DINAME 2017 - Proceedings of the XVII International Symposium on Dynamic Problems of Mechanics
A. T. Fleury, D. A. Rade, P. R. G. Kurka (Editors), ABCM, São Sebastião, SP, Brazil, March 5-10, 2017

Identification of a vocal tract configuration using genetic algorithms

Álvaro Campos Ferreira[1], **Gino Bertollucci Colherinhas**[1], **Marcus Vinicius Girão de Morais**[1]

[1] Universidade de Brasília, Campus Darcy Ribeiro, Faculdade de Tecnologia, Bloco G, Asa Norte, 70910-900 - Brasília/DF.
alvaro.campos.ferreira@gmail.com, ginobertollucci@hotmail.com, mvmorais@unb.br

Abstract: The complexity of the human voice has been subject to various acoustic studies over the last decades. It is possible to construct numeric models of the Vocal Tract (VT) and its complex geometry using experimental data from MRI procedures during the phonation of vowels. In this paper, analytical and numerical models for the acoustic behavior of the VT are constructed using an approximated geometry obtained by experimental measurements. An analytical solution using a transfer matrix method is used to construct the acoustic model. The parameters for the geometry are optimized using a Genetic Algorithm (GA) toolbox developed by the authors to approximate the Frequency Response Function (FRF) of the models to the experimental results. We restrict the length and the cross-section areas of the vocal tract geometry to approximate the frequency and the amplitude of the resonance peaks to the experimental results. The acoustic behavior of the optimized results are compared to BEM and FEM models.

Keywords: Acoustic, Genetic Algorithms, Transfer Matrix Method, Vocal Tract.

INTRODUCTION

The study of the acoustics of the voice is very important to many areas, including phonoaudiology, music and noise control. Many professionals use the voice as their main media to convey information and ideas, such as teachers, singers, orators and theater professionals. It is desirable to produce a resonant voice to reduce the strain and fatigue while speaking or singing. This resonant quality is already well known by bel-canto singers and theater professionals, but its production is still based in body sensations of the speaker (Titze, 2001).

Ekholm, Papagiannis, and Chagnon (1998) tries to fill the gap in the communication between voice pedagogues and voice scientists. Subjective ratings were related to objective measurements taken from acoustic analysis of the voice signal. Some acoustic phenomena correlations to critical perceptual parameters were identified, helping to bridge the terminology gap between vocal artists and scientists.

The study of the acoustic properties of the resonant voice and Vocal Tract (VT) during phonation are an essential step for the advance of vocal technology and vocal coaching for professionals which use spoken speech. A schematic model of the speech system is shown in Figure 1. The VT is the region contained between the glotis and mouth.

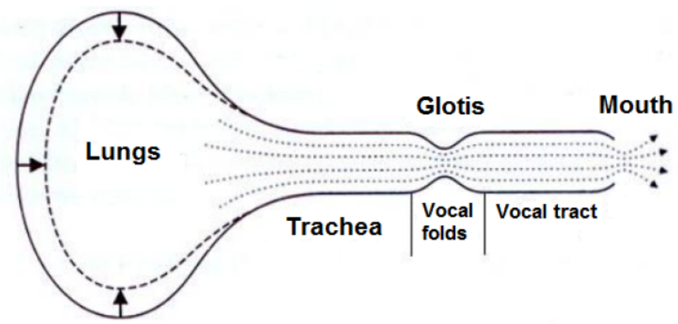

Figure 1 – Schematic model of the speech system (Cataldo, Sampaio, & Nicolato, 2004)

The acoustic resonance of the VT defines the quality of the vowel produced. Different VT configurations produce different modes and resonance frequencies, perceived by listeners as different vowels. A study of the resonance strategies of 22 singers shows that trained singers may use the resonance of the vocal tract to obtain certain desirable characteristics in the voice (Henrich, Smith, & Wolfe, 2011). The two first resonance frequencies of the VT were obtained experimentally. These results indicate that singers can repeatedly tune their resonances with precision, and that there can be considerable differences in the resonance strategies used by individual singers, particularly for voices in the lower ranges.

The aim of this study is to determine the best fitness function suitable to obtain the configuration of the VT given its acoustic behavior. This will be carried out using a Genetic Algorithm (GA) approach, in which the characteristics of the VT model will be selected based on the acoustic output produced by the model. The GA used is a floating point chromosome, which was used to maximize a fitness function defined by restricted parameters. Physical properties of the system such as resonance mode frequency and mode shapes were used to construct the fitness functions.

TRANSFER MATRIX METHOD (TM)

While studying cavity acoustics, one may encounter many types of long enclosures, such as pipes or VT. These cavities will behave like a one-dimensional system in the low frequency range, as longitudinal modes will possess a much lower frequency than transversal modes. In these cases, an interesting mathematical formulation of the problem is the TM (Gibert, 1988). This method is similar to other transfer matrix formulations as it relates the values of parameters in a physical point of the system with values in another point. For the acoustics of one-dimensional cavities, the acoustic pressure and the pressure variation are the variables to be determined in each end of the enclosure. The TM relates the values of these parameters at the beginning and at the end of the cavity.

Consider the homogeneous acoustic wave equation in the frequency domain 1.

$$\frac{\partial^2 p}{\partial x^2} = k^2 p \tag{1}$$

where $k = \omega/c$ is the wave number.

One solution of equation 1 may be written as shown in equation 2.

$$p(x,\omega) = A\sin kx + B\cos kx \tag{2}$$

It is possible to define the mass flux Q through a cylinder of area S as written in equation 3.

$$Q = \rho_f V_f = \rho_f S\dot{X}_f \tag{3}$$

where $V_f = S\dot{X}_f$ and $X_f = X_f(x,t)$ are the velocity and the displacement of a fluid particle through time.

From the dynamic balance of forces in a fluid particle in Δx, which tends towards zero,

$$\frac{\partial p}{\partial x} + \rho \ddot{X}_f = 0 \tag{4}$$

Replacing 3 and 2 in 4, is possible to write equation 5 after some algebraic manipulation.

$$q = \frac{iS}{c}(A\cos kx - B\sin kx) \tag{5}$$

From the relationships given by 2 and 5, applying the boundary conditions for points 1 ($x = 0$) and 2 ($x = L$), one obtains a relationship between the pressure (p) and the pressure variation (q) in the straight tube shown. This relationship is shown in equation 6.

$$\begin{Bmatrix} p_2 \\ q_2 \end{Bmatrix} = \begin{bmatrix} \cos\gamma & \frac{c}{iS}\sin\gamma \\ \frac{-iS}{c}\sin\gamma & \cos\gamma \end{bmatrix} \begin{Bmatrix} p_1 \\ q_1 \end{Bmatrix} \tag{6}$$

where c is the acoustic wave propagation speed, p_i and q_i are the values at point i of the pressure and the pressure variation, respectively; S is the cross-section area of the tube, L is the length of the tube, $k = \omega/c$ is the wavenumber and $\gamma = k \cdot L$ is the phase change over a distance L.

The TM method can also be used to approximate the behavior of concentric cavities, by coupling the cavities using the following procedure. Consider 3 cavities, each with length L_1, L_2 and L_3; and a constant cross-section area S_1, S_2 and S_3. If all the cavities are coupled, the relationship between points 1 and 3 are given by equation 7.

$$\begin{Bmatrix} p_3 \\ q_3 \end{Bmatrix} = A_{(3)(1)} \begin{Bmatrix} p_1 \\ q_1 \end{Bmatrix} \tag{7}$$

where $A_{(3)(1)}$ is the TM from point 1 to 3. Applying this same thought to points 1 to 2, this relationship can be written as shown in equation 8.

$$\begin{Bmatrix} p_3 \\ q_3 \end{Bmatrix} = A_{(3)(2)} \cdot A_{(2)(1)} \begin{Bmatrix} p_1 \\ q_1 \end{Bmatrix} \tag{8}$$

Generalizing for N cavities, equation 9 is obtained.

$$\begin{Bmatrix} p_N \\ q_N \end{Bmatrix} = A_{(N)(N-1)} \cdot A_{(N-1)(N-2)} \cdot A_{(N-2)(N-3)} ... A_{(3)(2)} \cdot A_{(2)(1)} \begin{Bmatrix} p_1 \\ q_1 \end{Bmatrix} \tag{9}$$

The boundary conditions are placed at points 1 and N. One may write equation 9 as 10 and use table 1 to determine the conditions necessary to obtain the desirable boundary conditions.

$$\begin{Bmatrix} p_N \\ q_N \end{Bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{Bmatrix} p_1 \\ q_1 \end{Bmatrix} \tag{10}$$

Table 1 – Relationships to determine the resonance frequency for different boundary conditions using TM

Boundary condition	Necessary condition
Open-Open	$A_{12} = 0$
Open-Closed	$A_{22} = 0$
Closed-Open	$A_{11} = 0$
Closed-Closed	$A_{21} = 0$

CASE STUDY DESCRIPTION

Using MRI visual procedures Clément et al. (2007) obtained geometric informations of the VT. Table 2 shows the cross-section areas of the vowel \\a\\.

Table 2 – Area A_i and standard deviation σ_i of the cross-section areas of the VT for the vowel \\a\\ in cm^2 (Clément et al., 2007, modified)

Cross-section area (cm^2)															
i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A_i	1,8	1,8	2,8	1,5	0,8	1,3	1,5	1,7	2,8	4,5	7,1	9,3	13,5	15,5	7,8
σ_i	0,1	0,1	0,3	0,2	0,1	0,1	0,1	0,2	0,2	0,3	0,2	0,4	0,3	0,4	0,2

The cross-sectional areas are essential dimensions to identify the acoustic behavior of the VT. These parameters will be investigated using a GA optimization. Ferreira (2014) created meshes of the VT and applied them into BEM, FEM and TM models. These models are compared to the GA optimization in the results section.

GENETIC ALGORITHM (GA) OPTIMIZATION

A Genetic Algorithm (GA) is a stochastic algorithm that mimics natural phenomena as operators. GAs are search techniques based on the processes of natural selection for survival through population genetics (Holland, 1975). For the GAs to start evolving, we can use the steps selection, recombination (crossover), mutation, and replacement, where the survival-of-the-fittest mechanism can be applied to the candidate solutions (Haupt & Haupt, 1998).

In this section we apply the GA MATLAB toolbox developed in other research (Colherinhas, 2016). The algorithm use float points, elitism, blend crossover (BLX-α) (Eshelman & Schaffer, 1993), and creep mutation. We define restrictions for cross-section areas to optimize the vocal tract. The initial population $C = [A_i]$ is created by restricting the variables in the following lanes:

$$0.50 \leq A_i \leq 17.00$$

where $i = 1, 2, \ldots, 14, 15$ are the corresponding sections.

Fitness Function

Three different fitness functions are analyzed to reach the experimental data of Table 2.

(a) Fitness function 1: Modal frequency

The first fitness function devised uses the resonance frequency of the first n_{freq} acoustic modes to restrict the cross-section area and select the individuals. The value of the resonance frequency is obtained by selecting the peaks in the FRF of the system. The fitness function created controls the cross-section areas by selecting the individuals with the maximum value of the function showed in equation 11. This function is the square root of the inverse of the square difference between the goal value of the resonance frequency ω^i_{goal} and the value obtained for individual r, ω_r^i for each mode i.

$$fit_1(r) = \left[\sum_{i=1}^{N} (\omega_{obj}^i - \omega_r^i)^2 \right]^{-\frac{1}{2}} \tag{11}$$

where N is the number of physical points being analyzed by the program.

(b) Fitness function 2: Pressure mode shape

One may use the value of the pressure along the tube to create a fitness function. The same operation of the Eqn. 11 is realized replacing the frequency into acoustic pressure values:

$$fit_2(r) = \left[\sum_{i=1}^{N} (p_{obj}^i - p_r^i)^2 \right]^{-\frac{1}{2}} \tag{12}$$

(c) Fitness function 3: Pressure variation mode shape

Similarly the value of the pressure variation at determined positions can be used to construct a fitness function defined by Eqn. 13

$$fit_3(r) = \left[\sum_{i=1}^{N} (q_{obj}^i - q_r^i)^2 \right]^{-\frac{1}{2}} \tag{13}$$

Fitness function evaluation

Preliminary results were conducted to test the convergence of the fitness in function of the cross-sectional areas. Two models with four areas of cross section $[1,1,1,1]cm^2$ and $[1,1,2,2]cm^2$ were performed. Table 3 and 4 shows the cross-sectional areas obtained using these three fitness functions.

Table 3 – Comparison of the cross-section areas for different fitness functions. Goal: [1,1,1,1] (cm^2)

	S_1	S_2	S_3	S_4
Goal	1.000	1.000	1.000	1.000
Fitness function 1	2.126	1.849	1.867	2.149
Fitness function 2	2.228	2.228	2.228	2.228
Fitness function 3	0.999	0.998	1.005	0.999

Table 4 – Comparison of the cross-section areas for different fitness functions. Goal: [1,1,2,2] (cm^2)

	S_1	S_2	S_3	S_4
Goal	1.000	1.000	2.000	2.000
Fitness function 1	0.802	0.899	1.815	1.619
Fitness function 2	0.938	0.939	1.877	1.878
Fitness function 3	1.001	0.998	2.003	1.991

A summary with convergence (OK) and non-convergence (X) is shown in Table 5

Table 5 – Summary of convergence of fitness functions

	Fitness 1	Fitness 2	Fitness 3
Modal frequency	OK	OK	OK
Pressure modal shape	X	OK	OK
Pressure variation modal shape	X	X	OK

Therefore the fitness function 3 is the best choice to be optimized.

RESULTS

Fitness function fit_3 provides the best results for the desired optimization. Therefore, this analysis establishes for the genetic toolbox the following parameters:

- $N_{ger} = 50000$, the number of generations;
- $N_{ind} = 150$, the number of individuals in the population;
- $p_c = 60\%$, crossover probability;
- $p_m = 4\%$, mutation probability;
- $p_{elit} = 2\%$, elitism probability;
- $p_{diz} = 20\%$, decimation probability;
- $N_{diz} = 100$, the step of generations for the occurrence of decimation.

The number of modes used in the optimization was $n_{freq} = 10$.
Figure 2 shows the graphical development of the best individuals and their mean over the generation.

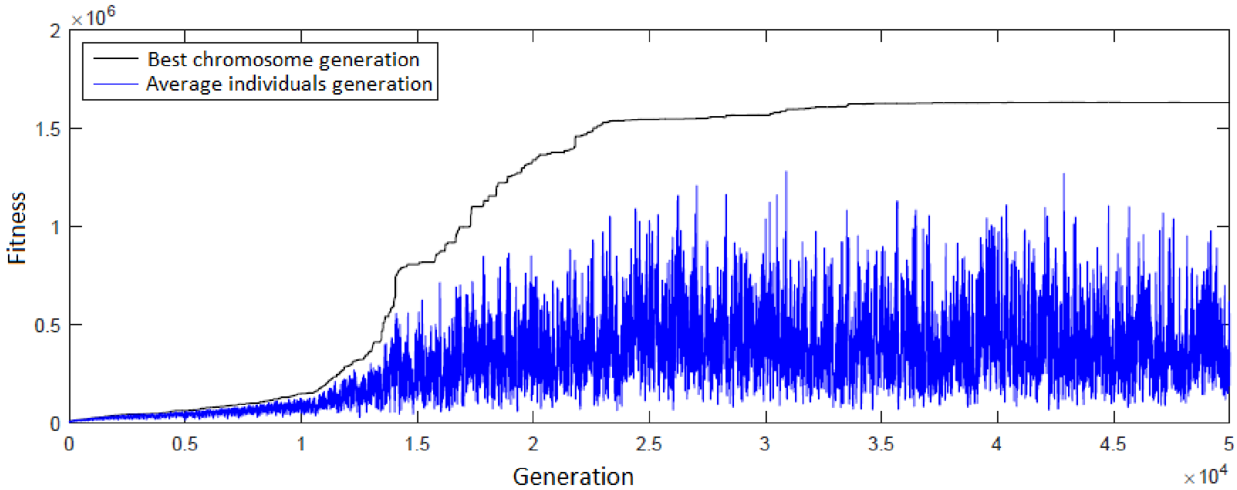

Figure 2 – Fitness function of the best chromosomes and their means over the generation

Table 6 presents the results obtained by the optimization and compares them to the fitness values, with a relative error. The largest relative error is 3,51%.

Table 6 – Cross-sectional areas A_i of the VT for the vowel \\a\\ in cm^2 (Clément et al., 2007) and the best chromosome founded by the GA optimization toolbox

Cross-sectional areas (cm^2)							
Section	1	2	3	4	5	6	7
A_{fit}	1,80	1,80	2,80	1,50	0,80	1,30	1,50
A_{AG}	1,77	1,78	2,79	1,53	0,82	1,33	1,54
error (%)	1,47	0,63	0,06	2,05	2,77	2,21	2,65

8	9	10	11	12	13	14	15
1,70	2,80	4,50	7,10	9,30	13,50	15,50	7,80
1,75	2,89	4,66	7,34	9,59	13,91	16,00	8,07
3,04	3,11	3,51	3,41	3,16	3,02	3,24	3,40

Figure 3 shows the FRF result obtained by GA optimization, compared to the goal produced by the TM. The frequency response of the acoustic behavior obtained by the GA is practically the same as the fitness function.

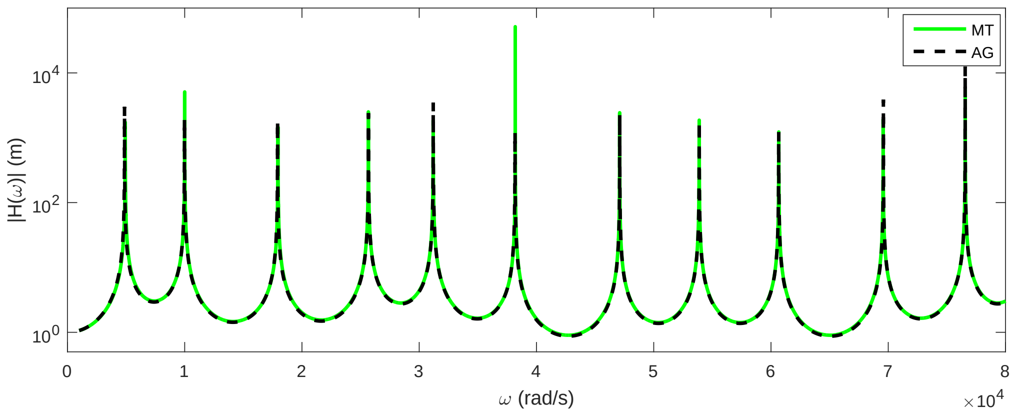

Figure 3 – FRF comparison between TM and GA optimization

Figure 4 shows the comparison between the objectives achieved by the TM and the optimization using pressure mode shape values.

Figure 5 shows the corresponding MAC values for these first 10 modal forms. You may notice that there is a strong correlation between the vibration modes of the diagonal elements with MAC values nearly equal to one. Other values outside the main diagonal have practically zero values. This indicates that there is a strong correlation between the pressure mode shapes optimization obtained by the GA and the goal. Between the second and third modes exists a poor orthogonality with MAC values equals to 0.2.

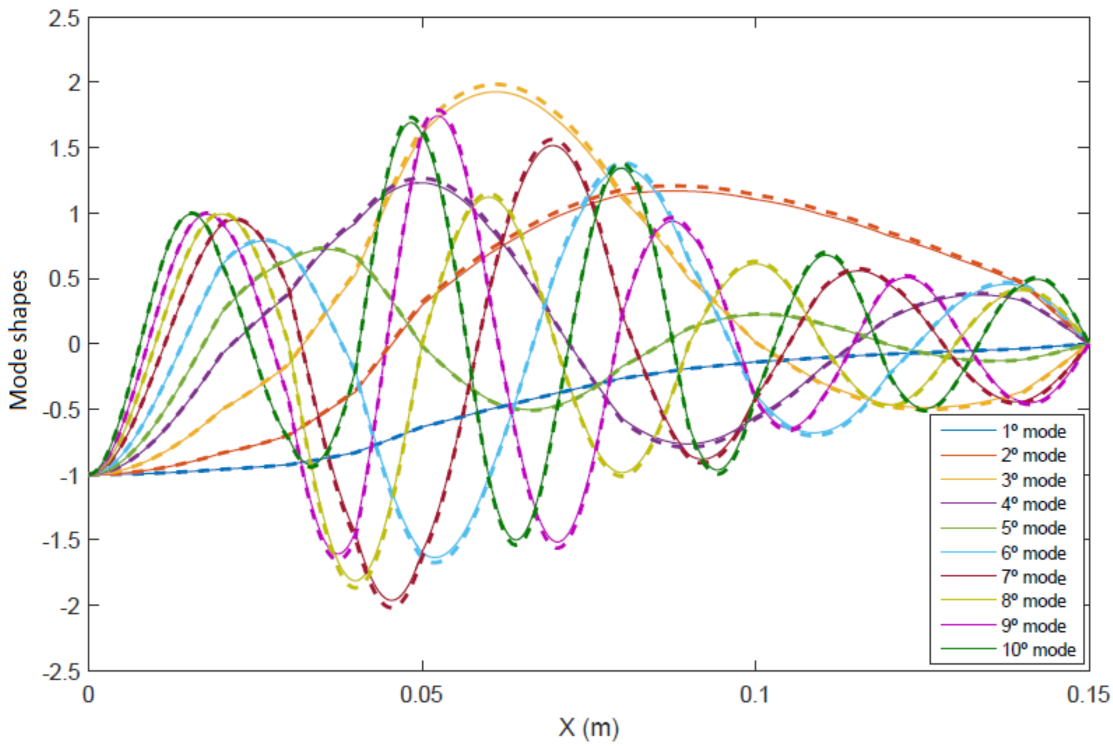

Figure 4 – Pressure mode shapes comparison between the TM and the GA optimization (dashed) for the firsts 10 mode shapes

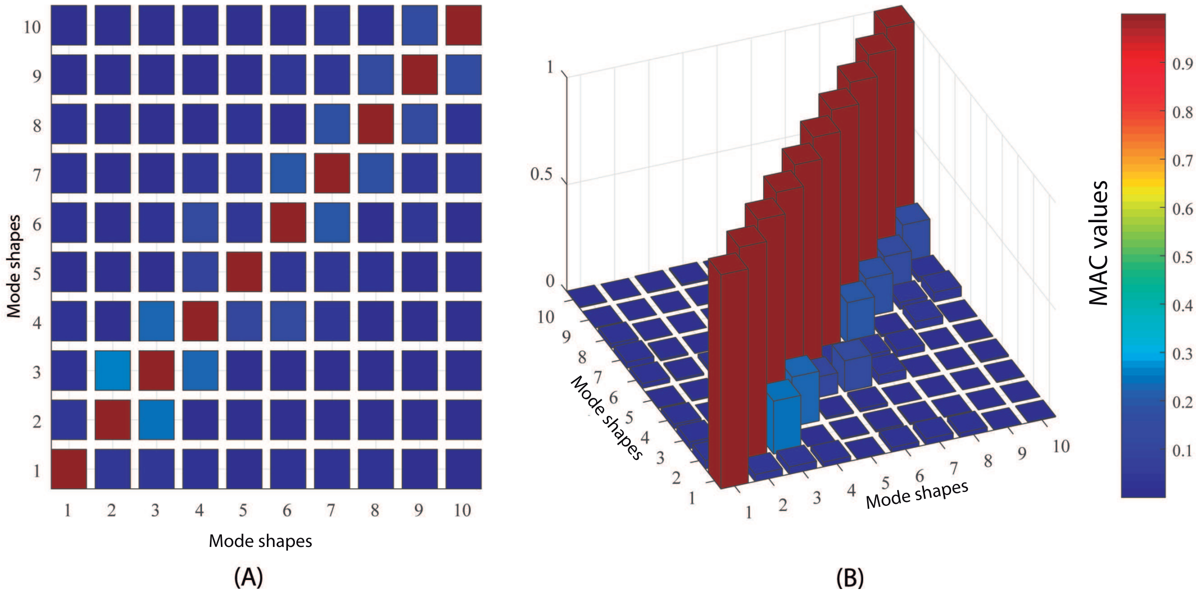

Figure 5 – MAC values for the firsts 10 pressure mode shapes

Figure 6 shows a comparison between the objectives achieved by the TM and using the GA toolbox for the pressure variation mode shapes.

Figure 7 presents the corresponding MAC values for the firsts 10 mode shapes.

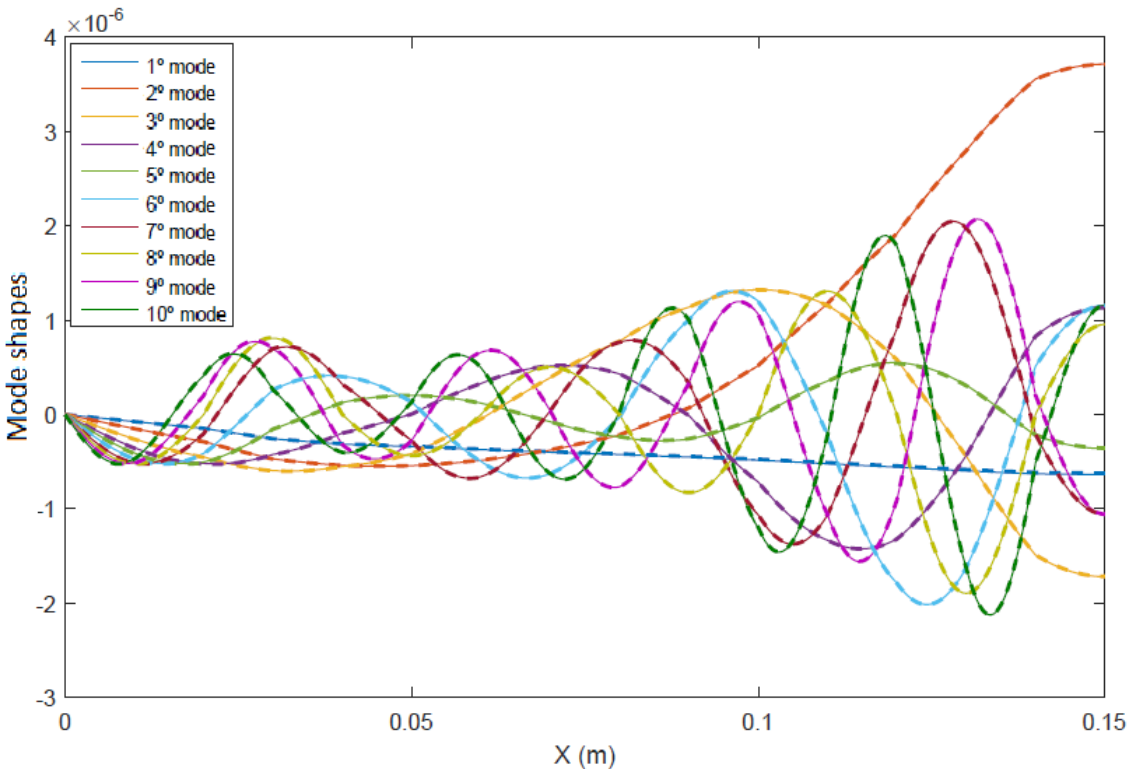

Figure 6 – Pressure variation mode shapes comparison between the TM and the GA optimization (dashed) for the firsts 10 mode shapes

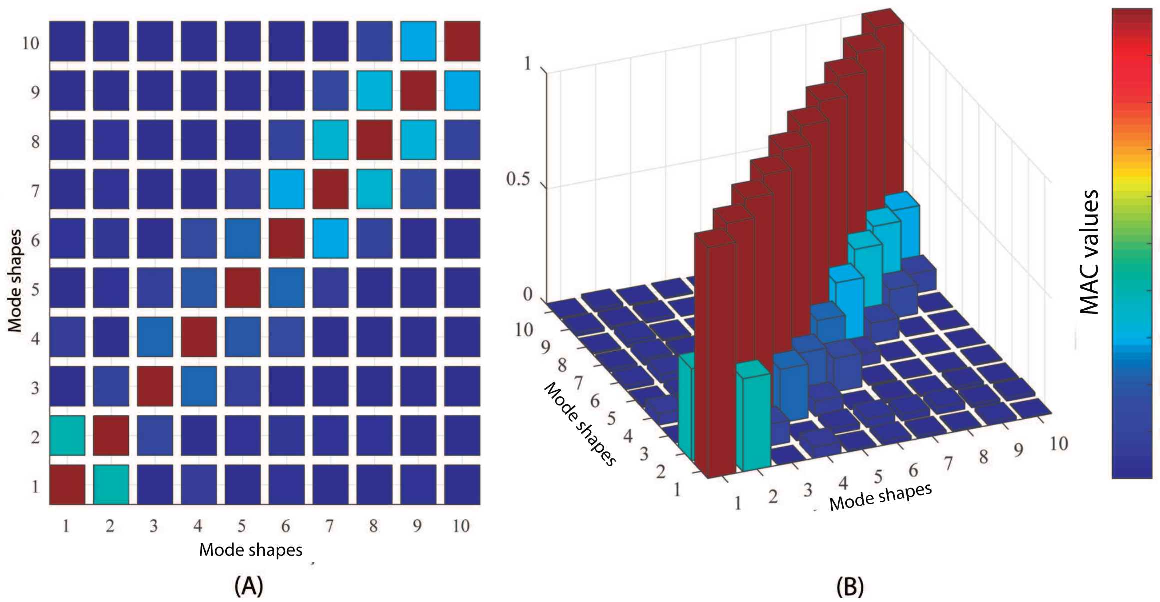

Figure 7 – MAC values for the firsts 10 pressure variation mode shapes

There is a strong correlation between the vibration modes of the diagonal elements, as they have MAC values almost equal to one. Other values have virtually nil, except for nearby vibration modes (modes 1 and 2, 2 and 3, ...). There average correlation between the first and second order of about 40%.

Comparison and discussion of the used methods

In this section we compared the Boundary Element Method (BEM) and Finite Element Method (FEM) with the GA optimization results for the firsts three normalized mode shapes in figures 8, 9 and 10.

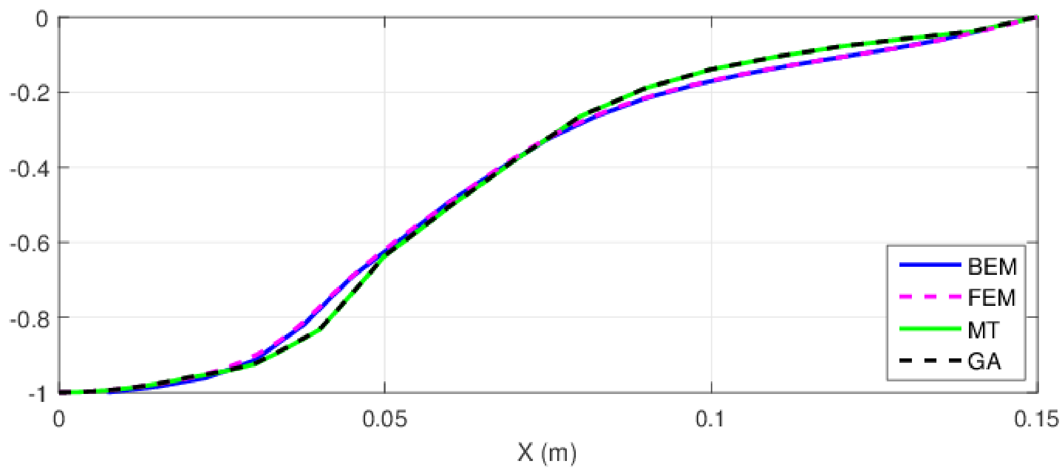

Figure 8 – First mode shape: comparison between BEM, FEM, TM and GA

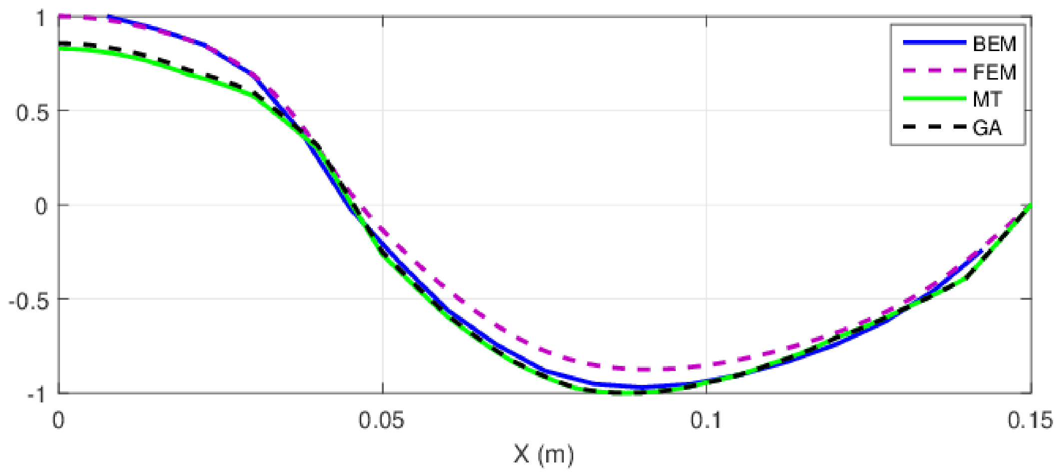

Figure 9 – Second mode shape: comparison between BEM, FEM, TM and GA

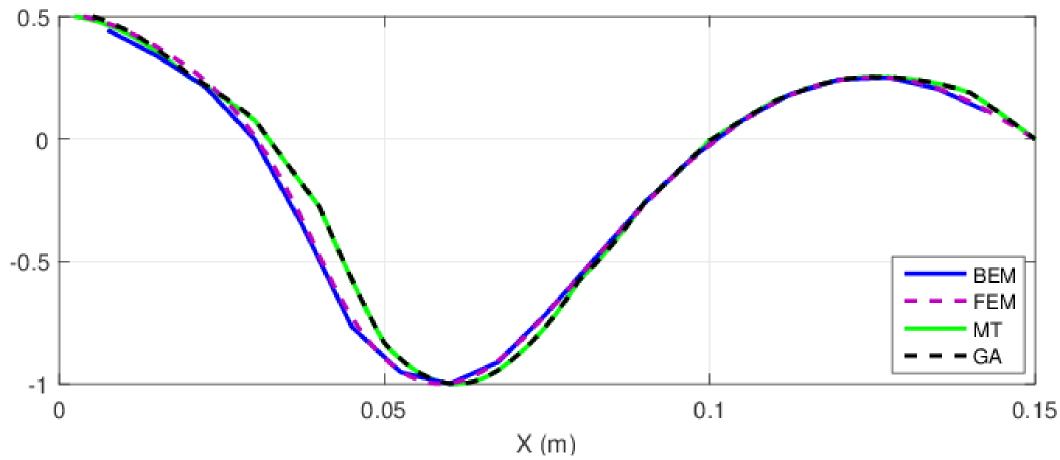

Figure 10 – Third mode shape: comparison between BEM, FEM, TM and GA

The TM results are similar to the BEM and FEM (Ferreira, 2015) results. The TM analytical formulation is a powerful tool that allows the GA implementation for achieving a specific goal.

The difference between the TM model in relation of the BEM and FEM models can be explained by the absence of three-dimensional effects, not implemented in the TM analytical model. There is almost no difference between the GA and TM graphical models, indicating the GA efficiency.

CONCLUSION

This paper presents an implementation of a genetic algorithm for one-dimensional acoustic tubes using the transfer matrix method analytic solution. An analysis of three fitness function using different parameters of the model was carried out. A fitness function using the pressure variation along the tube of both the goal and the individual tubes was used in the final genetic algorithm. The algorithm was able to reflect the cross section areas of an arbitrary tube with constant length. The use of pressure variation mode to construct the fitness function is adequate, as the pressure variation is a dual variable and has a direct influence from both the position in length and the cross section area of the tube. The use of the pressure mode to construct the fitness function does not yield the expected values as the pressure is a primal variable and describes a state of the section and not its dynamic. The use of the resonance frequency to construct the fitness function does not guarantees that the solution is unique, as many different acoustic configurations can generate the same FRF and yet show completely different acoustic behavior.

Two validation cases were carried out using three different fitness functions. The fitness function which obtained the best results used the pressure variation difference between the goal case and the individual. For these cases, only four variables were controlled by the genetic algorithm and the results were consistent. Afterwards, a vocal tract model based on anatomical data was created and the genetic algorithm was used to attempt to reproduce the cross section area of the vocal tract. This simulation used 15 variables of cross section area and a constant length. The cross section areas converged to the goal case with 3.4% relative error. MAC values guarantees the orthogonality between the optimization and TM analytical formulation for the pressure and pressure variation mode shapes. The difference between the TM model in relation of the BEM and FEM models can be explained by the absence of three-dimensional effects, not implemented in the TM analytical model. There is almost no difference between the GA and TM graphical models, indicating the GA efficiency.

ACKNOWLEDGMENTS

We like to thank CNPq for the research scholarship granted.

REFERENCES

Cataldo, E., Sampaio, R., & Nicolato, L. (2004). Uma discussão sobre modelos mecânicos de laringe para síntese de vogais.

Clément, P., Haus, S., Hartl, D. M., Maeda, S., Vaissière, J., & Brasnu, D. (2007). Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method.

Colherinhas, G. B. (2016). Dissertação de mestrado em ciências mecânicas: Ferramenta de otimização via algoritmos genéticos com aplicações em engenharia.

Ekholm, E., Papagiannis, G. C., & Chagnon, F. P. (1998). Relating objective measurements to expert evaluation voice quality in western classical singing" critical perceptual parameters. *Journal of Voice*, *12(2)*, -.

Eshelman, L. J., & Schaffer, J. D. (1993). Real-coded genetic algorithms and interval-schemata. , 187-202.

Ferreira, A. C. (2014). Dissertação de mestrado em ciências mecânicas: Análise harmônica de cavidades acústicas pelo método dos elementos de contorno direto.

Ferreira, A. C. (2015). Analytical and numerical modeling of vocal tract in vowel phonation.

Gibert, R. (1988). *Vibrations des structures: interactions avec les fluides, sources d'excitation aléatoires*. Eyrolles. Retrieved from `https://books.google.com.br/books?id=EO2bMQEACAAJ`

Haupt, R. L., & Haupt, S. E. (1998). *Practical genetic algorithm*. John Wiley G. Sons Inc.

Henrich, N., Smith, J., & Wolfe, J. (2011). Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. *journal of the acoustical society of america*, *129(2)*, -.

Holland, J. (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press. Retrieved from `https://books.google.com.br/books?id=JE5RAAAAMAAJ`

Titze, I. R. (2001). Acoustic interpretation of resonant voice.

RESPONSIBILITY NOTICE