted that although the same symbol is used to transcribe the sibilant fricatives in the three languages studied here, the phonemic inventory of each language may lead to slightly different realizations of these sounds. In particular, the English phoneme inventory contains the dental fricative /θ/, which although not a sibilant fricative, nevertheless may prevent a more dental production of English /s/. In Croatian, the literary standard contrasts a palatal affricate (written 'ć') with a post-alveolar affricate written 'č' – although the speaker in this study does not make this contrast, it is a prominent feature in descriptions of dialectal variation in Croatian, and may be expected to have an effect on production of /ʃ/, which is classified as a post-alveolar fricative (and written 'š').

2. ARTICULATORY DATA ACQUISITION AND PROCESSING

2.1 Recording setup

Three MRI sessions were performed (IRMaGe MRI facility, Grenoble, France), at 3T (Achieva 3.0T TX, Philips, Best, The Netherlands), using a 16 channel neurovascular coil. In order to minimize head movement, MRI foam cushions were wedged between the speaker's head and the MRI receiver coil. She had to sustain a fixed articulatory configuration during image acquisition. Static single slice (4 mm thickness) mid-sagittal images of the vocal tract were recorded, with an isotropic 1 mm in-plane resolution covering a 256×256 mm^2 field of view. Turbo Spin Echo mode was used, with 85% halfscan factor, no SENSE acceleration, 80° flip angle, shortest TR and TE, and minimum water-fat shift. In the first session, a TSE factor of 28 was used, with 6.3 s per image. In the remaining sessions, a TSE factor of 38 was preferred, with 6.9 s per image but slightly improved image quality.

2.1 Corpus, speaker and protocol

A single female speaker was recorded for this experiment (46 years old at recording time). Her native language is Australian English (AuE) from Melbourne; her heritage language is Croatian (CR), and she is highly fluent in French (FR) [1]. Following [2] for a bilingual study, the speaker was recorded on different days for AuE, FR and Croatian CR. For each session, care was taken that only the language of interest (except for CR) was spoken for the interactions with the operators, in order to keep the speaker in the right language mode. In addition, the speaker always tried to think of real words in the language before producing the particular sequences required.

The corpora were designed to be balanced and representative of the phonemic repertoires of the three languages (see [7], [17] and [11]). The AuE corpus contained 208 items: the 13 vowels /iː e æ a oː ɔ ʉː ʊ ɜː ə ɪ eː aː/ and the 15 consonants /p t k f θ s ʃ m n ŋ ʧ l ɹ h/ embedded in all the 13 symmetric Vowel Consonant Vowel (VCV) contexts. The CR corpus contained 133 items: the 7 vowels /iː eː ɛ oː ɔ uː a/ and 16 consonants /p t k f s ʃ m n ɲ ʦ ʧ j l ʎ ɾ h/ [2] in all 7 vowel contexts. The FR corpus contained 144 items: the 10 oral /i e ɛ a ɔ o u y ø œ/ and 3 nasal /ã ɛ̃ ɔ̃/ vowels, and the 10 consonants /p t k f s ʃ m n ʁ l/ in all 13 vowel contexts. Note that only voiceless counterparts were used for stops and fricatives (voicing being impossible to maintain for the up to 6.9 seconds of MRI acquisition).

For the consonants, the speaker was instructed to repeat the VCV sequences a few times in a natural manner, trying to think of real words, and to freeze the consonant in the last repetition; the scan was initiated as soon as the operator heard that the consonant position was being maintained (*e.g.* /asa asa as:::::::a/, with the scan being initiated at the onset of the very long /s:::::::/). This protocol ensured that the consonant was truly coarticulated with the vowel.

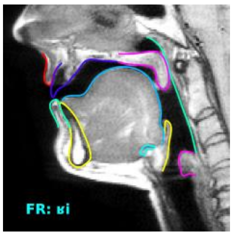
3. ARTICULATORY CONTOURS ANALYSIS

3.1 Semi-automatic contour segmentation

Automatic segmentation of each main speech articulators (jaw, lips, tongue, velum, hyoid, larynx, etc.) from the MRI images was performed according to the method described in [15]. The unwanted head movements of translation in the Head-Feet direction and of rotation around the Left-Right direction were determined and counterbalanced to realign the skull structures and in particular the hard palate, on the same reference, as proposed in [15] on a chosen reference image. The remaining tilt angle between skull and spine was determined and its influence on the shape of all articulators was compensated for. Note that sometimes the speaker's head was not fully aligned with the mid-sagittal plane, which produced parallax errors that could not be compensated for, and that resulted in slight deformations visible in the front region (nose, lips and incisors).

The next stage involved the manual segmentation of 60 images automatically chosen by an optimisation procedure to represent the whole corpus as faithfully as possible. The data were subsequently used to train an MLR-based algorithm that predicts contours from images; these contours serve next as seeds for a modified version of Active Shape Models that refines the results. A post-processing procedure aiming to prevent possible slight overlaps between articulators in contact was applied before all the 485 images were checked and manually corrected if needed. Results are illustrated in Figure 1. Note that all the contours are stored as x/y coordinates in cm units in a coordinate system attached to the reference image, and thus common to all images that have been aligned with this reference image. This allows reliable processing of contours for comparisons or modelling.

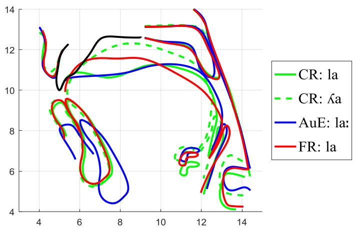
Figure 1: Articulator contours superimposed on a midsagittal image. Different colours represent different articulators, in a clockwise rotation along the vocal tract walls: upper lip, palate, velum, naso-oropharyngeal wall, back larynx, epiglottis, hyoid bone, tongue, jaw, and lower lip.



3.2 Analysis methods

Explicit articulatory modelling as performed by Hoole [13] or Badin *et al.* [2] is out of the scope of this study. However, the superposition of contours for a given set of phonemes can be useful for articulatory comparisons (*e.g.* [22]), as illustrated in Figure 2. This figure shows that unlike the English and Croatian alveolar laterals, the French lateral does not have secondary velarization. It also shows the very high tongue body position for the palatal lateral in Croatian.

Figure 2: Examples of contour superposition of laterals in /a/ context demonstrating different strategies between different languages (AuE /l/, CR /l/ and /ʎ/, FR /l/).



A more comprehensive way to present vocalic coarticulation effects on a consonant is to display the mean of articulations of the same consonant over all vocalic contexts of interest, and the associated dispersion of contour points, by means of dispersion ellipses drawn at ±2 standard deviations around the mean points, as illustrated further in Figure 3.

In order to perform inter-language comparisons of vocalic coarticulation effects on consonants, it is important to choose a set of vowels that can be considered as analogous. A series of comparisons of vowels contours for the three languages has shown that in most cases AuE /e/ is a bit closer to CR /eː/ and to FR /e/ than to AuE /ae/. AuE vowels /e/, /eː/ and /ae/ do not show great differences; pairs /ʉː/ - /ʊ/ and /iː/ - /ɪ/ are fairly close. It has therefore been decided to use AuE /ɪ eː e aː oː ɔ ʊ/, CR /iː eː ɛ a oː ɔ uː/ and FR /i e ɛ a o ɔ u/ as the common set of 7 vowels.

4. RESULTS

This database and this approach allow a great number of analyses and comparisons. The focus of the present study is on the stability of /s/ and /ʃ/.

For instance, Figure 3 shows that the tongue in /s/ in CR is somewhat more resistant to coarticulation than in AuE or FR. By contrast, the tongue in /ʃ/ is more resistant in FR than in AuE or CR. This seems to point to differences between languages, both in terms of language-specific mean articulation and of language-specific coarticulation level.

With regard to the questions posed in the Introduction section, it is clear that the front part of the tongue is very stable in the articulation of /s/, while the back part of the tongue shows greater variability according to vowel context. For /ʃ/, it appears that variability is more evenly distributed across the entire tongue, although there appears to be some language-specific variability (with Australian English having a more stable tongue back/root, while French has a more stable tongue body). It should also be noted that the tongue body is lower for /ʃ/ in Australian English than in the other two languages.

From Figure 4 we see that in CR the lingual /t/ closure of both affricates /ʦ/ and /ʧ/ is more resistant to coarticulation than the single /t/. The comparison between the second column of Figure 3 and the third column of Figure 4 shows that the tongue in /s/, and even more in /ʃ/, is more resistant to coarticulation in affricate contexts that in single fricatives. This seems to indicate that the coproduction of the closure and the fricative parts of affricates induces a stronger resistance to lingual coarticulation.

Figure 5 shows that, except for /u/, the tongue contour in the occlusive part of the affricate /t(ʦ)/ is close to that of the single /t/, and that the contour in the fricative part /s(ʦ)/ is close to that in the single /s/. This has been also observed on the average contours. On the other hand, tongue contours for /t/ and /s/ are slightly different, as expected, except in /u/ context, the laminal part of the tongue being slightly lower for /s/ than for /t/. The picture is different and less systematic for the affricate /ʧ/. The lingual constriction is more alveolar for /ʃ(ʧ)/ than for the post-alveolar /ʃ/, which has also been observed on the average contours. The lingual articulation is also slightly anterior for /t/ than for /t(ʧ)/, as found by Mair *et al.* [19] for one American English speaker.

Figure 3: Mean contours and dispersion ellipses for /s/ and /ʃ/ for AuE, CR and FR in the 7 common vowel contexts (ellipses are displayed for every contour point).

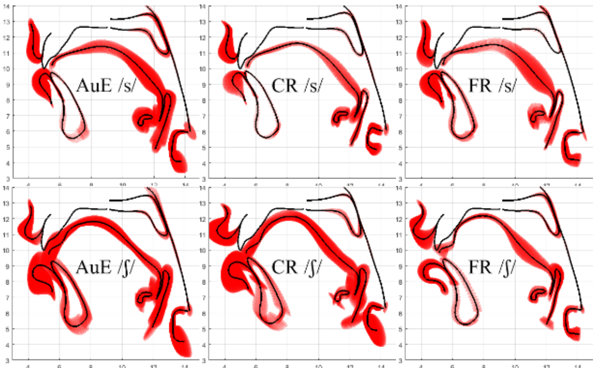


Figure 4: Mean contours and dispersion ellipses in CR for /t/, /t(ʦ)/ and /s(ʦ)/ (top), and for /t/, /t(ʧ)/ and /ʃ(ʧ)/ (bottom) in the 7 common vowels contexts.

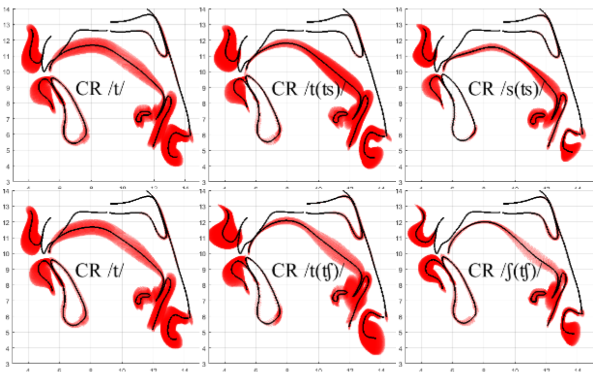
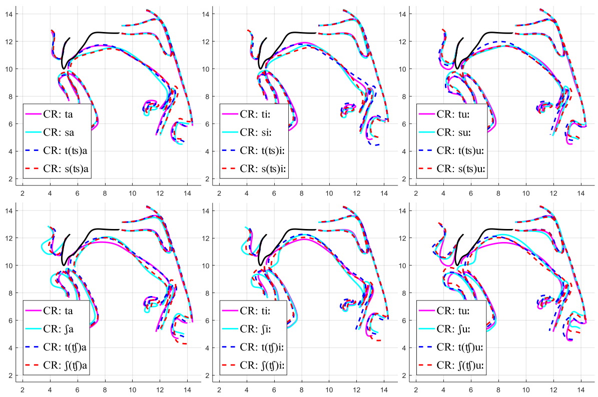


Figure 5: Superposition of CR contours for /t/, /s/, /t(ʦ)/ and /s(ʦ)/ (top), and for /t/, /ʃ/, /t(ʧ)/ and /ʃ(ʧ)/ (bottom), in /a/, /i/ and /u/ contexts.



[1] She is a native speaker of Australian English, and a heritage speaker of Croatian. Despite being born in Australia, she spoke Croatian before English, which she started from pre-school only. She is a fluent L2 speaker of French, which she studied at secondary school and at university, and she has also spent about 14 months working and studying in France.

5. DISCUSSION AND PERSPECTIVES

5.1 Discussion

These results show that, at least for English and Croatian, the front part of the tongue is much more stable for /s/ than for /ʃ/, suggesting that the very narrow constriction limits variability in this articulator. By contrast, results for the back part of the tongue vary according to the language – the back part of the tongue is more variable for /s/ in English and in French, but possibly less variable in Croatian. To what extent these differences are due to language repertoire and daily usage, and to what extent they reflect genuine differences in the languages, *i.e.* not related to the present speaker and her possibly different levels of proficiency, is not clear at this point. Moreover, the results are not necessarily in line with previous results using EPG, ultrasound or EMA, which suggest that /ʃ/ coarticulates less with vowels than does /s/. It is clear that a more careful consideration is needed as to what aspect of the articulation each technique is capturing.

5.2 Perspectives

The same approach will be used to study other coarticulatory and cross-language comparisons. For instance, the constriction in the pharyngeal cavity that would characterize rhotics according to several authors ([10], [5], [12], [26], [25], [4]) can be compared for AuE /ɹ/, CR /ɾ/ and FR /ʁ/ on the same speaker. Similarly, the degree of velarization / darkness of laterals could be compared across vowel context and languages. Another way to analyse the data is to submit them to articulatory modelling, as in [2] for instance. This would allow to compare models made for each language, and to display the phonemes in spaces of articulatory control parameters such as Jaw Height versus Tongue Body that can be useful to separate classes. Finally, we are planning to upload some of these data in an open source repository.

6. ACKNOWLEGMENTS

IRMaGe was partly funded by the French ANR grant "Infrastructure d'avenir en Biologie Santé - ANR-11-INBS-0006". Authors are very much indebted to Mathieu Labrunie for his help in running the semi-automatic segmentation software.

[2] Two recordings were actually made of the affricates: the speaker attempted first to freeze on the closure part of the affricates (/t(ʦ) t(ʧ)/ and then on the fricative part of it (/s(ʦ) ʃ(ʧ)/), in order to observe possible differences. This leads to a total of 18 consonants.

7. REFERENCES

[1] Badin, P., Mawass, K., Castelli, E. 1995. A model of frication noise source based on data from fricative consonants in vowel context. *Proc. 13th ICPhS*, Stockholm, Sweden.

[2] Badin, P., Sawallis, T.R., Crépel, S., Lamalle, L. 2014. Comparison of articulatory strategies for a bilingual speaker: Preliminary data and models. *Proc. 10th ISSP*, Cologne, Germany.

[3] Badin, P., Shadle, C.H., Pham Thi Ngoc, Y., Carter, J.N., Chiu, W., Scully, C., Stromberg, K. 1994. Frication and aspiration noise sources: contribution of experimental data to articulatory synthesis. *Proc. 3rd ICSLP*, Yokohama, Japan.

[4] Blackwood Ximenes, A., Shaw, J.A., Carignan, C. 2017. A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English. *J. Acoust. Soc. Am.* 142, 363-377.

[5] Boyce, S., Espy-Wilson, C. 1998. Coarticulatory stability in American English /r/. *J. Acoust. Soc. Am.* 104, 3741-3753.

[6] Byrd, D. 1994. Articulatory timing in English consonant sequences. *UCLA Working Papers in Phonetics* 86.

[7] Cox, F., Palethorpe, S. 2007. Australian English. *Journal of the International Phonetic Association* 37, 341-350.

[8] Dart, S.N. 1991. Articulatory and acoustic properties of apical and laminal articulations. *UCLA Working Papers in Phonetics* 79.

[9] Dart, S.N. 1998. Comparing French and English coronal consonant articulation. *Journal of Phonetics* 26, 71-94.

[10] Espy-Wilson, C., Boyce, S. 1994. Acoustic differences between ‘‘bunched’’ and ‘‘retroflex’’ variants of American English /r/. *J. Acoust. Soc. Am.* 95, 2823-2823.

[11] Fougeron, C., Smith, C.L. 1993. French. *Journal of the International Phonetic Association* 23, 73-76.

[12] Guenther, F.H., Espy-Wilson, C.Y., Boyce, S.E., Matthies, M.L., Zandipour, M., Perkell, J.S. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *J. Acoust. Soc. Am.* 105, 2854-2865.

[13] Hoole, P. 1999. On the lingual organization of the German vowel system. *J. Acoust. Soc. Am.* 106, 1020-1032.

[14] Keating, P. 1990. The window model of coarticulation: Articulatory evidence. in *Papers in laboratory phonology I: Between the grammar and physics of speech*, Beckman, M.E. and Kingston, J., Eds. Cambridge, UK: Cambridge University Press

[15] Labrunie, M., Badin, P., Voit, D., Joseph, A.A., Frahm, J., Lamalle, L., Vilain, C., Boë, L.-J. 2018. Automatic segmentation of speech articulators from real-time midsagittal MRI based on supervised learning. *Speech Communication* 99, 27-46.

[16] Ladefoged, P., Johnson, K. 2006. *A Course in Phonetics (6th ed.)*: Michael Rosenberg.

[17] Landau, E., Lončarić, M., Horga, D., Škarić, I. 2009. Croatian. *Journal of the International Phonetic Association* 25, 83-86.

[18] Laver, J. 1994. *Principles of phonetics*. Cambridge, U.K.: Cambridge University Press.

[19] Mair, S.J., Scully, C., Shadle, C.H. 1996. Distinctions between [t] and [tʃ] using electropalatography data. *Proc. 4th ICSLP*, Philadelphia, PA, USA.

[20] Narayanan, S.S., Alwan, A.A., Haker, K. 1995. An articulatory study of fricative consonants using magnetic resonance imaging. *J. Acoust. Soc. Am.* 98, 1325-1347.

[21] Peters, J., Heeringa, W.J., Schoormann, H.E. 2017. Cross-linguistic vowel variation in trilingual speakers of Saterland Frisian, Low German, and High German. *J. Acoust. Soc. Am.* 142, 991-1005.

[22] Proctor, M., Shadle, C.H., Iskarous, K. 2006. An MRI study of vocalic context effects and lip rounding in the production of English sibilants. *Proc. 11th AICSST*, Auckland, New Zealand.

[23] Recasens, D., Rodríguez, C. 2017. Lingual articulation and coarticulation for Catalan consonants and vowels: an ultrasound study. *Phonetica* 74, 125-156.

[24] Shadle, C.H. 1985. *The acoustics of fricative consonants*. MIT.

[25] Strange, W., Weber, A., Levy, E.S., Shafiro, V., Hisagi, M., Nishi, K. 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *J. Acoust. Soc. Am.* 122, 1111-1129.

[26] Zhang, Z., Espy-Wilson, C.Y., Boyce, S., Tiede, M.K. 2005. Modeling of the front cavity and sublingual space in American English rhotic sounds. *Proc. ICASSP'05*.