AN MRI STUDY OF SWEDISH FRICATIVES: COARTICULATORY EFFECTS

Olov Engwall[1] & Pierre Badin[2]
[1]Centre for Speech Technology (CTT), KTH, SE-100 44 Stockholm, Sweden.
[2]Institut de la Communication Parlée, UMR 5009, CNRS, INPG-Université Stendahl, Grenoble, France.
`olov@speech.kth.se, badin@icp.inpg.fr`

ABSTRACT

Five Swedish fricatives have been analysed in different vowel contexts using Magnetic Resonance Imaging (MRI). The fricatives were acquired for sustained articulations in VCV context both in the midsagittal plane and in full 3D for one speaker of Swedish. The MR images were used for three-dimensional reconstructions and extraction of the vocal tract outline in the midsagittal plane. This allowed for coarticulatory studies based on the area functions and the articulatory measures determined from the vocal tract contours. Coarticulatory effects have been found for the place of constriction, the protrusion, the jaw and larynx height and the cross-sectional area. Synergetic coupling are also found between the jaw advancing and lip protrusion and between the larynx height and lip protrusion.

1 INTRODUCTION

Coarticulation influences a number of articulatory and acoustic features and is hence of great importance in articulatory modeling and visual speech synthesis.

Ideally, coarticulation should be measured dynamically to capture temporal aspects. No existing method allows however for truly three-dimensional dynamic measurements, and coarticulatory effects on the vocal tract shape and the area function has thus to be studied on static articulations. Coarticulation due to vowel context can nevertheless be analysed even on sustained consonants using VCV sequences. This article presents such a coarticulation study using Magnetic Resonance Imaging (MRI). Earlier studies (e.g. [1]) have used dynamic MRI to study coarticulation, but they have been limited to the midsagittal plane. An exception is Shadle et al. ([2]) that studied coarticulatory effects on three-dimensional vocal tract movements. A large number of repetitions was however required for each phoneme in that study as in general, and static coarticulatory studies is the alternative if averaging over many repetitions is to be avoided.

2 MRI MEASUREMENTS AND RECONSTRUCTION

2.1 Data acquisition

The MRI data were collected at the Centre Hospitalier Régional Universitaire de Grenoble, France, using a Philips Gyroscan T10-NT that generates a static longitudinal magnetic field of 1.0 Tesla. The set up and protocols were as defined by Badin et al. [3], resulting in images of 256x256 pixels with a final resolution of 1 mm/pixel.

The subject was a 27-year-old male native speaker of Swedish with no record of speech or voice disorders and a perfect dentition. A database of artificially sustained vowels and consonants in VCV-context in one 3D set and one midsagittal set was collected during one three hour session, without subject repositioning.

The 3D set contained 13 vowels [ɑː, eː, æː, iː, yː, uː, ɯː, oː, œː, øː, a, θ, ɔ] and 10 consonants [p, t, k, l, r, f, s, ɧ, ʂ, ɕ] in vowel context provided by the point vowels [a, ɪ, u]. The midsagittal set consisted of 17 vowels [ɑː, eː, ɛː, æː, iː, yː, uː, ɯː, oː, œː, øː, a, ɪ, ʊ, ʏ, θ, ɔ] and 17 consonants [p, t, k, l, r, f, s, ɧ, ʂ, ɕ, ʝ, m, n, ŋ, ʈ, ɭ, ɳ] in four vowel contexts [a, ɪ, ɔ, u]. This article focuses on the five Swedish fricatives [s, f, ɕ, ʂ, ɧ] in VCV-context; refer to [6] for details on other parts of the corpus.

For each configuration in the 3D set, three 18-slice series of parallel slices were gathered. The series were oriented as first proposed by Demolin et al. [5] to give slices as orthogonal as possible to the vocal tract midline (see Fig. 1). The slices were 3.6 mm thick with a sampling interval of 4 mm, but with a partial overlap between slices from different series, to ensure that the entire vocal tract was covered. Each configuration was also imaged in the midsagittal plane [IMAGE 64.jpg shows /aɕa/, /iʂi/ and /uɧu/], with a seperate scan but during the same session.

The acquisition time was 11 seconds for the midsagittal set and 43 s for the 3D set. The subject made the initial VC-transition before the MR scan, then sustained the fricative steadily during the entire acquisition, breathing out slowly and finally produced the CV-transition after the scan. The speech signal preceding and following the scan was recorded on a DAT using a microphone placed inside the MRI tunnel.

In addition to the phonemes in the corpus, MR images were taken of the reference position with upper and lower incisors touching and aligned and of the subject's dental cast for use in the reconstruction processes.

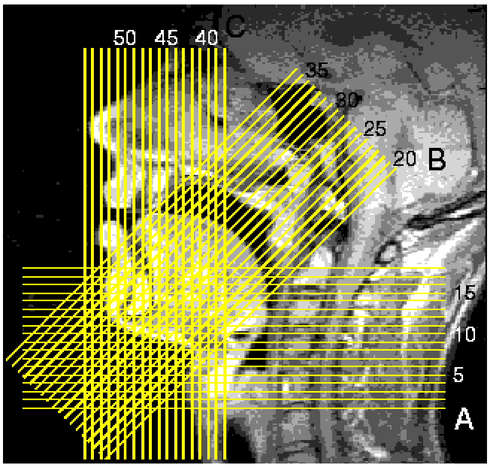
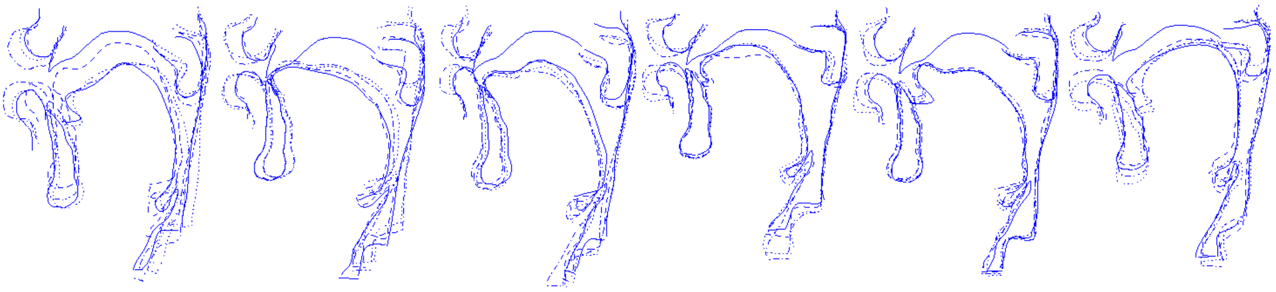


Figure 1. The 3D data acquisition grid, consisting of three series of 18 parallel slices: A) horizontal (axial) slices from below the glottis to the top of the pharynx. B) Slices inclined at 45° in the bent part of the vocal tract. C) Vertical (coronal) slices from the soft palate to in front of the lips. The grid is superimposed on the midsagittal image of /ɪsɪ/.



a) Vowels **b)** /s/ **c)** /f/ **d)** /ɕ/ **e)** /ʂ/ **f)** /ɧ/

Figure 2. Midsagittal contours of isolated vowels and of fricatives in vowel contexts /a/ (solid line), /ɪ/ (dashed), /o/ (dash-dotted) and /u/ (dotted). The main coarticulatory effects are the lip protrusion and the larynx height, but some context influence can also be seen on the tongue contour and the jaw height and advancing.

2.2 Image processing

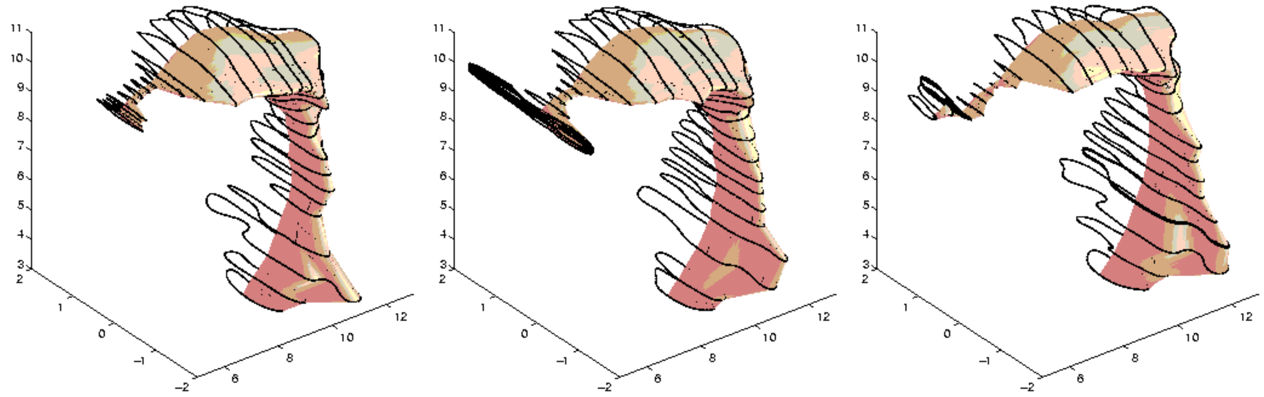
2.2.1 The midsagittal set The midsagittal contour for each configuration (see Fig. 2) was determined using a version of BtoC ([6]) adapted for MRI. The image was thresholded in binary mode to automatically detect the air-tissue boundary. Using this boundary as a guide, the 10 sub-contours of the upper lip, palate, velum, pharynx, back larynx, lower lip, jaw, tongue, front larynx and glottis were extracted manually using a Matlab-software developed at ICP. These sub-contours provided a number of articulatory parameters that have been studied and used in a linear midsagittal articulatory model ([4]). The results from the midsagittal parameter evaluation will be used in the KTH 3D vocal tract model ([7]).

2.2.2 The 3D set The contour of the vocal tract was extracted from each image using edge detection in thresholded images. Each contour was manually checked, and when needed, corrected by moving, adding or removing the Bezier control points of the contour. The classical MRI airway segmentation problems were handled as follows: 1) the piriform sinuses and the sublingual cavities were regarded in an image only when they were connected with the main air passage; 2) the volume occupied by the teeth was removed in the reconstruction process by fitting the vocal tract shape and the dental cast to the common reference system; 3) when the epiglottis was present, the outline of the vocal tract was delimited by the closed contour consisting of the epiglottis and the pharynx wall; 4) the vocal tract contour extraction at the lips were pursued as long as the lips determined a closed contour.

2.3 3D reconstruction

The vocal tract shape was reconstructed from the contours using a semi-polar partially dynamic grid of 37 gridlines as defined by Beautemps et al. [8]. It consists of a fixed central polar grid connected to two linear grids of variable length; one attached to the tongue tip and the other to the glottis. The grid hence follows movements of the larynx, the tongue tip and the lip protrusion dynamically. The hard palate of every configuration was aligned with a fixed common palate shape that serves as reference for the grid in the 3D reconstruction. Each planar contour was finally re-sampled with a fixed number of evenly distributed points. The contours were then rendered as shown in Fig. 3, allowing for comparisons between different vowel contexts.

Qualitative comparisons of vocal tract shape can be made between different fricatives and vowel contexts (refer to [4] for vocal tract shapes of other Swedish fricatives than /s/). As was evidenced also by a study of English fricatives [9], the difference in the vocal tract shape is not limited to the place and amount of constriction. The areas of minimum supraglottal constriction are smaller than for the English subjects in [9], but this cannot be attributed to language characteristics based on a single subject's data. Oral as well as pharyngeal volumes differ sometimes greatly between fricatives and vowel contexts, illustrating the value of 3D measurements and modeling.



a) /asa/ **b)** /ɪsɪ/ **c)** /usu/

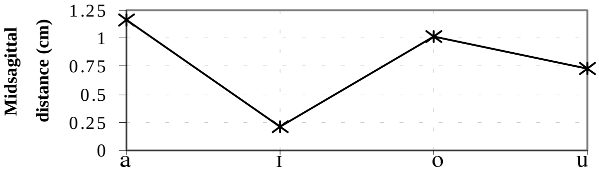
Figure 3. The vocal tract shape for /s/ in three different vowel contexts. Coarticulation influences the place and the length of the constriction, the lip protrusion and mouth opening, but also oral and pharyngeal volumes.

3 COARTICULATORY EFFECTS

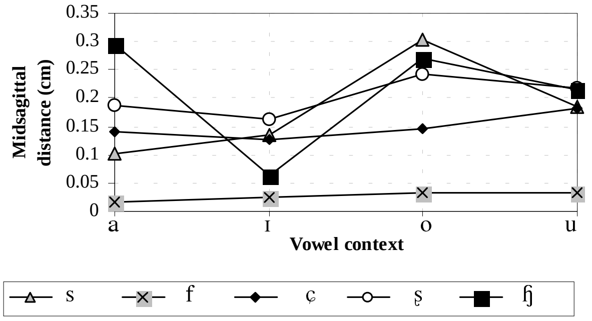
3.1 The midsagittal set

The articulatory measures of the jaw, tongue, lips and larynx positions or shapes were studied with respect to coarticulation.

3.1.1 Constriction The degree of constriction, i.e. the midsagittal distance at the narrowest air passage, is to some extent influenced by coarticulation. The openness of the surrounding vowel (Fig. 4a) determines the degree of constriction for the fricative, as shown by Fig. 4b). The constriction is narrower in /ɪ/ context than when the fricative is surrounded by open, rounded vowels. The range in midsagittal distance is 0.2-0.25 cm for /s/ and /ɧ/, whereas the range is smaller, but still observable for /ʂ/. For /ç/, the coarticulatory influence is rather on the place of constriction, the range being 0.5 cm from the front-most (/ɪ/) to the back-most (/u/) vowel context. The context also influences the place of constriction for /s/ and /ɧ/. For the alveolar fricative /s/, it is advanced in rounded vowel context, due to jaw protrusion. For the velar fricative /ɧ/ it is retracted for /o/ and /u/, following the tongue backing. As expected, Fig. 4 shows no coarticulatory influence for /f/.



a) Vowels



b) Fricatives

Figure 4. Midsagittal distance at the constriction for **a)** vowels and **b)** fricatives as a function of vowel context.

3.1.2 The Jaw The jaw height (Fig. 5) is relatively constant with respect to vowel context for the four front fricatives, indicating that a rather closed jaw is required to achieve the constriction. The velar fricative /ɧ/ allows for more coarticulation, since its place of constriction is less influenced by the jaw opening, and the jaw height consequently becomes much larger in open vowel context (more than 1 cm larger in /o/ than /ɪ/ context), as indicated also by Fig. 2f.

The jaw advancing (Fig. 5) is influenced by coarticulation in all contexts and two interesting phenomena are present. 1) The jaw moves in synergy with the lips to achieve lip protrusion in rounded vowels (a difference of 0.3-0.4 cm comparing /ɪsɪ/ to /oso/ or /ɪɧɪ/ to /uɧu/). 2) The coarticulation on /f/ is the opposite; the jaw is retracted in rounded vowel context in order to achieve tongue backing.

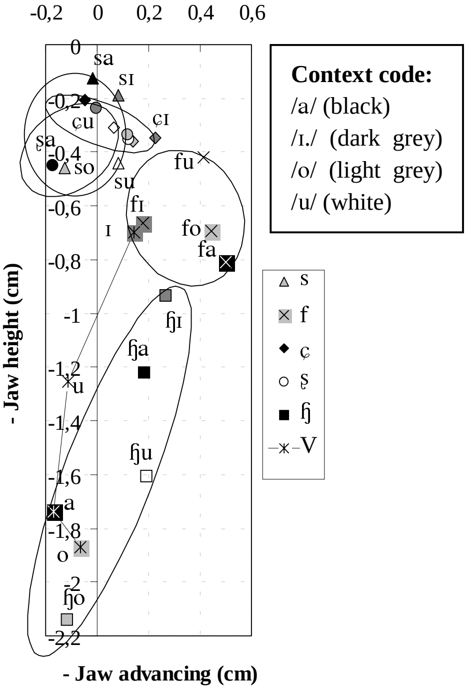


Figure 5. The jaw position defined by the horizontal and vertical coordinates of the lower incisor.

3.1.3 Vocal tract length: lip protrusion & larynx height The vocal tract is lengthened at both ends simultaneously in rounded vowel context (Fig. 6, solid lines), in agreement with the findings of e.g. Hoole & Kroos ([10]). The most important part of the lengthening is due to the coarticulatory effect on the lip protrusion (Fig. 6, unconnected series), with 0.6-1 cm larger protrusion in /o/ and /u/ context than for /ɪ/, except for /ofo/, that is constrained by the place of constriction. The coarticulatory contrast is increased by the larynx height (the larynx is 0.1-1.2 cm lower in /u/ than in /ɪ/ context), but compared to the lip protrusion it is influenced less by the vowel and more by the fricative. Note that the vocal tract lengthening due to larynx lowering in Fig. 6 is measured relative that of /ɪ/, having the highest larynx position.

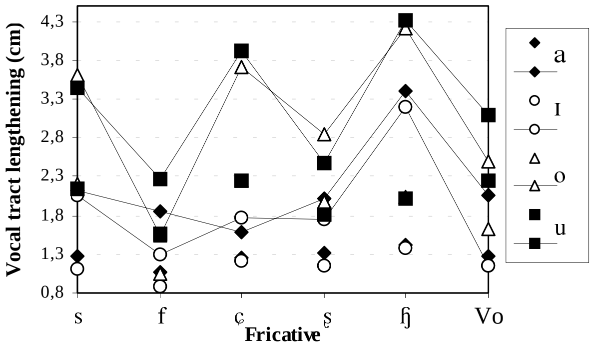


Figure 6. Vocal tract lengthening (*solid lines*), defined as the lip protrusion (*unconnected*) plus the larynx lowering relative the larynx position for /ɪ/. The lip protrusion is measured from the upper incisor to the front most point on the upper lip.

3.2 The 3D set

Generally, the vocal tract shape (Fig. 3) and area function (Fig. 7) differs most for /u/ compared to the other two contexts. The coarticulatory differences are largest in the oral cavity, but there are some differences in the pharynx as well. The coarticulation on the pharyngeal volumes is however uncertain, as the pharynx passage was found to be decreased due to the fact that the subject was whispering. Examples of the coarticulation in Figs. 3 and 7 are the decreased oral volume for /usu/, the broader mouth opening for /ɪsɪ/ and the very large front volume for /uɧu/, created by the retraction of the tongue body (Fig. 2f).

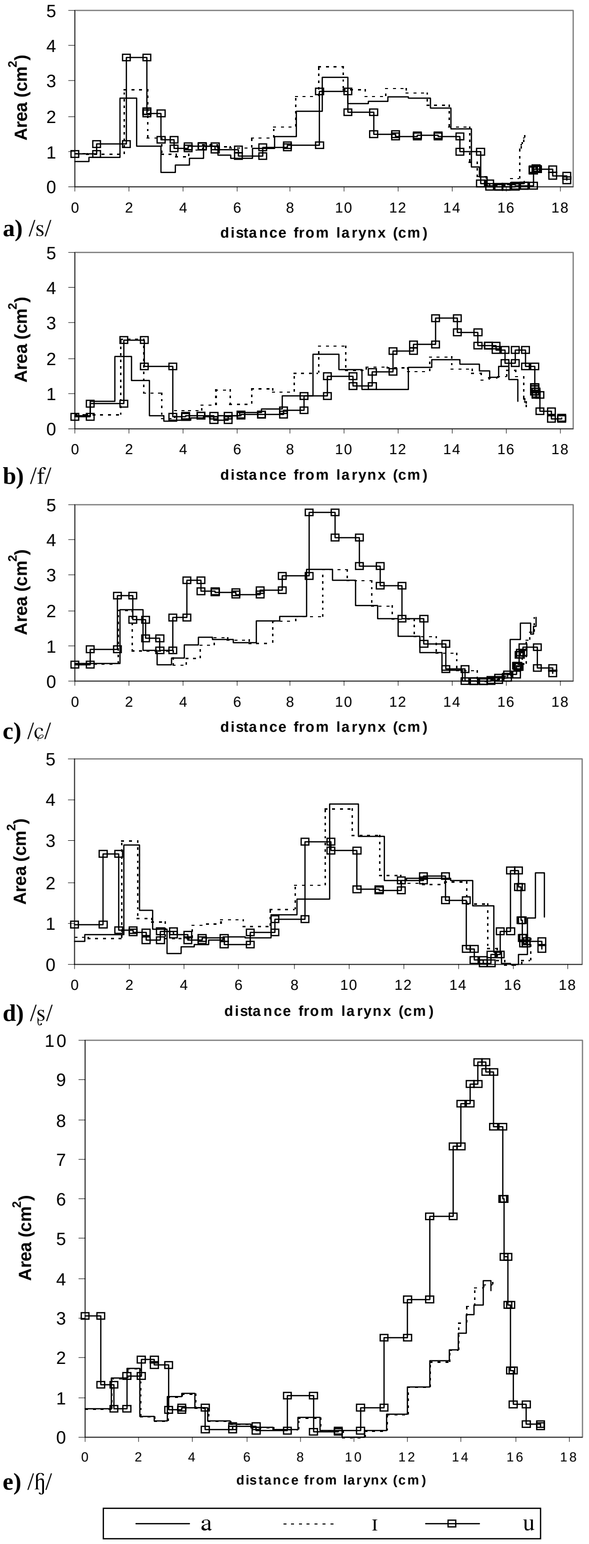


Figure 7. Coarticulatory effects on the area functions of the fricatives.

4 CONCLUSION

This MRI study on Swedish fricatives in symmetrical VCV context shows that the coarticulatory influence can be evidenced on sustained phonemes at the level of both 3D vocal tract shape and midsagittal contours. This work, in association with studies of temporal aspects of coarticulation, using e.g. articulography and EPG, will provide important data on coarticulation in 3D for the KTH 3D vocal tract model [7].

5 ACKNOWLEDGEMENT

This work is the result of a collaboration between the Centre for Speech Technology in Stockholm (CTT, sponsored by KTH, Nutek and Swedish industry), and the Institut de la Communication Parlée in Grenoble (ICP).

The MRI data acquisition was realised at the Grenoble University Hospital (CHRUG) by Christoph Segebarth, INSERM Unit U438. The French ARASSH (Association Rhône-Alpes pour les Sciences Humaines et Sociales), Telefonaktiebolaget LM Ericssons stiftelse för främjande av elektroteknisk forskning, the Swedish Institute and Ragnar & Astrid Signeuls fond, KTH, have given financial support.

6 REFERENCES

[1] Demolin D, George M, Lecuit V, Metens T, Soquet A, Raeymaekers H (1997). Coarticulation and articulatory compensations studied by dynamic MRI, *Proc Eurospeech97*, 43-46.

[2] Shadle C, Mohammad M, Jackson P, Carter J (1999). Multi-planar dynamic magnetic resonance imaging: New tools for speech research. *Proc ICPhS99*, 623-626.

[3] Badin P, Bailly G, Raybaudi M, Segebarth C (1998). "A three-dimensional linear articulatory model based on MRI data", *Proc 3rd ESCA/COCOSDA Int Workshop on Speech Synthesis*, pp. 249-254.

[4] Engwall O, Badin P (1999). Collecting and analysing a three-dimensional MRI corpus of Swedish sounds, KTH STL-QPSR 3-4/99, pp. 11-38.

[5] Demolin D, Metens T, Soquet A (1996). Three-dimensional measurements of the vocal tract by MRI, *Proc ICSLP96*, 1:272-275.

[6] Maeda S, Zerling JP, Simon P, Bothorel A, Wioland F (1993). Software for the digitalisation of labial and X-ray film data. In Speech Maps, Year 1 Report, Vol. III.

[7] Engwall O (1999). "Vocal tract modeling in 3D", KTH *TMH-QPSR 1-2/1999*: 31-38.

[8] Beautemps D, Badin P, Bailly G (in revision). Degrees of freedom in speech production: analysis of cineradio- and labio-films data for a reference subject, and articulatory modeling, Submitted to JASA.

[9] Narayanan S, Alwan A, Haker K (1995). An articulatory study of fricative consonants using magnetic resonance imaging, *JASA*, 98: 1325-1347.

[10] Hoole P, Kroos C (1998). Control of larynx height in vowel production, *Proc ICSLP'98* 2: 531-534.