Development of a velocity compensation factor for uniform Transmission Line Matrix Method (TLM) meshes applied to voice production

Alexandre de Souza Brandão, abrand@operamail.com
Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal Fluminense, Rua Passo da Pátria 156, São Domingos, Niterói, 24210-240, RJ, Brazil
Edson Cataldo, ecataldo@im.uff.br
Applied Mathematics Department, Universidade Federal Fluminense, Rua Mário Santos Braga S/N, Valonguinho, Niterói, 24020-140, RJ, Brazil.
Fabiana Leta, fabiana@ic.uff.br
Departamento de Engenharia Mecânica, Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal Fluminense, Rua Passo da Pátria 156, São Domingos, Niterói, 24210-240, RJ, Brazil

Abstract. *In this paper, single tube models are used to obtain a velocity compensation factor for uniform Transmission Line Matrix Method (TLM) meshes and a voice signal is generated by simulating the acoustic propagation of a glottal signal, obtained by inverse filtering of the voice signal, through a more realistic /a/ vowel shaped vocal tract TLM mesh. The meshes are constructed over the voxels (Volumetric Picture Elements) in segmented medical image sequences. The segmentation of the images is performed via neural network, island removal and some manual adjustments. The tools for segmentation, mesh extraction and the TLM itself were implemented as part of the ModaVox program, which is an open source application developed during this work. The developed velocity compensation factor provides results that match the analytical ones available in the literature and can also be applied to other structures like the two-tube model, used in voiced sounds generation, or even in a more realistic /a/ vowel shaped vocal tract. The methodology was validated, firstly, simulating the acoustic propagation of a probe signal through tube model meshes. Then, the same probe signal was used as the input for the vocal tract TLM mesh and the resonant frequencies obtained by simulation were compared with those of the real vocal tract. Finally, the glottal signal was used as input for the vocal tract mesh and the reflective boundary conditions of the vocal tract walls were scaled to reproduce the consistency of the soft tissues. The simulated voice signal was then compared with the original voice signal generated by the subject when his vocal tract was imaged. The results showed good realism and accuracy.*

Keywords: *Transmission Line Matrix, Vocal Tract, Numerical Model*

1 INTRODUCTION

The comprehension of the human voice production process, due to its complexity and flexibility, remains an open topic. Many researchers have studied the vocal tract acoustics over the years. In several works (Baer *et al.*, 1991; Story *et al.*, 1996; Takemoto *et al.*, 2006), Magnetic Resonant Images (MRI) have been successfully used to obtain information about the structure of the human vocal tract. However, anisotropic images can affect the results. For example, Clement *et al.* (2007) calculated the formants values for the French vowels /a/, /i/ and /u/, by measuring the cross-sectional areas of the vocal tract using MRI and setting the areas as input for a voice synthesizer. In all cases, the obtained formant values were significantly different from the real ones. The discrepancies were attributed to the use of 4 mm slice thickness.

Some finite element method (FEM) approaches (Motoki, 2002; Vohradnik *et al.*, 2003; Matsuzaki and Motoki, 2006) have used modal analysis to determine the formants. In (Matsuzaki and Motoki, 2006) the authors used the FEM to study the acoustic characteristics of the vocal tract in a 3D model for the Japanese /a/ vowel. The disagreement of the formant frequencies between simulated and real voice was attributed to the need of adjusting the soft tissue walls boundary condition to a more realistic one. Moreover, in FEM modelling, one solution for the acoustic field has to be calculated for each inspected frequency.

The transmission line matrix method (TLM) (Johns and Beurle, 1971) is easy to implement and to set boundary conditions for. It produces a time-domain output signal, which can be easily converted to the frequency domain using the Fast Fourier Transform algorithm. Moreover, in TLM, the frequency response of a given model can also be obtained by applying the Linear Predictive Coding (LPC) technique to the time-domain output signal. Perhaps, the first paper where TLM was applied to vocal tract modelling was the one written by El-Masri *et al.* (1998). In their work, a 3D vocal tract was built with several rectangular duct sections. The cross-sectional areas were obtained from MRI measurements in a real vocal tract, but the vocal tract was not reconstructed. In another work (Katsamanis and Maragos, 2008), the authors

investigated the potential of properly applying the 3D TLM to simulate the acoustic field in the vocal tract for the synthesis of fricatives in the higher end of the spectrum where the planar wave propagation assumption could not provide accurate results. The applied vocal tract geometry was determined from 3D MR images.

In this paper, the TLM is used to study the acoustic propagation through the vocal tract in the 0 to 10000 Hz range. The 3D vocal tract TLM mesh is constructed over a segmented MR volume image and a soft tissue walls boundary condition is applied. Moreover, the definition of a velocity compensation factor solved the problem of the apparent velocity, one of the major drawbacks in TLM. Perhaps, this is the first TLM model used to generate a real /a/ vowel sound directly from the simulation in a mesh constructed from MRI obtained from a real vocal tract.

This paper is organized as follows: Section 2 describes the details of the TLM implementation used herein; Sections 3 and 4 explain how the model meshes were constructed and the details of the simulations, respectively. In Sections 5 and 6 the TLM is applied to the tube models and to the vocal tract model meshes, respectively. Finally, Sections 7 and 8 present the simulations summary and the conclusions.

2 THE TRANSMISSION LINE METHOD (TLM)

The TLM was proposed by Johns and Beurle (1971) and it will be applied here to simulate the acoustic wave propagation through the vocal tract. In TLM the domain is subdivided in a transmission line network, in which the acoustic pulses propagate. A 3D TLM mesh consists of points (or nodes) and normal edges. The nodal values represent the solution $p(\mathbf{r}, t)$ for the acoustic wave equation in terms of the pressure p at a given position $\mathbf{r}(x, y, z)$ and time t. The incoming pulses at each node are the output pulses produced at the neighbor nodes at the previous time step. For each node, the relation between input and output pulses is given by the scattering matrix, for which several models can be found in (Salama, 1997). However, the scattering matrix model that will be used here is that of the lossless 3D shunt node, because it is very suitable and successful in acoustic problems (Cogan *et al.*, 2006). The relation between the transmission (τ) and reflection (ρ) coefficients for the lossless 3D shunt node (Eq. (1)) and the scattering matrix may be easily derived (Cogan *et al.*, 2006).

$$\tau = (1 - \rho) Z_T / Z \,. \tag{1}$$

where Z_T is the impedance at the ending of a given transmission line (or TLM mesh edge) and Z is the characteristic impedance of the same transmission line.

Each transmission line ending is considered in connection to other five transmission lines. Hence, Z_T is given by the summation in parallel of the impedances Z of these transmission lines, Eq. (2).

$$Z_T = Z/5 \tag{2}$$

For all the interior points in the TLM mesh, the characteristic impedances of the transmission lines are equal to the air characteristic impedance Z_0. Hence,

$$\rho = \frac{Z_T - Z}{Z_T + Z} = \frac{\frac{Z_0}{5} - Z_0}{\frac{Z_0}{5} + Z_0} = -2/3 \tag{3}$$

and $\tau = 1/3$, according to Eq. (1) with $Z_T = Z/5$. Since the TLM mesh is uniform, the scattering matrix is given by

$$\begin{bmatrix} V_1^O \\ V_2^O \\ V_3^O \\ V_4^O \\ V_5^O \\ V_6^O \end{bmatrix} = \begin{bmatrix} \rho & \tau & \tau & \tau & \tau & \tau \\ \tau & \rho & \tau & \tau & \tau & \tau \\ \tau & \tau & \rho & \tau & \tau & \tau \\ \tau & \tau & \tau & \rho & \tau & \tau \\ \tau & \tau & \tau & \tau & \rho & \tau \\ \tau & \tau & \tau & \tau & \tau & \rho \end{bmatrix} \begin{bmatrix} V_1^I \\ V_2^I \\ V_3^I \\ V_4^I \\ V_5^I \\ V_6^I \end{bmatrix} \tag{4}$$

where the subscripts represent the node port numbers for input pulses V^I and output pulses V^O.

The boundary values for ρ and τ are given according to the microwave theory analogy for acoustic waves propagation. For example, for rigid surfaces, $\rho = 1$ and $\tau = 0$ because $Z_T = \infty$. To describe a non-impeding opening to air the values used are $\rho = -1$ and $\tau = 0$ because $Z_T = 0$.

The discrete spatial step (Δx) is related to the wave propagation wavelength (λ) give by Eq. 5:

$$\Delta x / \lambda \leq 0.1 \,, \tag{5}$$

This condition ensures that both the numerical dispersion and propagation velocity variation between waves of different wavelengths will be kept small. Once Δx is defined, the time discretization, Δt, is defined by Eq. 6:

$$\Delta t = \Delta x / c \tag{6}$$

where c is the speed of sound in air.

Each TLM iteration step can be separated in three stages: scattering, connection and summation. In the scattering stage, for each mesh node, the scattering matrix at Eq. (4) is used to generate the output pulses from the input pulses. In the summation stage, the pressure value ($p(\mathbf{r}, t)$) is calculated by Eq. (7):

$$p(\mathbf{r}, t) = \frac{1}{3} \sum_{n=1}^{6} V_n^I \tag{7}$$

where V_n^I is the input pulse at the node port n. During the connection stage, each node input is given from the output values of the adjacent nodes.

3 MESH CONSTRUCTION

MRI can clearly show the soft tissues relevant to the voice production. The TLM meshes are constructed over the voxels (Volumetric Picture Elements) with the object label (or gray level) in the segmented MRI volumes. Consequently, the extracted TLM mesh will fit the segmented object as closely as possible, depending only on the size of the voxels. It is important to say that the 3D image must be isotropic, in order to obtain a uniform mesh.

3.1 Tube Models Meshes

The process of obtaining the tube meshes is performed in four steps:(i) The image sequence for the tube mesh generation is constructed from a single drawing (here, the Kolourpaint (Dang, 2006) image editor was used to draw the first slice of the tube); (ii) The single image is saved in the DICOM format with the Gimp (Mattis and Kimball, 2007) image editor; (iii) Several copies of the DICOM slice are created and then the ModaVox software (an open source application developed during this research) is used to open and append this slice sequence into a volume image with the desired tube length; (iv) The ModaVox is used again to segment the several slices of the created volume image, each one with a single disk object, appending them in a segmented 3D volume image from which the TLM mesh is extracted.

To construct a two-tube model mesh with sections of different diameters, two image sequences, as described above, should be created. The tube meshes have terminations for better modelling of the radiation effect. These terminations are tube sections with 7 pixels length and diameter 4 pixels greater than that of the tube to be terminated.

3.2 Vocal Tract Mesh

The process of generating the vocal tract mesh is performed in four steps: (i) The MRI sequence is extracted when an /a/ vowel is pronounced. This sequence was originally composed of 186 slices of 512×512 pixels, with pixel spacing $(0.9375, 0.9375)$ mm and 1 mm thickness, hence, an anisotropic image. (ii) The selection of the volume of interest (VOI) containing only the vocal tract reduces the original volume image, which is oversampled to remove the anisotropy. Hence, the final voxel dimensions were $(0.968498, 0.968498, 0.968498)$ mm. (iii) On this new isotropic 3D image, the segmentation is performed via neural network, island removal and some manual adjustments. (iv) Finally, the uniform TLM mesh is extracted from the segmented 3D image. The neural network, the island removal, the manual adjustment tools, the mesh extraction and the TLM are all implemented in the ModaVox program.

4 SIMULATION DETAILS

4.1 Mesh Attributes

In TLM meshes used here, each node has associated the pressure and the boundary values. The pressure values are the solutions of the corresponding wave equation at each node, at each time. The boundary values define what kind of boundary condition must be applied, setting specific values for the parameters ρ and τ at Eq. (4). The boundary values and their corresponding meanings are: (i) Value 0 - interior nodes ($\rho = -2/3$ and $\tau = 1/3$); (ii) Value 1 - reflective condition ($\rho = 1$ and $\tau = 0$); (iii) Value 2 - input nodes (also interior nodes); (iv) Value 3 - free space condition ($\rho = -1$ and $\tau = 0$).

4.2 Input Signals

The input signals represents pressure. One of the input signals consists of the summation of 200 frequency components, ranging from 50 to 10000 Hz, each one with unitary amplitude, as given by Eq. (8):

$$\text{input signal} = \sum_{n=1}^{200} \cos(2\pi f_n k \Delta t) \tag{8}$$

where k is the iteration step number and Δt is the TLM time descretization interval as given by Eq. (6). This signal was used as input either in the tubes or in the vocal tract model meshes to probe the frequency response in a single simulation. The other input signal used is the glottal signal, which was obtained by the inverse filtering of a real voice signal. This signal will be described in Section 6 and it was only used as input for the vocal tract mesh.

4.3 Output Signals

The ModaVox software enables to select the desired mesh nodes and to store their pressure values in separated files for later analysis. Hence, it is possible to see the output signals of the selected nodes, in time, after the simulation has been completed. FFT and standard LPC can be applied to these output signals to obtain spectral and frequency response information, respectively. The minimal duration considered here for the output signals was fixed in 0.2 seconds, because this time interval corresponds exactly to 10 periods of the component with the smallest frequency value (50 Hz) of the input signal at Eq. (8). This ensures that all frequency components will appear in the FFT plot and that the computing time of the simulation will not be too long.

4.4 Stability Condition

After several simulations, a stability condition with respect to the values of ρ and τ was fixed. This condition is given by Eq. (9):

$$\begin{cases} \tau = \dfrac{(1-\rho)}{5}, & \rho > -2/3 \\ \tau = 1 + \rho, & \rho \leq -2/3\,. \end{cases} \tag{9}$$

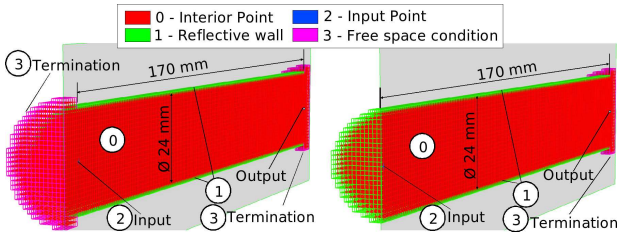
The first line of Eq. (9) is obtained by the substitution of $Z_T = Z/5$ in Eq. (1) and the second by the substitution of $Z_T = Z(1+\rho)/(1-\rho)$, also in Eq. (1).

5 TLM APPLIED TO TUBE MODELS

In voice and speech research, the tube models are frequently used to relate the shape configurations of the vocal tract to the observed formant frequencies (Rabiner and Schafer, 1978). Hence, before applying the TLM to the complex structure of the vocal tract, it will be applied to tube models, whose results can be compared with the analytical ones.

5.1 The sound velocity compensation factor

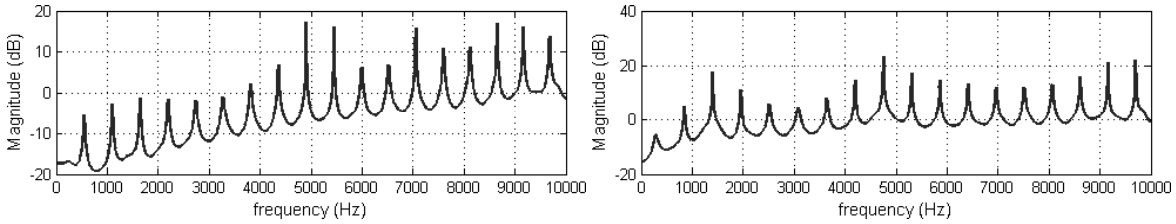
The first simulations were run on single tube model meshes, corresponding to the open and closed tubes with 24 mm diameter and 170 mm length (Fig. 1).



(a) (b)

Figure 1. Tubes with 24 mm diameter and 170 mm length. (a) Open. (b) Closed.

For the open tube model (Fig. 1-(a)), the frequency response calculated from the simulated output signal, considering the sound speed $c = 343.1$ m/s, gave the following resonant frequencies: $f_1 = 550.67$ Hz, $f_2 = 1101.34$ Hz, etc (Fig. 2-(a)).



(a) (b)

Figure 2. Frequency Responses (Simulations considering $c = 343.1$ m/s, $L = 170$ mm and $d = 24$ mm). (a) Open tube. (b) Closed tube.

However, the analytical values for the resonant frequencies are given by $f_1 = 1009.12$ Hz, $f_2 = 2018.23$ Hz, etc., according to Eq. (10):

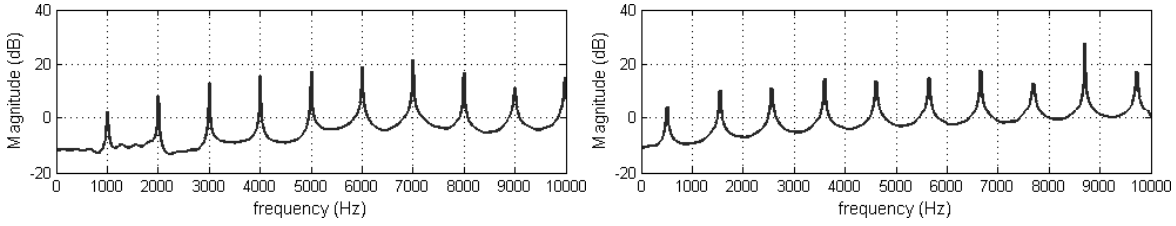
$$f_n = (nc)/(2\,L) \tag{10}$$

where $c = 343.1$ m/s is the speed of sound in air and L is the tube length. Hence, the simulation values were not compatible with the analytical ones, but compatible with a tube with a greater length. In this case, $L = 311,529$ mm. Then, another simulation was performed, but with a tube closed in one of its endings (Fig. 1-(b)). For this tube, the values obtained from the simulation were $f_1 = 275.335$ Hz, $f_2 = 826.005$ Hz, etc. (Fig. 2-(b)). However, the obtained resonant frequencies should be $f_1 = 504.55$ Hz, $f_2 = 1513.67$ Hz, etc., according to

$$f_n = [(2n-1)c]/(4\,L)\,. \tag{11}$$

Again, the simulation values for the resonant frequencies correspond to a longer tube, with $L = 311.529$ mm length. The conclusion was that, due to the orthogonal TLM mesh structure, the wavefronts travel at different velocities in diagonal directions. Hence, the tube seems to be longer because the velocity is lower. Then, a compensation factor is proposed: Using Eq. (10), the tube length L that gives the values found in the simulations is calculated. In this case, the value calculated is $L = 311.529$ mm, which differs from the true tube length (170 mm). The ratio between the two lengths is 1.8325. Hence, due to the linearity of Eq. (6), the sound velocity can be changed to $c = 343.1 \times 1.8325 = 628.7307$ m/s, maintaining the 170 mm tube length. Or, alternatively, it is possible to construct a new tube mesh with $170/1.8325 = 92.77 \simeq 93$ mm length, maintaining $c = 343.1$ m/s. Although both alternatives work, it is more simple to increase the velocity value than changing the mesh structure.

Then, considering a tube with 170 mm length, 24 mm diameter and compensating the sound velocity ($c = 343.1 \times 1.8325 = 628.7307$ m/s), the frequency responses obtained from the simulations are shown in Fig. 3.



(a) (b)

Figure 3. Frequency Responses (Simulations considering $c = 628.7307$ m/s, $L = 170$ mm and $d = 24$ mm). (a) Open tube. (b) Closed tube.

Now, the values of the resonant frequencies are compatible with those obtained analytically. In agreement with the theory, the resonant frequency values for the simulations for the 170 mm length single tube meshes with diameters 12 mm and 24 mm were the same shown in Fig. 3, even placing the source close to the tube walls.

5.2 Two-tube Model for the /a/ vowel

The analytic resonant frequencies for the two-tube model are estimated by calculating the acoustic impedance modulus $Z(z)$ at the junction between the two tube sections, Eq. (12):

$$Z(z) = \left| \frac{\rho c \cos\left(\frac{\omega L_1}{c}\right) \sin\left(\frac{\omega L_2}{c}\right)}{A_1 \sin\left(\frac{\omega L_1}{c}\right) \sin\left(\frac{\omega L_2}{c}\right) - A_2 \cos\left(\frac{\omega L_2}{c}\right) \cos\left(\frac{\omega L_1}{c}\right)} \right| \tag{12}$$

where L_1, L_2, A_1 and A_2 are, respectively, the lengths and cross-sectional areas of the tube sections 1 and 2, $\rho = 1.204$ kg/m^3 is the air density and $c = 343.1$ m/s is the sound velocity in air. For the /a/ vowel two-tube model, according to Fant (1970), the substitution of the dimensions $L_1 = 90$ mm, $L_2 = 80$ mm, $A_1 = 100$ mm^2 and $A_2 = 700$ mm^2, and also $\omega = 2\pi f$ into Eq. (12), yields a function that is plotted in the desired frequency range for comparison with the simulation results.

Figure 4 shows the extracted mesh for the two-tube model. The input signal is given by Eq. (8).

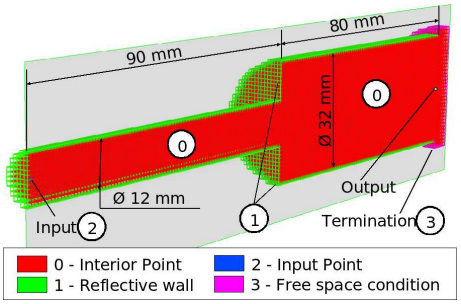
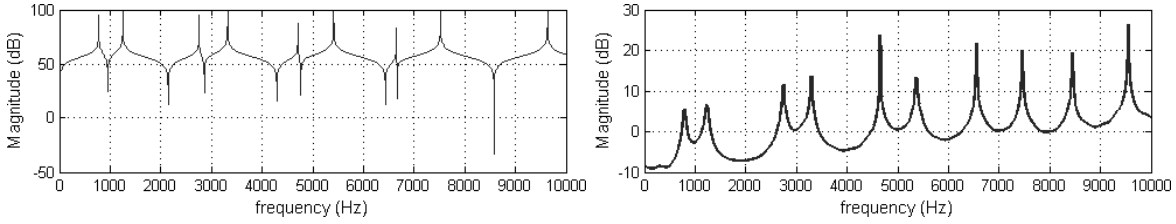


Figure 4. Two-tube Model for the /a/ Vowel.

For the construction of the two-tube model mesh for the /a/ vowel, the same dimensions of the analytic model were used. However, the diameters of tubes 1 and 2 (d_1 and d_2) were approximated to 12 mm and 32 mm, respectively (see Fig. 4), so that the ratio between the original values of d_2 and d_1, and, consequently, between the cross-sectional areas, remains, as closely as possible, the same, i.e.,

$$d_2/d_1 = 29.85410660/11.28379167 = 2.6457 \simeq 2.6666 = 32/12\,. \tag{13}$$

Figure 5 shows the acoustic impedance modulus, calculated by Eq. (12) in logarithmic scale and the frequency response simulated. In this figure, one can note that the values of the frequencies corresponding to the peaks are very similar, except for the resonant peak between 8000 and 9000 Hz, where the numerator of Eq. (12) also goes to zero.



(a) (b)

Figure 5. Two-tube Model for the /a/ Vowel. (a) Acoustic Impedance given by Eq. (12) in logarithmic scale. (b) Frequency Response (Simulation with $c = 628.7307$ m/s).

6 TLM APPLIED TO A VOCAL TRACT TO PRODUCE AN /A/ VOWEL UTTERANCE

In this section, the TLM simulation is run on a vocal tract TLM mesh whose shape conforms to that of a real vocal tract with the /a/ vowel shape. For this mesh, shown in Fig. 6, $\Delta x = 0.968498$ mm due to anisotropy removal (as discussed in Subsection 3.2), and $c = 628.7307$ m/s due to the use of the velocity compensation factor (as discussed in Subsection 5.1).

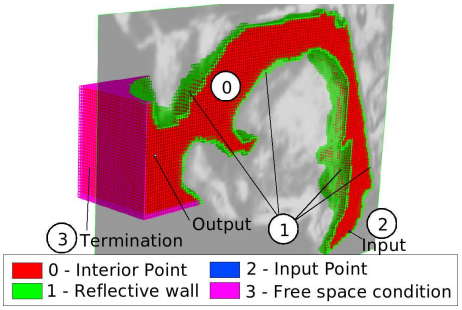


Figure 6. Vocal Tract Mesh corresponding to an /a/ vowel shape. Cut view showing the contour values. The cut plane shows a sagittal slice from the MRI sequence.

6.1 Vocal tract mesh - frequency response simulation

For this simulation, the input signal considered for the vocal tract mesh was the one given by Eq. (8). The validation of the results was performed by comparing the frequency response plot obtained from the TLM simulation with those obtained from the real /a/ vowel voice signal of the same subject, from whom the MRI sequence was extracted (Fig. 7).

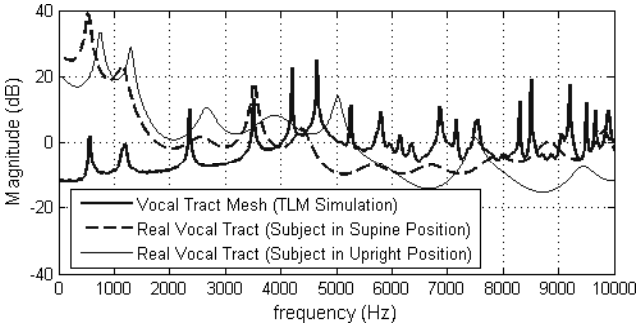


Figure 7. Frequency Responses of the vocal tract for the /a/ vowel (Real vs. TLM Simulation).

In Fig. 7, two real /a/ vowel utterances were recorded: one with the subject in supine position, similar to that maintained in the MRI equipment, and another with the subject in upright position. This was done because the subject position can affect the vocal tract shape (Kitamura *et al.*, 2005). As can be noted, the frequency response plot obtained from the TLM simulation output signal has its first four formant values very compatible with the ones of the real voice signal recorded in supine position. This result shows that the frequency response obtained from the TLM simulation is not only correct, but it could reproduce even the influence of gravity on the vocal tract due to the supine position when the MRI sequence was extracted.

However, the lip radiation effect is strong on the TLM simulation frequency response in Fig. 7. The same high-pass effect also happens in the results for the tube models, because as the input signal of Eq. (8) has uniform power distributed over its frequency components, it cannot compensate for the lip radiation effect on the output signal, as ilustrated in Fig. 8.

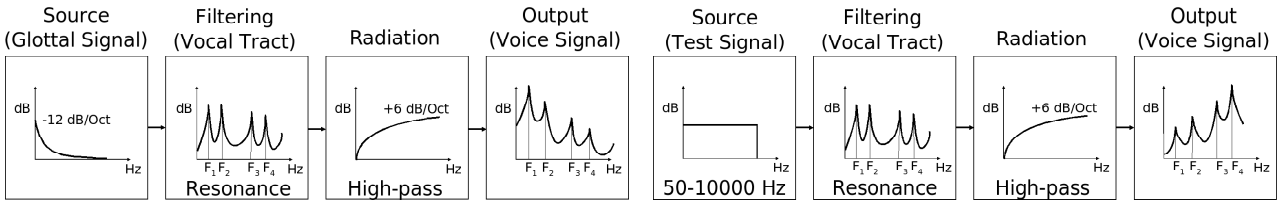


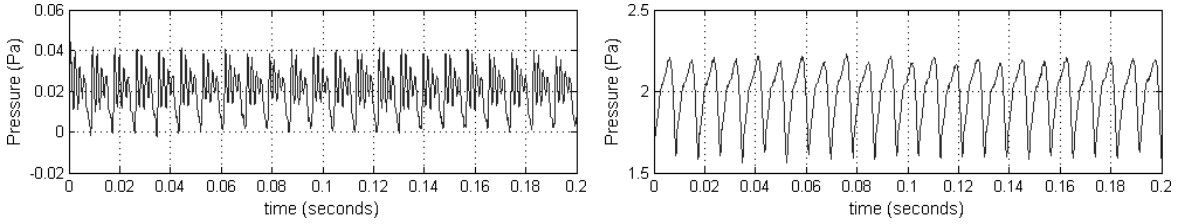
Figure 8. (a) Voice production Source-filter model block diagram. (Adapted from (Fant, 1970)). (b) Effect of the test signal on the source-filter model.

The glottal signal has an harmonic reduction ratio of 12 dB/Octave. This allows the compensation of the high-pass

filter effect caused by the lip radiation.

6.2 Vocal tract mesh /a/ Vowel simulation

To simulate the /a/ vowel production in the vocal tract mesh, a glottal signal was used as input. The glottal signal (Fig. 9-(b)) was generated with the function *iaif()*, from the TKK Aparat (Airas *et al.*, 2005, 2009) program, by the inverse filtering of the real voice signal (Fig. 9 top) produced by the same subject whose vocal tract mesh was extracted. After that, the generated glottal signal was oversampled so that its sampling rate become $1/\Delta t$, where $\Delta t = \Delta x/c$ is the TLM time discretization interval.



(a) (b)

Figure 9. (a) Real /a/ vowel voice signal. (b) Glottal signal obtained by inverse filtering of the signal in (a).

To simulate the human /a/ vowel sound, the boundary condition for the vocal tract walls was changed to a non-totally reflective condition. As the majority of the vocal tract structure consists in soft tissues, the reflection coefficient for the walls was approximated using the average value of the acoustic characteristic impedance of the human soft tissue (Hendee and Ritenour, 2002), which is given by 1630000 Ns/m^3. Hence, $Z_{walls} = 1630000$ Ns/m^3 and $Z_T = Z_{walls}/5$, according to Eq. (2), so that the value of the reflection coefficient (ρ) for the vocal tract walls is given by

$$\rho = \frac{Z_T - Z_0}{Z_T + Z_0} \tag{14}$$

where $Z_0 = 413.0923$ Ns/m^3 is the air acoustic impedance. Then, for the vocal tract walls $\rho = 0.9974688987$ and $\tau = 0.0005062203$ according to Eq. (9).

Figure 10 compares the frequency contents of the real voice signal in Fig. 9-(a), with that of the TLM simulation output signal, showing that its FFT is practically the same for the real signal in the voice frequency band, ensuring that a real /a/ vowel sound was generated by the developed model.

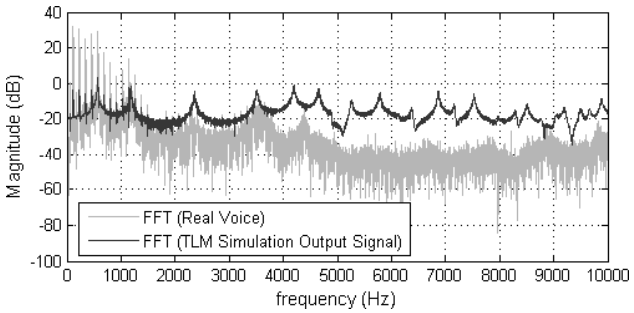


Figure 10. TLM simulation on vocal tract mesh with $c = 628.7307$ m/s, soft tissue walls, and the glottal signal at Fig. 9-(b) as input. FFT comparison between the real voice signal at Fig. 9-(a) and the TLM simulation output signal.

7 SIMULATIONS SUMMARY

Table 1 shows the summary of the presented simulations. The computer used to perform them has a 2.66 GHz processor and 1GB RAM.

Table 1. Summary of the presented simulations.

Mesh	Number of Points	Δx (mm)	c (m/s)	Computing Time	Sound Time (s)
Fig. 1-(a)	92929	1	343.1	8h 27'	0.2
Fig. 1-(b)	88274	1	343.1	7h 44'	0.2
Fig. 1-(a)	92929	1	628.7307	15h 34'	0.2
Fig. 1-(b)	88274	1	628.7307	14h 35'	0.2
Fig. 4	89666	1	628.7307	16h 10'	0.2
Fig. 6	148289	0.968498	628.7307	161h 14'	0.976

8 CONCLUSIONS

The numerical model developed in this work can be applied to the calculation of resonant frequencies from complex ducts. In the vocal tract case this model was capable of describing even the influence of gravity in the vocal tract shape due to the supine position of the subject when the MRI sequence was obtained. In fact, this is probably the first TLM mesh model to simulate real human voice, ensuring its accuracy.

The velocity compensation factor (1.8325) can be used for any uniform TLM mesh, as demonstrated with the simulation of the tube and vocal tract model meshes.

The developed simulation method produces an output signal in the time domain, from which the frequency response can be calculated with the standard LPC technique. The frequency range of the frequency response is the same of the input signal. Hence, just a single simulation is necessary to obtain the frequency response of the structure described by the mesh in the desired frequency range.

The boundary condition applied to the vocal tract walls has shown considerable importance on the sound quality of the generated voice sample. Future works should attempt to set tissue specific acoustical impedances allowing different reflection values according to different regions of the vocal tract walls.

9 ACKNOWLEDGEMENTS

The authors would like to thank Dr. Alair Augusto S.M.D. dos Santos from Hospital das Clínicas de Niterói for the valuable Radiology suggestions and for making the MR equipment of the ProEcho clinic available to this research. This work was supported by FAPERJ (Fundação de Amparo à Pesquisa no Rio de Janeiro - Programa Jovem Cientista do Nosso Estado), by CAPES (CAPES/COFECUB project N. 672/10) and by CNPq (Brazilian Agency: Conselho Nacional de Desenvolvimento Científico e Tecnológico).

10 REFERENCES

Airas, M., Pulakka, H., Bäckström, T. and Alku, P., 2005. "A toolkit for voice inverse filtering and parametrisation". In *Proceedings of the 9th European Conference on Speech Communication and Technology (Interspeech'2005 - Eurospeech)*. pp. 2145–2148.

Airas, M., Pulakka, H., Bäckström, T. and Alku, P., 2009. "Tkk aparat: Voice source analysis and parametrization toolkit". URL http://sourceforge.net/projects/aparat/.

Baer, T., Gore, J.C., Gracco, L.C. and Nye, P.W., 1991. "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels". *Journal of the Acoustical Society of America*, Vol. 90, pp. 799–828.

Clement, P., Hans, S., Hartl, D.M., Maeda, S., Vaissière, J. and Brasnu, D., 2007. "Vocal tract area function for vowels using three-dimensional magnetic resonance imaging. a preliminary study". *Journal of Voice*, Vol. 21, No. 5.

Cogan, D., O'Connor, W. and Pulko, S., 2006. *Transmission Line Matrix in Computational Mechanics*. CRC Press, Taylor & Francis Group.

Dang, C., 2006. "Kolourpaint is a free, easy-to-use paint program for kde". URL http://kolourpaint.sourceforge.net/.

El-Masri, S., Pelorson, X., Saguet, P. and Badin, P., 1998. "Development of the transmission line matrix method in acoustics. application to higher modes in the vocal tract and other complex ducts." *International Journal of Numerical Modelling*, Vol. 11, No. 3, pp. 133–151.

Fant, G., 1970. *The Acoustic Theory of Speech Production*. Mouton, The Hague.

Hendee, W.R. and Ritenour, E.R., 2002. *Medical Imaging Physics*. Wiley-Liss, New York.

Johns, P.B. and Beurle, R.L., 1971. "Numerical solution of two-dimensional scattering problems using a transmission-line matrix". In *Proc. IEEE*. Vol. 118, pp. 1203–1209.

Katsamanis, A. and Maragos, P., 2008. "A fricative synthesis investigations using the transmission line matrix method". *Journal of the Acoustical Society of America*, Vol. 123, p. 3741.

Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Shakudo, Y., Masaki, S., Kuroda, K., Oku-uchi, N. and Senda, M., 2005. "Difference in vocal tract shape between up right and supine postures: Observations by an open-type mr scanner". *Acoust. Sci. & Tech.*, Vol. 26, pp. 465–468.

Matsuzaki, H. and Motoki, K., 2006. "A finite-element method analysis of acoustic characteristics of the vocal tract with the nasal cavity during phonation of japanese /a/". *Journal of the Acoustical Society of America*, Vol. 120, No. 5, pp. 3371–3371.

Mattis, P. and Kimball, S., 2007. "Gnu image manipulation program". URL http://www.gimp.org/.

Motoki, K., 2002. "Three-dimensional acoustic field in vocal-tract". *Acoust. Sci. & Tech.*, Vol. 23, No. 4.

Rabiner, L.R. and Schafer, R.W., 1978. *Digital Processing of Speech Signals*. Prentice-Hall, Englewood Cliffs, New Jersey.

Salama, I., 1997. *TFDTLM: A New Computationally Efficient Frequency Domain TLM Based on Transient Analysis Techniques*. Ph.D. thesis, Virginia State University, Blacksburg.

Story, B.H., Titze, I.R. and Hoffman, E.A., 1996. "Vocal tract area functions from magnetic resonance imaging". *Journal of the Acoustical Society of America*, Vol. 100, pp. 537–554.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y. and Fujimoto, I., 2006. "Measurement of temporal changes in vocal tract area function from 3d cine-mri data". *Journal of the Acoustical Society of America*, Vol. 119, No. 2, pp. 1037–1049.

Vohradnik, M., Dedouch, K., Vokral, J. and Švec, J.G., 2003. "Finite element model of supraglottal space in cleft palate". In *International Federation of Otorhinolaryngological Societies, International Congress Series*. Vol. 1240, pp. 1145–1149.

11 Responsibility notice

The author(s) is (are) the only responsible for the printed material included in this paper.