

A Possible Optimisation Procedure for US and MRI Tongue Contours

Réka Trencsényi^a, László Czap^b

^aUniversity of Debrecen,
Department of Electrical and Electronic Engineering,
Debrecen, Hungary
trencsenyi.reka@science.unideb.hu

^bUniversity of Miskolc,
Institute of Automation and Infocommunication,
Miskolc, Hungary
czap@uni-miskolc.hu

*Proceedings of the 1st Conference on Information Technology and Data Science
Debrecen, Hungary, November 6–8, 2020
published at <http://ceur-ws.org>*

Abstract

The topic of the article is speech research. The main instruments of the study are US and MRI records of human beings, which were made during speech. In the dynamic records, primarily, the motion of the tongue is analysed and followed by automatic tongue contour tracking algorithms. The tongue contours are used to elaborate geometric transformations between US and MRI frames, which are the starting points of the optimisation of the matching of US and MRI tongue contours belonging to the same speech sound. As a result, the radial US geometry and the rectangular MRI geometry are embedded into each other in a biunique way.

Keywords: Data visualisation, computational linguistics, speech research, dynamic US and MRI records, automatic tongue contour tracking

AMS Subject Classification: 68N19, 68P05, 68T50, 68U10, 68W99

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Introduction

One of the fundamental tools of the study of speech production is the analysis of dynamic records of human speakers, made by ultrasound (US) [4, 9] and magnetic resonance imaging (MRI) [7] techniques. Investigating and processing these two-dimensional records created in the so-called sagittal plane resulting in a side view of the human body, relevant qualitative and quantitative information can be gained about the main features of articulation. Qualitative statements mainly refer to the relative position of the tongue and palate in the case of different speech sounds and sound transitions, while quantitative descriptions focus on the recognition and connection of the geometric parameters which have high importance in the understanding of the relationships between the acoustic and articulatory characteristics of speech. Quantitative analyses can be performed in several ways with a wide variety [3, 5, 6]. The starting points of the investigations of our present study are tongue contours fitted to the frames of US [10] and MRI [8] records by automatic algorithms [11]. The used US and MRI sources differ from each other in many details, such as the gender and nationality of the speakers, the geometry, resolution, and scale of the images, and the visually evaluable anatomic segments of the vocal tract. The aim of our research work is to match the US and MRI sources by elaborating, applying, and optimising the proper geometric transformations between the US and MRI tongue contours in a biunique way.

In the literature, several publications can be found that deal with the fusion of information arising from sources produced by different imaging techniques. The demand for automatic tongue contour tracking algorithms emerged even in the previous decade [2], confirming the necessity of fully automated procedures like our algorithm that does not require any manual actions, as it is based on dynamic programming [11]. Another benefit of our present results that we have been working with dynamic US and MRI records instead of static frames belonging to sustained sounds exclusively [1]. The US videos were made by the Micro system of the MTA-ELTE Lendület Lingual Articulation Research Group of the Hungarian Academy of Sciences, and the MRI videos made by fast MRI were downloaded from the website of the University of Southern California. Also, such studies appeared that aim to perform transformations between coordinate systems connected to US and MRI frames relying on the optimisation of distances measured between special points of the human head. [1]. In comparison with [1], it must be emphasised that our transformations relate directly to the tongue contours, and the transformation is carried out in one step without any intermediate coordinate system, so starting from the US contour, one gets to the MRI contour immediately. Furthermore, the optimisation procedure minimises the global distance between the linked US and MRI tongue contours in the case of more than one sounds simultaneously.

2. Transformation and Optimisation

2.1. The Geometrical Considerations and Mathematical Formulas of the Transformations for Tongue Contours

When writing the exact mathematical form of the transformation, we relied on the special geometry of the available US records. Namely, the imaging US head scans such a radial region of the oral cavity which is seen at an angle of 90° measured from a fixed centre C . Consequently, it is obvious to treat the US images and the points of the belonging tongue contours in such a polar coordinate system of C origin where the position of each pixel is given by radius r measured from point C and the signed angle φ measured from the central vertical axis of the image unambiguously. The aim of the transformation is to embed the radial geometry of the US frames to the rectangular geometry of the MRI records described by the two-dimensional Cartesian coordinates so that the US and MRI tongue contours assigned to the same sound should overlap with each other as much as possible. The transformation of the US tongue contours can include three basic operations: the scaling of the radial range, the scaling of the angular range, and the translation of the angular range. The three operations can be realised mathematically by the formulas

$$\begin{aligned}r' &= r \cdot R, \\ \varphi' &= FI \cdot \varphi, \\ \varphi'_0 &= \varphi_0 + FI_0,\end{aligned}\tag{2.1}$$

where the scale factors R and FI allow the normalisation of the radial and angular ranges, and the term FI_0 performs the translation of the initial angle of the angular range. The mathematical operations of (2.1) can be interpreted in the physical plane of the images in the following way: by scaling the radial range, the magnification of the tongue contours can be modified. The scaling of the angular range creates the possibility to change the width of the angular range covered by the tongue contours. The translation of the angular range means the rotation of the tongue contours in the plane of the image. Thus, relationships (2.1) fit the US tongue contour to the corresponding MRI frame. Applying the inverse of transformations (2.1), however, also the reverse conversion can be executed, i.e., by dint of the inverse operations

$$\begin{aligned}r &= \frac{r'}{R}, \\ \varphi &= \frac{\varphi'}{FI}, \\ \varphi_0 &= \varphi'_0 - FI_0,\end{aligned}\tag{2.2}$$

the MRI tongue contour can be mapped onto the corresponding US frame. The parameter set $\{R, FI, FI_0\}$ of the transformations performed in the directions US-MRI and MRI-US must necessarily be the same since, thereby, the maintaining of

the relative scale ratio of the US and MRI environment can be ensured independently of the direction of the conversion. During the investigations, we fixed the value of factor FI by $FI = 1$, which means that the transformation is conformal.

Transformations (2.1) and (2.2) become valid by the numerical determination of parameters R and FI_0 , to which the optimisation of the values of the parameters offers a possible way. During the optimisation procedure, using an algorithm elaborated by us, we find the parameter set, in the case of that the distance between the transformed US tongue contour and the MRI tongue contour serving as a reference curve is minimal. The calculation of the distance is carried out for all of the possible pairs of points of the two curves, then the average of the smallest distances assigned to each point of the US tongue contour is minimised [12]. To the successful transformation, however, not only the exact values of parameters R and FI_0 are needed but also centre C' designated in the MRI frame must be known, which is the image of centre C of the US record. Beyond these, during the construction of the optimisation algorithm, also the peak of the epiglottis can serve as a good reference point demonstrated by Figure 1, where the peaks of the epiglottis G and G' are marked by green circles in the US and MRI frames, and also centres C and C' are located by red cross signs.

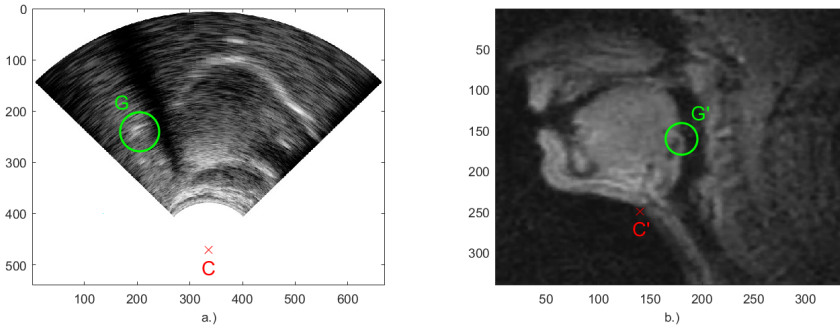


Figure 1. The peaks of the epiglottis G and G' , and the centres C and C' in the US (a.) and MRI (b.) frames.

The meaning of the parameter set $\{R, FI_0, C', G, G'\}$ can be understood by the geometrical considerations of Figure 2, where the left-side block depicts the points of the US frames (C and G), while the right-side block carries the points of the MRI frames (C' and G') in agreement with Figure 1. The radial distances R_1 and R_2 are measured between the centre of the images and the peaks of the epiglottis. The polar angles φ_1 and φ_2 are made by the central vertical axis of the image and radii R_1 and R_2 . Using these quantities, parameter R is interpreted as a magnification factor by

$$R = \frac{R_2}{R_1}, \quad (2.3)$$

and parameter FI_0 is produced by

$$FI_0 = \varphi_1 + \varphi_2, \quad (2.4)$$

which is actually a difference, since φ_1 is negative and φ_2 is positive, as the polar angle is related to the vertical direction in both frames.

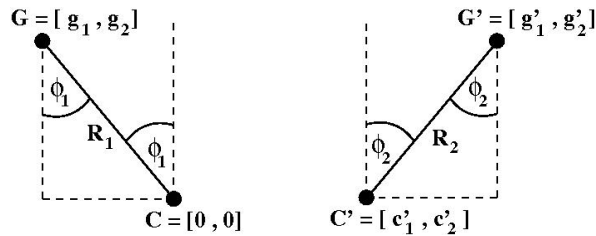


Figure 2. The graphical representation of the main parameters of the US (left) and MRI (right) frames.

2.2. The Optimisation Procedure for Minimising the Distance Between Tongue Contours

Proceeding along the geometrical features of Figure 2, we created an optimisation algorithm using MATLAB via such mathematical formulas which enable the simultaneous optimisation of the parameters $\{R, FI_0, C', G, G'\}$. Some details of the MATLAB scripts can be found in the appendix, where $[n1, n2]$, $[g1, g2]$, and $[g3, g4]$ are the coordinates of C' , G , and G' , respectively, while $FIKORR$ stands for FI_0 . The construction of the mathematical formulas for R_1 , R_2 and $FI1$, $FI2$ follows the geometrical structure of Figure 2. As explained in (2.3) and (2.4), the scale factor R is the ratio of the radial distances between the centre and the peak of the epiglottis, measured in the US and MRI frames separately, and the angular displacement FI_0 is given as the sum of the signed polar angles of the US and MRI frames, made by a radial section and the central vertical axis of the image. In the MATLAB scripts, $cu1$ denotes the actual US or MRI curve having points with complex coordinates to be transformed, $uhus1$ and $mriss1$ gives the US and MRI curves to be compared, and $nnu1 = nnum1$ provides the average of the smallest distances of all of the possible pairs of points of the two tongue contours. In the appendix, the codes are written only for one of the contours belonging to the investigated speech sounds and labelled by index 1, but to have the complete script, these program blocks must be repeated with the same structure for the other examined sounds with indices 2, 3, 4, ..., as well.

2.2.1. Results

We fulfilled the optimisation of the parameters $\{R, FI_0, C', G, G'\}$ for two speech sounds, k and t simultaneously, and we obtained the following numerical results:

$$\begin{aligned} R &= 0.4348, \\ FI_0 &= 0.1293 \text{ rad}, \\ C' &= (250, 140), \\ G &= (244, 230), \\ G' &= (158, 198). \end{aligned} \tag{2.5}$$

Based on the values in (2.5), it can be observed in Figure 3 that the optimised positions of the peaks of the epiglottis are very close to the static position when the voice box is at rest without speaking. In addition, the centre of the MRI frame is located under the jaw of the speaker.

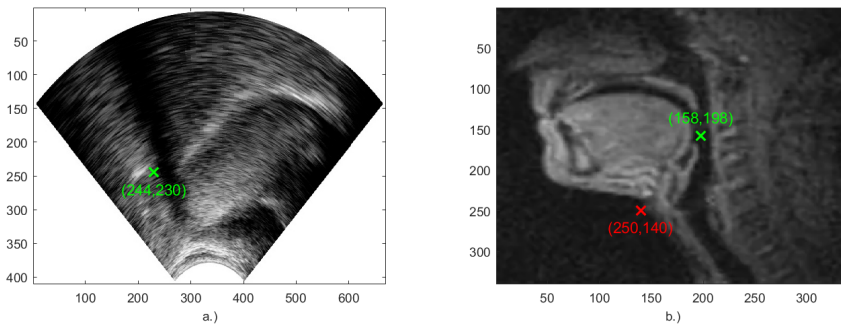


Figure 3. The results of the optimisation for the centre of the MRI frame (b.) and the positions of the peaks of the epiglottis in the US (a.) and MRI (b.) frames.

Using the parameters of (2.5), the transformation of the US and MRI tongue contours can be implemented in a bidirectional way according to Figure 4, where the tongue contours belonging to sound k can be seen. The green curves stand for the US, while the red curves represent the MRI tongue contours. In the MRI frame, also the contour of the palate is drawn by the yellow curve. The figures clearly show that the US and MRI tongue contours fit each other in an acceptable way, since it is demonstrated visually, as well, that the distance is minimised by the optimisation algorithm detailed in the appendix let the two curves mostly overlap with each other. At this level, it is a sufficient criterion for the acceptance of the results without any specified lower or upper limit for the distance between the two curves because we aimed to find the relative position of the two tongue contours, which corresponds to the minimal distance ensured by parameters (2.5). So, graphically, a realistic matching is obtained, and only this was expected. For instance, if the transformed tongue contour was out of the region of the oral cavity

represented by the US or MRI image or placed visually at an unrealistic distance from the reference curve then the optimisation would be surely false.

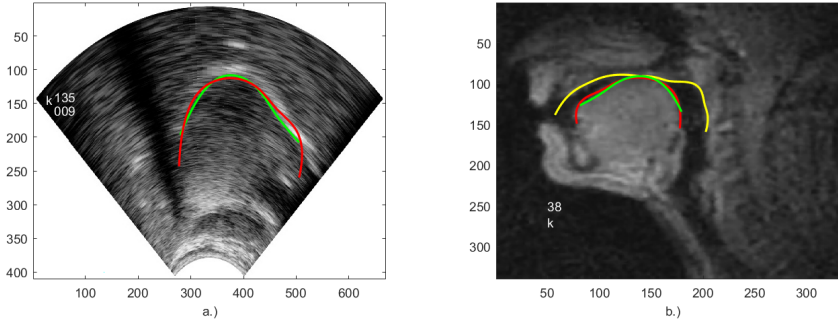


Figure 4. The results of the optimisation in the case of sound k by presenting the US (green) and MRI (red) tongue contours simultaneously. The contour of the palate is indicated by the yellow curve in the MRI frame.

2.2.2. Validation

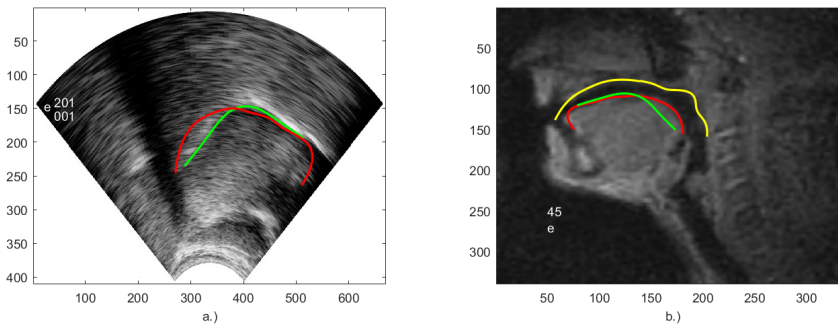


Figure 5. The results of the optimisation in the case of sound e by presenting the US (green) and MRI (red) tongue contours simultaneously. The contour of the palate is indicated by the yellow curve in the MRI frame.

In order to verify the results, we also checked the projection of such US-MRI pairs of contours onto each other, besides the setting of parameters (2.5) provided by the optimisation, which were not present in the set of sounds k and t . Accordingly, Figure 5 exemplifies our results in the case of sound e . It can be stated that the matching of the US and MRI tongue contours is approximately as good as in the case of sound k .

Further validation of the results is in progress currently, as well. To gain more experience about the harmonisation of US and MRI geometry and improve the fitting of the tongue contours, we aim to develop our research work in several directions. We wish to extend the optimisation procedure to more than two sounds to understand the connection between the result of the optimisation and the number of speech sounds. We would also like to investigate the optimisation in the dependence of different sound contexts and speakers of different nationalities and genders. Furthermore, advancing to a large number of speech sounds, we plan to involve machine learning algorithms, as well.

Acknowledgements. We would like to thank the MTA-ELTE Lendület Lingual Articulation Research Group for providing the recordings with the Micro system.

References

- [1] M. ARON, M.-O. BERGER, E. KERRIEN: *Multimodal fusion of electromagnetic, ultrasound and MRI data for building an articulatory model*, in: 8th International Seminar On Speech Production - ISSP'08, Dec 2008, Strasbourg, France. ffinria-00326290f, 2008, DOI: <https://hal.inria.fr/inria-00326290/document>.
- [2] J. CLELAND, A. WRENCH, J. SCOBIE, S. SEMPLE: *Comparing articulatory images: An MRI/Ultrasound Tongue Image database*, in: Proceedings of the 9th International Seminar on Speech Production, 2011, pp. 163–170, DOI: <https://eresearch.qmu.ac.uk/handle/20.500.12289/2477>.
- [3] S. G. DANNER, A. V. BARBOSA, L. GOLDSTEIN: *Quantitative analysis of multimodal speech data*, Journal of Phonetics 71 (2018), pp. 268–283, DOI: <https://doi.org/10.1016/j.jocn.2018.09.007>.
- [4] B. DENBY, M. STONE: *Speech synthesis from real time ultrasound images of the tongue*, in: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, 2004, pp. I-685, DOI: <https://doi.org/10.1109/ICASSP.2004.1326078>.
- [5] L. FULCHER, A. LODERMEYER, G. KAHLER, S. BECKER, S. KNIESBURGES: *Geometry of the vocal tract and properties of phonation near threshold: calculations and measurements*, Applied Sciences 9.13 (2019), 2755, DOI: <https://doi.org/10.3390/app9132755>.
- [6] A. OJALAMMI, J. MALINEN: *Automated segmentation of upper airways from MRI-vocal tract geometry extraction*, International Conference on Bioimaging 3 (2017), pp. 77–84, DOI: <https://doi.org/10.5220/0006138300770084>.
- [7] A. D. SCOTT, M. WYLEZINSKA, M. J. BIRCH, M. E. MIQUEL: *Speech MRI: morphology and function*, Physica Medica 30.6 (2014), pp. 604–618, DOI: <https://doi.org/10.1016/j.ejmp.2014.05.001>.
- [8] SPAN | SPEECH PRODUCTION AND ARTICULATION KNOWLEDGE GROUP: *the rtMRI IPA chart (John Esling)*, Accessed in 2020 May 9th, URL: https://sail.usc.edu/span/rtmri_ipa/je_2015.html.
- [9] M. STONE: *A guide to analysing tongue motion from ultrasound images*, Clinical Linguistics and Phonetics 19.6-7 (2005), pp. 455–501, DOI: <https://doi.org/10.1080/02699200500113558>.

- [10] K. XU, T. G. CSAPÓ, P. ROUSSEL, B. DENBY: *A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization*, Journal of the Acoustical Society of America 139.5 (2016), EL154–EL160, DOI: <https://doi.org/10.1121/1.4951024>.
- [11] L. ZHAO, L. CZAP: *Automatic tracking of tongue contours in ultrasound records*, Beszédtudomány – Speech Science 27.1 (2019), pp. 331–343, DOI: <https://doi.org/10.15775/Beszkt.2019.331-343>.
- [12] N. ZHARKOVA, N. HEWLETT: *Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/*, Journal of Phonetics 37.2 (2009), pp. 248–256, DOI: <https://doi.org/10.1016/j.wocn.2008.10.005>.

Appendix

Program code for the optimisation

```

FI = 1;
minimh = 1000;
for n1 = 250 : 260
for n2 = 200 : 210
for g1 = 220 : 230
for g2 = 100 : 110
for g3 = 150 : 160
for g4 = 190 : 200
R1 = sqrt((-g2)2 + g12);
R2 = sqrt((g4 - (340 - n2))2 + (n1 - g3)2);
R = R2/R1;
FI1 = atan(-g2/g1);
FI2 = atan((g4 - (340 - n2))/(n1 - g3));
FIKORR = FI1 + FI2;
uh2MRI;
minim1 = nnum1;
minim2 = nnum2;
minim12 = minim1 + minim2;
if minim12 < minimh
minimh = minim12;
minis = [R, n1, n2, g1, g2, g3, g4, FIKORR, minimh];
end
end
end
end
end
end
end
end

```

Program code for the transformation of the US tongue contour to the MRI frame

```
a = [471, 335];
b = [n1, n2];
clear px1 py1 x1 y1 firad_UH1 fideg_UH1 r_UH1 xtr1 ytr1 pxtr1 pytr1
for k = 1 : leu1
px1(k) = real(cu1(k));
py1(k) = imag(cu1(k));
x1(k) = px1(k) - a(2);
y1(k) = a(1) - py1(k);
firad_UH1(k) = atan(x1(k)/y1(k));
fideg_UH1(k) = 180 * atan(x1(k)/y1(k))/pi;
r_UH1(k) = y1(k)/cos(firad_UH1(k));
end

for k = 1 : leu1
xtr1(k) = R * r_UH1(k) * sin(FI * firad_UH1(k) - FIKORR);
ytr1(k) = R * r_UH1(k) * cos(FI * firad_UH1(k) - FIKORR);
pxtr1(k) = b(2) + xtr1(k);
pytr1(k) = b(1) - ytr1(k);
end
uh1 = 340 - pxtr1 + li * pytr1;
```

Program code for the transformation of the MRI tongue contour to the US frame

```
a = [471, 335];
b = [n1, 340 - n2];
clear px1 py1 x1 y1 firad_MRI1 fideg_MRI1 r_MRI1 xtr1 ytr1 pxtr1 pytr1
for k = 1 : leu1
px1(k) = real(cu1(k));
py1(k) = imag(cu1(k));
x1(k) = px1(k) - b(2);
y1(k) = b(1) - py1(k);
firad_MRI1(k) = atan(x1(k)/y1(k));
fideg_MRI1(k) = 180 * atan(x1(k)/y1(k))/pi;
r_MRI1(k) = y1(k)/cos(firad_MRI1(k));
end

for k = 1 : leu1
xtr1(k) = r_MRI1(k)/R * sin((firad_MRI1(k) - FIKORR)/FI);
ytr1(k) = r_MRI1(k)/R * cos((firad_MRI1(k) - FIKORR)/FI);
pxtr1(k) = a(2) + xtr1(k);
pytr1(k) = a(1) - ytr1(k);
```

```
end  
mri1 = 670 - pxtr1 + li * pytr1;
```

**Program code for the calculation of the distances between US and MRI
tongue contours**

```
K = 1;  
clear dc1 du1  
for q = K : lh1  
for qq = 1 : lm1  
b1 = imag(uhus1(q)) - imag(mris1(qq));  
b2 = real(uhus1(q)) - real(mris1(qq));  
dc1(qq) = sqrt(b1 * b1 + b2 * b2);  
end  
du1(q) = min(dc1);  
end  
nmu1 = mean(du1(K : length(du1)));
```