



Vocal-Tract Area Functions with Articulatory Reality for Tract Opening

Zhao Zhang¹, Ju Zhang¹, Jianguo Wei¹, Kiyoshi Honda¹, Tatsuya Kitamura²

¹College of Intelligence and Computing, Tianjin University, China

²Faculty of Intelligence and Informatics, Konan University, Japan

zhao-zhang@tju.edu.cn, khonda@sannet.ne.jp

Abstract

Vocal-tract area function is a one-dimensional representation of the vocal tract, in which speech signals are interpreted according to their place vs. area patterns. Recent work on deriving vocal-tract area functions from volumetric vocal-tract data is successful for the main tract part, whereas the region near the tract ends lacks accuracy due to the use of a planar grid system on the wedge-shaped tract opening. This study employs a special treatment on the anterior tract part using curved grid planes with a gradual evolution of convexity, which is applied to cross-sectioning the anterior tract regions including the post-incisor cavity, inter-dental channel, and lip tube. With the method, volumetric MRI data for vowels /a/ and /i/ were processed to describe the articulatory configuration in those regions. The results revealed that the anterior tract regions are observed as identifiable tract segments with a natural-shaped final opening. Thus, our proposed area function scheme promises nearly complete descriptions of articulatory configuration together with a smooth interface for sound radiation with minor modifications.

Index Terms: Vocal-tract area function, lip opening, curved grid planes, magnetic resonance imaging, articulatory geometry

1. Introduction

Vocal-tract area function is a one-dimensional (1D) representation of the main vocal tract comprising a uniaxial series of the cross-sectional areas of the tract. The area function has been used for various purposes such as acoustic modeling and synthesis of speech. It also offers a means for visual interpretations of speech articulation or a basis for acoustic to vocal-tract inversion as a geometrical platform, supplementing common frequency-domain analysis methods.

The ideas of vocal-tract area function originated from theoretical studies on vowels [1, 2]: Vowel quality is determined by air-column resonance in the vocal tract, and vowel spectra are reproduced by an electrical analogue of the area functions. In earlier work, a relatively small number of tubes were used to model two-dimensional (2D) vocal-tract geometry using lateral X-ray pictures as a reference. With the advent of magnetic resonance imaging (MRI) for three-dimensional (3D) visualization, studies in this century described area functions with a higher spatial resolution allowing finer illustrations of vocal-tract shapes.

Despite such refinements, the area function is limited as a one-dimensional model of the vocal tract with no side branches (i.e., nasal cavity, piriform fossa, interdental space, etc.). Moreover, the area functions are essentially incapable of describing the open end of the vocal tract because of the wedge-shaped lip opening [3]. This lip-end issue has been treated to have a short tube in front of the last segment with a closed circumference. This simple approximation, however, does not comply with spectral realization because of the lack of the natural ge-

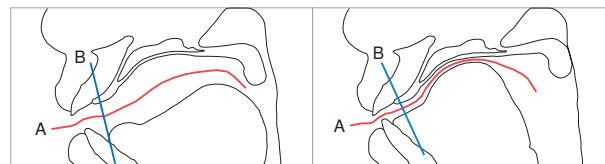


Figure 1: Midsagittal profiles of MRI-based anterior vocal tracts for /a/ (left) and /i/ (right). Curve A indicates a vocal-tract midline drawn by an inscribed-circle method. Line B corresponds to the coronal plane passing near the bilateral lip corners.

ometry determining vowel formants.

This study proposes a new 1D model to describe vocal-tract geometry in the form of area functions. Our approach is to handle the areas in the anterior vocal-tract region near the teeth and lips along with their curved shapes. This approach should allow us to illustrate acoustically-realistic functions of the vocal tract together with descriptions of articulatory configuration.

To attain this goal, we work on vocal-tract data in the ATR MRI Vowel Production Database obtained from a Japanese male speaker [4]. The volumetric images with the teeth for vowels /a/ and /i/ were used to derive new area functions to represent the whole tracts up to the true open end. The major difference from the past approaches is the use of curved grid planes of gradual changes in anterior convexity applied to the anterior vocal tract up to the tract opening at lip surfaces.

2. Issues on area functions at tract opening

Figure 1 illustrates the abovementioned problem by showing manual traces on the midsagittal oral cavity from the datasets for vowels /a/ and /i/. A line segment (labeled B) represents a conventional grid plane that passes near the bilateral lip corners where the perioral tissue is about to separate into the upper and lower lips. The orientation of the line is approximately perpendicular to the vocal-tract midline (labeled A), which is also manually drawn using a method similar to [5].

In the figure, from the glottis to plane B, a certain automatic procedure would obtain a series of cross-sectional areas along a vocal-tract midline. Beyond this plane toward the lip's outer surfaces, the procedure no longer works because of the open wall circumference, which needs certain artificial editing to make it closed. It is seen in this figure that the region to be edited includes the post-incisor cavity, anterior dental arch, and lip tissues, which show anterior convexity in the transverse plane. The curvature of those structures makes another problem: common planar grid planes are tangential to the curvature. The area functions obtained with planar grid planes would result in ambiguous tract shapes lacking landmark locations of the structures. In Fig. 1, for example, the pattern of a tract stricture

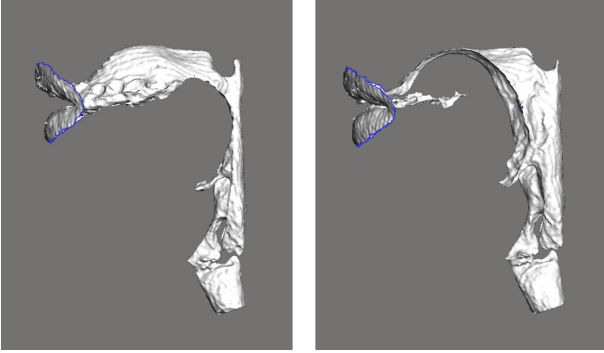


Figure 2: Lateral views of vocal-tract surfaces for vowels /a/ (left) and /i/ (right) showing the regions for computing area functions up to the lips' outer surface. The bilateral side branches, i.e., piriform fossa, interdental space, and vallecula, were eliminated for the subsequent analysis.

at the lower incisors followed by an abrupt widening in front is clearly visible on the image, while such a detail is hardly observable on common area functions known to date.

Past studies conducted various ways to handle the wedge-shaped lip region: some ignored the lips, and others added a lip tube of unclear geometry or ad hoc shapes. In recent studies with 3D MRI, the last measurable area is extended about half-way between the lip corners to the vermillion surface [6, 7]; or in another, the lips' wedge is edited by extending the measured lip area up to the place where the vertical lip aperture is minimum [8]. All those treatments for lip opening in the past work lack geometrical reality, which further makes it difficult to observe articulation of the tongue tip, jaw, and lips, or to apply adequate open-end corrections to simulate sound production from the tract.

In this study, as a first step to overcome the problems noted above, we use arc-like grid planes with a gradual change in anterior convexity from the post-alveolar region, which is applied to cross-section the anterior vocal tract using the method described in the following.

3. Materials and Methods

3.1. Vocal-tract Surface Data

The dataset with the teeth in the ATR MRI Vowel Production Database was used for this study. The same dataset has been documented and analyzed for obtaining vocal-tract transfer functions using acoustic experiments [4, 9] and finite-difference time-domain (FDTD) simulations [10]. The image processing technique used in those studies was to extract boundaries of the vocal tract and lower face from the MRI data, which was transform into a surface model of the STL format. In this study, the surface model was edited to have the whole tract surface from the subglottic region to the vermillion of the lips, as shown in Fig. 2.

3.2. Treatment on Anterior Vocal Tracts

One of the factors that obscures the open-end details is the anatomical curvature of the dental arch and lip tissue on a transverse view. If cross-sectional areas were obtained using curved grid planes adjusted to the shape of those structures, then the procedure would return plausible area functions showing artic-

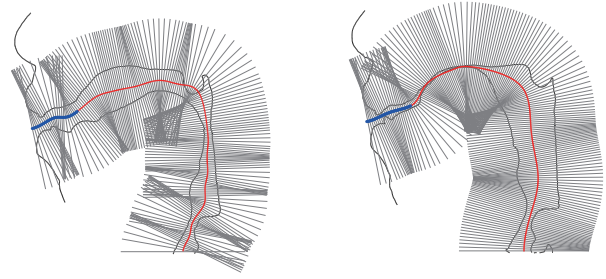


Figure 3: Vocal-tract midlines on midsagittal plane with grid plane positions for vowels /a/ (left) and /i/ (right).

ulatory details. With this expectation in mind, a series of curved grid planes that gradually augment anterior convexity was applied to the anterior region of the vocal tract along the midline. A series of quadratic planes was chosen for this purpose.

What may be concerned about this procedure are: 1) deviation from the basic principle of plane-wave propagation in the vocal tract, and 2) over-estimation of areas sectioned with curved grid planes. The former may partly be approved according to the acoustic measurement of pressure variations on the oral cavity for the vowel /a/ showing convex-forward wavefronts near the tract opening [11, 12]. The latter for area over-estimation can reasonably be corrected using a certain coefficient applied to area values to maintain the airway volume in the regions of interest.

3.3. Computation of Area Functions

The basic principle to compute the vocal-tract area function in this study follows the descriptions in [6]. Vocal-tract midlines were obtained by the region growing method, in which flat planes perpendicular to the midline were used to cross-section the vocal tract. Values of cross-sectional areas were measured by counting the number of pixels inside the wall circumference.

Our method employs the same basic steps to obtain vocal tract midlines and lines perpendicular to the midline to represent the location of cross-sections, as shown in Fig. 3. Instead of flat cross-sectional planes, our method applied curved grid planes to cross-section the anterior tract. As shown in Fig. 3, the vocal-tract midline regions (blue line) starting from a midline point near the alveolar ridge on the palate in vowels /a/ and /i/ were selected to apply the curved grid planes. Thus, the area functions with our method are composed of two parts: The area function of the main vocal tract (regions crossed by the red midline) was obtained by the basic method, and the area function of the open end region (regions crossed by the blue midline) was obtained by the proposed method.

With the clear division of the anterior region of the vocal tract, a group of curved grid planes was set up to follow the blue midlines depicted in Fig. 3 from the starting position to the lip end. The curvature of the grid planes gradually increases from flat to full toward the lip end so that the final cross-sectional curved plane matches the upper lip surface to achieve complete cross-sectional details at the open end. Figure 4(a) illustrates how the group of curved cross-sectional planes was arranged in top view, and the oblique view is shown in Fig. 4(b).

For each vowel, to quantify the grid curvature of each curved grid plane, the upper lip's edge contour was extracted from the MRI data first. A parabolic cylindrical function, $y = \alpha x^2 + C$, was used to model the upper lip shape, where the

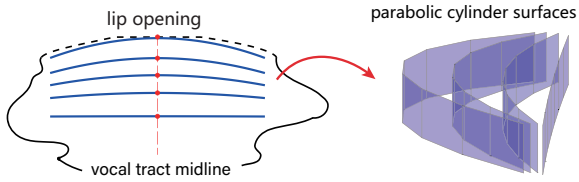


Figure 4: Curved grid plane system: (a) Parabolic cylindrical functions on the transverse plane. (b) Corresponding curved grid planes in a 3D view.

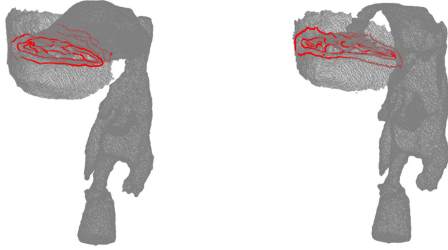


Figure 5: Cross-section result for curved grid planes on vowel /a/ (left) and vowel /i/ (right).

parameters α and C were estimated by the least-squares polynomial fit method [13] using the upper lip's edge contour as the target for fitting. With the obtained curvature parameters of the final cross-sectional plane, a group of quadratic functions was derived to represent the rest of the curved grid planes, and the curvilinear function at the i_{th} central points was defined by

$$f(x) = \begin{cases} y = \alpha_i x^2 + C_i, \\ \alpha_i = \alpha * i/N, \end{cases}$$

where α_i represents the grid curvature at i_{th} central points, C_i represents the offset of the i_{th} curved plane according to the corresponding midline point's location, and N represents the total number of the curved grid planes.

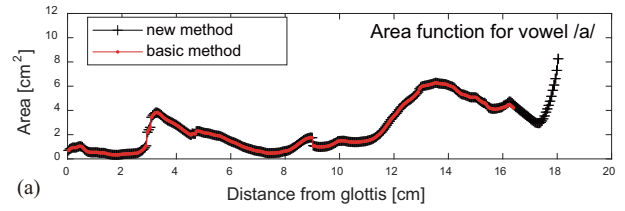
The group of curved surfaces as shown in Fig. 4(b) were generated based on the function above. The grid curvature and location were determined by parabolic cylindrical functions, and the grid tilt angles were adjusted by the slope of the midline. Then these cross-sectional planes were applied to the anterior vocal tract region, and the cross-sections obtained are illustrated in Fig. 5. It is clear to observe that the results (red points in Fig. 5) are capable of portraying the vocal tract cross-sectional shapes near the open-end region.

In the final step, the cross-sections were transformed into a 2D space to measure the area value for each slice. The full area function was generated by combining the area values from the new method for the anterior tract and those from the basic method for the remaining posterior regions.

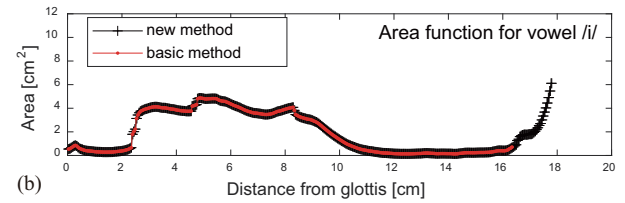
4. Results

The vocal-tract area functions measured for vowels /a/ and /i/ based on MRI-based digital vocal-tract models revealed what follows.

The comparison of the area functions with the proposed new method and basic method (with planar grids only) is shown in Fig. 6(a) and 6(b) for the whole tracts. The partial area functions for the anterior tract region obtained by the new method



(a)



(b)

Figure 6: Area functions of the whole tracts obtained for (a) vowel /a/ and (b) vowel /i/.

are shown in Fig. 7(a). The sectional slices at the lower and upper incisors are shown in Fig. 7(b) to identify the regions corresponding to the inter-dental channels on the partial tracts.

As shown in Fig. 6(a) and 6(b), the area functions from the new method depict the cross-sectional areas of the vocal tract toward the lip end, while the area functions with the basic method fail to capture the regions where the lip opening is wedge-shaped.

Observing Fig. 7(a) and 7(b), the area functions with the new method provide the following observations.

- The tract region for the inter-dental channel was observed as a relatively stable part in the area function in both vowels. From the channel anteriorly, the areas for the lip regions showed a consistent flaring-out pattern for both vowels toward the anterior-most lip tissues:
- In vowel /a/, the regional area behind the lower incisors showed a local tract widening, named the 'post-incisor cavity', which is derived from mild tongue-tip retraction for the Japanese mid open vowel.
- In vowel /i/, the post-incisor cavity was annihilated due to tongue-tip advancement reaching the lower incisors, showing a continuous extreme narrowing behind the tongue tip, which is likely to occur during MRI scans for ease of articulatory maintenance by speakers.

5. Discussions

The vocal-tract area functions obtained in this study show reasonable overall area patterns for vowels /a/ and /i/, while the tract open-end region provides better descriptions of anterior articulatory configurations. In comparison to the existing area functions from the same data [6], our area functions are more descriptive of the dental-arch regions and the labial vestibule of about 1.5-cm long in both vowels. Including the visibility of the post-incisor cavity, it can be stated that the area functions thus obtained carry meaningful information on vowel articulation.

Where does the vocal tract end? This question was raised by Ladefoged et al. [3] and reclaimed as an up-to-date topic. Accordingly, current researchers conduct studies to define the tract open end geometrically and acoustically. Acoustic experiments have been made with simple tube models with and without the lip's wedge, with the results suggesting that the loca-

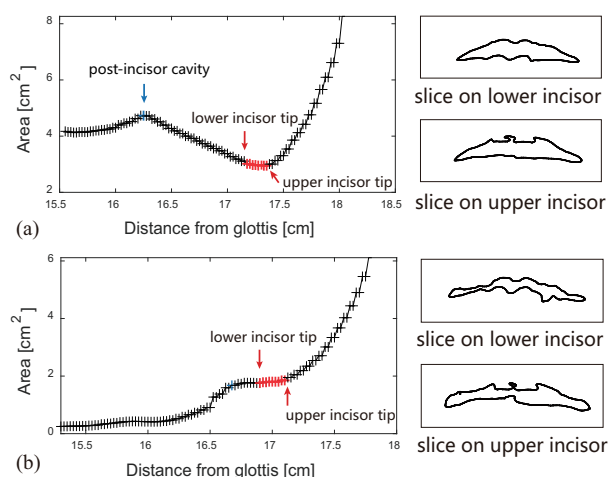


Figure 7: Anterior details obtained for (a) vowel /a/ and (b) vowel /i/ with curved cross-sections (right) on lower and upper incisors.

tion of the open end may be estimated at half-way between the apex and root of beak-like bifurcations on a straight tube wall [14, 15]. The current work is another approach to answer the same question by using a curved grid system.

Acoustical consequences of thus obtained area functions have not been examined yet. The present study deviates from our long-standing belief of plane-wave propagation in the vocal tract. The reason is given by study of sound pressure measurement in a vocal-tract replica: the wavefront is convex-forward near the open end in vowel /a/ [11]. With this observation as a hint, our approach employed a series of curved grid planes with a gradual evolution of anterior convexity. In comparison to area values measured by common planar grid planes, cross-sectioning with the curved grid plane results in greater area values. Whether the excess area values near the open end affect VT transfer functions or not will be examined in a future study.

A few other necessary considerations should be remarked here. The grid system of a linear progression of anterior curvature may need modification into a nonlinear progression of curvature changes so that the curved grids fit adequately on both dental arch and lip surface. Also, the finer-interval cross-sectioning by curved grid planes should be simplified to reduce the computational load together with a reduction of manual adjustment work to determine the midline's end-point and final grid plane being placed on the outer-most lip surfaces. Another distal work to be conducted will be to define both ends of the vocal tract by further answering where the vocal tract begins near the tract bottom with three small cavities.

6. Conclusion

This study attempted at developing a new method for obtaining vocal-tract area functions that represent realistic articulation near the lip end with wedge-like opening. Using a series of curved grid planes that change in anterior convexity from the post-alveolar region to the lip surface end, vocal-tract area functions thus obtained revealed consequences of realistic vowel articulation. From the area functions for vowels /a/ and /i/, tongue tip position can be estimated from the regional tract widening called the post-incisor cavity, jaw opening is observable as a stationary area region for the inter-dental arc channel, and lip

opening is represented by a flaring-out pattern portraying as an open-end flange. The improved area functions offer analytical descriptions of anterior articulatory details in addition to those on tongue articulation in previously reported area functions.

7. Acknowledgements

This work is supported by the National Key R&D Program of China (No. 2020YFC2004103), National Natural Science Foundation of China (No. U1936102), and Qinghai foundation research program No.2021-ZJ-609.

8. References

- [1] T. Chiba and M. Kajiyama, *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan Pub. Co., Tokyo, 1942.
- [2] G. Fant, *Acoustic Theory of Speech Production*. Mouton, The Hague, 1970.
- [3] P. Ladefoged, J. Wright, and W. Linker, "Where does the vocal tract end?" *The Journal of the Acoustical Society of America*, vol. 63, no. S1, pp. S32–S32, 1978.
- [4] T. Kitamura, H. Takemoto, S. Adachi, and K. Honda, "Transfer functions of solid vocal-tract models constructed from ATR MRI Database of Japanese vowel production," *Acoustical Science and Technology*, vol. 30, no. 4, pp. 288–296, 2009.
- [5] S. Maeda and Y. Laprie, "Vowel and prosodic factor dependent variations of vocal-tract length," in *Proc. Interspeech*, 2013, pp. 3196–3200.
- [6] H. Takemoto, K. Honda, S. Masaki, Y. Shimada, and I. Fujimoto, "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," *The Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 1037–1049, 2006.
- [7] P. Birkholz, "Enhanced area functions for noise source modeling in the vocal tract," in *10th International Seminar on Speech Production, Köln*, 2014, pp. 1–4.
- [8] M. Fleischer, A. Mainka, S. Kürbis, and P. Birkholz, "How to precisely measure the volume velocity transfer function of physical vocal tract models by external excitation," *Plos One*, vol. 13, no. 3, p. e0193708, 2018.
- [9] K. Honda, T. Kitamura, H. Takemoto, S. Adachi, P. Mokhtari, S. Takano, Y. Nota, H. Hirata, I. Fujimoto, Y. Shimada *et al.*, "Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 13, no. 4, pp. 443–453, 2010.
- [10] H. Takemoto, P. Mokhtari, and T. Kitamura, "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method," *The Journal of the Acoustical Society of America*, vol. 128, no. 6, pp. 3724–3738, 2010.
- [11] K. Motoki, N. Miki, and N. Nagai, "Measurement of sound-pressure distribution in replicas of the oral cavity," *The Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2577–2585, 1992.
- [12] M. Arnela, R. Blandin, S. Dabbaghchian, O. Guasch, F. Alías, X. Pelorson, A. Van Hirtum, and O. Engwall, "Influence of lips on the production of vowels based on finite element simulations and experiments," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2852–2859, 2016.
- [13] W. H. Beyer, "CRC Standard Mathematical Tables and Formulae," Boca Raton, 1991.
- [14] B. Lindblom, J. Sundberg, P. Branderud, and H. Djamshidpey, "On the acoustics of spread lips," *Proceedings of Fonetik, TMH-QPSR*, vol. 50, no. 1, pp. 13–16, 2007.
- [15] P. Birkholz and E. Venus, "Considering lip geometry in one-dimensional tube models of the vocal tract," in *11th International Seminar on Speech Production, Tianjin*, 2017, pp. 78–86.