



# **Modelling the piriform fossae using tuned two-dimensional vocal tracts in the finite element simulation of vowels**

Marc Arnela, David Urena

## **► To cite this version:**

Marc Arnela, David Urena. Modelling the piriform fossae using tuned two-dimensional vocal tracts in the finite element simulation of vowels. Forum Acusticum, Dec 2020, Lyon, France. pp.2169-2170, <10.48465/fa.2020.0366>. <hal-03235386>

**HAL Id: hal-03235386**

**<https://hal.science/hal-03235386v1>**

Submitted on 27 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# MODELLING THE PIRIFORM FOSSAE USING TUNED TWO-DIMENSIONAL VOCAL TRACTS IN THE FINITE ELEMENT SIMULATION OF VOWELS

Marc Arnela and David Ureña

GTM Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Barcelona, Spain

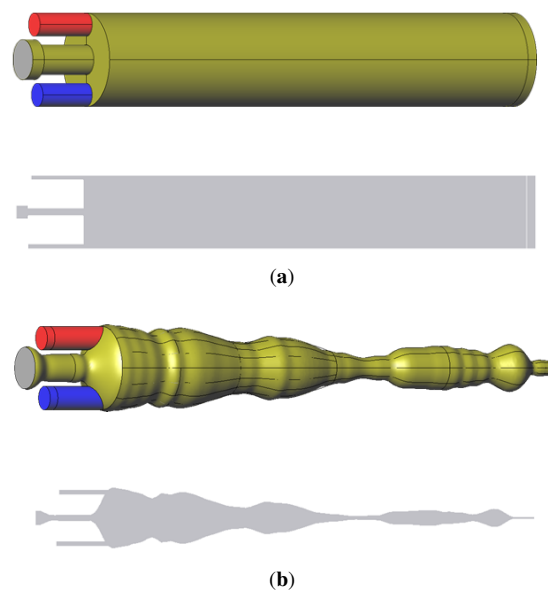
marc.arnela@salle.url.edu

## ABSTRACT

The piriform fossae are two small side branches located close to the laryngeal cavity that behave as quarter-wave resonators, producing two deep dips on speech spectrum around 4-5 kHz. Three-dimensional (3D) vocal tract acoustic models have shown that these side branches acoustically interact in the generation of these antiresonances. That is, if one piriform fossae is modified, it changes the frequency and amplitude of the two dips. However, 3D acoustic models, although very accurate, are computationally expensive. Two-dimensional (2D) approaches based on tuned 2D vocal tracts could be a possible alternative, since they provide a good balance between 1D and 3D in terms of accuracy and computational cost. This 2D methodology consists in four tuning steps in which the 2D vocal tract is modified to recover, to a large extent, the formant locations, bandwidths, and energies of a 3D vocal tract with circular cross-sections. This method has successfully been applied to generate vowels and diphthongs. However, side branches have still not been considered. In this work, we extend this 2D tuning methodology to incorporate side branches such as the piriform fossae in the numerical generation of vowels. The Finite Element Method (FEM) in the time domain is used to perform simulations in 3D and 2D. For comparison purposes, 1D simulations are also carried out, using a frequency domain approach based on the Transfer Matrix Method (TMM). Results show that, in contrast to 1D approaches, the presented 2D approach can deal with the acoustic interaction of the side branches, largely recovering 3D results, while reducing the computational costs.

## 1. INTRODUCTION

The acoustics of the vocal tract has typically been modeled using one-dimensional (1D) techniques. However, these approaches cannot correctly reproduce its 3D behaviour as they rely on the plane wave propagation assumption, which holds up to about 5 kHz. Three-dimensional (3D) models [1] could be used instead. However, they are more computationally expensive. Two-dimensional (2D) models might be a possible alternative. These models have successfully been applied for the simulation of vowels [2] and diphthongs [3], and they have shown to provide a good balance in terms of accuracy and computational cost. However, they still have not incorporated the piriform fossae,



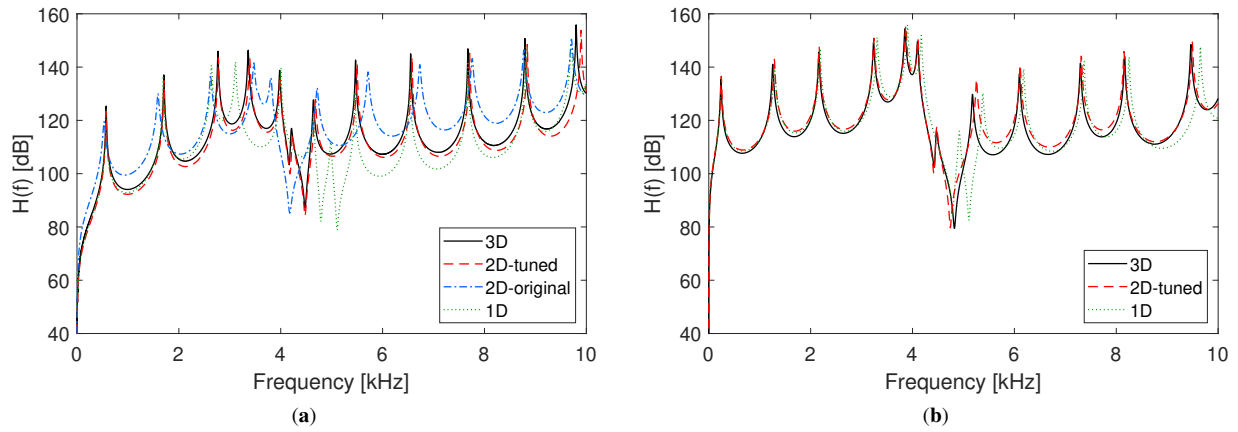
**Figure 1:** Vocal tract geometry in 3D and 2D for (a) the three-tube benchmark and (b) for the vowel /u/.

two small side branches, located close to the laryngeal cavity, that produce two prominent dips on the voice spectrum around 4-5 kHz [4]. The main goal of this work is to incorporate these two side branches in a 2D model.

## 2. METHODS

We have used the 2D model described in [2] and [3] for the simulation of vowels and diphthongs, respectively. This 2D model is based on a four-step tuning methodology that adapts the vocal tract shape, boundary admittance at the vocal tract walls, glottal source and radiation losses to mimic the behaviour of a straightened 3D vocal tract with circular cross-sections set in a sphere. The last step, however, has not been considered, since in this work radiation losses have been deactivated to better observe the effects of the piriform fossae.

The main purpose of this work, as said, is to incorporate the piriform fossae in the above 2D model. This means that the shape and wall losses of this new area have to be determined. To do so, we have followed a similar methodology than when tuning the main tract. That is, the expansion ratio when moving from the side branch to the main tract has



**Figure 2:** Vocal tract transfer function for (a) the three-tube benchmark and (b) for the vowel /u/ computed with the 3D model, the proposed 2D-tuned methodology, the 2D standard model, and the 1D model.

been set equal in 2D and 3D. This has lead to expressions of the shape and boundary admittance coefficient for the piriform fossae similar to those of the main tract.

The above methodology has first been tested with the three-tube benchmark in [4], which is made of three tubes conforming the main tract and two side branches that model the piriform fossae (see Figure 1a). Secondly, vowel vocal tracts have been generated with the area functions in [5] and combined with the side branches of [4] (see Figure 1b). Numerical simulations have next been performed using the custom FEM code described in [3]. These include not only simulations with the proposed 2D tuning methodology (2D-tuned), but also 2D cases without applying any correction (2D-original), the reference 3D configurations, and also 1D simulations, performed in this case with a frequency-based approach based on the transfer matrix method [3]. The vocal tract acoustic response  $H(f)$ , defined as the ratio between the acoustic pressure at the vocal tract exit and the volume velocity at the entrance, has finally been computed for each case.

### 3. RESULTS

Figure 2 shows the results obtained for the three-tube benchmark and the vowel /u/ obtained from 3D, 2D with the proposed tuning methodology (2D-tuned), 2D without applying any correction (2D-original), and 1D.

Focusing first in the three-tube benchmark (Fig. 2a), it can be observed that not only resonances show up, which are produced by longitudinal modes of the main tract, but also that two antiresonances appear because of the two piriform fossae. Note that the 2D-tuned curve perfectly matches the 3D configuration. In contrast, if we do not apply any correction (2D-original), results depart not only for the two antiresonances, but also for all formants. As far as 1D is concerned, it has correctly predicted formant frequencies, although the two antiresonances have shifted to higher frequencies. This can be attributed to the fact that 1D cannot consider the acoustic interaction between the main tract and the two side branches, whereas 3D and 2D do naturally take into account this effect. Similar occurs

for the vowel /u/ (Fig. 2b). Note in this case that the proposed 2D-tuned approach greatly follows the results obtained in 3D, whereas 1D again fails in predicting the frequencies of the piriform fossae.

### 4. CONCLUSIONS

A 2D tuning methodology has been proposed in this work to model the acoustics of vocal tracts with piriform fossae. It has been shown that 3D results can be recovered to a large extent while reducing the computational costs. Moreover, the accuracy of the proposed 2D approach has shown to be clearly higher compared to 1D, given that the latter cannot deal with the acoustic interaction between these two side branches and the main tract.

### 5. 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 Transactions on Audio, Speech, and Language Processing*, vol. 36, no. 2, pp. 2173–2182, 2019.
- [2] M. Arnela and O. Guasch, "Two-dimensional vocal tracts with three-dimensional behaviour in the numerical production of vowels," *J. Acoust. Soc. Am.*, vol. 135, no. 1, pp. 369–379, 2014.
- [3] M. Arnela and O. Guasch, "Finite element synthesis of diphthongs using tuned two-dimensional vocal tracts," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 10, pp. 2013–2023, 2017.
- [4] 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.
- [5] 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," *J. Acoust. Soc. Am.*, vol. 119, no. 2, pp. 1037–1049, 2006.