SIMPLIFICATION OF VOCAL TRACT SHAPES WITH DIFFERENT LEVELS OF DETAIL

Saeed Dabbaghchian[1], Marc Arnela[2], Olov Engwall[1]

[1]Department of Speech, Music, and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden
[2]GTM Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Barcelona, Catalonia
saeedd@kth.se, marnela@salleurl.edu, engwall@kth.se

ABSTRACT

We propose a semi-automatic method to regenerate simplified vocal tract geometries from very detailed input (e.g. MRI-based geometry) with the possibility to control the level of detail, while maintaining the overall properties. The simplification procedure controls the number and organization of the vertices in the vocal tract surface mesh and can be assigned to replace complex cross-sections with regular shapes. Six different geometry regenerations are suggested: bent or straight vocal tract centreline, combined with three different types of cross-sections; namely realistic, elliptical or circular. The key feature in the simplification is that the cross-sectional areas and the length of the vocal tract are maintained. This method may, for example, be used to facilitate 3D finite element method simulations of vowels and diphthongs and to examine the basic acoustic characteristics of vocal tract in printed physical replicas. Furthermore, it allows for multimodal solutions of the wave equation.

Keywords: vocal tract shape, simplification, voice production, acoustics, area function.

1. INTRODUCTION

The source-filter theory of speech production [4] models the vocal tract (VT) as a filter. In low frequencies, where we can assume plane wave propagation, the response of such a filter mainly depends on the area of the perpendicular cross-sections and their positions, which may also be represented by the vocal tract area function. Such area functions for different phonemes have been extracted from X-rays, e.g. in the early work of Fant [4] and more recently from Magnetic Resonance Imaging (MRI), by e.g., Story *et al.* [9],[10]. The benefit of the area function is its simplicity, as it compactly describes the vocal tract tube. However, the area function has its limitations as a representation of the VT and it is, for instance, not a valid representation in high frequencies (see e.g. [12]). For more accurate calculations of the acoustic output, additional properties, such as the shape of the cross-sections or the bending of the vocal tract, may be required. Using MRI, very detailed 3D surface meshes of the vocal tract can be generated [1], which could be used in 3D numerical simulations of the vocal tract acoustics such as the finite difference method [11] or the Finite Element Method [2], [13]. However, the level of detail in these meshes may be unnecessarily high, causing additional computational power and/or time to be required. Similarly, when experimenting with physical replicas of the vocal tract, it may be desirable to work with more uniform tube shapes, in order to examine the basic acoustic characteristics of the vocal tract [3]. In both cases, the simplifications are even more important when dynamic geometries are concerned, as the simplified geometries are a more viable alternative than using interpolation between intricate geometries with high levels of detail. There is hence a need for a simplification procedure that keeps the overall VT shape, while discarding the small details, which do not significantly contribute to the acoustic output. Some simplifications have been previously suggested by Motoki [7] converting cross-sections from MRI data to elliptical or rectangular shapes. In this work, we propose an alternative method based on a 3D VT surface mesh, rather than the original medical images, to determine simplified realistic, elliptic or circular cross-sections. It is hence applicable in 3D modelling of the vocal tract.

This work is an important first step in our on-going work on studying the effects on the acoustic output of the shape and number of cross-sections, the bending of the vocal tract and the representation of the vocal tract in the form of an area function or a 3D VT shape. Moreover, diphthong sounds could be easily generated. The simplification procedure is also a pre-requisite for on-going work where different biomechanical model meshes of the articulators should be connected to one common mesh in order to create a closed tube that can be used for acoustic simulations [14]. On the other hand, some of the regenerated simplified VT shapes could also be used by multimodal methods of the wave equation [3].

2. SHAPE ANALYSIS

In this work, we depart from the very detailed 3D meshes created by Aalto *et al.* [1], based on MRI measurements of one subject producing Finnish vowels, as exemplified by Fig. 1. We have adapted the mesh for our purposes, by removing additional parts of the face, neck and the sub-glottal tube, and by fixing minor problems of the mesh.

In this section, we describe the analysis of the VT shape that is carried out in order to extract its centreline and cross sections, which are used for calculating the area function and shape simplifications. From the acoustic point of view, supposing that the sound wave propagates along the VT centreline, the cross-sections should be perpendicular to the centreline in order to approximate the area of the wave propagation planes. This means that we, on the one hand, need the centreline in order to find cross-sections that are perpendicular to it; and on the other hand, the centreline is in fact the line through the centre of these cross-sections. In order to solve this circular reference, we follow the two-step method suggested by Kröger *et al.* [6] for calculating the centreline and the perpendicular cross-sections of the vocal tract from MRI images. In their first step, the user defines an initial semi-polar grid. The grid lines intersect the VT in the 2D mid-sagittal image and form the centreline as the line connecting the midpoints of the intersections. In the second step, this centreline is used to estimate the perpendicular cross-sections. We adapt this method to work in 3D space on a VT surface mesh, instead of on 2D MR images, as described further in sections 2.1-2.2.

2.1. Termination planes and grid

The first step in the procedure, summarized in Fig. 2, is to define the glottis and mouth termination planes, which is done by the user choosing one point near the mouth opening and one near the glottis, and specifying the normal vector of the termination planes by visual inspection. The two termination planes are defined as the lowest (glottis) and front-most (mouth) planes for which the intersection with the VT is a closed contour.

The next step, after calculating the termination planes, is to define a semi-polar grid by user interaction. As illustrated in Fig. 3, a semi-polar grid consists of two Cartesian sections connected by one polar. In 2D, it is a collection of horizontal, oblique, and vertical segment lines, while in 3D, planes parallel to these lines are used. The user specifies I) the origin of the grid, II) the position and the normal vector of the first and the last plane, matching the glottis and mouth terminations and III) the number of planes in each section (15, 10 and 15, respectively, in Fig. 3). The choice of the number of gridplanes affects the accuracy of the initial centreline. In this paper, we defined the grid with 80 planes (30 for Cartesian and 20 for polar section).

Figure 1: Initial 3D vocal tract surface mesh for the vowel [ɑ:] (16807 vertices, 33406 triangles).

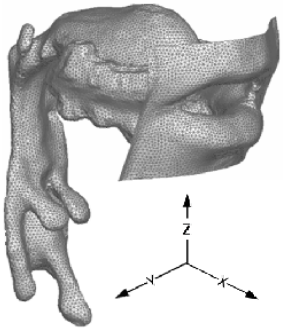


Figure 2: Flow chart for calculating the centreline and cross-sections. Shaded blocks indicate steps requiring user interaction.

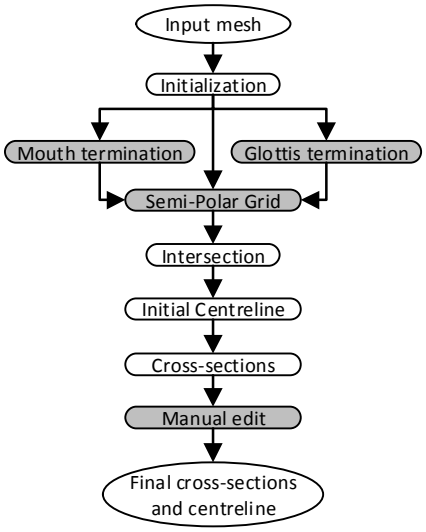


Figure 3: Initial semi-polar grid and vocal tract midsagittal boundary (intersection of VT with XZ plane) for the vowel [ɑ:].

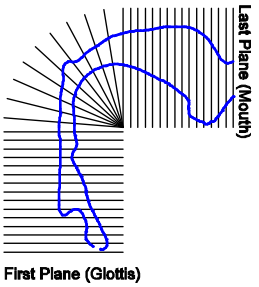
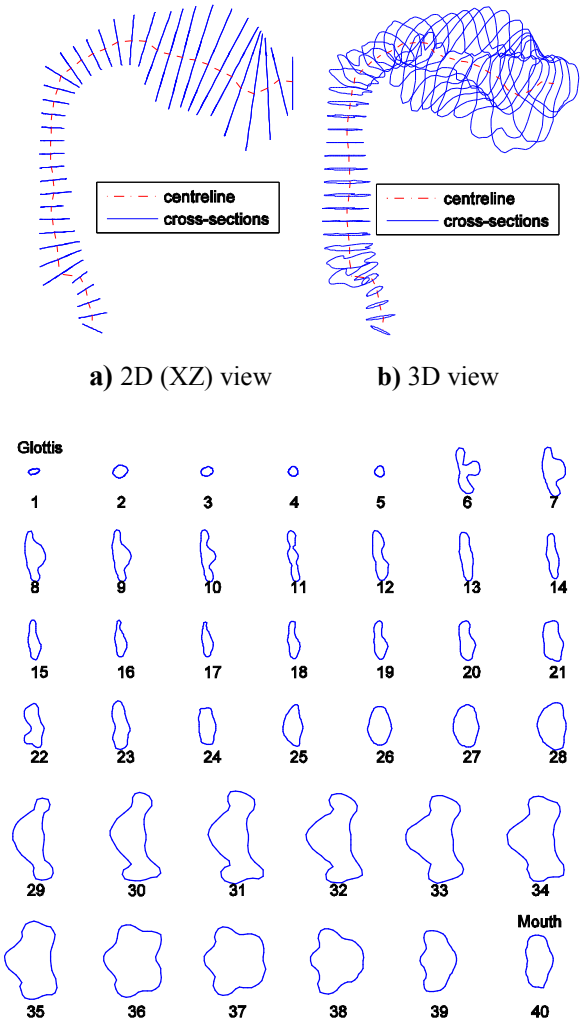


Figure 4: Centreline and cross-sections for [ɑ:].



a) 2D (XZ) view **b)** 3D view

c) Cross-sections from glottis to mouth

Table 1: Angle against the XY-plane (radians), centre (cm), area (cm^2), and lateral dimension (cm) of 40 cross-sections for vowel [ɑ:].

No.	Angle	Center			Area	Lateral
		x	y	z		
1	-0.43	-6.91	0.22	-8.49	0.20	0.39
2	0.20	-6.91	0.22	-8.05	0.60	0.84
3	0.34	-7.03	0.19	-7.61	0.36	0.62
4	0.56	-7.12	0.19	-7.14	0.30	0.62
5	0.65	-7.32	0.15	-6.72	0.34	0.71
6	0.57	-8.00	-0.08	-6.63	2.33	3.00
7	0.38	-8.08	-0.05	-6.12	2.95	3.25
8	0.19	-8.11	0.01	-5.64	2.75	3.34
9	0.06	-8.19	0.06	-5.19	2.42	3.43
10	0.02	-8.28	0.10	-4.74	1.94	3.44
11	0.05	-8.32	0.15	-4.30	1.52	3.37
12	0.09	-8.22	0.15	-3.84	2.17	3.31
13	0.11	-8.24	0.24	-3.39	2.06	3.23
14	0.08	-8.30	0.30	-2.94	1.66	2.98
15	0.04	-8.34	0.35	-2.48	1.40	2.59
16	-0.01	-8.33	0.23	-2.03	1.16	2.43
17	-0.05	-8.31	0.15	-1.58	1.08	2.31
18	-0.10	-8.29	0.21	-1.13	1.26	2.44
19	-0.17	-8.22	0.21	-0.68	1.51	2.52
20	-0.25	-8.09	0.15	-0.24	1.94	2.58
21	-0.55	-7.97	0.10	0.19	2.71	2.69
22	-0.96	-7.64	0.06	0.47	2.51	2.90
23	-1.03	-7.27	0.02	0.75	2.67	3.20
24	-1.12	-6.87	-0.05	0.96	2.20	2.43
25	-1.23	-6.45	-0.04	1.13	2.47	2.76
26	-1.41	-6.02	0.01	1.36	2.98	2.52
27	-1.65	-5.54	0.01	1.42	3.61	2.78
28	-1.86	-5.04	0.01	1.39	4.47	3.12
29	-1.92	-4.66	-0.17	1.09	7.65	5.41
30	-1.86	-4.24	-0.32	0.94	10.28	5.82
31	-1.83	-3.81	-0.28	0.83	12.30	6.10
32	-1.88	-3.35	-0.16	0.74	12.33	5.92
33	-1.97	-2.92	-0.10	0.59	12.90	5.81
34	-2.06	-2.50	-0.09	0.37	12.92	5.65
35	-2.06	-2.10	-0.08	0.18	11.98	5.19
36	-1.90	-1.78	-0.06	-0.22	13.21	4.80
37	-1.71	-1.34	-0.04	-0.42	13.01	4.51
38	-1.52	-0.87	-0.03	-0.23	10.07	4.32
39	-1.33	-0.48	-0.02	0.06	6.80	4.04
40	-1.57	0.00	0.00	0.00	4.81	3.49

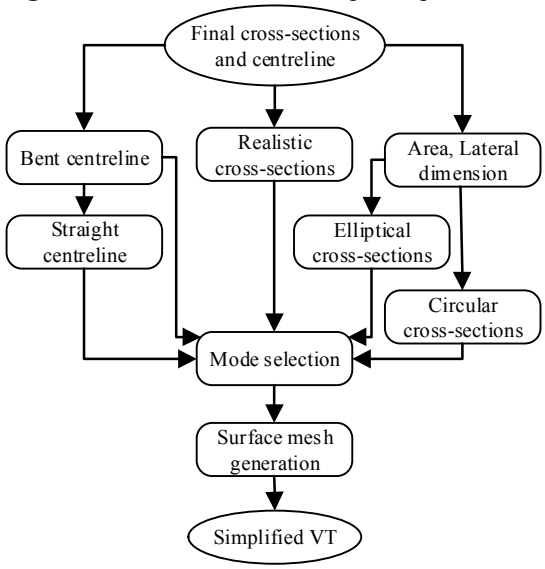
Vocal tract length: 18.39 cm

2.2. Centreline and cross-sections

The 3D grid intersects with the VT surface mesh and the centre of each cross-section is calculated. These centres are connected to form a 3D curve, which represents the initial centreline after Bézier smoothing.

This initial centreline is sampled at N_c points by dividing it into N_c-1 equal intervals, where N_c is the number of desired cross-sections (N_c =40 in this paper). N_c is independent of the number of gridplanes and the smaller N_c, the more details are eliminated. Tangent vector $\vec{T}$ at each sample point of this centreline is calculated. Although $\vec{T}$ indicates the normal vector of the planes which are perpendicular to the initial centreline, projection of $\vec{T}$ onto the XZ-plane (see Fig. 1 for the direction of X, Y, and Z axes) is used as the normal vector $\vec{N}$ to omit the variations of the initial centreline in the Y-direction. The planes passing through the sample points and with normal vectors $\vec{N}$ are used to find the cross-sections of the vocal tract. The final cross-sections are obtained after manual editing dedicated to removing some parts of the cross-sections and repairing errors. In this paper, we removed the parts belonging to the piriformi and the vallecula from the cross-sections, in order to avoid side branches (c.f. [5] and [11] for acoustic effects). The final centreline is constructed by connecting the centres of the final cross-sections. Fig. 4 shows the final centreline and cross-sections for the vowel [ɑ:] and Table 1 provides all data for generating simplified geometries. The variations in the Y-direction are omitted in the vocal tract length calculation, as proposed in [9].

Figure 5: Flow chart for VT shape simplification



3. SHAPE SIMPLIFICATION

Once the final centreline and cross-sections have been obtained, it is possible to perform shape simplification, as summarized in Fig. 5. The basis for the simplifications is the data provided in Table 1 and requiring that the area function should be maintained while simplifying the outline of the cross-sections and/or the centreline. We propose three different simplifications of the cross-sections: realistic, elliptic and circular. Realistic means that the cross-sections are regenerated to have the original shape, (as illustrated in Fig. 4c) but with an equal number of evenly spaced vertices in each cross-section outline (whereas the vertices on the original cross-sections vary in number between cross-sections and are non-uniformly distributed). In the case of elliptical and circular, cross- sections are regenerated with ellipse and circle outlines, while keeping the cross-sectional areas the same as for the original VT shape. For the elliptical shape, the lateral width is further set to correspond to that of the original shape. These simplified outlines may be combined with either a bent or straight centreline, as illustrated in Fig. 6. In the first case, the final centreline is used without any modification, and in the second, the final centreline is straightened, keeping the same length and distances between cross-sections.

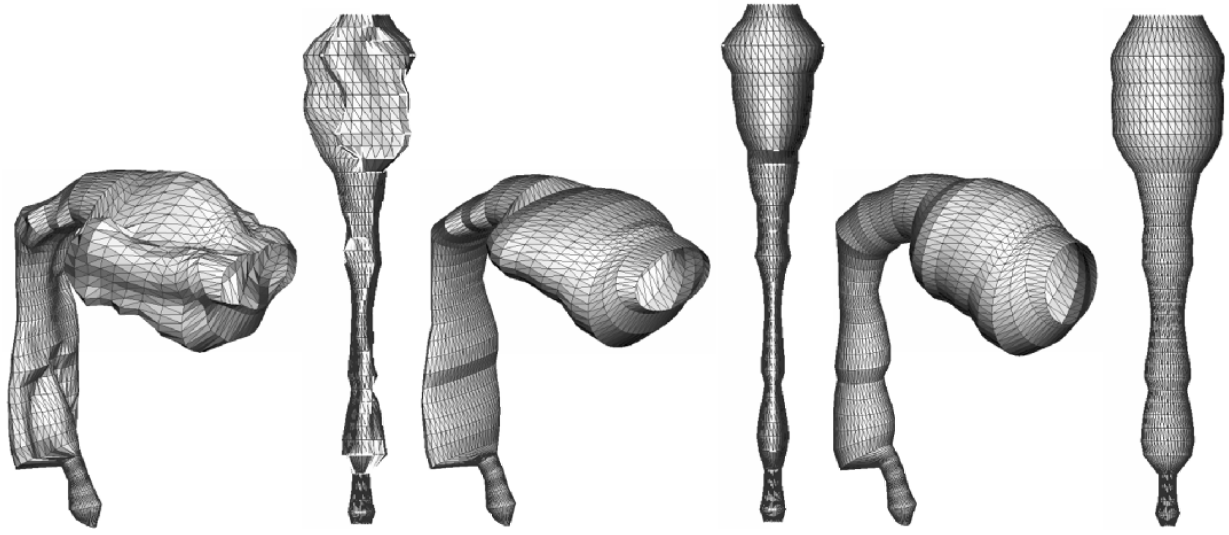
4. DISCUSSION AND CONCLUSIONS

In this paper, we have proposed a method to regenerate the VT surface mesh in simplified versions with different levels of details. These simplified geometries may decrease the computational cost, when they are used instead of the original intricate geometries. Our on-going work focuses on coupling this pre-processing step with aero-acoustic simulations and measurements in physical replicas to determine the acoustic significance of the simplifications. Future directions of the study could be considering other configurations, such as eccentric centreline, and interpolation between different simplified vowel shapes to simulate diphthongs. Another direction could be to modify parameters in the simplified geometry in order to examine the influence of small changes (geometrical perturbation) onto the acoustic output, similar to the earlier work by Motoki [8], but with a configuration that is more similar to the real VT shape.

5. ACKNOWLEDGEMENT

This research has been supported by EU-FET grant EUNISON 308874.

Figure 6: Simplified bent and straight VT shapes. From left to right: realistic, elliptical and circular cross-sections (1920 vertices, 3744 triangles).



6. REFERENCES

[1] Aalto, D., Aaltonen, O., Happonen, R. P., Jääsaari, P., Kivelä, A., Kuortti, J., et al. 2014. Large scale data acquisition of simultaneous MRI and speech. *Applied Acoustics* 83, 64-75.

[2] Arnela, M., Guasch, O., Alías, F. 2013. Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations. *The Journal of the Acoustical Society of America* 134(4), 2946-54.

[3] Blandin, R., Arnela, M., Laboissière, R., Pelorson, X., Guasch, O., Hirtum, A.V., Laval, X. 2015. Effects of higher order propagation modes in vocal tract like geometries. *Journal of the Acoustical Society of America* 137(2), 832-843.

[4] Fant, G. 1971. *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations* (2). Walter de Gruyter.

[5] Dang, J., Honda, K. 1997. Acoustic characteristics of the piriform fossa in models and humans. *The Journal of the Acoustical Society of America* 101, 456-65.

[6] Kröger, B. J., Winkler, R., Mooshammer, C., Pompino-Marschall, B. 2000. Estimation of vocal tract area function from magnetic resonance imaging: Preliminary results. *Proc. 5th Seminar on Speech Production* Germany.

[7] Motoki, K. 2002. Three-dimensional acoustic field in vocal-tract. *Acoustical Science and Technology* 23(4), 207-212.

[8] Motoki, K., Matsuzaki, H. 2004. Computation of the acoustic characteristics of vocal-tract models with geometrical perturbation. *Proc. INTERSPEECH.*

[9] Story, B. H., Titze, I. R., Hoffman, E. A. 1996. Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America* 100(1), 537-554.

[10] Story, B. H. 2008. Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. *Journal of the Acoustical Society of America* 123(1), 327-335.

[11] Takemoto, H., Mokhtari, P., Kitamura, T. 2010. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method. *Journal of the Acoustical Society of America* 128(6), 3724-38.

[12] Takemoto, H., Mokhtari, P., Kitamura, T. 2014. Comparison of vocal tract transfer functions calculated using one-dimensional and three-dimensional acoustic simulation methods. *Proc. INTERSPEECH* Singapore, 408-412.

[13] Vampola, T., Horáček, J., Švec, J. G. 2008. FE Modeling of human vocal tract acoustics. Part 1: Production of Czech vowels. *Acta Acustica United with Acustica* 94(3), 433-447.

[14] Widing, E., Ekeberg, Ö. Tailoring biomechanical model meshes for aero-acoustic simulations. *Submitted to Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization.*