ARCHIVES OF ACOUSTICS
36, 4, 695–712 (2011) *DOI: 10.2478/v10168-011-0047-3*

A New Approach to Parametric Modeling of Glottal Flow

Tahir Mushtaq QURESHI, Khalid Saifullah SYED

Bahauddin Zakariya University
Center for Advance Studies in Pure and Applied Mathematics
Bosan Road, 60800, Multan, Pakistan
e-mail: tahmush@hotmail.com

(*received August 7, 2010; accepted July 18, 2011*)

Glottal waveform models have long been employed in improving the quality of speech synthesis. This paper presents a new approach for modeling the glottal flow. The model is based on three control volumes that strike a one-mass and two-springs system sequentially and generate a glottal pulse. The first, second and third control volumes represent the opening, closing and closed phases of the vocal folds, respectively. The masses of the three control volumes and the size of the first one are the four parameters that define the shape, pitch and amplitude of the glottal pulse. The model may be viewed as parametric approach governed by second order differential equations rather than analytical functions and is very flexible for designing a glottal pulse. The glottal pulse generated by the present model, when compared with those generated by Rosenberg, LF and mucosal wave propagation models demonstrates that it appropriately represents the opening, closing and closed phases of the vocal fold oscillation. This leads to the validity of our model. Numerical solution of the present model has been found to be very efficient as compared to its analytical solution and two other well-known parametric models Rosenberg++ and LF. The accuracy of the numerical solution has been illustrated with the help of analytical solution. It has been observed that the accuracy improves by increasing the size of the first control volume and may decrease insignificantly with increase in the mass of any of the control volumes. Two experiments with the present model support its successful implementation as a voice source in speech synthesis. Thus our model renders itself as an efficient, accurate and realistic choice as a voice source to be employed in real-time speech production.

Keywords: control volumes, spring-mass system, vocal folds, Rosenberg glottal model, LF glottal model.

1. Introduction

Synthesizing high quality natural sounding speech has been an important topic of research for many years. There is a wide range of applications such as teaching aids, telecommunication, messaging systems, alternative access to information, augmentative communication, foreign language instruction and tutoring systems, first language instruction for hearing impaired, job training simulations, technical support for complex tasks, automated announcement systems, information and service access and speech-to-speech translation systems, in which speech synthesis systems are currently employed. All these areas of applications may benefit from achieving a natural sounding speech synthesis.

An efficient speech production model is fundamental for obtaining good performance in applications such as speech synthesis, coding. The vocal folds hold a central place in the description of such a model. The intelligibility of speech is mostly attributed to the configuration of vocal tract while the glottal source is essential in voice quality and speaker identity. For more accurate synthesis and analysis of speech, the observation of physical effects in the vocal folds and their inclusion in the model of vocal source are very important. Many approaches have been proposed in the literature for the modeling of vocal folds action. However, parametric and physical models are the most widely known approaches for this purpose.

In physical models, the oscillatory characteristics of the vocal folds are preformed through numerical simulations based on a physical description of the vocal fold system (Flanagan, Landgraf, 1968; Ishizaka, Flanagan, 1972; Liljencrants, 1991; Pelorson *et al.*, 1994; Herzel, Knudsen, 1995; Story, Titze, 1995; Berry, Titze, 1996; Lous *et al.*, 1998; de Vries *et al.*, 1999; Gunter, 2003). On the other hand parametric models parameterize the glottal flow or its first-time derivative in terms of piecewise analytical functions for different phases of the glottal cycle. To facilitate the analysis of the source signal, and to enable efficient characterization in terms of a small set of parameters, parametric source models have proven very useful.

Various parametric glottal models have been proposed in the literature (Rosenberg, 1971; Rothenberg *et al.*, 1975; Fant, 1979a; 1979b; 1982b; Fant *et al.*, 1985; Ananthapadmanabha, 1984; Hedelin, 1984; Fujisaki, Ljungqvist, 1986; Price, 1989; Klatt, Klatt, 1990; Schoentgen, 1993; Qi, Bi, 1994; Veldhuis, 1998). However, they share many common features, and they can generally be described by three to five parameters plus the fundamental frequency. One of the popular models to quantify the glottal-pulse is the Liljencrants-Fant (LF) model where the shape of the glottal-pulse is described by four parameters (Fant *et al.*, 1985). Unfortunately, its use in speech models is limited because of its computational complexity and inefficiency. This model involves a non-linear equation whose solution leads to the complex computation of generation parameters from the specification parameters. However, Raymond Veldhuis presents an alternative model for the LF model, which is derived from the Rosenberg model and calls it Rosenberg++ model (Rosenberg, 1971; Fant *et al.*, 1985; Veldhuis, 1998). The Rosenberg++ model has the same features as the LF model but it has the advantage over the later that it is computationally more efficient.

This work presents a four-parametric glottal model that simulates the pulse generated by the vocal folds action by a sequential interaction of three control volumes with a spring-mass system. The masses of these control volumes and the size of the first control volume are the controlling parameters that determine the shape, pitch and amplitude of the pulse. The present approach may be regarded as modeling the vocal folds oscillation and unsteady airflow collectively. This model sets a new dimension in the paradigm of parametric models as the glottal flow is represented by differential equations rather than by analytical functions, whose solution may be found numerically as well as analytically. Its validity, accuracy and efficiency have been established by various criteria. The effects of variation in masses and the size of the first volume on the accuracy of the numerical solution have also been shown. The present model may practically be used in real-time speech production because of its computational efficiency.

Four more sections follow the present section. In Sec. 2, we describe our proposed glottal model with geometrical view. We also develop its mathematical formulation, which results in a system of three initial value problems coupled with each other in a sequential manner. Section 3 describes how to obtain the numerical solution of this system. Section 4 is dedicated for the results and discussion. In this section, we present the glottal pulse generated by our model and demonstrate how the masses of the first and second control volumes influence its shape and the size of the first control volume governs its amplitude. The validity, accuracy and efficiency of the numerical solution are then illustrated. At the end of this section, we give a comparison of the original glottal flow, extracted from vowel /a/ by the inverse filter technique, with the model flow. The waveform of vowel /a/ generated from VOX software (Kob, 2004) having source as our proposed glottal model has also been shown. Section 5 is reserved for the conclusions.

2. Glottal model

This paper presents a new parametric approach for modeling the glottal pulse. The concept has been derived from the real physical process of measuring this pulse. In one cycle of vibration, the vocal folds chop the blowing air into three segments during its opening, closing and closed phases. In the segment of the opening phase, there is an increase in the glottal volume flow while a decrease in the glottal volume flow and negligible or zero glottal volume flow occur in the segments of closing and closed phases, respectively. We assume that the vocal tract is not coupled to the glottis thus allowing these three segments to enter into the space. In the space, the variation in volume flow rate generates pressure waves, which travel in the direction of flow. These traveling pressure waves strike the physical media like a microphone whenever it comes on its way, and impose its vibratory characteristics on it. The response of the physical media is just the measurement of the glottal pulse.

In the present model we generate the glottal pulse by a novel approach that is quite simple and very efficient. We consider three control volumes of constant densities and sizes that strike successively to a one-mass and two-spring system with their appropriate respective time periods and generate an approximated glottal pulse. These control volumes may be regarded as containing, in order, the air coming out of the vocal folds in their opening, closing and closed phases during one complete cycle. Although there will be density variations within the flow coming out during each of the opening and closing phases, lumping the phase-wise flow in three control volumes renders phase-wise constant densities so that each control volume is characterized by its own density and size. This is a great simplification as it no more requires unsteady airflow analysis. Moreover, the density gradients present in the glottal flow are consequences of the vocal folds oscillation and, in turn, determine the attributes of the glottal pulse. In the present model, these attributes are determined by the densities and sizes of the control volumes and we will see later that the masses of these control volumes control the pitch and shape of the pulse and the size of the first control volume determines its amplitude. These constitute the four parameters of our model that together with a spring-mass system characterize the vocal fold oscillation and determine the properties of the pulse generated by the present model. Thus the present approach may be regarded as modeling the airflow and fold oscillation collectively. The schematic diagram of the present model may be viewed in Fig. 1.

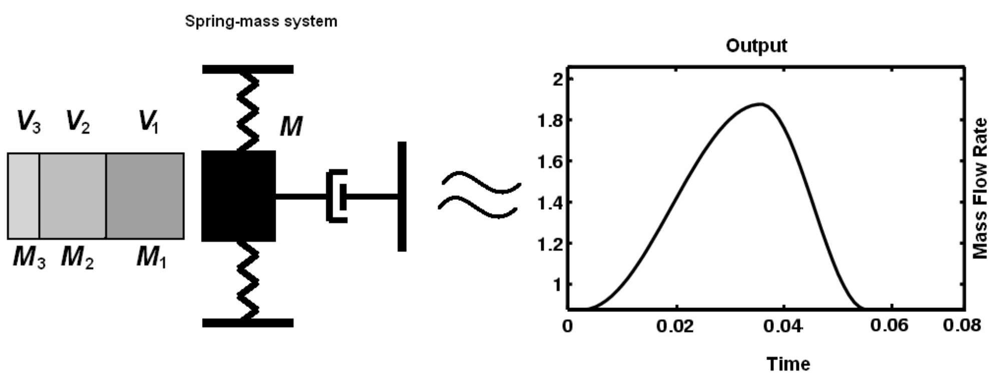

Fig. 1. Pulse generation in the present model.

In view of the above description, the present model is parametric in nature but it carries the physical sense of airflow in the form of three control volumes and that of the vocal folds oscillation due to the presence of the spring-mass system in the model. It simplifies and approximates the physical process of measuring the glottal flow as it encompasses the gradient of flow during one complete cycle of the vocal folds vibration into three control volumes and defines a new dimension for the investigation of speech production on the basis of control volumes.

Now we give formal description of our model followed by the presentation of certain parametric relations and the mathematical formulation.

2.1. Formal description

Let V_1, V_2 and V_3 be the three control volumes related to the opening, closing and closed phases of our glottal model, where M_1, M_2 and M_3 are the masses contained in these control volumes, respectively. Densities related to these control volumes may be expressed as $\rho_i = \dfrac{M_i}{V_i}$, where $i = 1, 2, 3$. We consider a spring-mass system consisting of two springs and one mass M together with three control volumes joined in series as shown in Fig. 1. We assume this mass to be negligible as compared to M_1, M_2 and M_3. Let K be the spring constant of this system and C be its damping constant. The interaction of three control volumes with the spring-mass system represents the glottal model as described below.

In the first step, the control volume V_1 strikes with the mass M of the spring-mass system and is attached to it so that the mass of the spring-mass system becomes $M + M_1$. Since M is negligible in comparison with M_1, we take the total mass of the system to be M_1. Let the density ρ_1 of the control volume V_1 be greater than the normal density ρ_0 of the fluid (air at temperature 20°C). As the volume V_1 strikes the spring-mass system, it exerts a force, in the direction of flow proportional to $\rho_1 - \rho_0$ on it which consequently starts oscillation with a time period, say, T_{01}. It may be noted that the mass of the system remains M_1 for the period $\left(\dfrac{T_{01}}{2}\right)$. The above description represents the opening phase of the glottal model. For the closing phase, we suppose that when the spring-mass system completes its half time period $\left(\dfrac{T_{01}}{2}\right)$, the control volume V_2 strikes the system and is attached to it resulting into oscillation of the spring-mass system with a different time period, say, T_{02}. In this phase, the mass of the system becomes M_2 and remains so for the period $\left(\dfrac{T_{02}}{2}\right)$. The force exerted in this phase is equal and opposite in sign to that exerted in the opening phase so that at the end of the closing phase, the system comes to its original position. We now assume that the control volume V_3 strikes the system when half the time period $\left(\dfrac{T_{02}}{2}\right)$ of the spring-mass system is completed. Since the third control volume V_3 is related to the closed phase of the glottal model in which the vocal folds are fully closed and there is no volumetric flow rate, the interaction of V_3 with the spring-mass system will not exert any force on the system. Under the condition of no force on the system, the damping factor may be assumed to be so high that it stops the oscillation of the spring-mass system abruptly. Physically, the situation of the closed phase may be conceived by taking $\rho_3 = \rho_0$. We denote the time period of this closed phase by T_{03}.

From the above description we note that the independent parameters of our model are M_1, M_2, M_3 and V_1. V_2 is not independent as it is so chosen that the force exerted by the second control volume on the spring-mass system is equal in magnitude and opposite in sign to that exerted by the first control volume. V_3 is determined by the equation $\rho_3 = \rho_0$. It may be noted that in terms of the number of parameters, our model matches those that require specification of only four parameters.

2.2. Parametric relations

We now present the parametric relations for the calculation of the time period and the force of each phase of the glottal model.

Let T_i be the half of the time period T_{0i} of the i-th phase of the model. Then

$$T_i = \frac{T_{0i}}{2} = \pi\sqrt{\frac{M_i}{K}}, \qquad i = 1, 2, 3. \tag{1}$$

It may be noted that the pitch period of glottal pulse becomes $\sum\limits_{i=1}^{3} T_i$.

Forces exerted by the three control volumes on the spring-mass system are in the transverse direction and may be expressed as

$$\begin{cases} F_i = (-1)^{i+1}\,(\rho_1 - \rho_0)\,RT & \text{for control volumes } V_i, \quad i = 1, 2, \\ F_3 = 0 & \text{for control volume } V_3. \end{cases} \tag{2}$$

We note that the forces F_1 and F_2, for the opening and closing phases of the model respectively, are equal in magnitude but opposite in sign to bring the system to its initial position after the completion of the closing phase. R is a gas constant and T is the temperature taken here as 20°C (293.15 Kelvin).

From Eq. (2), the size of the control volume V_1 defines the magnitude of the external force, which in turns determines the amplitude of the glottal pulse.

In order to determine the size of the control volume V_1 for the specific amplitude, we develop a relation between them as follows.

From Eq. (1), we have

$$F_1 = (\rho_1 - \rho_0)\,RT = \rho_1 RT - \rho_0 RT = \frac{M_1}{V_1}RT - \rho_0 RT, \qquad \text{for} \quad i = 1, \tag{3}$$

but from the spring-mass system, the force can be represented as

$$F_1 = \frac{K(X_{\max} - X_0)}{2}, \tag{4}$$

where K is a spring constant, X_0 and $X_{\max}$ represent extreme positions of the spring-mass system so that $(X_{\max} - X_0)$ defines its amplitude. A comparison of Eqs. (3) and (4) leads to the following required relation.

$$V_1 = \frac{2M_1RT}{(K(X_{\max} - X_0) + 2\rho_0RT)}. \tag{5}$$

2.3. Mathematical formulation

We now present the mathematical model of the spring-mass system in each of the three phases of our glottal model. The spring-mass equation for the opening phase may be written as

$$M_1\ddot{X}(t) + C\dot{X}(t) + K(X(t) - X_0) = F_1, \tag{6}$$

where M_1, C, F_1 and K are as described earlier, X_0 is the initial state of the opening phase and $X(t)$ is the displacement of the mass of the spring-mass system at time t. The initial condition for this phase may be taken to be

$$X(0) = X_0 = 0 \qquad \text{and} \qquad \dot{X}(0) = 0. \tag{7}$$

Since the first phase is completed at $t = T_1$, Eq. (6) is to be integrated subject to the above initial conditions given in Eq. (7) in the time interval $(0, T_1]$. As $t = T_1$ also marks the onset of the second phase of the glottal model, the condition of the system at $t = T_1$ determines the initial conditions of the second phase. The initial value problem for the second phase may, thus, be expressed as

$$M_2\ddot{X}(t) + C\dot{X}(t) + K(X(t) - X_0) = F_2, \qquad \forall \quad T_1 < t \leq T_1 + T_2, \tag{8}$$

where $X_0 = X(T_1)$. The initial conditions for this phase are

$$\dot{X}(T_1) = 0 \qquad \text{and} \qquad X(T_1) = X_0. \tag{9}$$

Since in the closed phase of the glottal model, there is no oscillation and the system is at rest with the half time period T_3, the state of the system in this phase may be represented by

$$M_3\ddot{X}(t) + C\dot{X}(t) + K(X(t) - X_0) = F_3, \tag{10}$$

such that

$$X(t) = 0, \qquad T_1 + T_2 \leq t \leq T_1 + T_2 + T_3. \tag{11}$$

Equations (6), (8) and (10) show that our proposed glottal model is a system of three, second order ODEs, where each ODE is related to the opening, closing and closed phases of the model having specific time periods, respectively.

The exact solution of the system has been given in Appendix.

3. Numerical solution

Our parametric model consists of three ODE's (6), (8) and (10) subjected to the initial conditions (7), (9) and (11) representing the three phases of glottal oscillation namely the opening, closing and closed phases, respectively. The solution of the above system may be found analytically as well as numerically. The analytical solution has been presented in Appendix. The complicated expression of the analytical solution may make this solution computationally inefficient and practically not suitable for real-time applications.

The numerical solution of the present model may be obtained by using a numerical integrator like the Euler method, Runge-Kutta method of order 4, etc. The Euler method is computationally the most efficient one but has the drawbacks of being less accurate and conditionally stable. On the other hand, the Runge-Kutta method of order 4 is much more accurate, unconditionally stable and easy to implement.

In the present work, we have computed the numerical solution of our model using the Euler method as well as the Runge-Kutta method of order 4 to establish the accuracy and efficiency of the numerical solution. Relevant results with discussion are presented in the next section.

4. Results and discussion

In the previous sections we have described our four-parametric glottal model and the procedure for finding its numerical solution. In this section, we describe the working of our model, its characteristics, efficiency and its application in speech production. The working of the model and its characteristics are described by specifying various constants involved in the model and by investigating the effect of four parameters, namely, the masses of three control volumes and the size of the first one, on the shape of the glottal pulse generated by our model. We will see that the mass of each control volume determines the shape of the pulse in its respective phase, while the size of the first control volume determines its amplitude. We also demonstrate the validity of our model by comparing its glottal pulse with those of three literature models (Rosenberg, 1971; Fant *et al.*, 1985; Drioli, 2002), the accuracy with the help of a real source and the efficiency by two literature models (Veldhuis, 1998; Fant *et al.*, 1985). A demonstration of the effects of variation in masses and the size of the first control volume on the accuracy of the numerical solution is also presented. For the application of our model we first extract the original glottal signal from vowel /a/ of a male speaker by using the Inverse Filter technique and fit our model pulse on it. This shows how well a real source may be approximated by our model pulse. Then we use our model pulse as a source in the VOX software (Kob, 2004) to synthesize the vowel /a/ of the word /had/.

For the working of the present model, we need to specify the following constants, $C = 1 \times 10^{-7}$, $\rho_0 = 0.001293$, $R = 0.28705$, $K = 0.1$, $T = 293.15$ Kelvin (20°C), where C is damping constant, ρ_0 (gm/cm^3) is the density of air at normal temperature 20°C, R is the gas constant, K is the spring constant and T is the temperature.

The four parameters of our model which control the shape of the pulse are the masses M_1, M_2 and M_3 of the control volumes respectively and the size of V_1. Various choices of the values of these parameters to be used in the present work are given in Table 1 and will be referred to as these are used. The time periods and forces may be determined from Eqs. (1) and (2), respectively. Equation (1) shows that the time period of a phase of the glottal model depends on the mass of the corresponding control volume. The greater the mass, the larger will be the time period. From Eq. (2), we note that the force exerted by a control volume on the spring-mass system is determined by its density. Therefore, by keeping the mass of a control volume constant and changing its size, we can adjust the corresponding force on the system.

Table 1. Various choices of volume sizes and masses of control volumes used in the numerical simulation.

	Control volumes [cm^3]	Masses [gram]		
	V_1	M_1	M_2	M_3
R1	2.8	0.028	0.01	0.0065
R2	2.8	0.1005	0.047	0.0823

Figure 2 shows the glottal pulse generated by the present model with the pitch period 70 for the sizes and masses of control volumes given in row R1 of

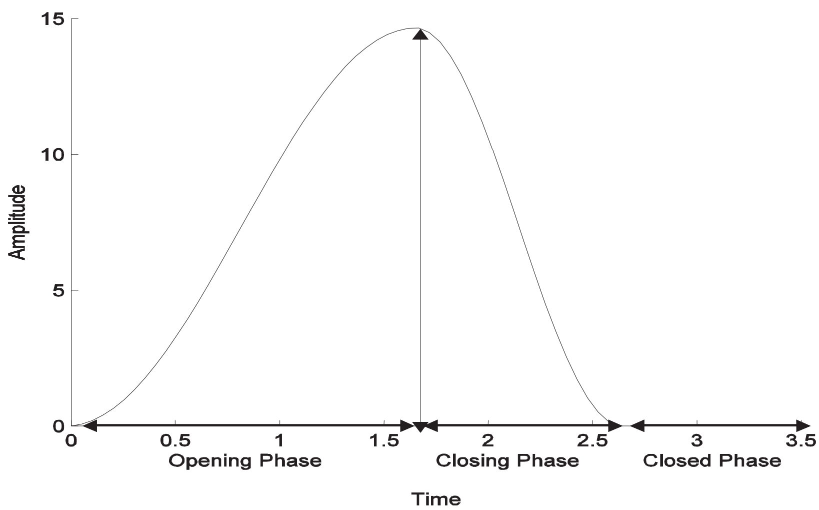

Fig. 2. Single glottal pulse shape, generated from the control volume based glottal model.

Table 1. We note from the figure that the time period of the opening phase is larger than that of the closing phase. In view of Eq. (1), this is due to the larger mass of the first control volume(M_1). Hence we may design pulses of different shapes and pitches by changing the masses of the control volumes. Figure 3a shows pulses generated by our model for various values of M_1 and Fig. 3b shows those generated for different values of M_2. The values of M_1 and M_2 used for these pulses are given in Table 2, while all the other parameters of the model are specified according to row R1 of Table 1.

We can observe from these figures the effects of variations in these masses on the time periods of the opening and closing phases, which then determine the shape and pitch of the pulses. We note that in Fig. 3a, the time period of the opening phase changes with time delay of the constant time period of the closing phase while, in Fig. 3b the time period of the closing phase changes without any time delay of the opening phase.

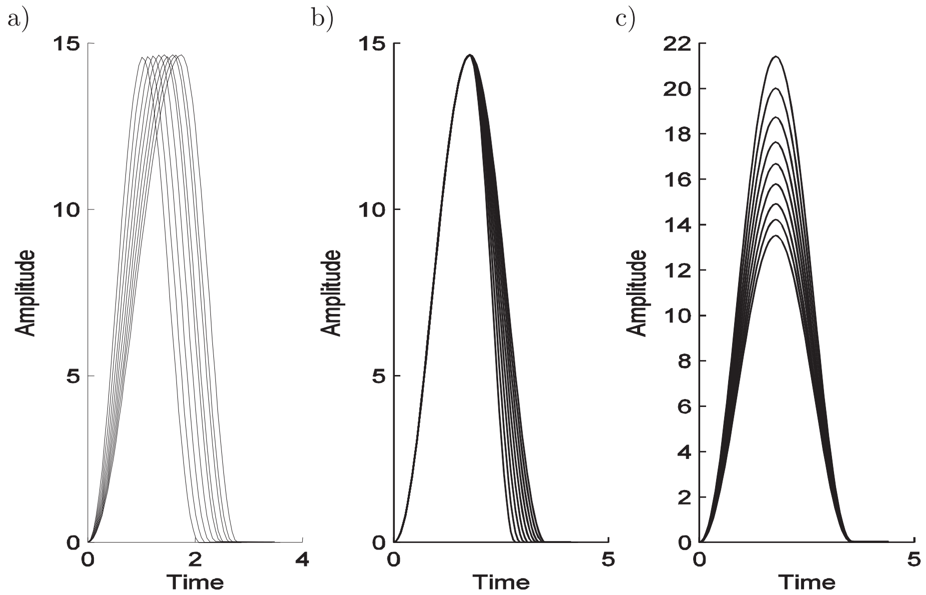

Fig. 3. a) Shape variation of the glottal pulse against variation in the mass of the first control volume; b) shape variation of the glottal pulse against variation in the mass of the second control volume; c) amplitude variation of the glottal pulse against variation in the size of the first control volume.

We now describe how the amplitude of the pulse generated by our model can be varied. We know that the amplitude of a spring-mass system depends on the force exerted on the system, which, in turn, depends on the density of the first control volume in view of Eq. (2). This implies that for a specified mass of the first control volume, we can change the amplitude of the pulse by changing the size of the control volume. Figure 3c shows the variation in the amplitude of the glottal pulse when the size of the first control volume is changed. The pulses shown in this Fig. correspond to the values of V_1 given in Table 2, all the other parameters being chosen according to R1 of Table 1. However, for a specific amplitude of the glottal pulse, Eq. (5) gives the corresponding size of the control volume V_1.

Table 2. Variation of masses contained in the first and second control volumes and that of the size of the first control volume for generating pulses of various shapes.

M_1 [gram]	M_2 [gram]	V_1 [cm^3]
0.0280	0.0100	3.00
0.0258	0.0123	2.87
0.0235	0.0145	2.75
0.0212	0.0168	2.62
0.0190	0.0190	2.50
0.0168	0.0212	2.37
0.0145	0.0235	2.25
0.0123	0.0258	2.12
0.0100	0.0280	2.00

We now establish the validity, accuracy and efficiency of our model by various criteria. The validity is proven by comparing the pulse generated by the present model with those of three well-known models, two of which are parametric and one is a physical one. The parametric models are the Rosenberg and LF (Rosenberg, 1971; Fant *et al.*, 1985), while the physical model is a mucosal wave propagation model (Drioli, 2002). Computational efficiency has been shown by comparing the CPU time taken in computing the present numerical solution, the present analytical solution and the solutions of two literature models Rosenberg++ and LF (Veldhuis, 1998; Fant *et al.*, 1985). The accuracy of the present model has been established by investigating how well it approximates a real voice source.

Figure 4 shows the pulses generated by the present model, the Rosenberg model, the LF model and the mucosal wave propagation model. These pulses have the same pitch period of 158 and are scaled to the same amplitude. Parametric values used for generating our pulse are given in R2 of Table 1. The glottal pulse of the Rosenberg model corresponds to closure time 0.6 and the positive/negative slope ratio of 0.5, while that of the LF model has been generated by taking the open phase parameter equal to 0.5, the positive/negative slope ratio parameter equal to 0.12 and the closure time constant/closed phase parameter equal to 0.2. For mucosal wave propagation model, the parameters and their values employed for the generation of its pulse are given in Table 6.

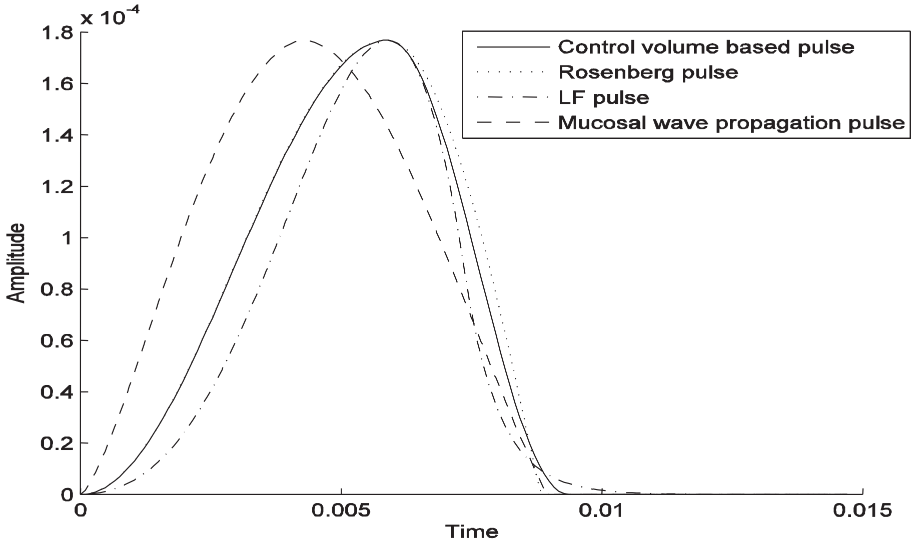

Fig. 4. Comparison among the control volume based glottal pulse, the Rosenberg glottal pulse, the LF glottal pulse and the mucosal wave propagation pulse.

We note from the figure that no two pulses match with each other over the whole time period. Therefore, this comparison may not be used for establishing the accuracy of the present pulse. The thing that matter is whether a glottal pulse represents opening, closing and closed phases or not. The present pulse represents these phases successfully. Moreover, the opening phase of our glottal pulse excellently matches with the opening phase of the Rosenberg glottal pulse, while the closing phase of our pulse rests between the closing phases of the Rosenberg and LF pulses.

Now we compare the computational efficiency of numerical and analytical solutions of our proposed model with the Rosenberg++ model and the LF model (Veldhuis, 1998; Fant *et al.*, 1985). Table 3 gives elapsed time taken by a P-IV computer (2.8 GHz processor, 512 MB ram and Visual C++ 6.0 computer language) in generating the pulses of our model based on the Euler method, the 4th order Runge-Kutta method and the analytical solution. This table also exhibits

Table 3. Elapsed Time taken by solutions of the proposed model based on numerical and analytical approaches in a pulse generation of pitch period 8×10^6 (time step = 0.001 and spring constant = 0.01).

Methods	Time [second]	Normalized Time
Euler numerical method (present model)	0.2626	1
4th order Runge-Kutta method (present model)	0.7412	2.8
Analytical solution (present model)	1.6559	6.3
Rosenberg++ model	1.7702	6.7
LF model	3.5340	13.5

the computational time taken by other parametric models such as the Rosenberg++ and LF ones. The second column of the table shows actual elapsed time, while the third column gives the normalized elapsed time, where normalization has been performed by the elapsed time of the Euler method. We set a long pitch period equal to 8×10^6 for each case to engage CPU for a long time. We note that the Euler method is about three times more efficient than the 4th order Runge-Kutta method and more than six times more efficient than the analytical solution.

The Euler method is usually less accurate than the 4th order Runge-Kutta method and is conditionally stable. However, it is the most efficient numerical method and makes a good choice when the efficiency is of great concern. The Runge-Kutta method of order four is more than two times more efficient than the analytical solution. The efficiency of the analytical solution of the present model is nearly equal to that of the Rosenberg++ model and is about 2.14 times more efficient than that of the LF model. This establishes the fact that the numerical solution of the present model is more efficient than its analytical solution as well as that of the well known Rosenberg++ and LF models. Therefore, it offers a good choice for its use as a voice source in real-time speech production.

Table 4 shows the maximum absolute error in the numerical solutions computed by the Euler and the Runge-Kutta methods for various values of the model parameters M_1, M_2 and V_1. Obviously, the Runge-Kutta method is more accurate than the Euler method with a minimum factor of 31.21. The table also demonstrates the effects of changes in the model parameters on our numerical solution. It has been examined that the accuracy improves by increasing the size of first control volume and may decrease insignificantly with increase in the mass of any of the control volumes.

Table 4. Maximum absolute error of the numerical solutions by taken the variation of the model parameters (time step = 0.001 and a spring constant = 0.01).

Parameters				Maximum absolute error	
M_1 [gram]	M_2 [gram]	M_3 [gram]	V_1 [cm^3]	Euler Method	Runge-Kutta Method
0.028	0.01	0.0065	2.8	0.00287	0.000055
0.019	0.01	0.0065	2.8	0.00195	0.000045
0.01	0.01	0.0065	2.8	0.00103	0.000033
0.028	0.019	0.0065	2.8	0.00208	0.000055
0.028	0.028	0.0065	2.8	0.00172	0.000055
0.028	0.01	0.0065	3	0.00267	0.000051
0.028	0.01	0.0065	2.5	0.00322	0.000061
0.028	0.01	0.0065	2	0.00402	0.000077

Now we investigate how well our model can approximate a real voice source. For this we recorded the vowel /a/ from a male speaker and the glottal flow obtained by the inverse filtering of voiced speech. The original glottal pulses and those generated by our model, for an appropriate choice of the sizes and masses of the control volumes, are shown in Fig. 5 for five pitch periods. The chosen values of the parameters are given in the second column of Table 5. The comparison is excellent. The global relative error in a single pulse of the present model is about 2%.

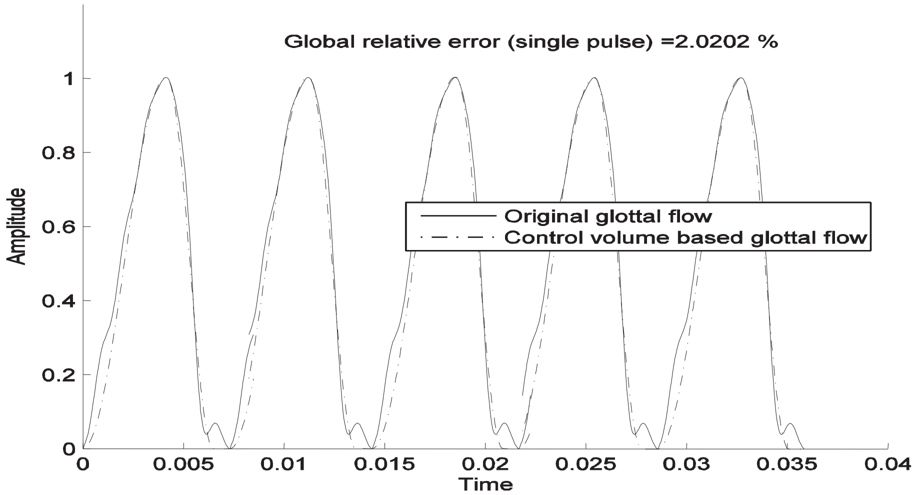

Fig. 5. Comparison of the control volume based glottal flow with inverse filtered glottal flow of a male speaker for vowel /a/.

Table 5. Parametric values for the present model used in the numerical simulation.

Parameters	Values	Units
M_1	0.00497	gram
M_2	0.00157	gram
M_3	0.00016	gram
V_1	0.1	cm^3
K	3000	gram-force/cm
Fs	22000	Hz

Table 6. Parametric values for the surface wave propagation vocal fold model.

Parameters	Values	Units
m	0.00017	kg
r	0.023	$N{\cdot}s{\cdot}m^{-1}$
L	0.014	m
ρ	1.15	$Kg{\cdot}m^{-3}$
k	34	$N{\cdot}m^{-1}$
Pl	3000	Pa (Pascal)
T	0.18	mm
x_0	0.05	mm
Sm	1.4×10^{-5}	m^2
Fs	11000	Hz

We performed another experiment with our model in connection with VOX software (Kob, 2004). The VOX software is based on physical modeling of human voice. For synthesis, it uses a multi-mass vocal fold model along with wave propagation through the vocal tract. It provides graphic environment for input and output of simulation data of the human vocal tract and vocal fold. We used our proposed glottal model as a source in VOX software, keeping the vocal tract set in a configuration for vowel /a/ of the word /had/, to synthesize the waveform of the vowel /a/. The simulation of the VOX software generated the waveform of the vowel /a/ in the word /had/ as shown in the Fig. 6.

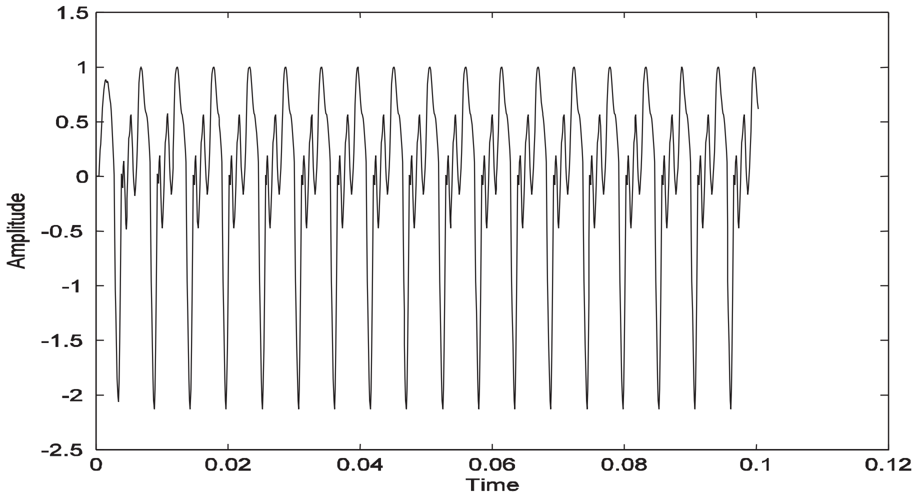

Fig. 6. Waveform of vowel /a/ of the word /had/ with proposed glottal model as a voice source in VOX software.

5. Conclusions

We have introduced a control volume based glottal model comprising three control volumes and one-mass and two-springs system. It has been derived from the idea that the flow of air through the glottis during the opening, closing and closed phases of vocal folds defines three basic control volumes, which have specific masses and sizes. These control volumes connected in series produce the shape of the glottal pulse when they interact with a spring-mass system. This model may be viewed as a parametric model and is controlled by four parameters. We have found that the shape of the glottal pulse is determined by the masses in these control volumes, while the amplitude varies by varying the size of the first control volume. The glottal pulse generated by the present model, when compared with those generated by the Rosenberg, LF and mucosal wave propagation models, demonstrates that it appropriately represents the opening, closing and closed phases of the vocal fold oscillation. Furthermore, a real voice source obtained by the inverse filtering technique is well approximated by our model indicating that it may be employed as a voice source. The present model, although parametric in nature, is governed by differential equations rather than analytical functions enabling us to find its solution analytically and/or numerically. In the present case, the numerical solution has been found to be more efficient than the analytical solution and the other parametric models Rosenberg++ and LF. It has been found that the Euler method of integration is about three times more efficient than the 4th order Runge-Kutta method, more than six times more efficient than the analytical solution and the Rosenberg++ model, and more than thirteen times more efficient than the LF model. The Rung-Kutta method of order four is more than two times more efficient than the analytical solution and the Rosenberg++ model, and about five times more efficient than the LF model. As far as the accuracy of the numerical solution is concerned, the Runge-Kutta method gives very accurate results and the accuracy improves by increasing the size of the first volume and may decrease insignificantly with increase in the mass of any of the control volumes. We conclude that the present model renders itself as an efficient, accurate and realistic voice source in real-time speech synthesis.

Glottal flow represented by the three control volumes with constant densities gives an opportunity to investigate the response of the vocal tract on these control volume for generating speech. We believe that the control volume based glottal model proposed in the present work may serve as a useful source model in speech synthesizers and will provide a new dimension for further investigations.

Appendix. Exact solution of the proposed model

$$G(t) = \begin{cases} \dfrac{-F_1}{\mu_1\sqrt{M_1 K}} e^{\left(\frac{-Bt}{2M_1}\right)} \cos\left(\mu_1 t - \tan^{-1}\left(\dfrac{B}{2\mu_1 M_1}\right)\right) + \dfrac{F_1}{K}, \\ \qquad\qquad \text{where } \mu_1 = \dfrac{\sqrt{4M_1K - B^2}}{2M_1} > 0, \quad \text{for } 0 \le t \le T_1, \\ \dfrac{F_1}{\mu_2\sqrt{M_2 K}} e^{\left(\frac{-B(t-T_1)}{2M_2}\right)} \cos\left(\mu_2 (t - T_1) - \tan^{-1}\left(\dfrac{B}{2\mu_2 M_2}\right)\right) + \dfrac{F_1}{K}, \\ \qquad\qquad \text{where } \mu_2 = \dfrac{\sqrt{4M_2K - B^2}}{2M_2} > 0, \quad \text{for } T_1 < t \le T_1 + T_2, \\ 0, \qquad\qquad \text{for } T_1 + T_2 < t \le T_1 + T_2 + T_3. \end{cases}$$

Acknowledgment

The authors of this paper acknowledge the financial support provided by the Higher Education Commission of Pakistan (HEC) under Indigenous Fellowship Scheme.

References

1. Ananthapadmanabha T.V. (1984), *Acoustic analysis of voice source dynamics*, Speech Transmission Laboratory-Quarterly Progress and Status Report, **2**, 3, 1–24.
2. Berry D.A., Titze I.R. (1996), *Normal modes in a continuum model of vocal fold tissues*, Journal of the Acoustical Society of America, **100**, 5, 3345–3354.
3. de Vries M.P., Schutte H.K., Verkerke G.J. (1999), *Determination of parameters for lumped parameter model of the vocal fold using a finite-element method approach*, Journal of the Acoustical Society of America, **106**, 6, 3620–3628.
4. Drioli C. (2002), *A flow waveform adaptive mechanical glotal model*, Speech, Music and Hearing Quarterly Progress and Status Report, 43, 69–79.
5. Fant G. (1979a), *Glottal source and excitation analysis*, Speech Transmission Laboratory-Quarterly Progress and Status Report, **20**, 1, 85–107.
6. Fant G. (1979b), *Vocal source analysis – a progress report*, Speech Transmission Laboratory-Quarterly Progress and Status Report, **20**, 3–4, 31–54.
7. Fant G. (1982b), *The voice source – acoustic modeling*, Speech Transmission Laboratory-Quarterly Progress and Status Report, **23**, 4, 28–48.
8. Fant G., Liljencrants J., Guang Lin Q. (1985), *A four-parameter model of glottal flow*, Speech Transmission Laboratory-Quarterly Progress and Status Report, **26**, 4, 1–13.
9. Flanagan J.L., Landgraf L.L. (1968), *Self-oscillating source for vocal tract synthesizers*, IEEE Transactions on Audio and Electroacoustics, **16**, 1, 57–64.
10. Fujisaki H., Ljungqvist M. (1986), *Proposal and evaluation of models for the glottal source waveform*, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1605–1608, Tokyo.
11. Gunter H.E. (2003), *A mechanical model of vocal-fold collision with high spatial and temporal resolution*, Journal of the Acoustical Society of America, **113**, 2, 994–1000.
12. Hedelin P. (1984), *A glottal LPC-vocoder*, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 161–164, San Diego.
13. Herzel H., Knudsen C. (1995), *Bifurcations in a vocal fold model*, Nonlinear Dynamics, **7**, 53–64.
14. Ishizaka K., Flanagan J.L. (1972), *Synthesis of voiced sounds from a two mass model of the vocal cords*, Bell System Technical Journal, **51**, 6, 1233–1268.
15. Klatt D.H., Klatt L.C. (1990), *Analysis synthesis, and perception of voice quality variations among female and male talkers*, Journal of the Acoustical Society of America, **87**, 2, 820–856.
16. Kob M. (2004), *Aachen University, Dep. of Phoniatrics, Pedaudiology, and Communication Disorders, Aachen, Germany*, http://www.kunst.psychosomatik.ukaachen.de/go/show?ID=3668027&DV=1&COMP=page&ALTNAVID=1251124&ALTNAVDV=1.
17. Liljencrants J. (1991), *A translating and rotating mass model of the vocal folds*, Speech Transmission Laboratory-Quarterly Progress and Status Report, **32**, 1, 1–18.
18. Lous N.J.C., Hofmans G.C.J., Veldhuis R.N.J., Hirschberg A. (1998), *A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design*, Acta Acustica (united with Acustica), **84**, 1135–1150.

19. Pelorson X., Hirschberg A., van Hassel R.R., Wijnands A.P.J. (1994), *Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two mass model*, Journal of the Acoustical Society of America, **96**, 6, 3416–3431.

20. Price P.J. (1989), *Male and female voice source characteristics: Inverse filtering results*, Speech Communication, **8**, 261-277.

21. Qi Y.Y., Bi N. (1994), *Simplified approximation of the 4-parameter LF model of voice source*, Journal of the Acoustical Society of America, **96**, 1182–1185.

22. Rosenberg A. (1971), *Effect of glottal pulse shape on the quality of natural vowels*, Journal of the Acoustical Society of America, **49**, 2, 583-590.

23. Rothenberg M., Carlson R., Granström B., Lindqvist–Gauffin J. (1975), *A three-parameter voice source for speech synthesis*, In G. Fant (Ed.) Proceedings of the Speech Communication Seminar, vol. **2**, pp. 235-243, Stockholm.

24. Schoentgen J. (1993), *Modelling the glottal pulse with a self-excited threshold autoregressive model*, Proceedings of Eurospeech'93, pp. 107–110, Berlin.

25. Story B.H., Titze I.R. (1995), *Voice simulation with a body-cover model of the vocal folds*, Journal of the Acoustical Society of America, **97**, 2, 1249–1260.

26. Veldhuis R. (1998), *A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation*, Journal of the Acoustical Society of America, **103**, 1, 566–571.