

Circuit-Level Design of Human Voice Source

R. K. Sharma and Nikhil Raj

Abstract—Communication is heavily dependent on ideas expressed through speech. The ideas and tonal qualities during vocal expression give voice an idea about person characteristics and personality. The problem of representing speech events with robust and compact signals that describe the salient features of speech is an important area of speech communication. The Glottal Source is an important component of voice as it can be considered as the excitation signal to the voice apparatus. Nowadays, new techniques of speech processing such as speech recognition and speech synthesis use the glottal closure and opening instants. The pitch synchronous analysis that is used in several area of speech processing often requires robust detection of the instants of glottal closure and opening. Current models of the glottal waves derive their shape from approximate information rather than from exactly measured data. General method concentrate on assessment of the glottis opening using optical, acoustical methods, or on visualization of the larynx position using ultrasound, computer tomography or magnetic resonance imaging techniques. In this work, an experimental integrated circuit of human glottis using MOS is presented by exploiting fluid volume velocity to current, fluid pressure to voltage, and linear and nonlinear mechanical impedances to linear and nonlinear electrical impedances. The glottis modeled as current source includes linear, non-linear impedances to represent laminar and turbulent flow respectively, in vocal tract. The MOS modeling and simulation of glottis were carried out on TSMC 0.18 micrometer technology.

Index Terms—Alveolar pressure, MOS resistor, voice source, vocal fold.

I. INTRODUCTION

Speech synthesis has been a topic of special interest among the researchers. In speech codec and synthetic speech systems, an efficient representation of speech and naturalness of generated speech are important requirements. An emerging approach to improve the naturalness of synthetic speech is to exploit bio-inspired models of speech production. In the real human voice production mechanism, the excitation of voiced speech is represented by the glottal volume velocity waveform generated by the vibrating vocal folds. This excitation signal, that is, glottal waves has naturally attracted interest of researchers in the area of speech synthesis. Glottis is the space (opening) between the vocal folds. Many techniques have been proposed to mimic the glottal source of natural speech. Glottal wave estimates are needed in vocal

fold function analysis, speaker identification, and natural sounding speech synthesis. Artificial models for the glottal source have been used in order to improve the quality of the synthesis. However, current models for the glottal source are oversimplified as well, and the resulting quality of the synthesis has not been satisfactory. To overcome the problems due to oversimplified glottal source models, the idea of utilizing glottal flow pulses extracted directly from voicing source has been proposed. The proposed concept is based on analog (electrical) model of glottis.

Measurements of glottal flow provide benchmark data for voice source models. Voice and speech production is the end result of linguistic operation performed by muscles of diaphragm, the larynx, tongue, lips. The neural impulses of brain coordinate the right amount of diaphragmatic pressure through larynx while vocal fold serve as flow converter, that is, the vocal fold oscillate changing steady stream of airflow in many small individual puffs of air in pharynx. Moreover it can be stated that larynx act as DC to AC flow converter. This airflow conversion comprises of rapid compression and rarefaction causing minute change in pressure in air. The lungs and respiratory muscles act as vocal power supply. Voiced speech is produced by air expelled from the lungs causing the vocal folds to vibrate in non-linear periodic fashion which is approximated by a relaxation oscillator. The ejected air stream flows in form of pulses which further get modulated by the vocal tract. In unvoiced speech, sounds are created by passing the stream of air through a narrow constriction in the tract. The pulses can also arise by making a complete closure, building up pressure behind it, and then followed by an abrupt release. In the first case, a turbulent flow is produced while in the second case, a brief transient excitation occurs. The puffs of air are shaped into sound waves of speech and eventually, radiated from the lips or nose. The periodic signal generated by the vibrating motion of the vocal folds is called the glottal flow, glottal volume velocity waveform, or simply the voice source. The rate at which the vocal folds vibrate defines the fundamental frequency of the speech. In normal speech, fundamental frequency changes constantly, providing linguistic information about emotional content, such as differences in speaker mood. In addition, the fundamental frequency pattern determines naturalness of utterance production.

The organization of this paper is as follows. In Section II, electric equivalent circuit model of glottis is presented. Section III comprises simulation results and discussion followed by conclusion in section IV.

II. CIRCUIT MODEL OF GLOTTIS

The vocal tract can be assumed as a non-uniform acoustic tube, with time-varying cross-sectional areas; terminated by

Manuscript received December 20, 2009.

R.K.Sharma is with the National Institute of and Technology, Kurukshetra, Haryana, 136119, India (e-mail: mail2drks@gmail.com).

Nikhil Raj is with the Electronics and Communication Engineering Department, National Institute of and Technology, Kurukshetra, Haryana, 136119, India (e-mail: nikhilquick@gmail.com).

the vocal folds at one end, while the other end by lips and nose. Vocal fold vibration produces a periodic interruption of the air flow from the lungs to supraglottal vocal tract based on the principle of Bernoulli Effect. It has been found that at most frequencies of interest, the glottal source has high acoustic impedance compared to the driving point impedance of the vocal tract. Consequently, a current source may be used as the electrical analog that approximates the volume velocity source, i.e. U_{gl} at the glottis [1]. Alternatively, the constriction at the glottis is represented by variable impedance $Z_{gc}(t)$ that serves to model the constrictions created by the opening and closing of the vocal folds in the glottis and thus model turbulent and laminar flow in the vocal tract. The glottal impedance is modulated by a glottal oscillator to model the opening and closing of the vocal folds. Based on these approximations, the electrical equivalent of glottis is shown in Fig. 1.

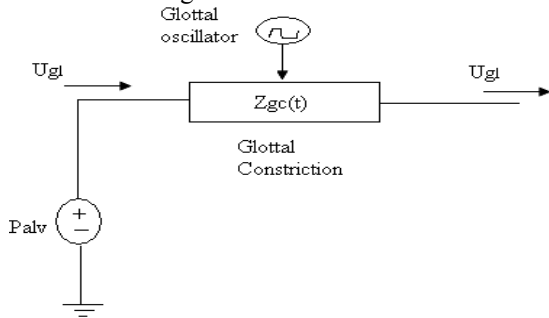


Figure 1. Electrical model of Glottis

A. Lung Modelling

The basic source (power supply) of voice production is lung so the analysis of how lung is affected by muscular pressure involved during breathing and correspondingly change in lung volume with respect to pressure becomes important. Pressure and volume affect each other as they are interrelated (i.e. $work = pressure \times Vol.$).

Lung Pressure

Lung models are of great importance in determining respiratory mechanics. Model of the respiratory system should incorporate the dynamic nature of the respiration in order to track all rapid changes in the parameters but also consider the time-varying nature of the model parameters. Thus appropriate model of the respiration system should be composed of the time varying components. Lung, diaphragm and abdominal control system are the source muscles which takes part in voice production. The inspiratory and expiratory movement of lungs is carried out by diaphragm and abdominal muscles. During inhalation (inspiration), the volume of the thorax increases due to the lowering of the diaphragm. In addition, the ribs are raised and moved outward by the contraction of the external intercostals muscles. Such increase in thoracic volume reduces alveolar pressure, causing air to move into the lungs while during exhalation (expiration), the diaphragm and external intercostals muscles relax, reducing the thoracic volume. Such reduction of thoracic volume raises alveolar pressure and forces air out of the lungs.

During partial ventilatory support, the motion of the respiratory system can be represented by the single 1st order differential equation [2] as

$$P_{aw}(t) + P_{mus}(t) = R \times \dot{V}(t) + E \times \Delta V(t) \quad (1)$$

where $V(t)$ and $\Delta V(t)$ are the instantaneous flow and volume displacement respectively with R and E represent resistance and elastance of respiratory system whereas the term on left side of equation (1) represents total driving pressure applied to the respiratory system, that is, $P_{aw}(t)$ and $P_{mus}(t)$ which shows time varying airway pressure and the patient-generated pressure respectively. $P_{aw}(t)$ is simulated by assuming that the ventilator is triggered as soon as inspiratory effort succeeds in either generating inspiratory flow or reducing the P_{aw} below the base-line pressure level (assume zero). Once the ventilator is triggered, P_{aw} is assumed to exponentially increase to the P_{ps} (maximum ventilator set pressure support) level with a ventilator time constant (τ_v) and then maintain that level until the termination of inspiration. The expression of $P_{aw}(t)$ is

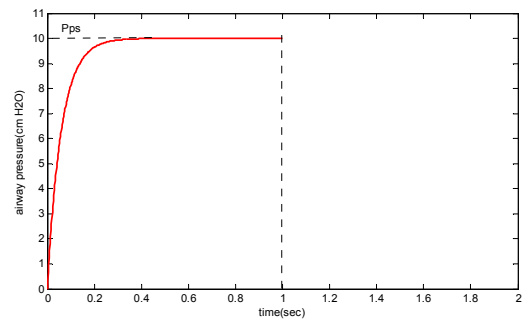
$$P_{aw}(t) = P_{ps} (1 - e^{-t/\tau_v}) \quad (2)$$

where $t \geq 0$. The time varying P_{mus} for inspiration is approximated by 2nd order polynomial function [2] as

$$P_{mus}(t) = -d \cdot (t - T_I)^2 + dT_I^2 = -dT_I^2 \left(\frac{t}{T_I} - 1 \right)^2 + dT_I^2 \quad (3)$$

$$P_{mus}(t) = -P_{mus \max} \left(\frac{t}{T_I} - 1 \right)^2 + P_{mus \max} \quad (4)$$

where d is a constant, $P_{mus \max} (= dT_I^2)$ represents maximal inspiratory muscle pressure, T_I and T is neural inspiratory time and time lapsed in one complete cycle of respiration. In case of expiration, P_{mus} decays exponential at faster rate with steep slope. At the end of patient neural inspiration, the decline in P_{mus} after neural inspiration augments the flow decay so as to decelerate flow to reach the threshold subsequently.



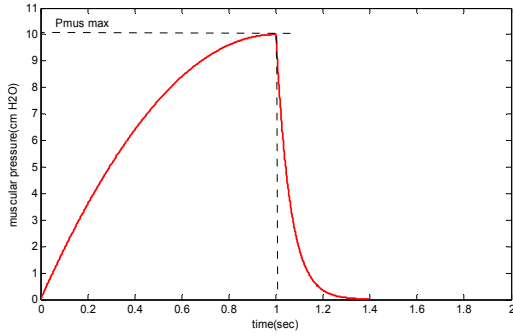


Figure 2. Airway pressure P_{aw} (top) and muscular pressure P_{mus} (bottom)

Thus, the mathematical form of inspiratory muscles $P_{mus}(t)$ is approximated as

$$P_{mus}(t) = \begin{cases} -P_{mus\ max} \left(1 - \frac{t}{T_i}\right)^2 + P_{mus\ max}, & 0 \leq t \leq T_i \\ P_{mus\ max} e^{(-t/\tau_m)}, & T_i \leq t \leq T \end{cases} \quad (5)$$

where τ_m determine expiratory asynchrony in assist ventilation. Physiologically, the end of P_{mus} is not instantaneous rather the inspiratory muscle activity generally extends into the expiratory phase, resulting in residual inspiratory P_{mus} during neural expiration. The values of P_{ps} and $P_{mus\ max}$ is set to 10 cmH_2O and τ_v/T_i is set to 0.06. The patient neural inspiration time (T_i) is set to 1.0 s which results in τ_v to 0.06 [3] as shown in Fig. 2.

Lung Volume

Lung volume contributes to muscle pressure, which, in turn, determine respiratory power. Previous literature showed that subglottal pressure, larynx height, glottal compliance varies with lung volume. It affects the tracheal pull which widens the glottis and thus may affect phonation. Since one of the largest contributors to the work of breathing is elastic work, a change in the initial position of the lung can have a significant effect on total energy cost. Human lung volume during a breathing cycle (inhalation and exhalation) at rest begins and ends at the functional residual capacity (FRC). At this point, the respiratory and glottis muscles are totally relaxed so sometimes it is often referred as relaxation volume. Within this volume, gas exchange continues to occur throughout the respiratory cycle resembling a continuous circulation of blood through lungs. If this volume is not maintained then there might be chance that there is continuously reabsorbing of carbon dioxide. The value of FRC is determined by the compliance of pressure. FRC is composed of two volumes, the expiratory reserve volume (ERV) and the residual volume (RV), that is, volume in lung after maximum expiration. In adult, the value of RV is around 1 liter. From [4], the mathematical relation between lung volume and pressure is:

$$\frac{VC}{V - RV} = 1.00 + \exp(b - cp) \quad (6)$$

Rearranging (6),

$$V = \frac{VC}{1.00 + \exp(b - cp)} + RV \quad (7)$$

where V is lung volume (m^3), VC represents vital capacity (m^3), p is pleural pressure (N/m^2) and RV is residual volume (m^3) whereas b , c are coefficients.

Using values of [5],

$$V = \frac{0.0048}{1.00 + \exp(1.00 - 0.0018p)} + 0.0012 \quad (8)$$

$$= 2.49 \times 10^{-3} m^3 \text{ at } p = 0$$

This lung volume is the value of FRC as shown in Fig. 3.

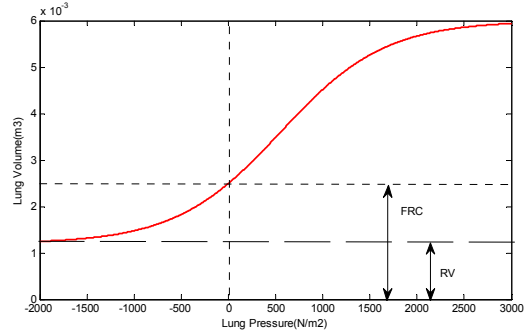


Figure 3. Lung pressure-volume curve

The slope of the curve gives compliance which becomes small at extreme values of lung volume. Most pressure-volume characteristics of biological organs and tissues, becomes flatter as volume extremes are reached. Variation of volume in such fashion provides less muscular pressure during exhalation compared to inhalation. If inspiratory and expiratory pressures become equal then there is less power needed for respiration which is often seen in case of adult.

Subglottal Pressure Measurement

Commonly used unit for pressure measurement in speech is cmH_2O . The subglottal pressure (lung pressure) is measured in terms of cmH_2O , quite similar to that of displacement occurring in U-tube manometer. Yet another device, specialized pressure transducer, is also used to measure speech related pressure. Subtly, Worth and Sakuda used a semiconductor strain gauge device of about 10mm in diameter and 3mm in thickness for lung pressure measurement [6]. This was pasted to palate to sense intraoral pressure during speech. With assumption of linear system, using linear regression analysis to translate amplifier output voltage within range to equivalent air pressure; they derived an expression to convert cmH_2O pressure in dc voltage equivalent. The expression takes the form:

$$1cmH_2O = 1.27 \times P_{alv} + 5.94 \quad (9)$$

where P_{alv} is the alveolar pressure approximated as dc voltage equivalent. For ordinary speech, lung pressure ranges from 7-10 cmH_2O , for loud speech it ranges from 10-12 cmH_2O and for shouting around 40 cmH_2O .

B. Vocal Fold Approximation

Human phonation is produced when expiratory air flows through the vocal tract and causes the vocal folds to undergo self-sustained vibration as they exhibit elastic and viscous behavior. The fundamental frequency of human phonation is

the fundamental frequency of vocal fold vibration; and the intensity of voice is closely related to the amplitude of vocal fold vibration [7]. Different models of vocal fold vibration have been described [8]-[9]. In [8], focus is on production of periodic motion using two coupled masses. In other models, attention is focused on a detailed representation of the distributed viscoelastic properties of the vocal folds [9]. Yet, another model [10] incorporates mechanisms for independent control of the body and the cover of the vocal folds. In general, multimass models of vocal folds are useful to describe main behavior observed in human voicing but their principle of functioning, based on harmonic oscillation, may appear complex. A labile nonlinear oscillator exhibits a rhythmic burst when excited by an appropriate input stimulus, i.e. lung pressure. Such oscillations can be obtained by relaxation oscillator [11]. A relaxation oscillator is an oscillator in which a capacitor is charged gradually and then discharged rapidly. The electrical output of a relaxation oscillator is a sawtooth wave. As shown in Fig. 1, it modulates the value of $Z_{gc}(t)$ in a periodic fashion to produce a volume velocity waveform U_{gl} .

C. Glottal Constriction Modelling

Constriction at the glottis can be approximated as a narrow cylindrical duct. To implement glottal constriction resistance, it is modeled as a series combination of linear and nonlinear resistor to represent losses occurring at the glottis due to laminar and turbulent flow, respectively. To design such electronically tunable resistors, it is preferably implemented using MOS. There are two glottal constrictions connected in series as shown in Fig. 4. The reason behind is that the upper and lower folds abduct and adduct with a time lag between them [12].

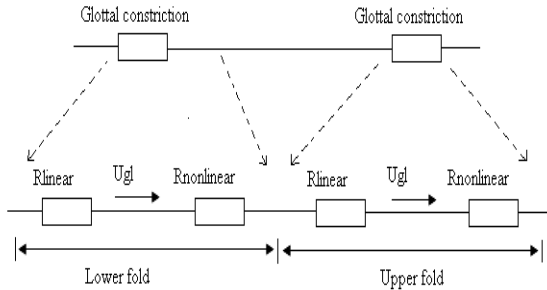


Figure 4. Glottal constrictions to model lower and upper fold

The impedance of each glottal constriction is varied by a glottal oscillator in a corresponding manner to represent vibration of upper and lower folds. Electronically tunable linear resistors are highly versatile circuit elements. They find various applications like in variable gain amplifiers, oscillators, balanced resistive bridges and analog filters. A combination of linear and nonlinear resistances is often useful in creating building blocks in electrical models of physical systems. MOS transistors are generally used for resistor modeling as when it is operated under triode mode behaves as a resistor controlled by its gate terminal voltage. MOS being a four terminal device offers two control parameters that is gate and bulk terminal to control resistor value. Through electronic tuning of gate terminal voltage of MOS transistor correspondingly electronic control on

resistance can be achieved.

In the past, MOS resistors with approximately linear I-V characteristics were obtained by operating the transistor in the ohmic (triode) region of strong inversion to exploit the resistive nature of the channel. Generally, these approaches were limited by the small ohmic region and its intrinsic non-linearities. Various techniques have been proposed to minimize the nonlinear effects associated with MOS in ohmic strong inversion regime with good results [13]-[15]. In regard to this, a MOS resistor is used that does not require triode operation [16]. In addition, this can be applied to produce linear as well as nonlinear resistances. For non-linear I-V characteristics, it uses translinear circuits which incorporate functions like square-root and square. As shown in Fig. 5, the MOS transistor M_R act as a tunable resistor and for tuning its resistive value, a capacitor is connected at its gate terminal. To maintain resistive nature of M_R , a feedback network is configured at its gate terminal. As the two OTAs, OTA1 and OTA2 have same inputs, V_X and V_Y connected to their input terminals in alternative fashion, they are biased by the same current source I_{gm} . The potential difference ($V_X - V_Y$) across the MOS device M_R is sensed and converted into a current $I_{out, gm}$ using a wide linear range OTA [17] for which the output current equation is of the following form.

$$I_{out, gm} = G_M V_{XY} = G_M (V_X - V_Y) \quad (10)$$

where G_M ($G_M = I_{gm}/V_L$) is the transconductance of OTA while I_{gm} and V_L are the biasing current and linear range of the OTA, respectively. These two OTAs are configured in conjunction with diode connected transistors M1 and M3 to produce two half-wave rectified currents that are proportional to $|V_{XY}|$, that is, voltage across the source-drain terminals of M_R . The rectified output currents get mirrored via M2 and M4 to create a full wave rectified current I_{in} . This I_{in} further serves as input current to translinear block and correspondingly output current I_{out} is generated at output end of translinear block as a function of I_{in} i.e. $I_{out} = f(I_{in})$.

The translinear block consists of current domain circuits which implement functions like linear, square-root and square. The saturation currents I_{Xsat} and I_{Ysat} of M_R is proportionally replicated by sensing V_g , V_W , V_X and V_Y on the gate, well, source and drain terminals of M_R , buffered via source followers and applying potentials V_{gX} and V_{gY} across the gate-source terminals of transistors M_X and M_Y . Transistors M7-M13 serve to compute $I_{Xsat} - I_{Ysat}$ or $I_{Ysat} - I_{Xsat}$ and transistors M14-M17 compare $|I_{Xsat} - I_{Ysat}|$ with a mirrored version of the translinear output current using M6. Any difference between these two currents causes the capacitor C to charge or discharge tuning the gate bias voltage V_g which equilibrates at a point where the two are nearly equal via negative feedback action.

The MOS resistor can be easily extended to implement a nonlinear resistor which shows behavior of form compressive ($I \propto \sqrt{V}$) or expansive ($I \propto V^2$) depending on appropriate choice of translinear circuit. For a linear MOS resistor, a translinear circuit with the following input-output relationship $I_{out} = I_{in}$ is used. Likewise, for compressive resistor and for expansive resistor a translinear circuit with relation $I_{out} = \sqrt{I_{in} I_{ref}}$ and $I_{out} = I_{in}^2 / I_{ref}$ is used respectively, where I_{ref} is a reference current.

To overcome loading effect on terminals of MOS transistor M_R , source follower is employed shown within dotted lines marked as SF in Fig. 5. The source follower has the capability to source and sink large output currents. Its primary use is to buffer signals and provide low output impedance to drive resistive loads while, at the same time handle large output voltage swing and obtain low harmonic distortion. Traditional source have load drive capability limited to the quiescent current in the buffer. In addition traditional source followers require too much power for many applications. To reduce power dissipation (and area), composite source follower is used [18]. The composite source follower comprises a current source, Msf3 configured to provide a (relatively) constant current to the rest of the circuit, a source follower NMOS (Msf0, Msf2) configured to receive an input signal, a folded cascade device Msf4 connected to sense the drain current of the source follower, and a current mirror device Msf1 connected to multiply the sensed drain current for application to an output load connected at the source follower output. It provides a four-fold increase in transconductance which offer perfect tracking of input by output having no level shift problem as compared to common voltage buffers. Being less complex circuitry, it is most effetyly used in field of low power architectures.

Modified OTA design

The OTA used in tunable resistor architecture has been modified using modified Wilson mirror architecture which not only enhances its linearity but too increases its dc gain. Such OTA along with current reference circuit to generate bias current in range of nanoamperes is shown in Fig. 6. The PMOS transistors, W_1 and W_2 act as input differential pair whose bulk is connected to input terminals V^- and V^+ respectively. Two PMOS transistors S_1, S_2 provide source degeneration feedback. The bump transistors B_1, B_2 are used to reduce effects of parasitic bipolar transistors which cause bump-shaped characteristics on output when input goes below 1 volt. Transistors $M_{n16} - M_{n19}$ and $M_{p13} - M_{p15}$ along with R_s comprise current generator circuit. The rest of transistors are configured as Wilson topology. The output current equation is given as

$$I_{OUT} = I_B \tanh\left(\frac{x}{V_L}\right) \quad (11)$$

where I_{out} is output current, I_B is the bias current and V_L is the linear range of OTA. The linear range of OTA is expressed as

$$V_L = 2V_T / g \quad (12)$$

where g is the overall transconductance of OTA which is decreased by a feedback factor introduced by Wilson mirror. From equation (12), when transconductance g decreases, then V_L increases. If V_L is made sufficiently high then it become possible to neglect cubic order term in the tanh expansion in (11) resulting I_{out} to be a linear function of I_B . In this case, the linearity obtained is about 2 volt. Compared to simple current mirrors Wilson circuit provides much higher output impedance r_{out} , which is one of the best choices to use in OTA being a VCCS device. Wilson current mirror circuit described in [19] is shown in Fig. 7. It consists of a simple current mirror and a current-to-voltage converter connected in the feedback loop. The output resistance is increased through the use of negative current feedback, that is, if output current I_{out} increases then the current flowing through M_{n2} also increases. However, the mirroring action of M_{n1} and M_{n2} causes the current in M_{n1} to increase. If I_{in} is constant and assuming that there is some resistance seen from gate of M_{n3} to ground then the gate voltage of M_{n3} is forced to decrease if current I_{out} increases.

The small-signal output resistance is approximated as below

$$r_{out} = r_{ds3} + r_{ds2} \left(\frac{1 + r_{ds3} g_{m3} (1 + \eta_3) + g_{m1} r_{ds1} g_{m3} r_{ds3}}{1 + g_{m2} r_{ds2}} \right) \quad (13)$$

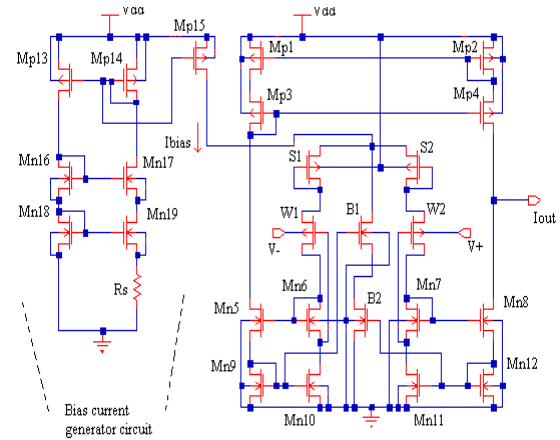


Figure 6. Modified OTA with bias current generator circuit

According to [20], the frequency response of the Wilson current mirror with all transistors identical causes a peak. It contains two pole and one zero, and a peak caused due to zero (equal to real part of pole) in their bode plot which leads to undesirable great overshoot in time domain. The modified Wilson current mirror is shown in Fig. 7.

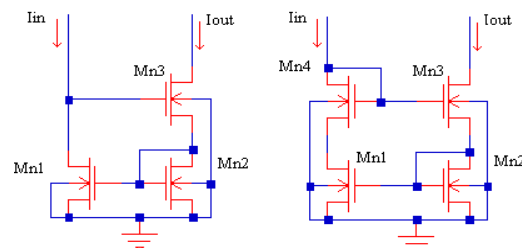


Figure 7. Wilson mirror (left) and modified Wilson mirror (right)

It consists of one extra transistor M_{n4} which leads to one more pole, making it a third order system and shows sufficient increase in output resistance. In this configuration, the bulk of each transistor is connected to lowest bias voltage, i.e. ground; which makes it a low threshold device and introduces g_{mb} (source-to-bulk transconductance) effect. This g_{mb} causes unequal shift for zero and real part of poles resulting zero on left of real part of pole. This overall effect reduces peak in bode plot. This structure also increases the value of resistance r_{out} .

For bias current generation, as the source-to-gate voltage of M_{p13} and M_{p14} are equal their corresponding currents will be equal, i.e. $I_{D14} = I_{D13}$ (assume $\lambda = 0$). Furthermore, it can be noted that $I_{D18} = I_{D13}$ and $I_{D19} = I_{D14}$. The equation for drain current of nMOS transistor is

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{Tn})^2 \quad (14)$$

Solving for V_{GS}

$$V_{GS} = \sqrt{\frac{2I_D}{\mu_n C_{ox} (W/L)}} + V_{Tn} \quad (15)$$

Using, KVL in current bias circuit (Fig. 6).

$$V_{GSn18} = V_{GSn19} + I_{Dn19} R_S \quad (16)$$

From (15)

$$\sqrt{\frac{2I_{Dn18}}{\mu_n C_{ox} (W/L)_{Mn18}}} = \sqrt{\frac{2I_{Dn19}}{\mu_n C_{ox} (W/L)_{Mn19}}} + I_{Dn19} R_S \quad (17)$$

Rearranging above expression and equating equivalent current yields

$$I_{D14} = \frac{2}{\mu_n C_{ox} (W/L)_{Mn18}} \frac{1}{R_S^2} \left(1 - \sqrt{\frac{(W/L)_{Mn18}}{(W/L)_{Mn19}}} \right)^2 \quad (18)$$

The output current I_{bias} , that is, I_{D15} is now the function of I_{D14} . By adjusting the aspect ratio of M_{p15} relative to M_{p14} , desired I_{bias} can be obtained. The W/L ratio of M_{p15} is kept four times lower than M_{p14} , which results in output current $I_{D15} = I_{bias} = I_{D14}/4$.

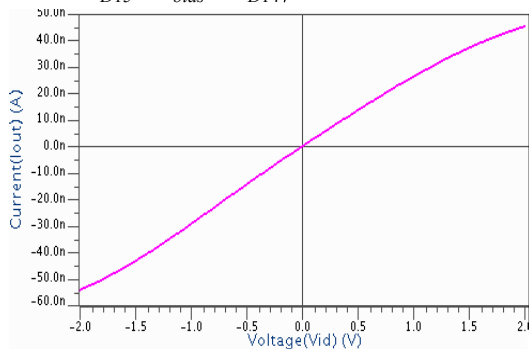


Figure 8. Linear response of modified OTA

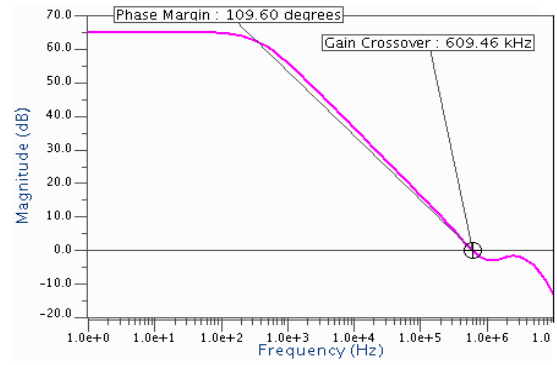


Figure 9. AC response of modified OTA

For I_{bias} of 65nA, the value of R_S is kept $10K\Omega$ and the supply voltage V_{dd} is kept about 0.9 volt. Fig. 8 shows the linear response of OTA. The linearity extends to range about ± 2 volt. Fig. 9 shows the open loop dc gain and UGB of 65 dB and 609.46 kHz respectively. Its high gain supports it for use in biomedical instruments. The aspect ratios of transistors used for OTA implementation is summarized in table 1.

III. RESULTS AND DISCUSSION

As discussed earlier glottal constriction is a combination of two series connected resistances to model lower and upper vocal folds under laminar and turbulent airflow. With such approximation, the lower fold is modeled as combination of linear and square-root circuit while the upper fold is modeled as combination of linear and square circuit as shown in Fig. 10.

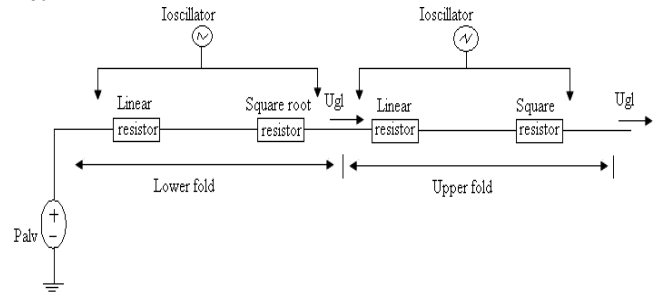


Figure 10. Complete glottal circuit

The oscillator is provided as an input current pulse which also controls bias current of OTAs used in tunable MOS resistor architecture. As oscillation is approximated by relaxation oscillator, the output is sawtooth wave in the form of current pulse where its rise is kept much higher than that of fall time. For simulation purpose, the pulse duration is kept about 8ms to model vibration of fold in adult man. The time lag of 1ms is kept between successive oscillator input to model lower and upper vocal fold oscillations. A normal voiced speech creates a pressure of about 7-10 cm H_2O within lungs which in turn exerts force on vocal folds which make them to vibrate and in turn produces sound. To model this lung pressure, it is converted in dc voltage equivalent using equation (9), which acts as input voltage P_{dv} for the glottal circuit. For normal voice, a nominal pressure of about 10 cm H_2O is generated and correspondingly the dc equivalent voltage P_{dv} is approximately 3 volt. The simulated output of glottal circuit is in form of periodic

current pulses shown in Fig. 11, along with its derivative. The frequency range between successive glottal pulses is around 125Hz which satisfies the fundamental frequency generated during voiced speech in normal mode by male speakers. According to the source-filter theory of speech production, lip radiation is represented by the derivative of the produced acoustic signal which means voice source is actually derivative of glottal pulse.

Thus, the intensity of the produced acoustic wave depends rather on the derivative of the glottal flow signal than the amplitude of the flow itself. In other way, the derivative is the effective excitation of the vocal tract [21]. The principal acoustic excitation of vocal tract occurs at the discontinuity of derivative pulses. For analysis, a single pulse of glottal circuit along with its derivative is taken as shown in Fig. 12. More the negative value, higher is the excitation. Generally, the inertia of air in the glottis and supraglottal airways prevent the occurrences of the abrupt discontinuities that occur at time of vocal fold closure. With the assumption of complete glottal closure, there will be discontinuity in first derivative at the endpoints of the open phase of the glottal wave. For voiced speech, the glottal flow derivative consists of two phases, that is, open phase and closed phase. During the open phase, vocal folds are progressively displaced from their initial state due to increasing subglottal pressure. When the elastic displacement limit of folds is reached, they suddenly return to the position of closing phase. There is slight time-gap between vocal folds separation, primarily due to the inertia of the vocal tract air below and above the glottis. Likewise, as the vocal folds come together during each oscillation, the inertia of the air supports and maintains a high flow until the closing of the glottis finally forces the flow to zero (assuming a total glottal closure).

During the most closed portion of the glottal cycle, where the flow is minimum, the waveform is relatively flat. This air flow during the most closed portion of the glottal cycle, when the waveform is relatively flat, is probably due to an incomplete closure of the vocal folds between the arytenoids cartilages. In general, for a normal voiced glottal cycle there must be significantly long period in which the glottis is either closed, or sufficiently closed so that the glottal impedance is high enough to satisfy this condition. From simulations, it can be observed that the glottal waveform admittance before and after the glottal pulse is zero. It is often desirable to monitor the degree of abduction or adduction of the vocal folds during voiced speech, both for steady voicing and during abductory or adductory movements. During pressed condition pressure in lungs is about 10-12 cm H_2O . For 12 cm H_2O , the equivalent P_{av} is 3.8 volt and corresponding glottal pulse is shown in Fig. 13. In such condition pressure rises slightly to higher value but closing phase takes longer time and the intensity of sound is approximately equal to normal phonation. Due to this high pressure created, vocal fold opens quickly for short interval of time. Similarly, in case of breathy voice pressure in lungs is too high about 40 cm H_2O and its corresponding glottal pulse output is shown in Fig. 14. It is observed that there is continuous flow of air which leads to continuous smooth glottal wave with almost

zero time closed phase. Vocal fold vibration is so high that there is instant opening and closing of folds.

IV. CONCLUSION

The Glottal Source is an important component of voice as it can be considered as the excitation signal to the voice apparatus. Its modelling increases the parametric flexibility of the system and permits to transform voice characteristics of the speech. Using MOS model of glottis, drastic reduction of power consumption can be achieved which could be useful in portable speech processing systems of moderate complexity, like in cell phones, digital assistants, and laptops. The glottal model can be useful in pathology detection or the biometric characterization of the speaker.

ACKNOWLEDGMENT

The authors would like to thank to Mr. Wee, Mr. Turicchia and Mr. Sarpeshkar for their meaningful discussion on vocal tract model.

TABLE 1. ASPECT RATIOS OF TRANSISTORS USED IN MODIFIED OTA

MOS	W (μm)	L (μm)	MOS	W (μm)	L (μm)
S1	2.22	1.6	Mn8	1.75	2
S2	2.22	1.6	Mn9	2.25	2
W1	1.6	1.6	Mn10	2.25	2
W2	1.6	1.6	Mn11	2.25	2
B1	2.24	1.44	Mn12	2.25	2
B2	2.24	1.44	Mp13	25.2	1.4
Mp1	1.75	2	Mp14	25.2	1.4
Mp2	1.75	2	Mp15	6.16	1.37
Mp3	2.25	2	Mn16	3.8	1.52
Mp4	2.25	2	Mn17	5.52	1.38
Mn5	1.75	2	Mn18	3.8	1.52
Mn6	1.75	2	Mn19	5.52	1.38
Mn7	1.75	2			

REFERENCES

- [1] K. H. Wee, L. Turicchia and R. Sarpeshkar, "An Analog Integrated Circuit Vocal Tract," *IEEE Trans. on Biomedical Circuits and Systems*, vol. 2, no. 4, pp. 316-327, Dec. 2008.
- [2] Y. Yamada and H. L. Du, Analysis of the Mechanisms of Expiratory Asynchrony in Pressure Support Ventilation: A Mathematical Approach, *J. Appl. Physiol.*, vol. 88, 2000, pp 2143-2150.
- [3] Nellcor Puritan Bennett. Nellcor Puritan Bennett 840 Ventilator *Operating Manual* (version A). St. Louis, MO: Mallinkrodt, 1997.
- [4] "Multidimensional curve fitting program for biological data," *Comp. Prog. Biomed.*, vol. 18, pp. 259-264, 1984.
- [5] "Biomechanics and Exercise Physiology." New York: Wiley and Sons, 1991.
- [6] *Clinical Measurement of Speech & Voice*, second edition, by Ronald J. Baken, Robert F. Orlikoff, Edition: 2, Published by Cengage Learning, 2000.
- [7] Kai Zhanga, Thomas Siegmunda, and Roger W. Chanb, "Modeling of the transient responses of the vocal fold lamina propria," *Journal of the Mechanical Behavior of Biomedical Materials*, Vol. 2, pp. 93-104, Jan. 2009.
- [8] K. Ishizaka and D. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Sell Syst. Tech. S.*, 51, pp. 1233-1268, 1972.
- [9] R. Titze, "On the mechanics of vocal fold vibration," *S. Acoust. Soc. An.*, 60, pp. 1366-1380, 1976.
- [10] O. Fujimura, "Body-Cover theory of the vocal fold and its phonetic implication," in *Vocal Fold Physiology*, K.N. Stevens and M. Hirano, Eds., Tokyo: University of Tokyo Press, Chapter 19, 1981.

- [11] B. L. Bardakjian, T. Y. El-Sharkawy and N. E. Diamant, "On a population of labile synthesized relaxation oscillators," *15FF trans. Biomedical Engineering*, BME-30, Nov. 1983.
- [12] K. N. Stevens, *Acoustic Phonetics*, vol. 30, Cambridge, Mass.: MIT Press, 1998, pp. 607.
- [13] K. Nay and A. Budak, "A voltage-controlled resistance with wide dynamic range and low distortion," *IEEE Trans. Circuits and Systems*, vol. 30, no.10, pp. 770-772, Oct. 1983.
- [14] J. Ramirez-Angulo, M. S. Sawant, R.G. Carvajal and A. Lopez-Martin, "Linearisation of MOS resistors using capacitive gate voltage averaging," *Elec. Letters*, vol. 41, no. 9, pp. 511-512, Apr. 2005.
- [15] C. Popa, "Linearized CMOS active resistor independent on the bulk effect," *Proc. 17th Great Lakes Symposium on VLSI*, 2007.
- [16] K. H. Wee and R. Sarpeshkar, "An electronically tunable linear or nonlinear MOS resistor," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 9, pp. 2573-2583, Oct. 2008.
- [17] R. Sarpeshkar, R. F. Lyon and C. Mead, "A low-power wide-linear-range transconductance amplifier," *Analog Integrated Circuits and Signal Processing*, vol. 13, pp. 123-151, 1997.
- [18] U. S. Patent Publication No. 2005/6924674 B2 (Jalaleddine et al.)
- [19] P. E. Allen and Douglas D. Holberg, *CMOS Analog Circuit Design*, 2nd Edition, Chap. 4, pp. 140-141, Oxford university Press 2004.
- [20] G. Palumbo, "Frequency behaviour of the Wilson and improved Wilson MOS current mirrors: analysis and design strategies," *Microelectronics Journal*, vol. 27, pp. 79-85, 1996.
- [21] G. Fant, J. Liljencrants and Q. Lin, "A four-parameter model of glottal flow," *STL QPSR*, no.4, 1985.

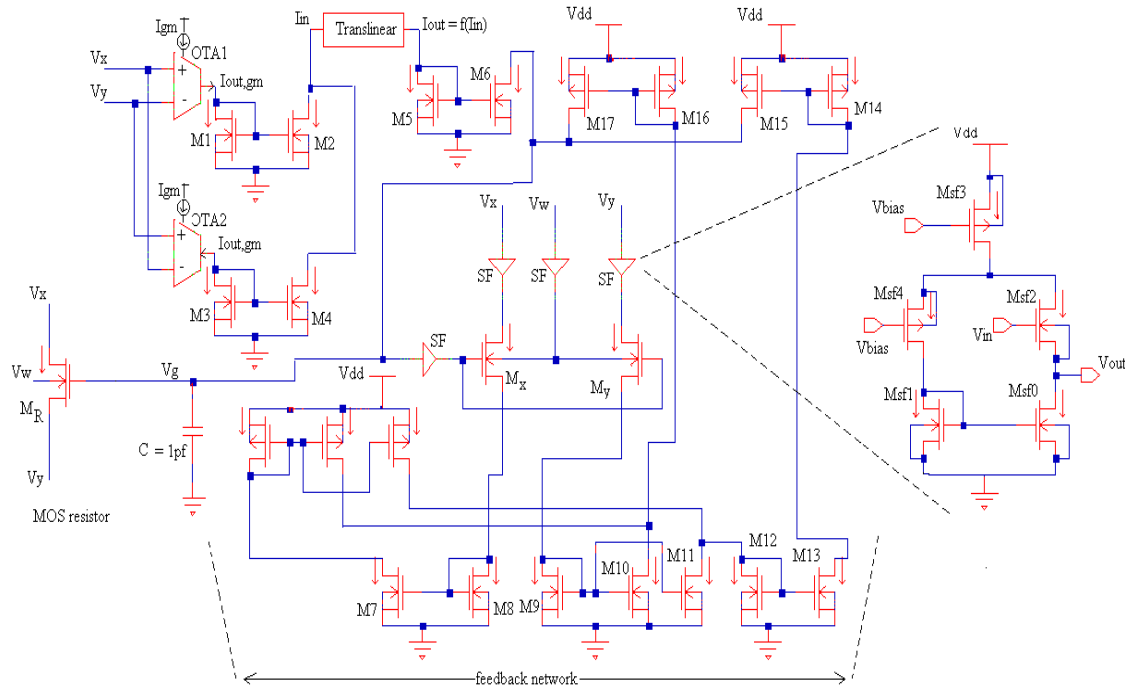


Figure 5. Electronically tunable MOS resistor with feedback circuit

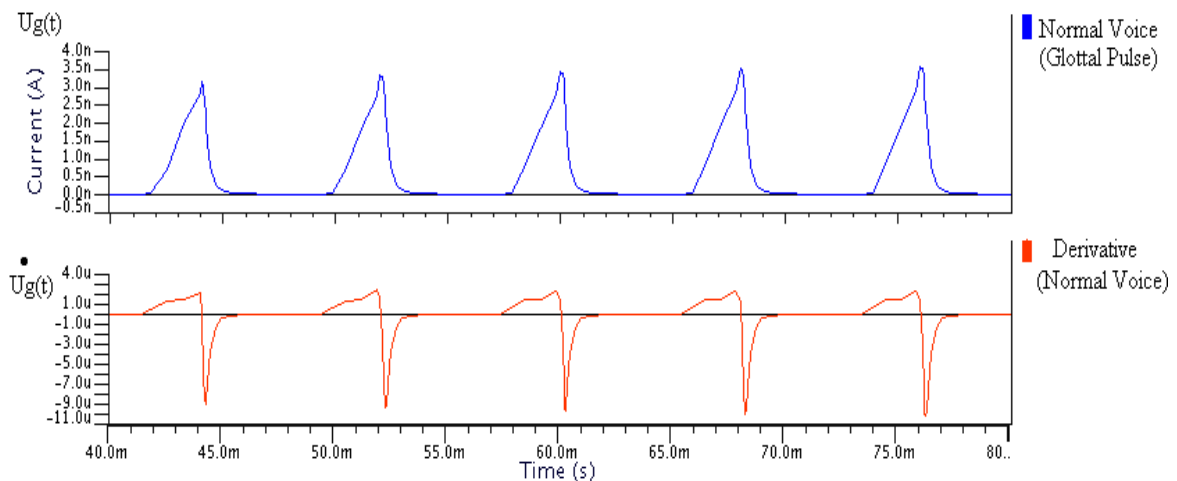


Figure 11. Glottal pulse and its derivative for normal voice

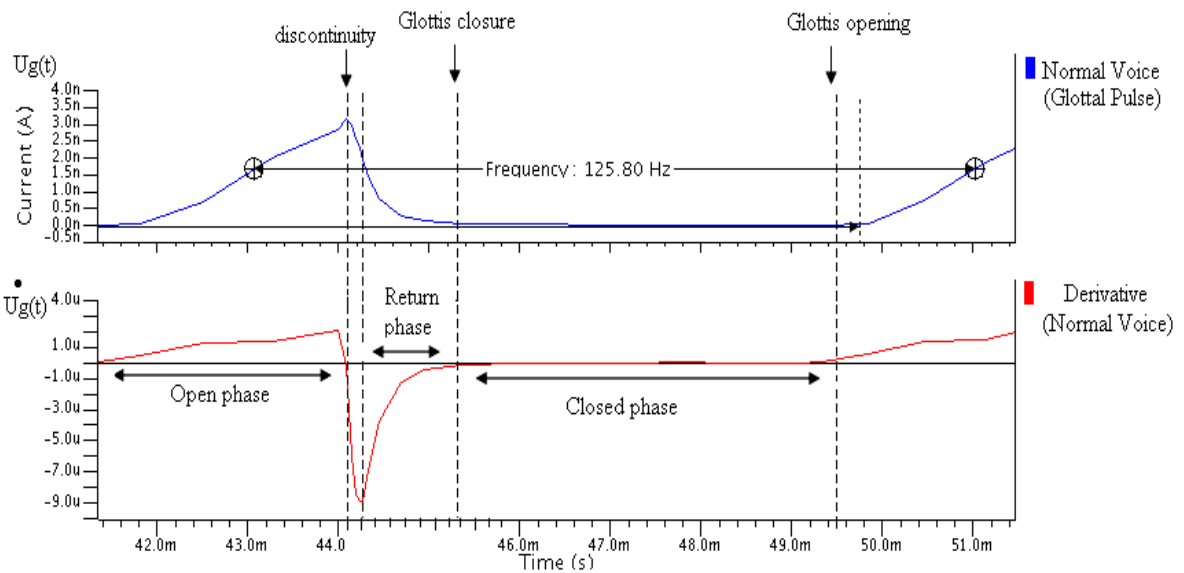


Figure 12. One single pulse of glottal circuit and its derivative for normal voice

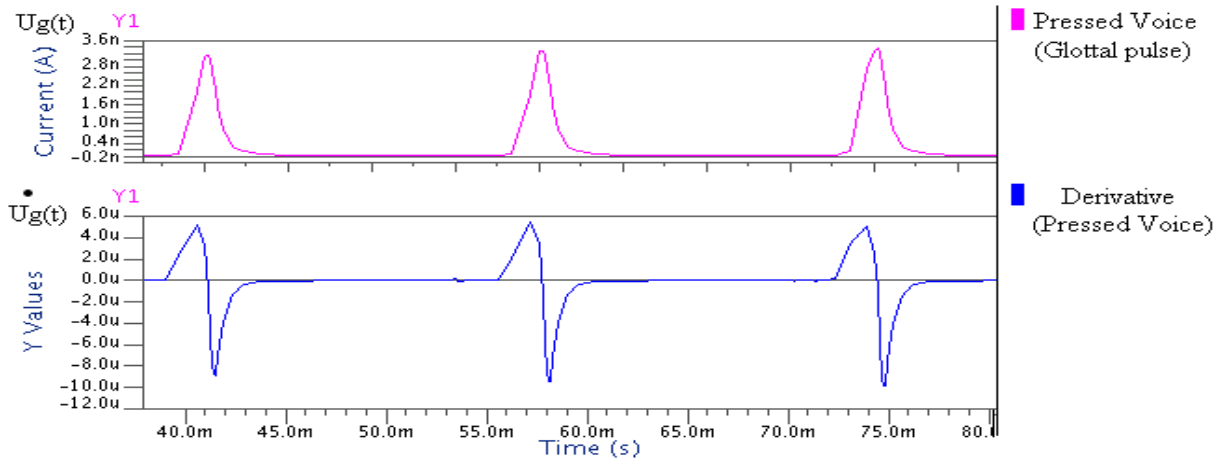


Figure 13. Glottal pulse and its derivative for pressed voice

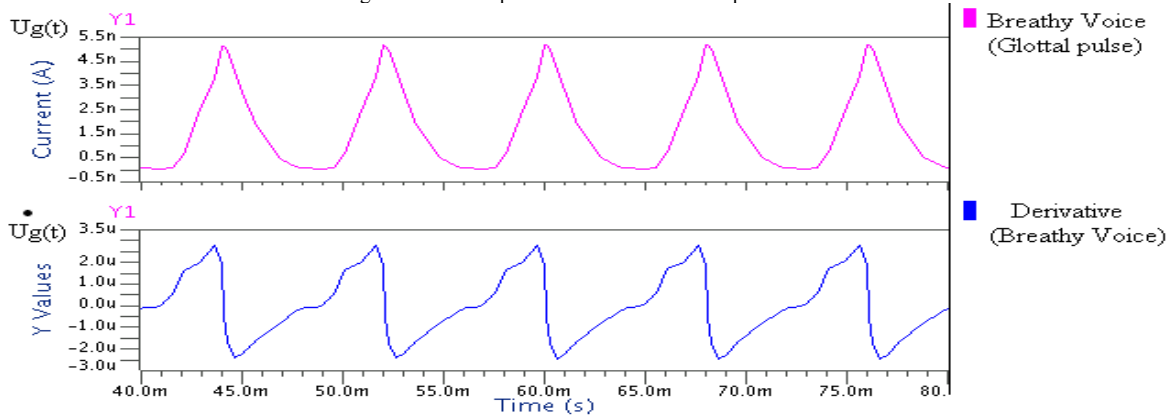


Figure 14. Glottal pulse and its derivative for breathy voice