DOI: 10.61782/fa.2023.0400

forum acusticum 2023

INFLUENCE OF A REALISTIC HUMAN HEAD ON THE DIRECTIVITY OF VOWEL /A/ BASED ON THREE-DIMENSIONAL FINITE ELEMENT SIMULATIONS

Marc Arnela* **Leonardo Pereira-Vivas** **Jorge Egea**

Human-Environment Research (HER) group, La Salle, Universitat Ramon Llull, Barcelona, Catalonia, Spain

ABSTRACT

Three-dimensional (3D) finite element simulations can be used to accurately simulate the human voice. These simulations involve the generation of sound waves from the vocal cords, their 3D propagation through a detailed 3D vocal tract, usually generated from Magnetic Resonance Imaging (MRI), and their emission outward from the lips. Most studies have focused on internal aspects of the vocal tract, which allowed them to simplify the human head as a spherical baffle or a flat plane. However, it is not clear to what extent this simplification affects the directivity of the voice. This work aims at examining 3D directivity effects of vowel sounds in the horizontal and vertical planes by means of finite element simulations. A detailed geometry is generated for this purpose, consisting of a 3D MRI-based vocal tract connected to a realistic human head designed from scratch. Preliminary results are presented for the vowel /a/, showing that large variations occur in the vocal tract acoustic response as the orientation and frequency increases for both the horizontal and vertical planes. These variations are especially notorious above 5 kHz, since in this frequency range not only planar modes but also higher order modes propagate through the vocal tract.

Keywords: *vocal tract acoustics, voice directivity, finite element method, vowels*

**Corresponding author*: *marc.arnela@salle.url.edu.*
Copyright: *©2023 Marc Arnela et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

1. INTRODUCTION

The human voice can be simulated using three-dimensional (3D) acoustic models, which emulate the physical phenomena involved during voice production. That is, from vocal fold vibration to 3D acoustic wave propagation through the vocal tract and posterior emission outwards from the lips. The most extended approaches are probably those based on finite elements [1], finite differences [2], multimodal methods [3, 4], and digital waveguide models [5]. 3D acoustic models allow for an accurate representation of the 3D vocal tract acoustics, allowing not only the propagation of planar modes, but also higher order ones involving the transverse direction. The latter have shown to have a large impact above 3-5 kHz in the voice spectrum [6, 7].

Typically the interior of the vocal tract has been analysed, focusing on aspects such as the influence of some side branches located in the main vocal tract (e.g., the piriform fossae [8]), the vocal tract shape [7], or the lips [9]. 3D voice radiation is, however, also gaining interest. 3D voice directivity has been examined by means of 3D simulations [3, 10] and also by experiments in mechanical replicas [11], which allow for a detailed analysis of the radiation mechanism compared to voice recordings (see e.g., [12]). However, none of the above works considered a detailed geometry of the human head and vocal tract, but rather rough approximations of the human head, such as a spherical or a flat baffle.

This work aims at examining the influence of a realistic human head on vowel directivity, presenting some preliminary results for the production of vowel /a/. A 3D vocal tract geometry based on Magnetic Resonance Imaging (MRI) [7] is used for this purpose. This geometry is

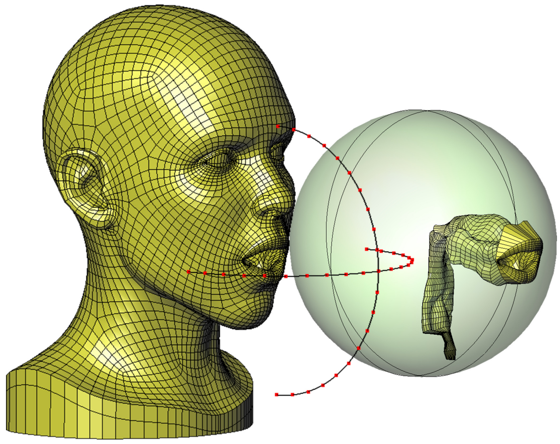


Figure 1. Vocal tract geometry of vowel /a/ set on a realistic human head and on a sphere. Tracking points (in red) are located 0.1 m from the mouth in the vertical and horizontal planes.

attached to a 3D human head model designed and constructed from scratch. For comparison, the vocal tract is also attached to a spherical baffle. The 3D acoustic model in [13] based on the finite element method is used to perform all simulations. This model solves the acoustic wave equation for the acoustic pressure and incorporates a Perfectly Mathced Layer (PML), which allows to simulate free-field conditions around the human head. The acoustic pressure is tracked at several points located in the radiation domain to obtain the directivity in the vertical and horizontal planes.

This paper is organized as follows. Section 2 presents the methodology followed to generate the human head and vocal tract of vowel /a/ and gives details about the finite element simulations. Section 3 shows the results obtained for the vowel /a/, comparing the configuration with a realistic head against that with a spherical baffle. Section 4 closes the paper with the conclusions.

2. METHODOLOGY

2.1 Head and vocal tract geometry

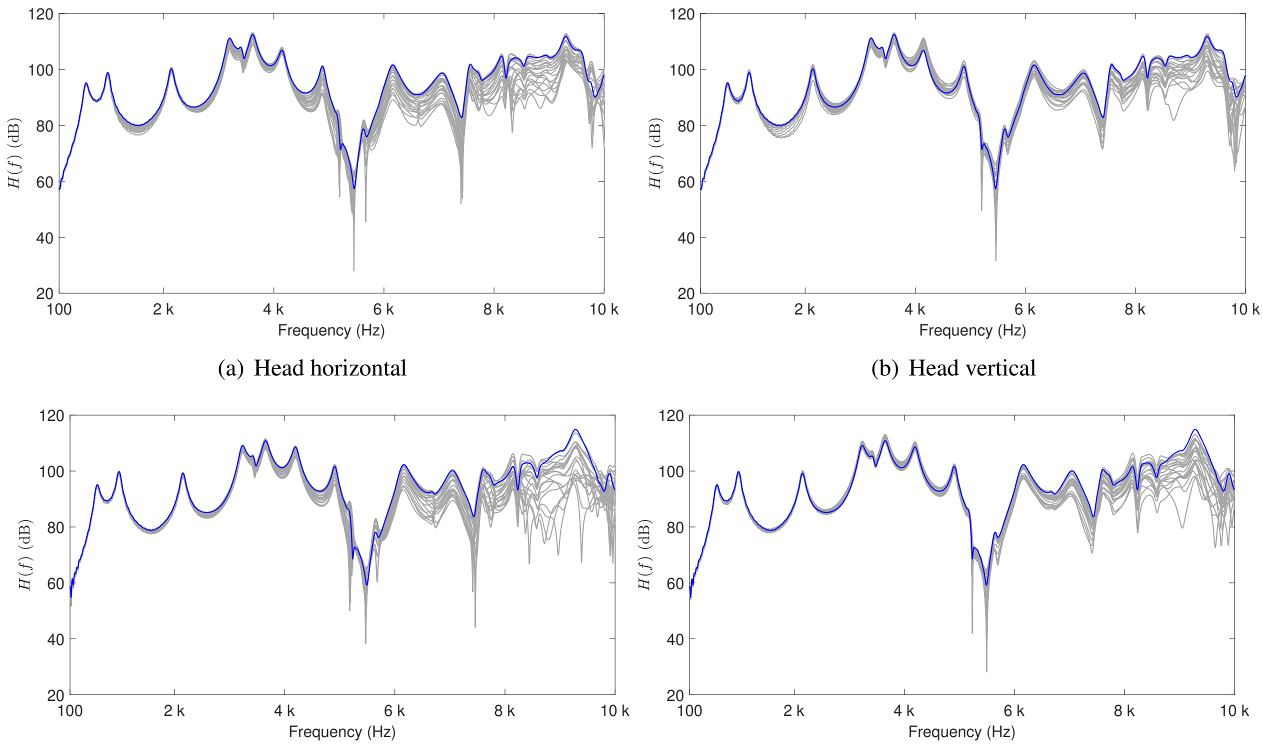
The simplified version of a realistic head model presented in [14] has been used for the finite element simulations of this work. The model preparation consists of the unification of the 3D vocal tract generated from MRI to the face and the simplification of the head. First, the vocal tract is aligned with respect to the mouth. To do this, we follow the procedure described in [13] and project two 45º planes from the outermost points of the upper and lower lip. The intersection of these planes determines the location of the vocal tract exit. The vocal tract geometry is a realistic representation of an MRI-based vocal tract consisting of 40 cross-sections [7]. To fit this geometry with the head, the vocal tract is simplified by eliminating the last two cross-sections. The geometries need to match the number of vertices, 48 in total, with the number of vertices inside the lip of the model, to perform a merge of these vertices to obtain a unified polygon. For the simplification of the model, the Retopologize tool of Autodesk Maya is used, which allows us to lower the number of polygons by choosing the number of faces and the tolerance between them. As a result, we get a much lighter model for the simulation than that used for animation purposes. The original model had a total of 113,984 vertexes, after applying the retopology and remeshing with the Maya tool, the model ended up with 67,104 vertexes, 58% less. Taking the simplified geometry of the head, the vocal tract and lips have been separated and placed inside a sphere of radius 0.105 m (sphere radius from [11, 13]), in order to compare the results with both geometries (see Fig. 1).

2.2 Finite element simulations

Finite element simulations are next performed as in [13]. In a nutshell, a custom finite element code numerically solves the wave equation for the acoustic pressure combined with a Perfectly Matched Layer (PML) to absorb sound waves and emulate free-field propagation. The computational domain is meshed with tetrahedra of size of about 0.004 m within the vocal tract, 0.006 m in the radiation domain, and 0.008 m in the PML. A time step of 2.5e-8 s is used to discretize a time event of 25 ms. In this occasion, however, the acoustic pressure is tracked not only in front of the mouth, but also at 21 points in the vertical plane and 21 points in the horizontal plane, all of them set at a constant distance of 0.1 m from the mouth exit (see Fig. 1). The vocal tract transfer function at the i-th point ($i = 1 \dots 42$) is computed as

$$H_i(f) = \frac{P_i(f)}{Q(f)}, \quad (1)$$

where $P(f)$ is the Fourier transform of the acoustic pressure tracked at the i-th point, and $Q(f)$ the Fourier transform of a Gaussian pulse introduced at the vocal tract entrance.



(a) Head horizontal (b) Head vertical

(c) Sphere horizontal (d) Sphere vertical

Figure 2. Vocal tract transfer functions $H(f)$ obtained for the configurations with the head (top row) and the sphere (bottom row) in the horizontal (left column) and vertical (right column) planes. Blue colour denotes curves in front of the mouth exit.

3. RESULTS

Figure 2 shows the vocal tract transfer functions obtained in the horizontal and vertical planes. Blue color is used to indicate the results in front of the mouth exit, whereas curves at other points are plotted in gray. As expected, some resonances (planar modes or formants in the speech community) appear below 5 kHz, the first two being located at the typical frequencies of an /a/. Note, however, that above this frequency the spectrum becomes more complex. This is because higher order modes involving the transverse direction also propagate in this range, which can only be captured by 3D acoustic models.

As can be observed in Fig. 2, sound pressure levels decay for frequencies below 5 kHz when moving from the axis position, being more important as the frequency increases. This observation agrees with current knowledge on voice directivity. However, note that above 5 kHz not only levels decrease but also the location of the different modes that are present in this region change, resulting in a more complex spectrum. Grey curves also show that the acoustic pressure is less uniform in the horizontal than in the vertical plane, regarthless the baffle configuration. Comparing now the results obtained between the head and the spherical baffle, observe that larger differences are produced between the models as the frequency increases. This is especially notorious above 7-8 kHz in the blue and grey curves for both the horizontal and vertical planes. Note also that the spherical baffle tends to reduce the dynamical range between grey curves compared to the termination with the head. This is observed not only for high frequencies, but also in the low frequency range. For instance, below 3 kHz, all curves are very close to each other in the spherical baffle, for both the horizontal and vertical planes. However, this is not the case for the head, which shows larger variations especially in the valleys between resonances.

4. CONCLUSIONS

In this work finite element simulations have been performed to examine the directivity of vowel /a/. An MRI-based vocal tract geometry articulating this phoneme has been attached to a realistic human head and also to a spherical baffle for comparison. Results show large variations between the two models as the frequency increases, being more notorious above 4-5 kHz when higher order modes also propagate. Significant changes have also been observed at low frequencies, especially between the valleys of resonances. Further vowel sounds need to be examined, but these preliminary results suggest that the head may have a significant contribution to voice directivity in the whole frequency range, tending to extend the dynamic range compared to a spherical baffle configuration.

5. ACKNOWLEDGMENTS

This work has been partially funded by the Spanish Ministry of Science and Innovation - Agencia Estatal de Investigación through the project FEMVoQ (PID2020-120441GB-I00 / AEI / 10.13039/501100011033) and by the Catalan Government through the project Anivoice (IdC Seed No. 2021 LLAV 00043). The authors thank the support of the Catalan Government (Departament de Recerca i Universitats) for the grant 2021 SGR 01396 given to the HER group.

6. REFERENCES

[1] M. Arnela, S. Dabbaghchian, O. Guasch, and O. Engwall, "MRI-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 12, pp. 2173–2182, 2019.

[2] H. Takemoto, P. Mokhtari, and T. Kitamura, "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method," *J. Acoust. Soc. Am.*, vol. 128, no. 10, pp. 3724–3738, 2010.

[3] M. Brandner, R. Blandin, M. Frank, and A. Sontacchi, "A pilot study on the influence of mouth configuration and torso on singing voice directivity," *J. Acoust. Soc. Am.*, vol. 148, no. 3, pp. 1169–1180, 2020.

[4] R. Blandin, M. Arnela, S. Félix, J.-B. Doc, and P. Birkholz, "Efficient 3D acoustic simulation of the vocal tract by combining the multimodal method and finite elements," *IEEE Access*, vol. 10, pp. 69922–69938, 2022.

[5] A. J. Gully, H. Daffern, and D. T. Murphy, "Diphthong synthesis using the dynamic 3D digital waveguide mesh," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 2, pp. 243–255, 2018.

[6] R. Blandin, M. Arnela, R. Laboissière, X. Pelorson, O. Guasch, A. Van Hirtum, and X. Labal, "Effects of higher order propagation modes in vocal tract like geometries," *J. Acoust. Soc. Am.*, vol. 137, no. 2, pp. 832–843, 2015.

[7] M. Arnela, S. Dabbaghchian, R. Blandin, O. Guasch, O. Engwall, A. Van Hirtum, and X. Pelorson, "Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds," *J. Acoust. Soc. Am.*, vol. 140, no. 3, pp. 1707–1718, 2016.

[8] H. Takemoto, S. Adachi, P. Mokhtari, and T. Kitamura, "Acoustic interaction between the right and left piriform fossae in generating spectral dips," *J. Acoust. Soc. Am.*, vol. 134, no. 4, pp. 2955–2964, 2013.

[9] 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," *J. Acoust. Soc. Am.*, vol. 139, no. 5, pp. 2852–2859, 2016.

[10] R. Blandin, A. Van Hirtum, X. Pelorson, and R. Laboissière, "Influence of higher order acoustical propagation modes on variable section waveguide directivity: Application to vowel [ɑ]," *Acta Acust. united Ac.*, vol. 102, no. 5, pp. 918–929, 2016.

[11] R. Blandin, J. Geng, and P. Birkholz, "Experimental investigation of speech directivity mechanisms," in *16ème Congrès Francais d'Acoustique*, 2022.

[12] M. Brandner, M. Frank, and A. Sontacchi, "Horizontal and vertical voice directivity characteristics of sung vowels in classical singing," in *Acoustics*, vol. 4, pp. 849–866, MDPI, 2022.

[13] M. Arnela, O. Guasch, and F. Alías, "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations," *J. Acoust. Soc. Am.*, vol. 134, no. 4, pp. 2946–2954, 2013.

[14] M. Arnela, L. Pereira-Vivas, and J. Egea, "An animated realistic head with vocal tract for the finite element simulation of vowel/a," *Proc. IberSPEECH 2022*, pp. 6–10, 2022.