ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

ARE STATIC MRI MEASUREMENTS REPRESENTATIVE OF DYNAMIC SPEECH? RESULTS FROM A COMPARATIVE STUDY USING MRI, EPG AND EMA

Olov Engwall

olov@speech.kth.se
Centre for Speech Technology (CTT), KTH, Drottning Kristinas väg 31, SE-100 44 Stockholm, Sweden

ABSTRACT

As the acquisition times of static MRI have diminished, the focus of MRI studies has shifted from isolated phonemes to e.g. vowel-consonant sequences. The articulations are hence somewhat closer to running speech and measurements of contextual influence can be made. The acquisition in the vast majority of the MRI studies still requires artificially sustained articulations, however, and it is hence not evident that the measurements are representative of the subject's normal speech. In order to assess the influence of the artificial sustaining, results from two coarticulatory studies, using static MRI and EMA-EPG, respectively, are compared.

The differences between the two studies are substantial concerning jaw position, lip protrusion and tongue contours. However, rather than showing non-representative articulations, the articulations in the MRI data are judged to represent a case of hyperarticulated speech, and should still be valid if considered as that.

1. INTRODUCTION

In the advent of Magnetic Resonance Imaging of speech production, acquisition times were very long indeed, amounting to up to several minutes ([1]) and MRI studies focused on sustained and isolated vowels (e.g. [1]) or consonants (e.g. [2]). As the acquisition time has diminished, it has become possible to focus not merely on isolated articulations, but on phoneme sequences as well.

With the exception of a few studies ([3]-[5]) employing different versions of newly developed dynamic MRI techniques, the majority of MRI studies still employ static MRI, as this is necessary for full three-dimensional imaging and the image quality in static MRI is far superior that of dynamic MRI.

The question is however if the static MR Images can be considered as good examples of the subject's normal speech, as MRI has a number of possible error sources, the most important being the supine position and the long acquisition time.

Tiede ([6]) recently conducted an X-ray microbeam study of sitting and supine position and concluded that "area functions derived from supine MRI data are not invalid in principle" (p. 28). The effect of the artificially sustaining of the articulations rest however to be taken into account. This study is an evaluation of to what extent static MRI measurements can be considered as representative of dynamical aspects, such as coarticulation, of speech production.

Engwall & Badin ([7]) recently showed that contextual influence of surrounding vowels can be evidenced in Swedish fricatives measured with static MRI. To assess to what extent the coarticulatory effects found in [7] give a true picture of real-time coarticulation, the same fricative corpus was collected using simultaneous electromagnetic articulography (EMA) and electropalatography (EPG).

2. DATA ACQUISITION & PROCESSING

2.1 The Subject and Corpus

The subject, a 27 year-old male native speaker of Swedish, produced the five fricatives /s, f, ʂ, ɕ, ʃ/ in VCV context with V= /a, ɪ, ʊ/. /ʂ, ɕ, ʃ/ are the variants of /ʃ/ that the subject use, in a manner representative of mid-Swedish speakers. The articulations were acquired once using MRI and five times using EMA-EPG, in random order between each reading and without any carrier phrase.

2.1 MRI data

Three-dimensional and midsagittal MRI data was collected at the Centre Hospitalier Régional Universitaire de Grenoble, France. The 3D set consisted of three 18-slice series of parallel slices (cf. Fig. 1): a coronal stack, an oblique stack tilted at 45° and an axial stack. The images were collected at 4 mm inter-slice interval between centres, with a final resolution of 1 mm/pixel. A midsagittal image (cf. Fig. 1) was also collected of each articulation, with a seperate scan but during the same session. Details on the MRI acquisition can be found in [8].

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.

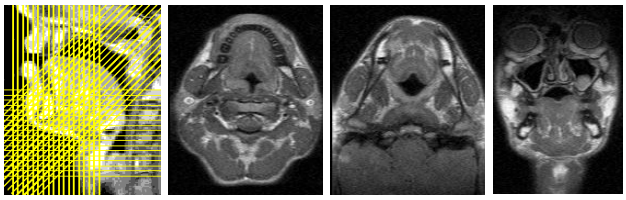

Figure 1. The 3D data acquisition grid on the midsagittal image of /ɪsɪ/ and images from the axial, oblique and coronal stack.

2.1 EMA & EPG data

The EMA data was collected with the Movetrack system ([9]) at the Department of Linguistics, University of Stockholm. Six receivers were used in this study: one on the upper and lower incisor respectively, one on the upper lip and three on the tongue (at 1.1 cm, 3.6 cm and 5.5 cm from the tongue tip). The receiver on the upper incisor served as reference to minimise variations due to head movements, the one on the lower teeth measured jaw motion in the midsagittal plane and the upper lip coil the protrusion. The EPG data was collected with a Reading 62 electrode system. A more detailed survey of the EMA-EPG acquisition will be presented in [10].

The EMA data will be put in relation with the midsagittal MRI data of [7], whereas the EPG measurements will be used to assess the 3D MRI data.

1. ARTICULATORY MEASURES

3.1 Jaw Position

The jaw position, measured as the vertical (JawHei) and horizontal (JawAdv) displacement of the lower incisor relative the lower edge of the upper incisor is shown in Fig. 2 (ellipses for the EMA data have grouping purposes only).

Several differences between the EMA and MRI data are important. /ɧ/ is significantly more open in all contexts in the static condition. /ɪɧɪ, ɪçɪ, ɪʂɪ, ɪsɪ/ are more retracted and /ɪfɪ/ more advanced. /ɪʂɪ, ɪçɪ/ are less open. /ʊsʊ/ is more open and /ʊçʊ/ is more retracted, whereas /aʂa/ is both more open and advanced. The difference between contexts and to some extent between fricatives is over all larger in the MRI data, with larger distinctions in jaw height and advance.

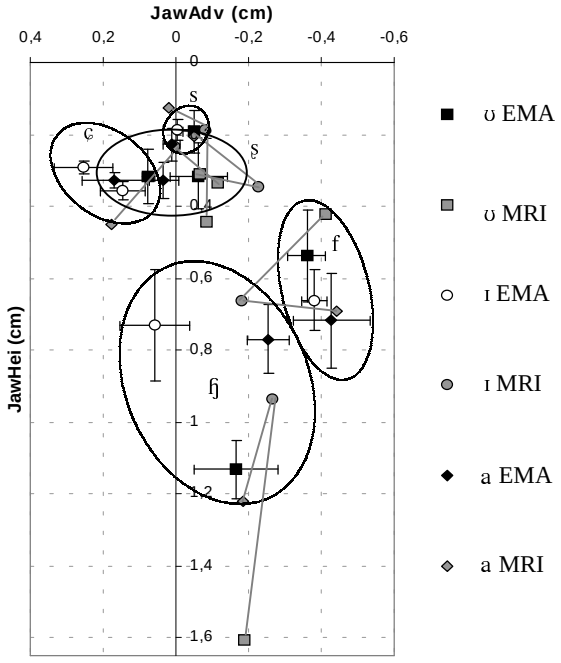

Figure 2. Jaw position measured with EMA and MRI.

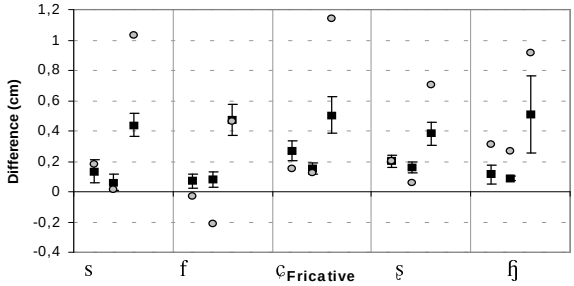

Figure 3. Lip protrusion measured with EMA (black) and MRI (grey). Vowel context from left to right /a, ɪ, ʊ/.

3.2 Lip Protrusion

The protrusion of the upper lip (cf. Fig. 3) shows the same pattern in the EMA and MRI data. The static lip protrusion is however more extreme, with larger protrusion in /ʊ/ context and for /ɧ/, and smaller in /ɪ/ context (except for /ɪɧɪ/, where the protrusion is enlarged due to the fricative).

3.3 Linguopalatal distance

The openness at the second tongue coil was investigated for contextual differences. As the palatal outline in the EMA data was only estimated it was considered better to compare MRI and EMA data with respect to the increase in openness in /a, ʊ/ relative /ɪ/ context for each data set, rather than the absolute values of linguopalatal distances.

Fig. 4 hence shows the difference in linguopalatal distances within each data set. The relative openness for /a, ʊ/ is larger in the EMA measurements, as the fricatives are more constricted at the tongue body in /ɪ/ context compared to the MRI measurements (cf. Fig. 5).

The static MRI tongue positions are more neutral and less coarticulated, except for /ʊsʊ/, /aça, ʊçʊ/ and /ʊfʊ/. The tongue body is raised in the first case and lowered in /ɪ/ context in the second (cf. Fig. 5) in the static condition, thus showing a coarticulatory influence that is the opposite of that in the dynamical measurements. For /ʊfʊ/, the coarticulation is larger in the static condition, as the entire tongue was lowered in the labiodental.

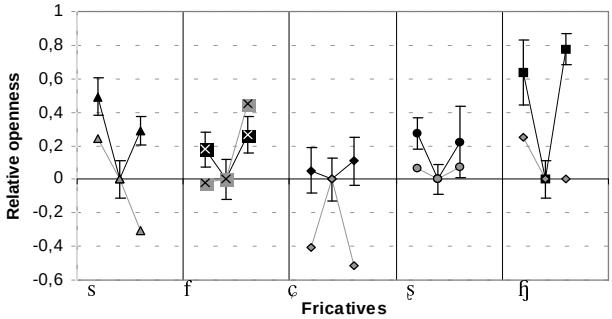

Figure 4. The openness at the second tongue coil in /a, ʊ/ context relative /ɪ/ context for EMA (black) and MRI (grey). Vowel context from left to right: /a, ɪ, ʊ/.

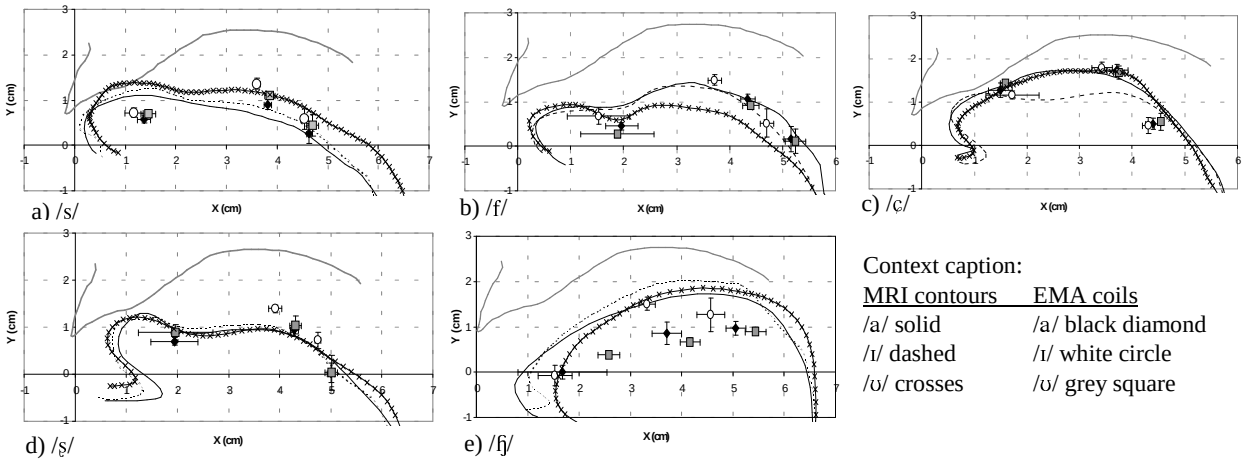

Figure 5. Midsagittal tongue and palatal contours from MRI in relation to mean EMA coil positions in the fricative.

3.4 Tongue Contour

Two contrasting effects appear regarding the tongue contours in Fig. 5; whereas the tongue shape is more neutral in static conditions with respect to coarticulation, it is more extreme with respect to the fricative. The tongue body is more retracted and raised for /ʃ/, it is lowered for /ʂ/, contrasting more with the raised tongue tip in the retroflex and the tongue blade is pressed more against the alveolar ridge in /s/.

3.5 Constriction Place and Width

The centre of gravity (COG) of the linguopalatal contact pattern, calculated as $COG=(8*R_1+7*R_2+\ldots+2*R_7+R_8)/(R_1+\ldots+R_8)$ ([11]), is shown in Fig. 6. The statistically significant advancing of /ɪʂɪ/ vs. /aʂa, ʊʂʊ/ and /ɪʃɪ/ vs. /ʊʃʊ/ (t=-2.7346, 2.9443, 2.6214 respectively, DF=8, $p<0.05$) in the COG index is reflected also in the MRI data (cf. [7] for details). The MRI area functions of /ʃ/ and /ʂ/ show an advancing of the constriction, and the midsagittal contours have a more frontal and narrower constriction for /ɪçɪ, ɪʂɪ, ɪʃɪ/, in accordance with the EPG data.

Fig. 7 shows the frequency of activation of each row, defined as the sum of electrode contacts in that row divided by the product of the number of repetitions and the number of columns in the row (6 for the first row and 8 for the remaining seven). The contrast of linguopalatal contacts in Fig. 7 can be put in relation with contrasts in the area functions, as summarised by Table 1. The contrasts between /ɪ/ and /ʊ/ are present in both data sets, whereas the contrast between /ɪ/ and /a/ found in the EPG data is found in the area functions only for /s/.

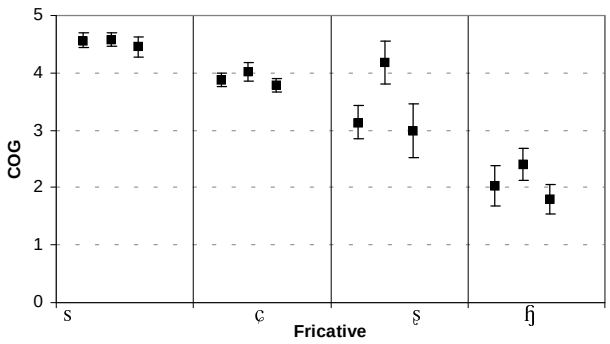

Figure 6. The centre of gravity index calculated from the EPG data. Vowel context from left to right: /a, ɪ, ʊ/.

The coarticulatory influence on the tongue was less in the static conditions, with a more neutral position in /ɪ/ context. Note that the distinction in openness between /ɪ/ and /ʊ/ found in both data sets is only partially due to the active positioning of the tongue, as the jaw height is an important factor in the contrast.

Table 1. Correspondence of contrasts found in the EPG and MRI data. ✓ indicates that the distinction is found in the MRI data as well and (*) that the distinction is found between /ɪ/ and /ʊ/ only, adv=advanced, ±cons=more or less constricted.

	Front (row 1-3)		Middle (R 4-5)		Back (R 6-8)	
	EPG	MRI	EPG	MRI	EPG	MRI
s			I +cons	✓	U +cons	✓
ɕ	I +cons U –cons	I +adv -				
ʂ	I +cons I +adv	✓ (*) ✓ (*)	I –cons	✓(*)		
ʃ					I +cons U –cons	✓ (*) ✓

4. CONCLUSIONS

The static MRI measurements gave indications of coarticulation that were also found in the dynamical EMA and EPG measurements. The MRI data should nevertheless be considered with caution in relation to running speech, as the articulations were both more extreme and the coarticulatory influence on the tongue more limited. The main divergences between the static and dynamic data were significantly more pronounced lip protrusion and jaw height, whereas the tongue position became more neutral when it was constrained to a static position.

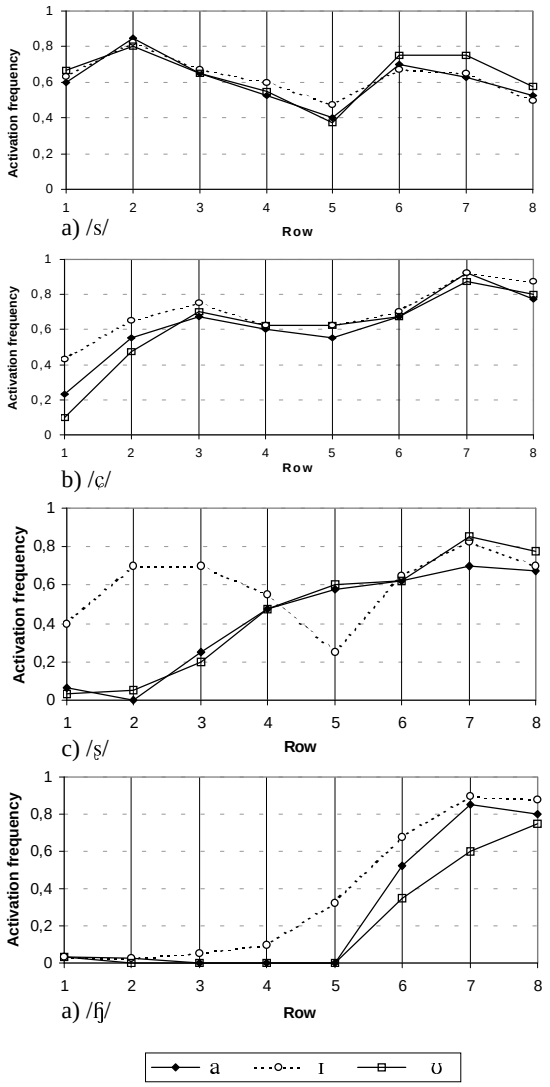

Figure 7. Frequency of electrode activation as a function of row and vowel context.

The above results really boil down to accord with the hypothesis of hyper- and hypo-speech ([12]), the articulations in the MRI data being clearly hyperarticulated, explaining the observed patterns in protrusion, jaw position and tongue articulation. The hyperarticulation in the artificially sustained articulations is a quite natural consequence of the subject aiming to produce as clear examples as possible of each articulation, thus enlarging the important distinctions and reducing coarticulatory effects at the tongue contour.

The conclusion on the divergences between the MRI and EMA-EPG data sets is hence not that the static MRI data is non-representative of running speech, but rather that it is a case of very hyperarticulated speech (note that the EMA-EPG data in itself is carefully articulated 'lab speech'). MRI data used in an articulatory model should hence be considered as articulatory targets and a hyper/hypoarticulation control theory is needed to determine to what extent the targets are reached, in order to produce a running-speech-like output from the model.

5. ACKNOWLEDGEMENT

This work is supported by the Centre for Speech Technology (CTT) at KTH. The MR image acquisition was done by Christoph Segerbarth, Centre Hospitalier Regional Universitaire, Grenoble, with the assistance of Pierre Badin, ICP, INPG, Grenoble. Elisabet Eir Cortes, Peter Branderud and Hassan Djamshidpey of the Department of Linguistics, Stockholm University assisted with the EMA acquisition. Robert Espesser of the Laboratoire Parole et Language, Aix-en-Provence, provided scripts for transferring EPG data to raw matrix format.

6. REFERENCES

[1] Baer, T., Gore, J.C., Gracco, L.W. and Nye, P.W. "Analysis of the vocal tract shape and dimensions using Magnetic Resonance Imaging: Vowels", *JASA*, 90: 799-828, 1991.

[2] Narayanan, S., Alwan, A. and Haker, K. "An articulatory study of fricative consonants using Magnetic Resonance Imaging", *JASA*, 98: 1325-1347, 1995.

[3] Shadle, C.H., Mohammad, M., Carter, J.N. and Jackson, P.J.B. "Multi-planar dynamic magnetic resonance imaging: new tools for speech research", *Proc ICPhS'99*: 623-626, 1999.

[4] Demolin, D., Metens, T. and Soquet, A. "Real time MRI and articulatory coordinations in vowels", *Proc 5th Speech Prod. Seminar*, 93-96, 2000.

[5] Stone, M., Dick, D., Douglas, A., Davis, E. and Ozturk, C. "Modelling the internal tongue using principal strains", *Proc 5th Speech Prod. Seminar*, 133-136, 2000.

[6] Tiede, M. "Contrasts in speech articulation observed in sitting and supine conditions". *Proc 5th Speech Production Seminar*, 25-28, 2000.

[7] Engwall, O. and Badin, P. "An MRI study of Swedish fricatives: coarticulatory effects", *Proc 5th Speech Prod. Seminar*, 297-300, 2000.

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

[9] Branderud, P. "Movetrack – a movement tracking system", *Proc of the French-Swedish Symposium on Speech*, Grenoble, 113-122, 1985.

[10] Engwall, O. Dynamical aspects of coarticulation in Swedish fricatives: a combined EMA & EPG study, to appear in *KTH STL-QPSR*, 2000.

[11] Nguyen, N. "A MATLAB toolbox for the analysis of articulatory data in the production of speech". Accepted for publication in *Behaviour Research Methods, Instruments and Computers*, 2000.

[12] Lindblom, B. "Economy of speech gestures". In MacNeilage P.F. (ed.) *The Production of Speech*, Springer, New York, 217-245, 1983.