PAN
ARCHIVES OF ACOUSTICS
Vol. **44**, No. 2, pp. 287–300 (2019)
DOI: 10.24425/aoa.2019.128492

Real-Time Vocal Tract Model for Elongation of Segment Lengths in a Waveguide Model

Tahir Mushtaq QURESHI*, Muhammad ISHAQ

Department of Mathematics
COMSATS University Islamabad
Vehari campus, Vehari (61100), Pakistan
*Corresponding Author e-mail: tahirmushtaq@ciitvehari.edu.pk

(*received April 14, 2017; accepted January 21, 2019*)

A vocal tract model based on a digital waveguide is presented in which the vocal tract has been decomposed into uniform cylindrical segments of variable lengths. We present a model for the real-time numerical solution of the digital waveguide equations in a uniform tube with the temporally varying cross section. In the current work, the uniform cylindrical segments of the vocal tract may have their different lengths, the time taken by the sound wave to propagate through a cylindrical segment in an axial direction may not be an integer multiple of each other. In such a case, the delay in an axial direction is necessarily a fractional delay. For the approximation of fractional-delay filters, Lagrange interpolation is used in the current model. Variable length of the individual segment of the vocal tract enables the model to produce realistic results. These results are validated with accurate benchmark model. The proposed model has been devised to elongate or shorten any arbitrary cylindrical segment by a suitable scaling factor. This model has a single algorithm and there is no need to make section of segments for elongation or shortening of the intermediate segments. The proposed model is about 23% more efficient than the previous model.

Keywords: digital waveguide; vocal tract; elongation of cylindrical segment.

1. Introduction

The acoustic theory of the speech production system has a long history, beginning from the earliest works by Helmholtz (1863), Fant (1971), and Rabiner and Schafer (1978). The early mathematical models were presented by the authors (Gold *et al.*, 2011). Voice production system of the human being is built on the theory of the source-filter model in which the action of the source is independent of the filter (Gunnar, 1960). The vocal folds are two symmetric soft-tissue structures fixed between the thyroid cartilage and arytenoid cartilages and considered as the source of sound in the human being. The vocal tract is the aero-acoustic cavity between the vocal folds and the open surface at the position of the lips, which acts as a filter in the source-filter theory. Due to lungs pressure, the self-excited motion of the vocal folds generates a train of pulses, which is further modulated by the resonance of the vocal tract.

Vocal folds have been modelled with different degrees of complexity and a lot of work has been dedicated to this field in the literature (Avanzini *et al.*, 2001; Flanagan, Landgraf, 1968; Ishizaka, Falanagan, 1972; 1977; Maddox *et al.*, 2014; Qureshi, Syed, 2011a; 2011b; Shimamura, Tokuda, 2016; Titze, Titze, 2014). Parallelly, for modelling of the vocal tract, several approaches have been used in (Birkholz *et al.*, 2010; Kelly, Lochbaum, 1962; Mullen *et al.*, 2003; Story, 2013; Välimäki, Karjalainen, 1994; Vampola *et al.*, 2015; Wang *et al.*, 2012b). The variation in the movement of the tongue, jaw, and lips forms different shapes of the vocal tract. This results in variation in the cross-sectional area along the length of the vocal tract. Hence, the vocal tract is considered as a function of the cross-sectional area that varies over time. This leads to a special class of speech production models that depend on the area function of the vocal tract (Hoefer, 1985; Kelly, Lochbaum, 1962; Smith, 1992; Van Duyne, Smith, 1993b). A waveguide is a bidirectional delay line in which motion of the wave can be considered by a one-dimensional wave equation (Morse, 1981; Smith, 2002). The connec-

tions of these waveguides in a grid lead to a higher-dimensional space (Savioja *et al.*, 1994; Speed *et al.*, 2013; Van Duyne, Smith, 1993b). The delay between two nodes of the grid is one unit long and each node on the grid represents a junction in which scattering of incoming waves occurs. Kelly and Lochbaum (1962) were first to present a one-dimensional waveguide model of the vocal tract. Transmission line matrix (Hoefer, 1985; Johns, Beurle, 1971), finite-difference time-domain methods (Karjalainen, 2003; Välimäki *et al.*, 2006; Wang *et al.*, 2012a; 2012b), real-time waveguide model (Mathur *et al.*, 2006), and wave digital filters (Fettweis, 1971) were developed in accordance with the idea of Kelly-Lochbaum model (Kelly, Lochbaum, 1962). An extension of a one-dimensional digital waveguide (digital waveguide model) was first introduced by Smith (1985; 1992) and Van Duyne, Smith (1993a; 1993b), and is being used in the modelling of the vocal tract (Cooper *et al.*, 2006; Mullen *et al.*, 2003; 2006; 2007; Qureshi, Syed, 2015; Speed *et al.*, 2013; Wang *et al.*, 2012b). Qureshi and Syed (2015) introduced a novel approach to the development of two-dimensional featured one-dimensional waveguide model of the vocal tract that has comparable formant frequencies with the standard two-dimensional waveguide, but its efficiency is comparable with that of a one-dimensional waveguide model. Digital waveguides are very popular for realistic, high-quality sound generation in real time and are successfully employed for physical modelling of sound synthesis.

The Kelly-Lochbaum model uses fixed-length cylindrical segments of different cross-sectional areas to approximate the vocal tract (Kelly, Lochbaum, 1962), while Mathur *et al.* (2006) model uses the cylindrical segments of the vocal tract with variable lengths. In their paper, the elongation of any segment is achieved by concatenation of two extra segments. One of the extra segments represents the fractional part of the acoustic tube while another one is used as the fictitious tube. In the elongation of the intermediate segment, the segments of the vocal tract are divided into two sections. The scattering on each section is evaluated independently and they are coupled with the help of the delay line. This implies that the number of sections increases with the increment to the number of elongated segments. The problem arises when a large number of elongated segments is present in the vocal tract segmentation. For example, there are N intermediate cylindrical segments to be elongated in the segmentation of the vocal tract. We have to make $N+1$ different sections of the cylindrical segments for the evaluation of wave scattering in the Mathur's work (Mathur *et al.*, 2006). The evaluation of wave scattering is taken on each section separately and it is completed in $N+1$ steps. Finally, they are coupled again with the help of delay lines in N steps. This implies that the previous approach completes one iteration of wave scattering in $2N+1$ steps which makes the previous model complex and inefficient. In the present work, we propose an extension to the work of Mathur *et al.* (2006). The proposed model addresses the issue of making many sections of the cylindrical segments of the vocal tract to accommodate the wave scattering. The current model evaluates wave scattering at all cylindrical segments of the vocal tract in a row without making the sections of the cylindrical segments. The new approach makes the current model about 23% more efficient than the previous model. The proposed model has been developed in such a way that it works for both elongation and shortening of any cylindrical segment of the vocal tract.

The present section is followed by four more sections. In Sec. 2, we describe a basic waveguide model of the vocal tract and present its mathematical formulation. Section 3 describes the algorithm for the elongation of the segments of the vocal tract. Section 4 is reserved for results and discussion. Section 5 is for the conclusions.

2. Basic vocal tract model and fractional delay

The vocal tract is assumed to be constructed of a certain number of co-axial uniform cylindrical segments. It is assumed that the cross-sectional area is constant within each segment of the vocal tract and the total vocal tract length is quantised to an integer multiple of the segment length. In this case, the time delay for the wave scattering is the same for each segment. Figure 1 describes the structure of the vocal tract by the concatenation of total ten uniform cylindrical segments of the same length. The first segment S_1 and last segment S_{10} are representing larynx and lips of the vocal tract, respectively, in the current figure while others segments from S_2 to S_9 are called intermediate segments. Any two consecutive cylindrical segments with different cross-sectional areas are called a junction. For example, the concatenation of segments S_2 and S_3 forms a junction, while the concatenation of segments S_3 and S_4 forms another junction. A change in the cross-sectional area at the junction of two cylindrical segments leads to a change in the wave impedance. In such a junction, a part of the travelling wave is transmitted while the other is reflected back. This phenomenon is called scattering.

The segment of the index of an even number is called an even-number segment and the segment of the index of an odd number is an odd-number segment. The propagation of a sound wave through each cylindrical segment of the vocal tract takes the time that depends on the length of each cylindrical segment of the vocal tract. If the cylindrical segments are of the same in length, the time taken by the sound wave in each segment will be the same. In the current example

given in Fig. 1, all cylindrical segments are uniform and of the same length, the time taken by the sound wave to propagate through each cylindrical segment in an axial direction is the integer multiple of each other. In such a case, simulation of the wave propagation in the vocal tract is easy and straightforward. However, some more realistic speech sounds need a fractional change in the total length of the vocal tract so that some segments of the vocal tract are not quantised to an integer multiple. In other words, some segments are elongated with a fractional length and time delays in these segments are the fractional delays. In such cases, interpolators are used to approximate the fractional delay. For example, Fig. 2 shows the elongation of the segments S_1, S_6, and S_{10} with a factor of 1.5. The measurement of wave scattering at the junction formed by the segments S_1 and S_2 is not accurate because the delay time of the wave in the segment S_1 is more than that of the segment S_2. For accurate measurement of the wave scattering in the current case, the delay time in the segment S_1 is approximated with the help of interpolation. Similarly, this is the same case for the junctions formed by the segments S_6 and S_7, and the segments S_9 and S_{10}.

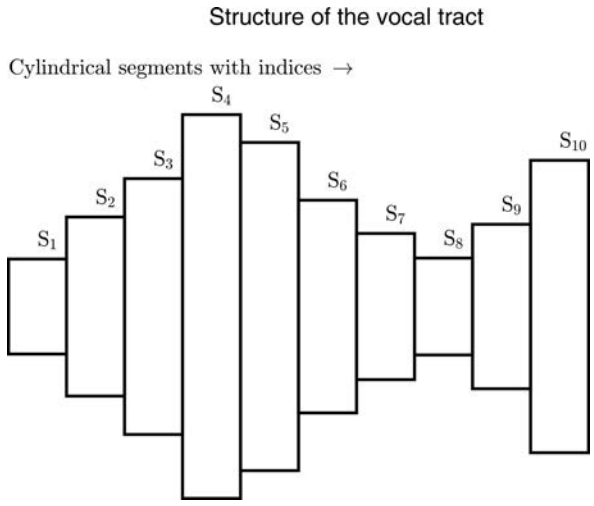

Fig. 1. Cylindrical segments of the vocal tract model.

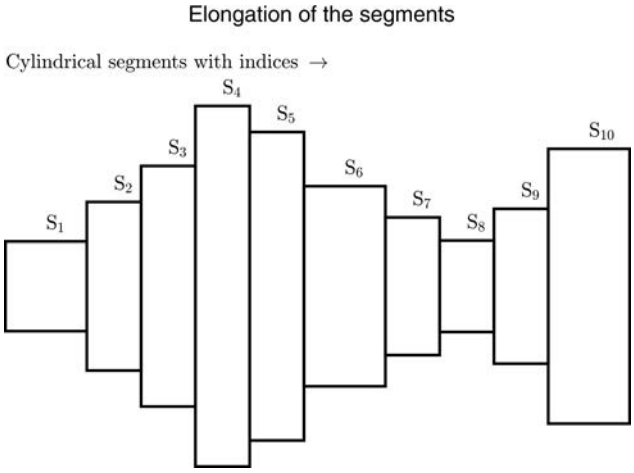

Fig. 2. Elongation of the segments of the vocal tract.

First of all, we derive the equations for the scattering of the wave at the junction of two successive segments S_i and S_{i+1} as shown in Fig. 3. In the current junction, the arrows represent the flow direction of the wave components while curved arrows represent the reflected back of the wave components. The components p^+ and p^- are representing the right and left travelling wave components of the pressure. Both wave components of the pressure p^+ and p^- split into two parts from the point where the cross-sectional area changes. At this point, one part is moved forward and another part is reflected back. Within a uniform tube, the relationship between velocity and pressure can be described by a wave equation (Markel, Gray, 1976; Rabiner, Schafer, 1978). D'Alembert's solution of the wave equation is the sum of the left and right travelling-wave components. By solving the well-known momentum equation and mass continuity equation (Välimäki, 1995) for the i-th segment of the vocal tract, we obtain:

$$u_i(x,t) = \frac{1}{Z_i}\left[p_i^+(t-x/c) - p_i^-(t+x/c)\right], \quad (1)$$

$$p_i(x,t) = \left[p_i^+(t-x/c) + p_i^-(t+x/c)\right], \quad (2)$$

where Z_i is the characteristic impedance of the i-th tube section and other terms are described as earlier.

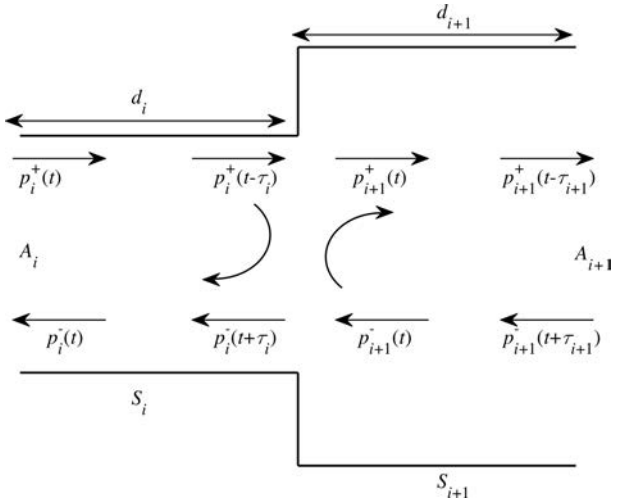

Fig. 3. Flow representation in the junction of two successive segments S_i and S_{i+1}.

Let l_i be the length of the i-th cylindrical tube with different length. Under the above assumptions, we have the following boundary conditions at the junction of the i-th and $(i+1)$-th cylinders (Välimäki, 1995),

$$p_i(l_i,t) = p_{i+1}(0,t), \quad (3)$$

$$u_i(l_i,t) = u_{i+1}(0,t). \quad (4)$$

Using Eqs (1) and (2) into Eqs (3) and (4), we have

$$p_i^+(t-\tau_i)+p_i^-(t+\tau_i) = p_{i+1}^+(t)+p_{i+1}^-(t), \tag{5}$$

$$\frac{1}{Z_i}\left[p_i^+(t-\tau_i)-p_i^-(t+\tau_i)\right] = \frac{1}{Z_{i+1}}\left[p_{i+1}^+(t)-p_{i+1}^-(t)\right], \tag{6}$$

where $\tau = l_i/c$, is the time required to travel the cylindrical tube.

By solving Eqs (5) and (6), we have (Välimäki, 1995),

$$p_i^-(t+\tau_i) = r_i p_i^+(t-\tau_i) + (1-r_i)p_{i+1}^-(t), \tag{7}$$

$$p_{i+1}^+(t) = (1+r_i)p_i^+(t-\tau_i) - r_i p_{i+1}^-(t), \tag{8}$$

where

$$r_i = \frac{Z_{i+1}-Z_i}{Z_{i+1}+Z_i} = \frac{A_i - A_{i+1}}{A_i + A_{i+1}}.$$

By rewriting Eqs (7) and (8), we obtain (Välimäki, 1995):

$$p_i^-(t+\tau_i) = p_{i+1}^-(t) + w(t), \tag{9}$$

$$p_{i+1}^+(t) = p_i^+(t-\tau_i) + w(t), \tag{10}$$

where

$$w(t) = r_i\left[p_i^+(t-\tau_i) - p_{i+1}^-(t)\right].$$

The current work is based on the elongation of an arbitrary segment of the vocal tract. In such a case, the delay in an axial direction is necessarily a fractional delay (Laakso *et al.*, 1996; Mathur *et al.*, 2006; Samadi *et al.*, 2004; Välimäki, 1995). In the present work, the fractional delay is also approximated by the Lagrange interpolator (Laakso *et al.*, 1996; Mathur *et al.*, 2006; Samadi *et al.*, 2004; Välimäki, 1995). Lagrange interpolation is a type of FIR filter which is popular for easy and fast calculation of the filter coefficients. It has a very good magnitude and phase response at low frequencies with the magnitude response never exceeding one (Välimäki, 1995).

Suppose that the total length of the vocal tract is l, the length of each cylindrical segment is d, M is any positive integer, and αd is the fractional change in the length of the vocal tract due to the movement of the articulators (Mathur *et al.*, 2006), then

$$l = (M+\alpha)d, \qquad \alpha \in [0,1]. \tag{11}$$

In a full sample delay waveguide model, there is only a forward wave in every other tube at each instant of time and a backward wave in every other. To make it efficient, we consider the half-sample delay waveguide model in the current work (Lim, Lee, 1993; Mathur *et al.*, 2006). This means that two consecutive steps of the full-sample model are combined into a single step which leads to the efficiency of the model. Thus, the sampling frequency of the half-sample delay waveguide model is obtained as

$$F_s = c/2d, \tag{12}$$

where c is the velocity of sound and d is described earlier.

Let's define Pf for forward pressure component and Pb for backward pressure component. Then interpolation of Pf and Pb may be written as,

$$Pf_{\text{interp}} = \sum_{k=0}^{NFL} h(k+1)Pf(q+k), \tag{13}$$

$$Pb_{\text{interp}} = \sum_{k=0}^{NFL} h(NFL+1-k)Pb(q+k), \tag{14}$$

where q is the index of the segment of the vocal tract, NFL is the filter order, and h is the impulse response for the NFL order Lagrange filter (Laakso *et al.*, 1996).

3. Model for the variation in the length of any segment of the vocal tract

In the current work, we present the model for the elongation or shortening of any cylindrical segment of the vocal tract. Our model is an extension of the work presented by Mathur *et al.* (2006). In the previous work, the elongation of any segment is achieved by adding two extra segments. However, one of the extra segments represents the fractional part of the acoustic tube while another one is for the fictitious acoustic tube. In the elongation of the intermediate segment, the segments of the vocal tract are divided into two sections. The scattering on each section is evaluated independently and they are coupled later with the help of the fractional delay line. This indicates that the number of independent sections increases with the increment to the total number of segments elongation. The problem of complexity arises when a huge number of elongated-segments is present in the vocal tract segmentation. The proposed model addresses the issue of making many sections of the cylindrical segments of the vocal tract to simulate the wave propagation. The current model simulates wave propagation on all cylindrical segments of the vocal tract in a row without making the sections of the cylindrical segments. The new approach makes the current model about 23% more efficient than the previous model. The proposed model has been developed in such a way that it works for both elongation and shortening of any cylindrical segment of the vocal tract.

We take an example to show the difference between the modelling of the previous and proposed works. Let us elongate two intermediate segments S_4 and S_6 in the example given in Fig. 1. According to the previous approach, there is a need to make three separate sections of the segments for the simulation of wave propagation in the vocal tract, as shown in Fig. 4. In the current figure, the extra segments are represented by the hatch style segments. The first section

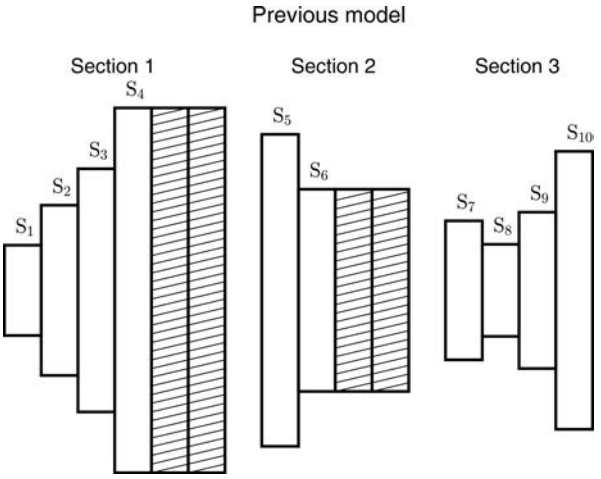

Fig. 4. Previous approach to elongation of two segments.

starts from segments S_1 to S_4 with two extra segments which are concatenated at the end of the segment S_4 to accommodate its elongation. The second section has segments S_5, S_6, and two extra segments that lead to total four segments in this section. There is a total of four segments in the third section starting from segment S_7 to S_{10}. In the previous work, scattering for the wave propagation is taken separately in these sections and then these three sections are rejoined with the help of the delay line to complete the wave propagation in the vocal tract. However, the proposed approach for the modelling of the vocal tract has been shown in Fig. 5. There is no need to make the sections of segments in the current approach and all segments including extra segments are concatenated in a row. This implies that the scattering for wave propagation in the vocal tract can be easily taken in a row, which leads to the efficiency of the proposed model.

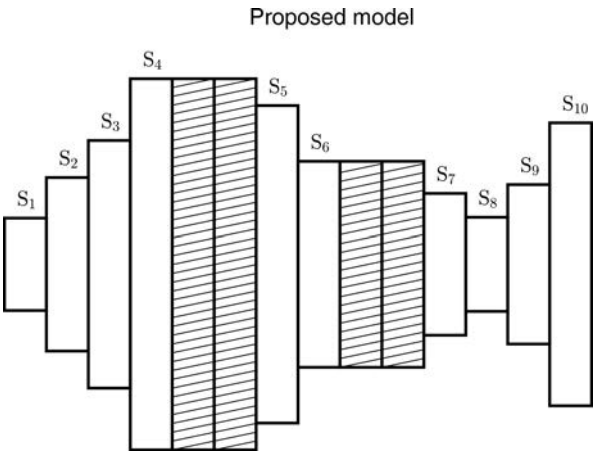

Fig. 5. Proposed approach to elongation of two segments.

In the current work, the elongated segments are assumed to have the fractional increments in their lengths with respect to the length of the smallest segment. To accommodate elongation or shortening of any cylindrical segment, we suppose that l is the total length of the vocal tract and d_i is the length of the i-th cylindrical segment S_i that has cross-sectional A_i as shown in Fig. 3, then

$$l = \sum_{i=1}^{M} d_i, \tag{15}$$

where M is the total number of cylindrical segments of the vocal tract.

We find the cylindrical segment of the minimum length such that

$$d_{\min} = \min_{1 \le k \le M} \{d_k\}, \tag{16}$$

$$d_{f_i} = \frac{d_i}{d_{\min}} - 1, \qquad i = 1, 2, ..., M, \tag{17}$$

where $d_{\min}$ is the length of the smallest segment of the vocal tract and d_{f_i} is representing the fraction delay value of the i-th segment of the vocal tract.

Equation (17) gives the values of d_{f_i} within the interval $[0, 1)$. If the value of d_{f_i} is zero then the i-th cylindrical segment has no fraction delay. However, the $i-$th cylindrical segment has a fraction delay when the value of d_{f_i} is greater than zero and less than 1. This approach helps us model elongation or shortening of any cylindrical segment. In our work, the division of the shortened cylindrical segment leads to a normal state or elongation of the other cylindrical segments. This implies that both cases become the problem of elongation of the cylindrical segments. However, the sampling frequency for the half-sample delay turns out to be higher for the case of the shortened cylindrical segment. In the current work, the sampling frequency can be written as

$$F_s = \frac{c}{2d_{\min}}, \tag{18}$$

where c is the velocity of sound.

As described earlier, the two extra cylindrical segments are added successively to the original elongated segment so that the number of segments remains even. The addition of extra segments to the original segment is of two types depending on the position of the segment. If the position of the elongated segment is even then the extra segment will be added after the elongated segment, otherwise, the extra segment will be added before the elongated segment.

In this way, the total number of segments of the vocal tract increases with the addition of positive even numbers. Let N be the total number of segments after the addition of extra segments, then,

$$N = M + 2K, \tag{19}$$

where K is the total number of elongated segments.

During the addition of extra segments, we also define the type of these segments in our work. The original segments whether they are elongated or non-elongated, may be named as normal and they are assigned

the value of 1. The extra segment adjacent to the original elongated segment may be called as fractional-delay segment and it is assigned a value of its fractional delay. However, the last extra segment is assumed as fictitious and it is assigned a value of zero.

The type of each segment may be found by the equation:

$$\text{Type}(S_i) = \begin{cases} 1 & \text{if } S_i \text{ is a normal segment,} \\ d_{f_i} & \text{if } S_i \text{ is a fractional-delay segment,} \\ 0 & \text{if } S_i \text{ is a fictitious segment.} \end{cases} \tag{20}$$

Further, the fractional delay is approximated with the help of Lagrange interpolation (Mathur *et al.*, 2006). In the model, the approximation of the fractional delay is different for an even and an odd elongated segment of the vocal tract. For an odd elongated segment, the backward pressure component of the fractional delay segment is approximated with the help of Eq. (14), while the forward pressure component is approximated with the help of Eq. (13) in the case of an even elongated segment. The scattering equations of the first and the last segments are different from the intermediate segments. The following equation describes basic three blocks of the scattering equations for the non-elongated segments of the vocal tract:

$$\text{Scatt}(S_i) = \begin{cases} Pf_i = \dfrac{u\,\rho\,c}{A_1} + r_g Pb & \text{if } i = 1, \\ \begin{pmatrix} \Delta = r_i(Pf_i - Pb_{i+1}) \\ Pf_{i+1} = Pf_i + \Delta \\ Pb_i = Pb_{i+1} + \Delta \end{pmatrix} & \text{if } 0 < i < M, \\ \begin{pmatrix} Pb_i = r_l\, Pf_i \\ P = (1 + r_l)\, Pf_i \end{pmatrix} & \text{if } i = M, \end{cases} \tag{21}$$

where u is the volume velocity, ρ is the density of air, c is the velocity of sound, A_1 is the cross-sectional area of the first segment, r_g is the glottal reflection coefficient, r_l is the lips reflection coefficient, M is the total number of segments, P is output pressure and r_i is the reflection coefficient defined in the Eq. (22):

$$r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}}, \tag{22}$$

where A_i is the cross-sectional area of the segment S_i.

In our model, we define four blocks for the algorithm of scattering equations. In the first block, the algorithm checks the elongation of the first segment. If elongation of the first segment is found, then backward pressure component of the fractional delay segment is approximated as Pb_{interp} with the help of Eq. (14) and is used in the Eq. (21) when $i = 1$. The second block is reserved for intermediate segments and it has further two sub-blocks. The first sub-block is used for the scattering of even-number segments, while the scattering of odd-number segments is performed in the second sub-block. In the third block, the adjustment of the scattering for the fractional delay segments is accomplished by using Eqs (13) and (14) if it is required. For elongation of the odd-number segment, backward pressure component Pb is approximated as Pb_{interp} with the help of Eq. (14), while forward pressure component Pf is approximated as Pf_{interp} by using Eq. (13) in the case of the even-number elongated segment. In the final block, the forward pressure component Pf is approximated as Pf_{interp} in the case of the last-elongated segment of the vocal tract. Figure 6 presents a complete algorithm of the model.

1. Read M cross-sectional areas in array A and their corresponding delay lengths in array dA.
2. Read the values of glottal and lips reflection coefficients r_g and r_l respectively.
3. Evaluate the Eqs (16) and (17) for fractional delay length d_{f_i} of each segment S_i.
4. For each $0 < d_{f_i} < 1$, insert two segments (fictitious and fractional delay segments) at position i of the array A with the same cross-sectional area A_i of the segment S_i. These two segments are inserted before the position i if the segment number is odd, otherwise they are inserted after the position i. This step changes the array A to A' with the length N such that $N \geq M$. Define also an array of name Type for the status of each segment S_i according to Eq. (20).
5. Compute reflection coefficients r_i, $i = 1, ..., N - 1$ by using Eq. (22).
6. Define the total number of samples such as $tSamples = C_0$ and volume velocity $u = u_0$.
7. Define two arrays Pf and Pb of the length N for pressure forward and pressure backward components, respectively, and initialise with values of zero.
8. Define the order of Lagrange Filter $NFL = 3$ and define an array h of impulse response with the length $NFL + 1$ for fractional delay.
9. While $k \leq tSamples$, repeat the steps 10–16.

[Fig. 6.]

10. The first segment S_1 represents the larynx segment. Evaluation of this segment depends on the condition of non-elongation or elongation.
    If $\text{Type}(S_1) = 0$ and $0 < \text{Type}(S_2) < 1$ (if elongation is true):
    Interpolate the pressure backward component as Pb_{interp} using Eq. (14) with the value of $\text{Type}(S_2)$ that is the fractional delay value and put $q = 1$,
    $$Pf(2) = \frac{u\rho c}{A'(1)} + r_g Pb_{\text{interp}}.$$
    Else
    $$Pf(1) = \frac{u\rho c}{A'(1)} + r_g Pb(1).$$
11. Put $u = 0$.
12. For even junction scattering, take $i \in \{2, 4, 6, ..., N-2\}$ and evaluate such as:
    If $\text{Type}(S_i) > 0$ and $\text{Type}(S_{i+1}) \neq 0$ (non-elongated segments),
    $$\Delta = r_i[Pf(i) - Pb(i+1)], \qquad Pf(i+1) = Pf(i) + \Delta, \qquad Pb(i) = Pb(i+1) + \Delta.$$
    Else if $\text{Type}(S_i) = 0$ and $\text{Type}(S_{i+1}) = 0$ (if even-odd segments are elongated):
    Interpolate pressure backward component as Pb_{interp} using Eq. (14) with the value of $\text{Type}(S_{i+2})$ that is the fractional delay value and put $q = i + 1$,
    $$\Delta = r_i[Pf(i-1) - Pb_{\text{interp}}], \qquad Pf(i+2) = Pf(i-1) + \Delta, \qquad Pb(i-1) = Pb_{\text{interp}} + \Delta.$$
13. For odd junction scattering, take $i \in \{1, 3, 5, ..., N-1\}$ and evaluate such as:
    $$\Delta = r_i[Pf(i) - Pb(i+1)], \qquad Pf(i+1) = Pf(i) + \Delta, \qquad Pb(i) = Pb(i+1) + \Delta.$$
14. Adjustment of fractional delayed scattering wave components is done in this section.
    For even junction scattering, take $i \in \{2, 4, 6, ..., N-2\}$ and evaluate such as:
    If $\text{Type}(S_i) = 0$ and $\text{Type}(S_{i+1}) \neq 0$ (if single even elongated):
    Interpolate the pressure forward component as Pf_{interp} using Eq. (13) with the value of $\text{Type}(S_{i-1})$ that is the fractional delay value and put $q = i - 3$,
    $$\Delta = r_i[Pf_{\text{interp}} - Pb(i+1)], \qquad Pf(i+1) = Pf_{\text{interp}} + \Delta, \qquad Pb(i-1) = Pb(i+1) + \Delta.$$
    Else if $\text{Type}(S_i) = 0$ and $\text{Type}(S_{i+1}) = 0$ (if even-odd elongated):
    Interpolate the pressure forward component as Pf_{interp} using Eq. (13) with the value of $\text{Type}(S_{i-1})$ that is the fractional delay value and put $q = i - 3$,
    $$\Delta = r_i[Pf_{\text{interp}} - Pb(i+2)], \qquad Pf(i+2) = Pf_{\text{interp}} + \Delta, \qquad Pb(i-1) = Pb(i+2) + \Delta.$$
    For odd junction scattering, take $i \in \{1, 3, 5, ..., N-1\}$ and evaluate such as:
    If $i > 1$ and $\text{Type}(S_i) = 0$ and $\text{Type}(S_{i-1}) \neq 0$:
    Interpolate the pressure forward component as Pb_{interp} using Eq. (14) with the value of $\text{Type}(S_{i+1})$ that is the fractional delay value and put $q = i$,
    $$\Delta = r_{i-1}[Pf(i-1) - Pb_{\text{interp}}], \qquad Pf(i+1) = Pf(i-1) + \Delta, \qquad Pb(i-1) = Pb_{\text{interp}} + \Delta.$$
15. The last segment S_N represents the lips segment. Evaluation of this segment also depends on the condition of the non-elongation or elongation.
    If $0 < \text{Type}(S_{N-1}) < 1$:
    Interpolate the pressure forward component as Pf_{interp} using Eq. (13) with the value of $\text{Type}(S_{N-1})$ that is the fractional delay value and put $q = N - 3$,
    $$Pb(N-1) = r_l Pf_{\text{interp}}, \qquad P(k) = (1 + r_l) Pf_{\text{interp}}.$$
    Else
    $$Pb(N) = r_l Pf(N), \qquad P(k) = (1 + r_l) Pf(N).$$
16. Increment the value of k by 1 and go to step 9.

Fig. 6. Algorithm for the modelling of the vocal tract with the non-uniform length of segments.

4. Results and discussion

In the previous sections, we have described the mathematics of the basic vocal tract model and presented the model for elongation of any segment of the vocal tract. In this section, we describe the working of our model, its characteristics and comparison with the benchmark model. In the current work, ABCD matrix model is assumed as a benchmark model (Sondhi, Schroeter, 1987). The ABCD matrix model contains chain matrices of the order 2×2. It accurately models the frequency response for any cylindrical segment

and may be used for validation of any other models. We also validate our model with ABCD matrix model. We choose cross-sectional areas of the vocal tract for the vowel in the literature (Mathur *et al.*, 2006; Story, Titze, 1998). This vowel has almost equal spaced formant frequencies in the frequency domain named as the neutral vowel. It has 44 cylindrical segments for the vocal tract of the length 17.5 cm. Each segment of the vocal tract has the length d such that $d = 17.5/44 = 0.397$ cm. For demonstration of our work, we elongate some segments of the vocal tract with 1.5% of the original length in such a way that the $d_{\min}$ in Eq. (16) becomes the original length of the segments, i.e. $d_{\min} = 0.397$. The density of air ρ and velocity of sound c are taken as 0.00114 g/cm^3 and 35000 cm/s, respectively. Thus, the sample frequency can be obtained as 44.1 kHz with the help of Eq. (18). The glottal and lips reflection coefficients are chosen as $r_g = 0.999$ and $r_l = -0.999$, respectively.

The frequency profiles of the vocal tract can be found with the help of impulse response. In the present work, the first value of the impulse response is set as 100, while other values are kept zeros. The Fast Fourier transformation is applied to the output of the model to get frequency profiles of the vocal tract model. The peaks in the frequency profiles are called formant frequencies which represent the frequencies of the particular vowel. The validity and accuracy of the proposed model are based on the comparison of these formant frequencies with that of the benchmark model. In the later discussion, we will use these formants frequencies as a reference in the figures.

Figure 7 represents the formant frequencies of the neutral vowel up to 5000 Hz. In the current case, the 44 segments of the vocal tract have the same length. The dotted line shows the frequency profile obtained by Mathur's model, the dashed line represents the frequency profile generated by the algorithm in Fig. 6, and the solid line shows the frequency profile generated by ABCD matrix model. The current figure shows that the formants frequencies (peaks) of the current model are very close to that of ABCD matrix model. Table 1 gives error comparison of the first five formant frequencies of the current model with ABCD matrix model.

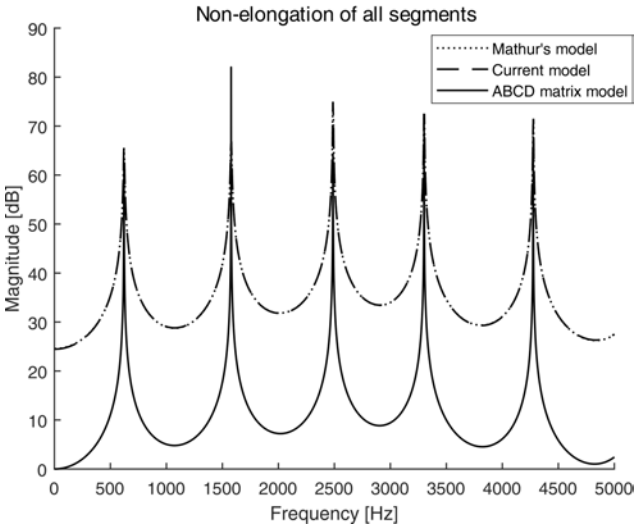

Fig. 7. Comparison of formant frequencies of the current model and Mathur's model with ABCD matrix model for non-elongated segments.

First, second, third, fourth, and fifth formant frequencies are represented by F1, F2, F3, F4, and F5, respectively, in the current table. The second, third, and fourth columns of the tables show the measurement of the formant frequencies from ABCD matrix model, Mathur's model, and proposed model, respectively. The absolute relative errors of the Mathur's model and the present model with respect to ABCD matrix model are mentioned in the fifth and sixth columns of the table, respectively.

It is noted from the current table that Mathur's model and the proposed model have the same relative errors as shown in the fifth and sixth columns of the table. The maximum relative error occurs at the second formant frequency with the value of 0.06%. From the table, it is noted that the maximum relative error of Mathur's model occurs at the second formant frequency with the value of 0.12%, while it has a maximum value of 0.09% at the fifth formant frequency in the case of the present model. This shows that the present model and Mathur's model generate the same formant frequencies in the current case and both models are very close to ABCD matrix model.

Table 1. Relative errors of the current model and Mathur's model compared with ABCD matrix model for non-elongated segments.

Formant frequency	ABCD matrix model [Hz]	Mathur's model [Hz]	Proposed model [Hz]	Relative error in Mathur's model	Relative error in the proposed model
F1	621	621	621	0.00	0.00
F2	1578	1579	1579	0.06	0.06
F3	2487	2487	2487	0.00	0.00
F4	3301	3301	3301	0.00	0.00
F5	4275	4277	4277	0.05	0.05

In the second case, we elongate the first segment S_1 that represents the larynx of the vocal tract. The increment in this segment is taken as $d_1 = 1.5d_1$. Figure 8 also shows the first five formant frequencies up to 5000 Hz. Table 2 represents the error comparisons of formant frequencies for the elongated segment S_1. In this case, the formants frequencies of the current model and Mathur's are also very close to that of ABCD matrix model. Table 2 depicts that the proposed model and Mathur's model have same relative errors with ABCD matrix model. The maximum relative error occurs at the first formant frequency with the value of 0.16% which is insignificant. This implies that the present model and Mathur's model generate the same formant frequencies in the current case and both models are very close to ABCD matrix model in the view of the fifth and sixth columns of the table.

In the next case, the elongation of lips is carried out by extension in the length of the last segment S_{44} by choosing $d_{44} = 1.5d_{44}$. Figure 9 represents that the formants frequencies of the current model and Mathur's model are nearly equal to that of ABCD matrix model. Maximum relative errors in the proposed model and Mathur's model are the same with the values of 0.17%, as shown in Table 3. The negligible values of the relative errors in the fifth and sixth columns of the current

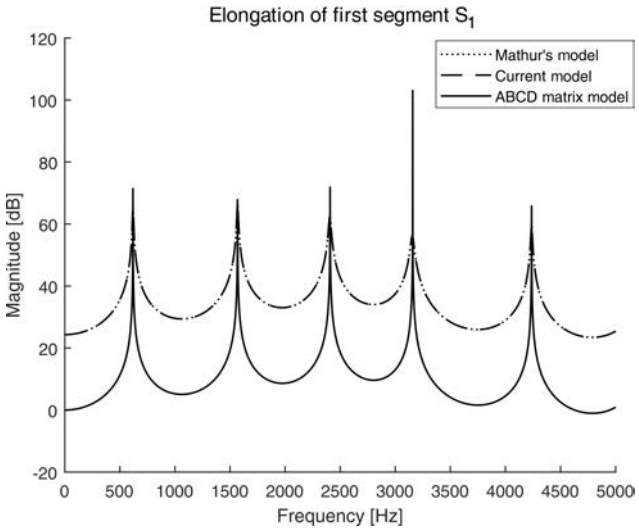

Fig. 8. Comparison of formant frequencies of the current model and Mathur's model with ABCD matrix model for the elongated segment S_1.

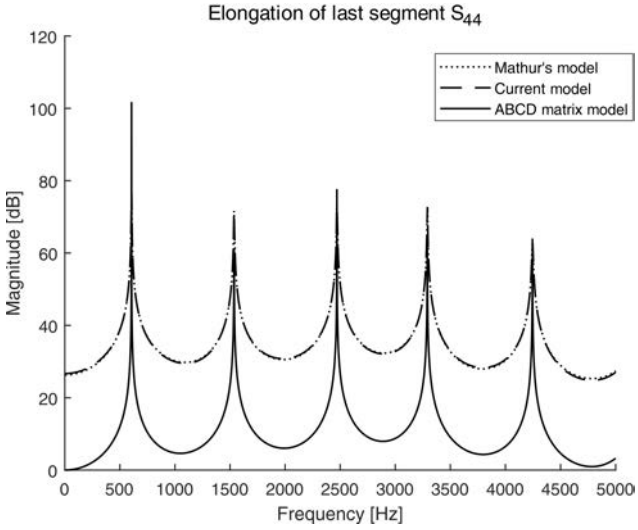

Fig. 9. Comparison of formant frequencies of the current model with ABCD matrix model for the lips elongation.

Table 2. Relative errors of the current model and Mathur's model compared with ABCD matrix model for elongated segment S_1.

Formant frequency	ABCD matrix model [Hz]	Mathur's model [Hz]	Proposed model [Hz]	Relative error in Mathur's model	Relative error in the proposed model
F1	619	620	620	0.16	0.16
F2	1568	1567	1567	0.06	0.06
F3	2408	2407	2407	0.04	0.04
F4	3158	3155	3155	0.09	0.09
F5	4237	4238	4238	0.02	0.02

Table 3. Relative errors of the current model and Mathur's model compared with ABCD matrix model for elongated segment S_{44}.

Formant frequency	ABCD matrix model [Hz]	Mathur's model [Hz]	Proposed model [Hz]	Relative error in Mathur's model	Relative error in the proposed model
F1	606	607	607	0.17	0.17
F2	1537	1536	1536	0.07	0.07
F3	2470	2469	2469	0.04	0.04
F4	3290	3292	3292	0.06	0.06
F5	4245	4247	4247	0.05	0.05

table imply that the formants frequencies of the current model and Mathur's model are very close to those of the ABCD matrix model in the present case.

Figure 10 represents the formant frequencies of the elongation of even number segment S_{14} with the same scaling factor 1.5. This figure exhibits the similarity of formants frequencies of the current model and Mathur's model with that of ABCD matrix model. Due to elongation of segment S_{14}, Table 4 depicts that the maximum relative errors of Mathur's model and the proposed model relative to the ABCD matrix model are measured as 33% and 16%, respectively. The negligible errors of the proposed model and Mathur's model establish that both models have approximately the same formant frequencies as those of ABCD matrix model.

Figure 11 shows the formant frequencies of the elongation of odd number segment S_{31} with the scaling factor 1.5. The close matching of the proposed model and Mathur's model with that of ABCD matrix model have also been observed in this figure. In this case, the present model and Mathur's model have a maximum error of value 0.33%, as shown in Table 5.

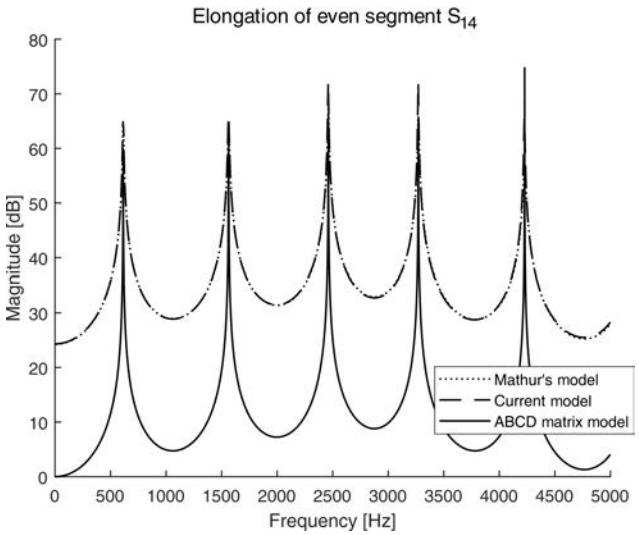

Fig. 10. Comparison of formant frequencies of the current model and Mathur's model with ABCD matrix model for the elongated segments S_{14}.

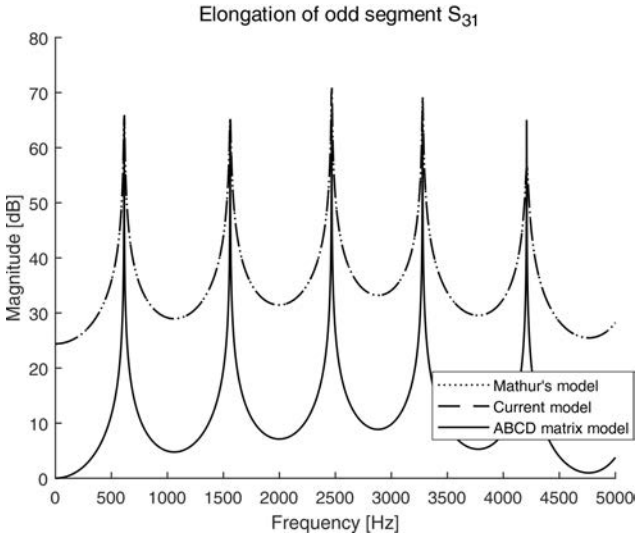

Fig. 11. Comparison of formant frequencies of the current model and Mathur's model with ABCD matrix model for the elongated segments S_{31}.

Table 4. Relative errors of the current model and Mathur's model compared with ABCD matrix model for elongation of an even-number segment S_{14}.

Formant frequency	ABCD matrix model [Hz]	Mathur's model [Hz]	Proposed model [Hz]	Relative error in Mathur's model	Relative error in the proposed model
F1	614	612	613	0.33	0.16
F2	1563	1562	1562	0.06	0.06
F3	2461	2460	2460	0.04	0.04
F4	3271	3270	3270	0.03	0.03
F5	4228	4229	4227	0.02	0.02

Table 5. Relative errors of the current model and Mathur's model compared with ABCD matrix model for elongation of an odd-number segment S_{31}.

Formant frequency	ABCD matrix model [Hz]	Mathur's model [Hz]	Proposed model [Hz]	Relative error in Mathur's model	Relative error in the proposed model
F1	614	616	616	0.33	0.33
F2	1561	1562	1562	0.06	0.06
F3	2467	2469	2469	0.08	0.08
F4	3278	3279	3279	0.03	0.03
F5	4208	4210	4210	0.05	0.05

We also consider two special cases of elongation for the even-odd and odd-even consecutive segments. Figures 12 and 13 demonstrate the formant frequencies of the elongation of consecutive even-odd and odd-even numbered segments, respectively, with the scaling factor of 1.5. In the present work, we take the elongation of both consecutive segments $S_{10}-S_{11}$ in the first case, while we consider the elongation of both consecutive segments $S_{37}-S_{38}$ in the second case. Both figures show that the formants frequencies of the current model and Mathur's model are also very close to that of ABCD matrix model. Tables 6 and 7 show the calculated formant frequencies with respect to the Figs 12 and 13, respectively. From Tables 6 and 7, the proposed model and Mathur's model have maximum the same relative errors 0.16% with ABCD matrix model. In both cases, the negligible errors show that the formants frequencies of the current model and Mathur's model are very close to that of ABCD matrix model.

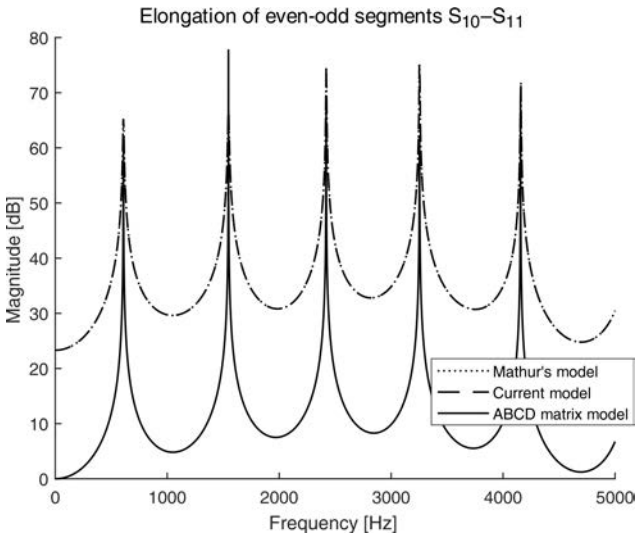

Fig. 12. Comparison of formant frequencies of the current model and Mathur's model with ABCD matrix model for even-odd number segments elongation.

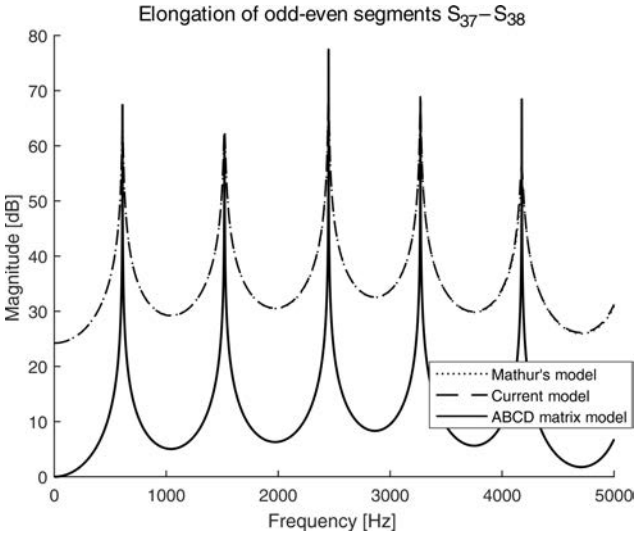

Fig. 13. Comparison of formant frequencies of the current model and Mathur's model with ABCD matrix model for odd-even number segments elongation.

From Figs 7–13 and Tables 1–7, it is concluded that the present model and Mathur's model have approximately the same formants frequencies as those of the

Table 6. Relative errors of the current model and Mathur's model compared with ABCD matrix model for even-odd-number segments S_{10} and S_{11}.

Formant frequency	ABCD matrix model [Hz]	Mathur's model [Hz]	Proposed model [Hz]	Relative error in Mathur's model	Relative error in the proposed model
F1	609	608	608	0.16	0.16
F2	1547	1548	1548	0.06	0.06
F3	2420	2420	2420	0.00	0.00
F4	3254	3253	3253	0.03	0.03
F5	4156	4158	4158	0.05	0.05

Table 7. Relative errors of the current model and Mathur's model compared with ABCD matrix model for odd-even-number segments S_{37} and S_{38}.

Formant frequency	ABCD matrix model [Hz]	Mathur's model [Hz]	Proposed model [Hz]	Relative error in Mathur's model	Relative error in the proposed model
F1	609	610	610	0.16	0.16
F2	1521	1519	1519	0.13	0.13
F3	2450	2451	2451	0.04	0.04
F4	3271	3270	3270	0.03	0.03
F5	4176	4176	4176	0.00	0.00

ABCD matrix model. Despite the different approach used in the proposed model, our model generates exactly the same frequency profile as that of Mathur's model. However, the main difference between the proposed model and Mathur's model is the efficiency of the current model.

Table 8 presents the computational efficiency of the current model compared with the Mathur's model and ABCD matrix model. We used a high-level computer language Matlab 2016 for the development of computer codes for all the models. To compute the maximum elapsed time, all the segments of the vocal tract have been elongated with a factor of 1.5% except for the first segment. We worked on a laptop with Windows 10 Professional as an operating system. The specifications of the laptop are: i7-7500U processor, 8 GB RAM, and dual-core architecture. The actual average elapsed time of twenty iterations of the algorithms has been shown in the second column of the table and the normalised elapsed time has been mentioned in the third column, where normalisation was performed by the elapsed time of the proposed method. It may be noted from the table that the present model is 1.23 times more efficient than the Mathur's model and more than four times more efficient than the ABCD matrix model.

Table 8. Elapsed time taken by the proposed, Mathur's, and ABCD matrix models.

Methods	Time [s]	Normalized time
Proposed model	0.7556424	1.00
Mathur's model	0.9317220	1.23
ABCD matrix model	3.1517223	4.17

5. Conclusions

The vocal tract of the length 17.5 cm has been chosen in the present work. The vocal tract may be divided into N number of segments. The new model has been devised to elongate or shorten any arbitrary cylindrical segment by a suitable scaling factor. This model has a single algorithm and there is no need to make the sections of the segments for the elongation or shortening of the intermediate segments. The fractional delay has been approximated with linear Lagrange interpolator. Many cases include elongation of the first segment, last segment, intermediate even segment, intermediate odd segment. Intermediate even-odd segments and intermediate odd-even segments have been tested with the current model and the results have been validated with a more accurate ABCD matrix model. In all cases, we conclude that:

- Despite the different approach used in the current model, the formants frequencies generated by the proposed model are equal to that of Mathur's model.
- The formants frequencies of the proposed model and Mathur's model are found to be very closely matched with that of benchmarked ABCD matrix model.
- The proposed model is about 23% more efficient than Mathur's model.
- The proposed model is about 400% more efficient than the benchmark ABCD matrix model.

Hopefully, the proposed model in the present work may serve as a useful vocal tract model in speech synthesizers.

Acknowledgment

The authors of this paper acknowledge the financial support provided by the Higher Education Commission of Pakistan (HEC) under Start-up Research Grant Program (SRGP).

References

1. Avanzini F., Alku P., Karjalainen M. (2001), *One-delayed-mass model for efficient synthesis of glottal flow*, [in:] 7th European Conference on Speech Communication and Technology, "INTERSPEECH", pp. 51–54.
2. Birkholz P., Kröger B.J., Neuschaefer-Rube C. (2010), *Articulatory synthesis and perception of plosive-vowel syllables with virtual consonant targets*, [in:] Proceedings of the 11th Annual Conference of the International Speech Communication Association, pp. 1017–1020, Makuhari, Chiba, Japan.
3. Cooper C., Murphy D., Howard D., Tyrrell A. (2006), *Singing synthesis with an evolved physical model*, IEEE Transactions on Audio, Speech, and Language Processing, **14**, 1454–1461.
4. Fant G. (1971), *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*, Walter de Gruyter.
5. Fettweis A. (1971), *Digital filters related to classical structures*, AEU: Archive für Elektronik und Übertragungstechnik, **25**, 78–89.
6. Flanagan J., Landgraf L. (1968), *Self-oscillating source for vocal-tract synthesizers*, IEEE Transactions on Audio and Electroacoustics, **16**, 57–64.
7. Gold B., Morgan N., Ellis D. (2011), *Speech and audio signal processing: processing and perception of speech and music*, John Wiley & Sons.
8. Gunnar F. (1960), *The acoustic theory of speech production*, s'Gravenhage, Mouton.
9. Hoefer W. (1985), *The transmission-line matrix method theory and applications*, IEEE Transactions on Microwave Theory and Techniques, **33**, 882–893.
10. Ishizaka K., Falanagan J.L. (1972), *Synthesis of voiced sounds from a two-mass model of the vocal cords*, Bell System Technical Journal, **51**, 1233–1268.

11. Ishizaka K., Flanagan J. (1977), *Acoustic properties of longitudinal displacement in vocal cord vibration*, Bell System Technical Journal, **56**, 889–918.
12. Johns P.B., Beurle R. (1971), *Numerical solution of 2-dimensional scattering problems using a transmission-line matrix*, [in:] Proceedings of the Institution of Electrical Engineers, Vol. 118, pp. 1203–1208, IET.
13. Karjalainen M. (2003), *Mixed physical modeling: DWG+ FDTD+ WDF*, [in:] Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on, pp. 225–228. IEEE, New Paltz, NY.
14. Kelly J.L., Lochbaum C.C. (1962), *Speech synthesis*, [in:] Proceedings of the Stockholm Speech Communications Seminar, RIT, Stockholm, Sweden, pp. 1–4.
15. Laakso T.I., Valimaki V., Karjalainen M., Laine U.K. (1996), *Splitting the unit delay: tool for fractional delay filter design*, IEEE Signal Processing Magazine, **13**, 30–60.
16. Lim I.-T., Lee B.G. (1993), *Lossless pole-zero modeling of speech signals*, IEEE Transactions on Speech and Audio Processing, **1**, 269–276.
17. Maddox A., Oren L., Khosla S., Gutmark E. (2014), *Prediction of pressure distribution between the vocal folds using Bernoulli's equation*, The Journal of the Acoustical Society of America, **136**, 2126–2126.
18. Markel J.E., Gray A.H. (1976), *Linear prediction of speech*, Springer-Verlag, New York.
19. Mathur S., Story B.H., Rodriguez J.J. (2006), *Vocal-tract modeling: Fractional elongation of segment lengths in a waveguide model with half-sample delays*, IEEE Transactions on Audio, Speech, and Language Processing, **14**, 1754–1762.
20. Morse P. (1981), *Vibration and Sound*, Acoustical Society of America.
21. Mullen J., Howard D.M., Murphy D.T. (2003), *Digital waveguide mesh modeling of the vocal tract acoustics*, [in:] Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop, pp. 119–122, IEEE.
22. Mullen J., Howard D.M., Murphy D.T. (2006), *Waveguide physical modeling of vocal tract acoustics: flexible formant bandwidth control from increased model dimensionality*, IEEE Transactions on Audio, Speech, and Language Processing, **14**, 964–971.
23. Mullen J., Howard D.M., Murphy D.T. (2007), *Real-time dynamic articulations in the 2-D waveguide mesh vocal tract model*, IEEE Transactions on Audio, Speech, and Language Processing, **15**, 577–585.
24. Qureshi T., Syed K. (2011a), *A one-mass physical model of the vocal folds with seesaw-like oscillations*, Archives of Acoustics, **36**, 1, 15–27.
25. Qureshi T.M., Syed K.S. (2011b), *A new approach to parametric modeling of glottal flow*, Archives of Acoustics, **36**, 4, 695–712.
26. Qureshi T.M., Syed K.S. (2015), *Two dimensional featured one dimensional digital waveguide model for the vocal tract*, Computer Speech & Language, **33**, 47–66.
27. Rabiner L.R., Schafer R.W. (1978), *Digital processing of speech signals*, Prentice-Hall.
28. Samadi S., Ahmad M.O., Swamy M. (2004), *Results on maximally flat fractional-delay systems*, IEEE Transactions on Circuits and Systems I: Regular Papers, **51**, 2271–2286.
29. Savioja L., Rinne T.J., Takala T. (1994), *Simulation of room acoustics with a 3D finite difference mesh*, [in:] Proceedings of International Computer Music Conference, pp. 463–466, Aarhus, Denmark.
30. Shimamura R., Tokuda I.T. (2016), *Effect of level difference between left and right vocal folds on phonation: Physical experiment and theoretical study*, The Journal of the Acoustical Society of America, **140**, 3393–3394.
31. Smith J.O. (1985), *A new approach to digital reverberation using closed waveguide networks*, [in:] Proceedings of International Computer Music Conference, pp. 47–53, Vancouver, Canada.
32. Smith J.O. (1992), *Physical modeling using digital waveguides*. Computer Music Journal, **16**, 74–91.
33. Smith J.O. (2002), *Principles of digital waveguide models of musical instruments*. [in:] M. Kahrsand, K. Brandenburg [Eds.], Applications of digital signal processing to audio and acoustics, pp. 417–466, Kluwer Academic Publishers, Boston, Dordrecht, London.
34. Sondhi M., Schroeter J. (1987), *A hybrid time-frequency domain articulatory speech synthesizer*, IEEE Transactions on Acoustics, Speech and Signal Processing, **35**, 955–967.
35. Speed M., Murphy D., Howard D. (2013), *Three-dimensional digital waveguide mesh simulation of cylindrical vocal tract analogs*, IEEE Transaction on Audio Speech, and Language Processing, **21**, 449–454.
36. Story B.H. (2013), *Phrase-level speech simulation with an airway modulation model of speech production*, Computer Speech & Language, **27**, 989–1010.
37. Story B.H., Titze I.R. (1998), *Parameterization of vocal tract area functions by empirical orthogonal modes*, Journal of Phonetics, **26**, 223–260.
38. Titze I.R., Titze I.R. (2014). *One glottal airflow – Two vocal folds*, The Journal of the Acoustical Society of America, **136**, 2163–2163.
39. Välimäki V. (1995), *Discrete-time modeling of acoustic tubes using fractional delay filters*, Helsinki University of Technology.
40. Välimäki V., Karjalainen M. (1994), *Improving the Kelly-Lochbaum vocal tract model using conical tube sections and fractional delay filtering techniques*, [in:] Processings of the International Conference on Spoken Language Processing (ICSLP), Vol. 2, pp. 615–618, Yokohama, Japan.
41. Välimäki V., Pakarinen J., Erkut C., Karjalainen M. (2006), *Discrete-time modelling of musical instruments*, Reports on Progress in Physics, **69**, 1–78.
42. Vampola T., Horáček J., Laukkanen A.-M., Švec J.G. (2015), *Human vocal tract resonances and*

the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement, Logopedics Phoniatrics Vocology, **40**, 14–23.

43. Van Duyne S.A., Smith J.O. (1993a), *The 2-D digital waveguide mesh*, [in:] Applications of Signal Processing to Audio and Acoustics. Final Program and Paper Summaries, 1993 IEEE Workshop, pp. 177–180, IEEE, New Paltz, NY.

44. Van Duyne S.A., Smith J.O. (1993b), *Physical modeling with the 2-D digital waveguide mesh*, [in:] Proceedings of the International Computer Music Conference, pp. 40–40, International Computer Music Association, Tokyo, Japan.

45. von Helmholtz H. (1863), *On the sensations of tone as a physiological basis for the theory of music* [in German: *Die Lehre von den Tonempfindungen ais physiologische Grundlage für die Theorie der Musik*], Braunschweig.

46. von Helmholtz H. (1866), *Treatise on physiological optics* [in German: *Handbuch der physiologischen Optik*], Leopold Voss, Leipzig.

47. Wang Y., Wang H., Wei J., Dang J. (2012a), *Acoustic analysis of the vocal tract from a 3D physiological articulatory model by finite-difference time-domain method*, [in:] Proceeding of international conference on Automatic Control and Artificial Intelligence, pp. 329–333, IET, Xiamen, China.

48. Wang Y., Wang H., Wei J., Dang J. (2012b), *Mandarin vowel synthesis based on 2D and 3D vocal tract model by finite-difference time-domain method*, [in:] Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific, pp. 1–4, IEEE, Hollywood, CA.