Indian Journal of Science and Technology, *Vol 10(18), DOI: 10.17485/ijst/2017/v10i18/110786, May 2017*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Investigations of the Effect of Nonlinearly Generated Excitations on the Quality of the Synthesized Alaryngeal Speech

Romilla Malla Bhat[1]*, Jang Bahadur Singh[2] and Parveen Kumar Lehana[2]

[1]Government Gandhi Memorial Science College, Jammu - 180001, Jammu and Kashmir, India; romillamalla@gmail.com
[2]D. S. P. Lab, Department of Electronics, University of Jammu, Jammu – 180006, Jammu and Kashmir, India; sonajbs@gmail.com, pklehana@gmail.com

Abstract

Objectives: To explore the use of ultrasonic waves based glottal excitation for reducing the distracting background noise present in the conventional artificial larynx. **Methods/Analysis:** The persons, whose larynx is removed, mostly use artificial larynx to generate the excitation, which produces background noise. Our hypothesis is that if the glottal excitation is generated by using two ultrasonic waves, the background noise may be reduced as the ultrasonic frequencies are inaudible. The audible excitation is generated within the vocal tract by nonlinear interaction of the external high frequencies focused around the larynx. Vocal tract parameters were extracted from the recorded normal and alaryngeal speech segments using Linear Predictive Coding (LPC) technique using analysis window of 20 ms with overlap of 10 ms and order of 500. **Findings:** Five types of excitations, sinusoid of the desired pitch frequency, sinusoid of high frequency, sinusoid of high frequency plus pitch, amplitude modulated high frequency and non-linearly generated excitation are used. The analysis of the results shows that non-linearly generated excitation may be used for reducing the background noise along with enhanced intelligibility and Mean Opinion Score (MOS) estimated for the cardinal vowels using non-linear excitation is 2.1, 4.0 and 3.7 with laryngeal vocal tract parameters, for the three cardinal vowels as compared to the MOS of 1.4, 1.0 and 2.3 with alaryngeal vocal tract parameters. **Novelty:** In the proposed technique, use of ultrasonic frequencies for generating the glottal excitation reduces the distracting audible background noise present in conventional techniques resulting in enhanced alaryngeal speech.

Keywords: Alaryngeal Speech, LPC, LPC Residual, Non-Linear Excitation, Vocal Fold Parameters, Vocal Tract Parameters

1. Introduction

Human beings interact with each other by speaking their thoughts. This act of speaking produces an acoustic signal called speech signal. Speech is an exceedingly non-stationary signal changing vigorously and rapidly in time[1,2]. It is produced by well-organized signals from brain to diaphragm, laryngeal muscles, vocal folds, vocal tract and nasal tract. If the speech segment is produced due to the vibrations of the vocal folds, the speech segment is called voiced, otherwise, unvoiced. Unvoiced speech segments are produced by perturbations within or outside the vocal tract. The frequency of vibrations of the vocal folds is called the pitch frequency. Typical pitch frequency of a male speech is approximately 85-155 Hz and for the females is approximately 165-255 Hz. There is an overlap in the pitch frequency band between the two genders. In case of singers, the vocal range is from bass to soprano and the range of pitch frequency is 80 Hz-1100 Hz. Therefore, speech is generated when the air stream from the lungs, which acts as DC energy source, is constricted by the vocal folds in the larynx or some other disturbance

*Author for correspondence

within the system to ac flow. The vocal folds enveloped by the larynx situated at the upper end of the trachea shape the continuous air flowing from the lungs into puffs of air, which while passing through the entire vocal tract and the articulators in the mouth, gives a certain spectral shape to form the speech signal[1,2] shown in Figure 1(a). The speech production mechanism is shown in Figure 2.

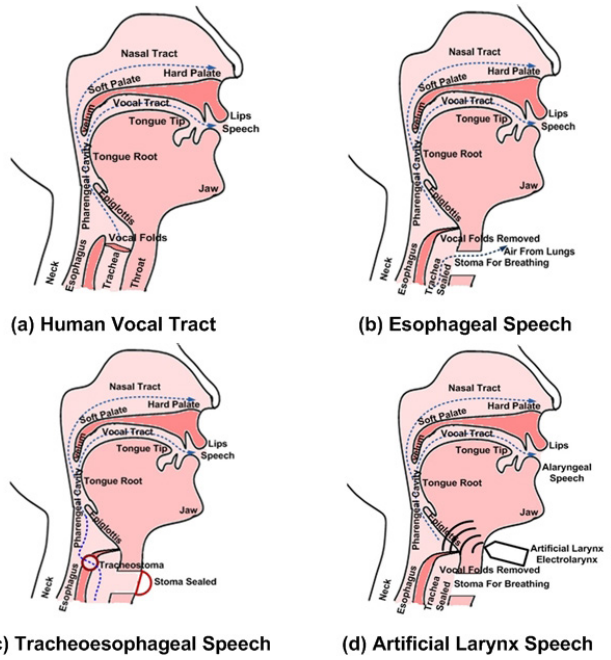

Figure 1. Normal and alaryngeal human vocal tract.

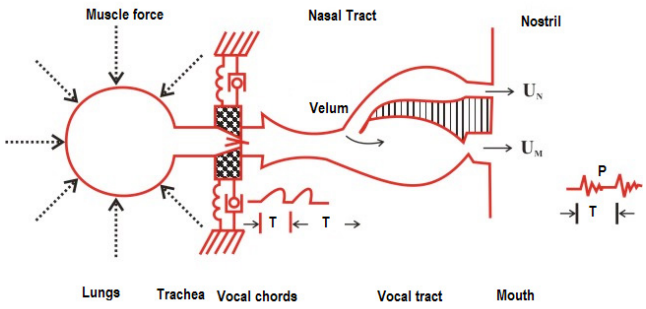

Figure 2. Speech production mechanism.

In India and world over, all the total head and neck surgeries performed, about 30-40 % persons have to undergo laryngectomy. Laryngectomy is a surgical procedure where the larynx along with the vocal folds, supporting cartilage and muscles are removed and the person who undergoes this procedure is called laryngeotomee. The most prevalent cause for undergoing this operation is laryngeal cancer. The predisposing causes are smoking and chewing tobacco or some pulmonary disease such as obstruction or carcinoma. Its prevalence increases if the person is consuming non-vegetarian diet along with alcohol[3–5]. A laryngeotomee, in consultation with a surgeon and a speech therapist, can be trained to produce speech which is called as alaryngeal speech. Various methods have been researched and improvised upon to increase the quality of life of a laryngeotomee. The most common methods are the esophageal speech, tracheoesophageal speech and artificial larynx speech. The first method is the esophageal speech, where injection method is utilized by the laryngeotomee to produce speech shown in Figure 1(b). The laryngeotomee uses the tongue as a driving mechanism to trap air between the tongue, hard and soft palates and the pharynx. The compressed air is then forced backward towards the stomach. Since the passage towards the lungs is sealed, the trapped air goes into the hypopharynx and the esophagus, where it gets entrapped for phonation[6]. This air thus trapped in the esophagus is then expelled out forcibly towards the mouth. While moving out it causes vibration of the vocal tract muscles giving rise to phonation[7]. Using the normal processes of resonation and articulation, the esophageal speaker is thus able to produce intelligible speech.

The second method is tracheoesophageal speech as shown in Figure 1(c) and is preferred in comparison to other methods. The reason is that it allows pulmonary air to be used for phonation and thus voiced and unvoiced sounds can be produced. Tracheoesophageal speech is generated using a transesophageal puncture made in posterior wall of the trachea and the anterior wall of the esophagus just behind the tracheostoma. A one-inch valve or prosthesis is placed in the puncture. Speech is produced by blocking the tracheostoma either with a finger or an adjustable valve. The air expelled from the lungs is directed through the prosthesis into the esophagus, providing excitation to the vocal tract. The main problems associated with this method are bleeding disorders, anxiety disorders, dementia, poor vision and poor manual dexterity, which get further aggravated with increase in age of the patient.

The third method to produce alaryngeal speech is by using an artificial larynx generally called as electrolarynx as shown in Figure 1(d). There are numerous models of artificial electrolarynxes available for the laryngeotomees which aid to produce voiced speech[8–10]. In order to produce consonants without any pulmonary airflow, the

laryngeotomees have to undergo training period under a speech therapist. The electrolarynx is a handheld device, which has a vibrator at one end and the other end is placed firmly against the neck in the area of the hypopharynx. When the device is switched on, it produces a mechanical sound, which when coupled with the upper part of throat, allows the sound vibrations to travel through the throat tissue to the vocal tract. These vibrations are given spectral shape together by the vocal and nasal tract and result in audible speech. A proper coordination between switching on and off of the device must be maintained to avoid the noise generation between the utterances[6]. Researchers observe esophageal speech and artificial larynx speech as attuned to offer the laryngectomee with a viable solution for ready speech source for all occasions[6,11]. The process of laryngeal and alaryngeal speech production is shown in Figure 1.

The artificial larynx has some inherent drawbacks, due to which it does not allow intelligible and natural speech, thereby impacting the quality of life of a laryngeotomee. The first drawback is the spectral distortion introduced by the throat tissues through which the vibrations generated by the artificial larynx travel[12]. The second drawback is loss in signal energy due to low frequency attenuation by the improper coupling of the artificial larynx with the skin as the transmission loss is inversely proportional to frequency. Sometimes the vibrations may not propagate through the medium when the neck muscles have thickened due to the radiation, generally given after the laryngectomy operation[12]. Inefficient coupling of the device to the body results in the deficiency of low frequency. Finally, there is continuous background noise emanating from the device even during the unvoiced speech and stops. Thus, the quality of the alaryngeal speech so produced is unintelligible and annoying.

A multipronged strategy has been adopted by speech researches and developers to address the problem of EL speech. One strategy is to modify the alaryngeal speech[13,14], other is to enhance the EL speech[15–19] and another is to modify the design of electrolarynx[20–22]. Literature survey indicates that there are huge gaps in EL speech and normal speech and further research needs to be done to improve the quality and intelligibility of EL speech. Recently researchers have employed the non-linearity of the medium for propagation and interaction of ultrasonic waves for transmitting speech in restricted zones[23–25]. The idea of generating alaryngeal speech using interaction of two ultrasonic waves was first given in[26–29]. The same idea was further investigated using 3D based simulation of Austin man model[30]. In this paper, the concept has been explored for synthesizing the speech for laryngeotomee by using the audible excitation generated by non-linear interaction of ultrasonic signals. Use of ultrasonic frequencies avoids the audible background noise present in conventional artificial larynxes. The theory of non-linear propagation is presented in Section 2 and methodology of investigations carried out in Section 3. Results are discussed in Section 4. The conclusion and future scope has been discussed in Section 5.

2. Non-Linear Acoustic Theory of Sound Propagation

Sound is an acoustic wave and may be treated as a pressure wave for deriving its equations of propagation[31]. For comparatively large materials in comparison to the wavelength, the propagating wave may be treated as a ray[32]. Therefore, the concepts of reflection, refraction, diffraction, Doppler effect[33], principal of superposition[34], Fermat's[35], and Huygens principle[33] can be used for describing the wave-material interactions[36,37]. This is called geometric acoustics[38] as the dual to wave acoustics[39]. All acoustic waves, audio or ultrasonic obey the laws of geometric acoustics[40]. Since a sound wave propagates through a medium as a localized pressure change, increasing the pressure in the medium increases its temperature, the speed of sound in a particular area in a compressed material increases with temperature, due to which the wave travels faster during the high pressure phase of the oscillation than during the lower pressure phase. This affects the wave's spectral shape. The wave which is in the beginning a plane sine wave comprising of a single frequency, cumulatively becomes more like a saw tooth wave. The wave distorts itself and new frequency components are introduced which can be described by the Fourier series. This phenomenon, where the medium exerts its properties on the wave propagation, is called non-linear behavior. This effect always comes into existence, but the effect of geometric spreading and of absorption usually overcomes this self-distortion, thereby allowing the linear behavior to prevail over far field area. Linear acoustic theory is valid for small amplitude waves and holds good for many vibro-acoustic interactions but is invalid in sound fields with high sound pressure levels. However, the non-linear acoustic behavior sustains when the waves are having finite amplitude and close to the

source of generation of waves. Finite amplitude waves can be generated in interior fields when resonance occurs[41], in the far-field of atmospheric and underwater explosions[42], in tire noise generation[43] and in many aero-acoustic sources such as sonic booms[44]. Nonlinear effects increase with the frequency of the waves, and thus the study of nonlinear acoustics has also become important in high-frequency applications such as ultrasound[45]. For example, when two finite amplitude acoustic waves having different frequencies interact with one another in a medium, secondary waves, whose frequencies correspond to the sum and difference of the primary waves, may be produced as a result of nonlinear interaction of the incoming waves[46,47]. Additionally, a wave of different amplitudes also generates different pressure gradients, contributing to the non-linear effect. The pressure changes within a medium cause the wave energy to transfer to higher harmonics. Attenuation increases with frequency, as a counter effect exists, that changes the behaviour of nonlinear effect over distance. The level of nonlinearity of materials is described by a nonlinearity parameter (B/A). The values of A and B are the coefficients of the first and second order terms of the Taylor series expansion equation, relating the material's pressure to its density. The Taylor series has more terms and hence more coefficients but the coefficients C, D. etc. are seldom used. Linear modelling computes approximate estimate of the behavior of the system, on the other hand, the exact behaviour of the underlying systems may only be determined by non-linear modelling. Real world problems in coastal and ocean engineering usually have been solved numerically by non-linear formulations. If the effect of convective and constitutive nonlinearities is neglected, the partial differential equation for a linear acoustic wave may be written as:

$$\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} - \Delta\phi = 0 \tag{1}$$

Where ϕ is the velocity potential and c the speed of sound. Velocity potential and particle velocity (u) are related by relation $\phi = \nabla u$. This equation has very rare applications as most of the time the propagation of the acoustic wave is non-linear. The amount of non-linearity depends upon the relation between pressure, density, and ratio of specific heats (γ). If air compression (or expansion) is isothermal (taking place at constant temperature T), then, according to the ideal gas law $PV = nRT$, the pressure P would simply be proportional to density ρ giving the value of γ as 1. However, heat diffusion is much slower than audio acoustic vibrations due to which air compression and expansion is much closer to isentropic or constant entropy S in normal acoustic situations. An isentropic process is also called as reversible adiabatic process. This means that when air is compressed by shrinking its volume V, not only does the pressure P increases but the temperature T also increases. In a constant-entropy compression or expansion, temperature changes are not given time to diffuse away to thermal equilibrium, instead, they remain largely frozen in place. Compressing air heats it up and relaxing the compression cools it back down, thereby, introducing nonlinearities in the wave. Accordingly the nonlinear isentropic equation of state for air can be written as follows:

$$\frac{P}{P_0} = \left(\frac{\rho}{\rho_0}\right)^{\gamma} \tag{2}$$

Where, P and P_o are the total and reference pressures, ρ and ρ_o are the current and reference densities. For air, the value of γ is about 1.4 for air under normal conditions of pressure and temperature. Equation (2) may be expanded using Taylor series at system entropy $s = s_0$[44]:

$$p = P - P_0 = \left(\frac{\partial P}{\partial \rho}\right)_{s0,\rho 0}(\rho-\rho_0) + \frac{1}{2}\left(\frac{\partial^2 P}{\partial \rho^2}\right)_{s0,\rho 0}(\rho-\rho_0)^2 + \cdots \tag{3}$$

This can be written as

$$p = A\left(\frac{\rho-\rho_0}{\rho_0}\right) + \frac{B}{2}\left(\frac{\rho-\rho_0}{\rho_0}\right)^2 + \cdots \tag{4}$$

Where, $A = \rho_0\left(\frac{\partial P}{\partial \rho}\right)_{s0,\rho 0} \equiv \rho_0 c^2$,

$B = \rho_o^2\left(\frac{\partial^2 P}{\partial \rho^2}\right)_{s0,\rho 0}$, and $\left(\frac{\partial P}{\partial \rho}\right)_{s0,\rho 0} = c^2$

The parameter $\frac{B}{A} = \frac{\rho_0}{c_0^2}\left(\frac{\partial^2 P}{\partial \rho^2}\right)_{s0,\rho 0}$ accounts for the amount of nonlinearity[44]. For linear approximation the higher order terms may be neglected, giving:

$$p = A\left(\frac{\rho-\rho_0}{\rho_0}\right) = c_0^2(\rho-\rho_0)$$

Which can be written as $\frac{p}{(\rho-\rho_0)}=c^2$ (6)

Equation (6) shows that stiffness or bulk modulus $\frac{p}{(\rho-\rho_0)}$ is simply the square of the linear speed of sound. Incorporating the effect of conservation of mass and momentum in (4) up to second order terms, the resulting equation is[41,47–49]:

$$\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}-\Delta\phi-\frac{1}{c^2}\frac{\partial}{\partial t}\left(b(\Delta\phi)+\frac{B/A}{2c^2}\left(\frac{\partial\phi}{\partial t}\right)^2+(\nabla\phi)^2\right)=0 \quad (7)$$

The first two terms in (7) are the same as in (1), but the fourth and fifth terms are due to incorporating non-linear behavior of the system. The third term is actually a linear absorption term, but it is usually grouped with the nonlinear terms to indicate deviation from the linear wave equation. The parameter b is for absorption in the fluid due to viscosity and thermal conductivity. Putting $\phi=\frac{1}{\rho}\psi$ in (7):

$$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}-\Delta\psi-\frac{1}{c^2}\frac{\partial}{\partial t}\left(b(\Delta\psi)+\frac{B/A}{2\rho c^2}\left(\frac{\partial\psi}{\partial t}\right)^2+\frac{(\nabla\psi)^2}{\rho}\right)=0 \quad (8)$$

The range of validity of nonlinear wave equations is typically given in terms of acoustic mach number, defined as the ratio of particle velocity and the velocity of sound. Because of the non-linear propagation, whenever, two high frequency sinusoidal waves travel through a medium, waves having frequencies as the sum and difference of the propagating frequencies, along with their harmonics are generated along the direction of propagation. The higher frequency components get attenuated relatively at higher rates as compared to lower frequencies[50,51]. The resultant primary and secondary pressure waves may be given as[52];

$$p_0(\vec{r},t)=\frac{1}{2i}\left[a_{01}(\vec{r})e^{i2\pi f_1 t}+a_{02}(\vec{r})e^{i2\pi f_2 t}\right]+p_c(\vec{r},t)\ , \quad (9)$$

$$p_s(\vec{r},t)=\frac{1}{2i}\left[a_{s1}(\vec{r})e^{i4\pi f_1 t}+a_{s2}(\vec{r})e^{i4 f_2 t}+a_{+}(\vec{r})e^{i2\pi(f_1+f_2)t}+a_{-}(\vec{r})e^{i2\pi(f_1-f_2)t}\right] \quad (10)$$
$$+p_c^{'}(\vec{r},t),$$

Where p_0 primary wave consisting of two frequencies f_1 and f_2. The resultant secondary waveform p_s consists of sum- and difference frequencies long with the constituent frequencies f_1 and f_2. Here, $p_c(\vec{r},t)$ and $p_c^{'}(\vec{r},t)$ are remaining constituent terms.

The investigations of interaction of two frequencies, both having high frequencies were analyzed. The effect of source separation, the crossing angle on the propagating ultrasonic wave shapes and different high frequencies have been studied[24]. The frequencies of the ultrasonic waves were taken as 40.5 kHz, 39.5 kHz, 41 kHz, 39 kHz, 41.5 kHz, 38.5 kHz, 42 kHz and 38 kHz. The beam crossing angle was fixed at 3° and the source separation at 0.15 m. The investigations showed that the amplitude of the difference frequency generated was proportional to the square of the difference frequency. Investigations of the source separation showed that on increasing the source separation, the sound spot with small area could be generated. Also, the maximum value of the difference frequency decreases with the increase of the source separation and its position of the sound spot gradually shifts further away from the origin. As the crossing angle increases, the maximum sound amplitude is closer to the origin along with the decrease in the sound zone. The valid length in the direction of propagation also becomes smaller with increasing interaction angle.

The observations may be used for designing an artificial larynx, using two ultrasonic vibrators separated by fixed around the neck and directing their beams along the vocal tract interacting at an crossing angle of about 60 to 80 degree, giving the spot of low frequencies near the centre of the laryngeal space, which may act the source of audio excitation to the vocal tract.

3. Methodology

The methodology of the investigations may be categorized into six sub-sections: Speech material preparation, estimation of coefficients of quadratic non-linear equation, estimation of ultrasonic frequency, speech analysis and speech synthesis and speech evaluation, as follows;

3.1 Speech Material Preparation

The recording of 6 speakers in Hindi (3 male and 3 females) and 6 alaryngeal speakers (4 male and 2 female) was carried out using high quality recording system. The speakers were of the age group of 20 to 30 years, all university students pursuing Post-graduate studies in sciences. Although the first language of the speakers was Dogri but they were able to speak Hindi fluently. The speakers were asked to utter cardinal vowels /*a*/, /*i*/ and /*u*/. The recording was carried out in an acoustically treated room. The recorded speech was segmented into isolated vowels and

labeled accordingly. The recording was carried out at a sampling frequency of 16000 Hz, keeping quantization at 16 bits. The recorded speech was up-sampled at 300000 Hz to enable the experimentation at ultrasonic frequencies.

3.2 Estimation of Non-Linear Parameters

The estimation of non-linear parameters *a*, *b*, *c*, *d*, and *e* was done using genetic algorithm by taking a population of 10 DNAs and 200 iterations. Spectral error between the expected vowel and the synthesized vowel using the excitation obtained from the non-linear relation was taken as the cost function. The quadratic non-linear equation used to synthesize the excitation is given below:

$$y = ay_3^2 + by_2^2 + cy_3y_2 + dy_3 + ey_2 \quad (11)$$

Where, y_2 and y_3 are described in Table 1.

Table 1. Synthesized speech using different excitations

S. No.	Excitation	Synthesized speech
1.	y_1: Sine wave of desired pitch frequency.	The speech is synthesized using y_1 and vocal tract parameters of cardinal vowels spoken by normal/ alaryngeal speakers
2.	y_2: Sine wave of high frequency (12 kHz or 40 kHz).	The speech is synthesized using y_1 and vocal tract parameters of cardinal vowels spoken by normal/ alaryngeal speakers
3.	y_2: Sine wave of high frequency (12 kHz or 40 kHz) plus pitch of the speaker.	The speech is synthesized using y_3 and vocal tract parameters of cardinal vowels spoken by normal/ alaryngeal speaker.
4.	y_4: Modulated high frequency sine wave. The modulating signal is a sine wave having frequency equal to average pitch of the speaker.	The speech is synthesized using y_4 and vocal tract parameters of cardinal vowels spoken by normal/ alaryngeal speaker.
5.	y_5: Non-linear combination of y_2 and y_3.	The speech is synthesized using non-linear combination of y_2 and y_3, and vocal tract parameters of cardinal vowels spoken by normal/alaryngeal speaker

3.3 Estimation of Optimum Ultrasonic Frequency

Frequency of ultrasonic waves and frequency differences were selected on the basis of heating of biological tissues present in the vocal tract by an ultrasonic wave. The model was developed using comsol, wherein a tissue phantom was selected and subjected to different high frequencies and the corresponding rise in temperature of the tissue was noted. It was found that frequencies in and around 40 kHz range are comfortable for investigations to be carried out. Further ultrasonic transducers in 40 kHz range are also economically available. The difference frequency generated is also of enough amplitude to produce an audible wave.

3.4 LPC Analysis

Vocal tract parameters were extracted from the recorded normal and alaryngeal speech segments using Linear Predictive Coding (LPC) technique[53] shown in Figure 3. In this technique, the parameters are estimated in the form of LPC coefficients, vocal tract area functions, Line Spectral Frequencies (LSFs) along with fundamental frequency F_o, bandwidth *BW*, intensity of the sound and residual signal representing the vocal tract excitation[54–56]. In this model, the relation between speech $s(n)$ signal and excitation signal $u(n)$ can be written as:

$$s(n) = \sum_{k-1}^{p} a_k s(n-k) + Gu(n) \quad (12)$$

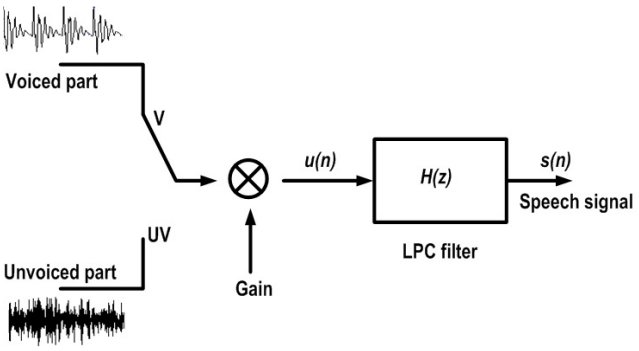

Figure 3. LPC model of speech.

Where G is gain and $a_k's$ are the LPC coefficients representing the vocal tract shape during the production of the speech segment under consideration. Features were extracted with window size 20 ms, overlap of 10 ms, and order of 500. It may be noted that the LPC order has been

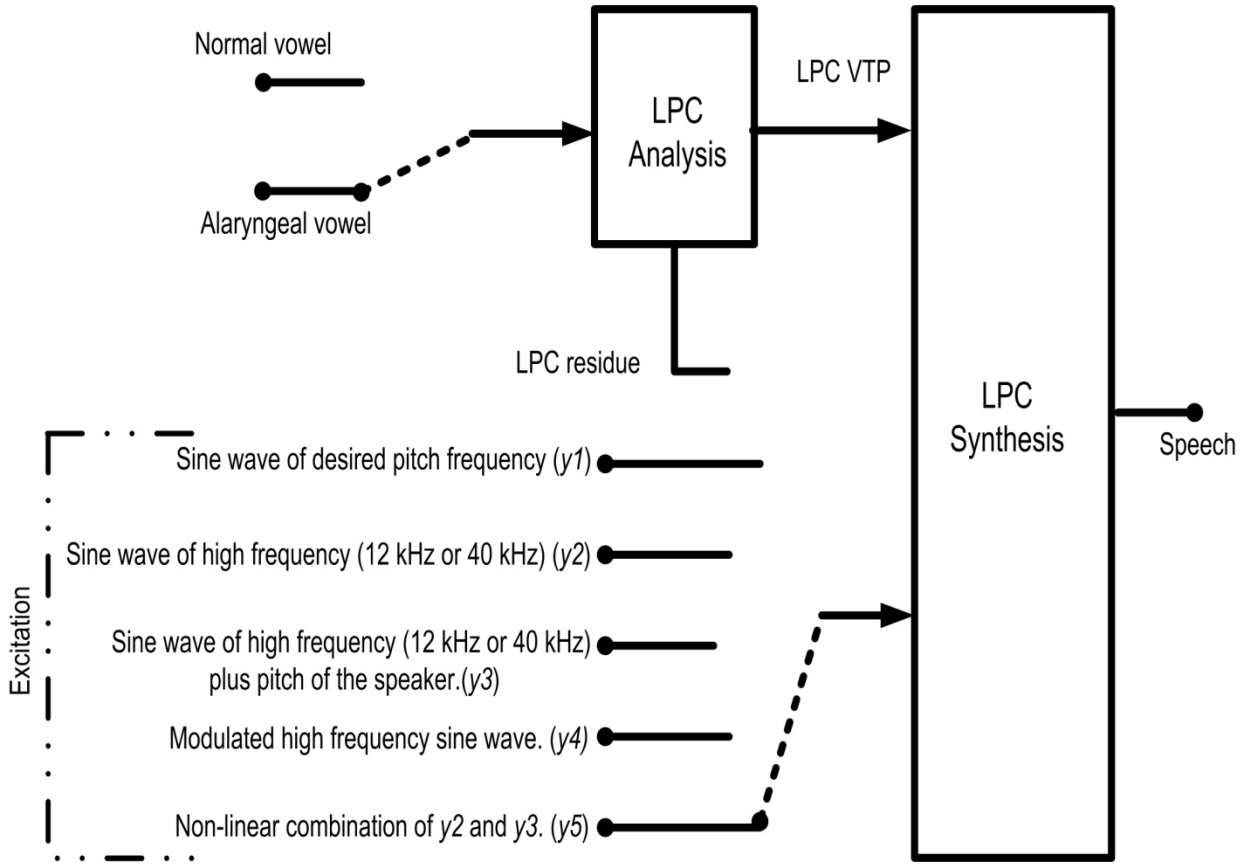

Figure 4. Schematic of the investigations.

taken on higher end because of the high sampling frequency of the signal available for experimentation. The vowels of six normal and six alaryngeal speakers were taken for the analysis. The scheme for the analysis is shown in Figure 4.

3.5 Speech Synthesis

The synthesized speech was generated with different excitations to enhance the alaryngeal speech. As explained in Section 2, ultrasonic generated waves create audible sound in air by means of the nonlinear interaction of ultrasonic waves. This phenomenon of sound generation due to nonlinear interaction of ultrasonic waves has been exploited in this experiment; the detail is shown in Figure 4. The figure shows different types of excitations such as y_1, y_2, y_3, y_4 and Non-linear combination of y_2 and y_3 along with the excitations obtained from the LPC analysis of normal and alaryngeal speech samples. The excitation y_1 is a sinusoid of the desired pitch frequency. The excitation y_2 is a high frequency (12 to 40 kHz) sine wave. The excitation y_3 is the sum of a high frequency (12 to 40 kHz) sinusoid and the desired pitch frequency. The excitation y_4 is a high frequency sinusoid modulated by pitch frequency and excitation y_5 is a quadratic non-linear function of excitations y_2 and y_3. The pitch frequency selected is 200 Hz, 350 Hz and 550 Hz to account for the varying human speech pitch frequency range. The coefficients of the non-linear quadratic equation were optimized using genetic algorithm. For synthesis, the vocal tract parameters were taken either from normal speech or alaryngeal speech along with the excitation taken from y_1 to y_5, using one type of excitation at a time. The different combination of the output speech is listed in the Table 1. For further processing, the speech was down-sampled to the sampling frequency of 16 kHz. .

3.6 Speech Evaluation

The evaluation of the synthesized speech was carried out using informal listening tests, visual analysis of the spectrograms and MOS based subjective evaluation. It is expressed in number from 1 to 5, 1 being the worst and 5 the best. MOS is quite subjective as it is based on figures that result from what is perceived by people during tests. However, there are software applications that measure

MOS on networks as we see below. A perceptual evaluation was done to compare the quality of the synthetic and the original vowels. Listeners generally preferred to listen to the synthesized words, indicating that alaryngeal speech enhancement was achieved. It is expressed in number from 1 to 5, 1 being the worst and 5 the best.

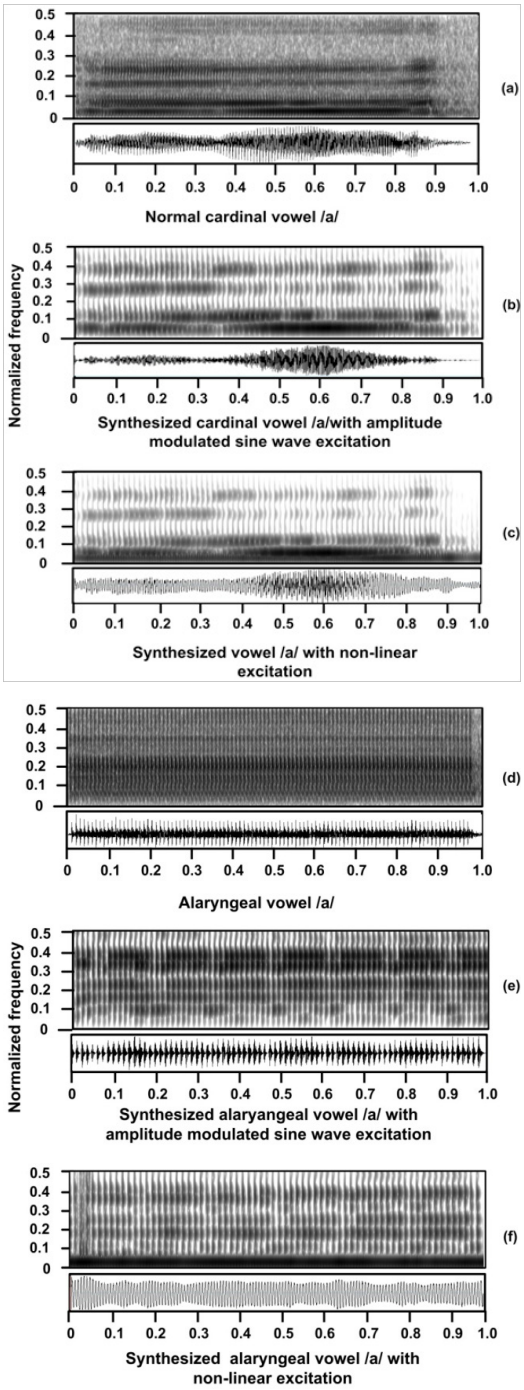

Figure 5. Speech synthesis for cardinal laryngeal vowel /a/ and alaryngeal vowel /a/ with (a, d) recorded (b, e) amplitude modulated and (c, f) non-linear excitations.

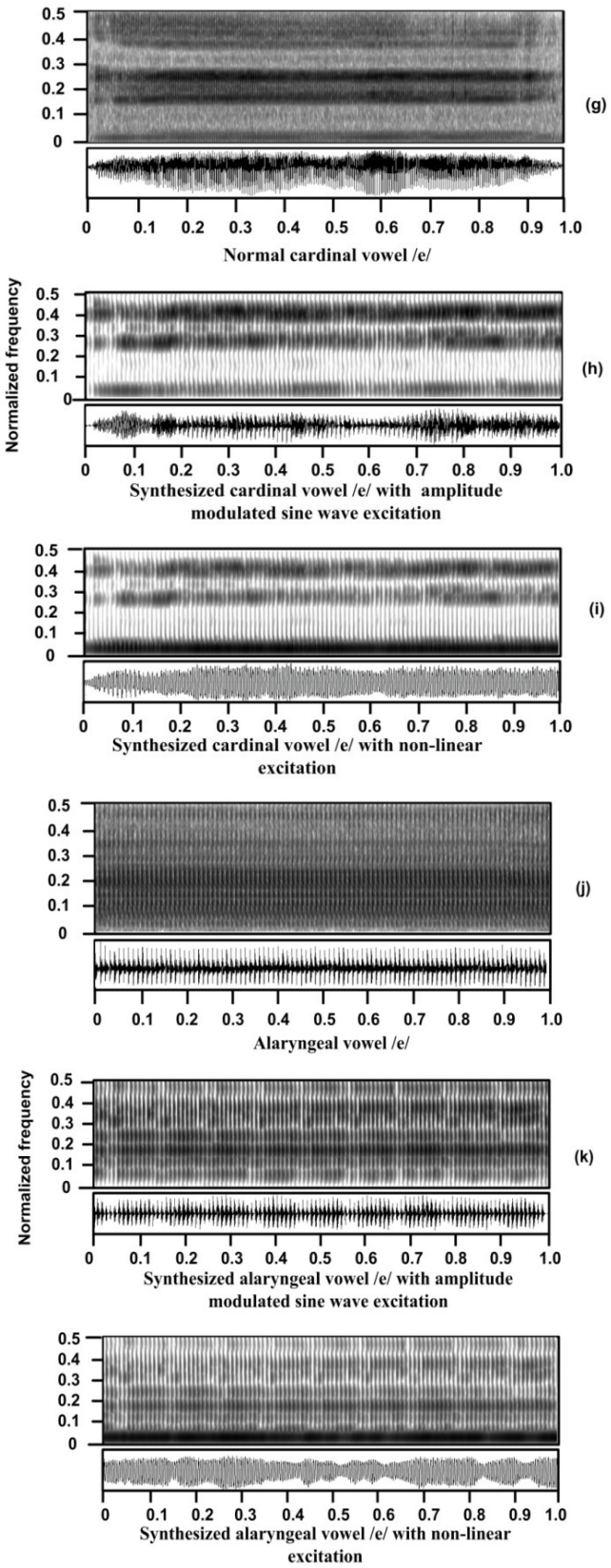

Figure 6. Speech synthesis for cardinal laryngeal vowel /i/ and alaryngeal vowel /i/ with (g, j) recorded (h, k) amplitude modulated and (i, j) non-linear excitations.

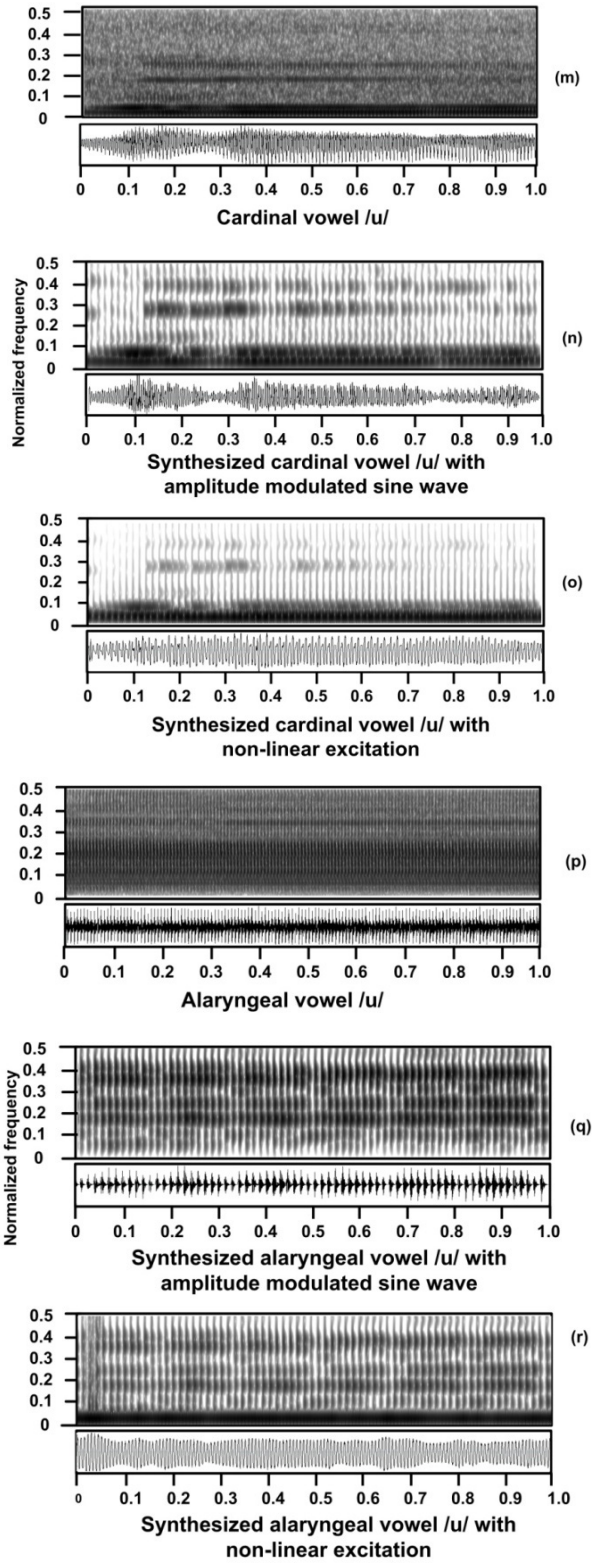

Figure 7. Spectrograms of speech synthesized for cardinal laryngeal vowel /u/ and alaryngeal vowel /u/ with (m, p) recorded (n, q) amplitude modulated and (o, r) non-linear excitations.

4. Results

Synthesized speech generated using different excitations, as mentioned in Table 1 was evaluated using visual, objective and subjective tests. For estimating the values of the coefficients of the non-linear Equation (11), the cardinal laryngeal vowels were analyzed, using LPC for extracting glottal excitation. One cycle of excitation for each was obtained from the beginning, middle and ending section of the glottal waveform. The full excitation was synthesized by repeating these segmented cycles. These excitations were used as reference signals for tuning the parameters of the genetic algorithm, finally giving the values of the coefficients as a = 0.7329, b = -0.5448, c = -0.6170, d = -0.8093 and e = 0.4988.

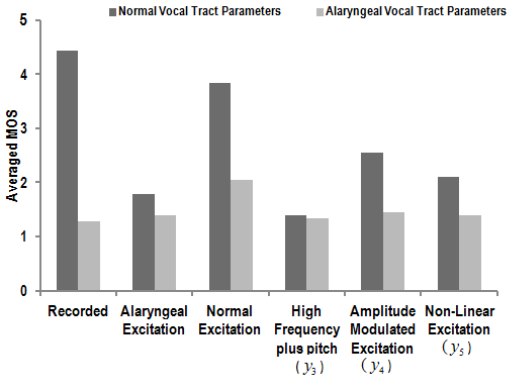

Figure 8. Histographic representations of averaged MOS for the cardinal vowel /a/.

Visual analysis was carried out using spectrograms; some of them are shown in Figure 5 to Figure 7. The scope of the analysis in the present paper has been confined to only cardinal vowel /*a*/, /*i*/ and /*u*/. The analysis and synthesis was carried out using LPC platform. The formants structure of the synthesized speech was comparable to that of the recorded speech. The speech synthesized with non-linear excitation has smooth formant structure as compared to the speech synthesized with other excitations.

Informal listening tests indicated that the quality of the speech synthesized with amplitude modulated excitation (y_4) was slightly better than the speech synthesized by using high frequency sine wave as excitation (y_2 and y_3). The speech synthesized with non-linear combination of y_2 and y_3, represented as y_5, showed even better naturalness.

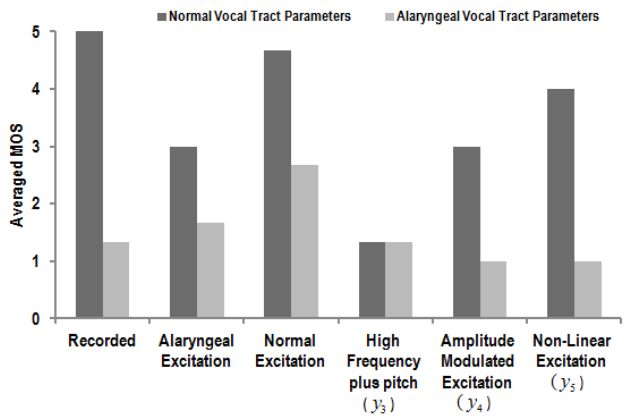

Figure 9. Histographic representations of averaged MOS for the cardinal vowel /i/.

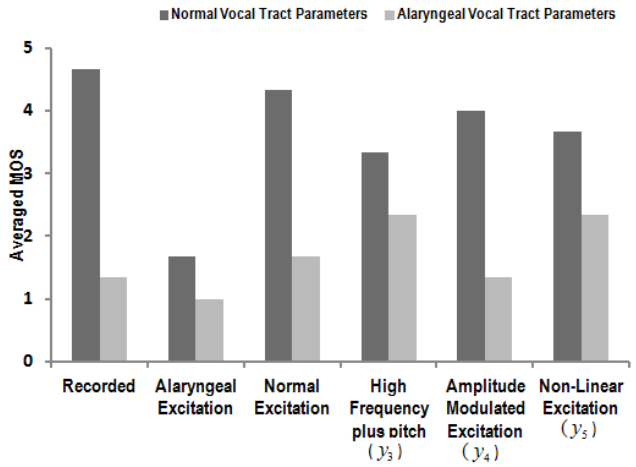

Figure 10. Histographic representation of averaged MOS for the cardinal vowel /u/.

Subjective evaluation using MOS was conducted for six listeners and the results are shown in Figure 8 to Figure 10. Three listeners were post graduate science students and other three were scholars doing research in speech signal processing. The speech synthesized using nonlinear excitation and alaryngeal vocal tract parameters reduced background noise to very large extent, improving the intelligibility and naturalness of alaryngeal speech. The speech synthesized for vowel /*i*/ was found to be more intelligible and natural than other vowels (/*a*/ and /*u*/). The averaged MOS of recorded laryngeal speech for /a/, /*i*/, and /*u*/ were calculated as 4.4, 5 and 4.7, respectively. On the other hand, averaged score for the recorded alaryngeal vowels was 1.3, 1.3, and 1.3. The averaged MOS for y_3 excitation was obtained as 1.4, 1.3, and 3.3, for /*a*/, /*i*/ and /*u*/ with normal vocal tract parameters, and 1.3, 1.3, and 2.3 with alaryngeal vocal tract parameters. The average MOS for the speech synthesized using amplitude modulated excitation (y_4) was 2.6, 3.0, and 4.0 with normal vocal tract parameters and 1.4, 1.0, and 1.3 with alaryngeal vocal tract parameters. The average MOS for non-linear excitation was 2.1, 4.0, and 3.7 with normal vocal tract parameters and 1.4, 1.0, and 2.3 with alaryngeal vocal tract parameters. It can be seen that the averaged MOS score for speech synthesized with non-linear excitation for /*i*/ and /*u*/ is comparatively better than that of /*a*/ . The overall quality of speech synthesized with non-linear excitations is intelligible and natural with negligible background noise.

5. Conclusion

Investigations were carried out to explore the use of non-linear propagation of ultrasonic waves to generate audible excitation for reducing background noise for alaryngeal speech. LPC was used for extracting vocal tract parameters and excitation. For investigations, cardinal laryngeal and alaryngeal vowels were used. Five types of excitations, sinusoid of the desired pitch frequency, sinusoid of high frequency, sinusoid of high frequency plus pitch, amplitude modulated high frequency and non-linearly generated excitation were used. The coefficients of the non-linear equation were obtained using genetic algorithm. The investigations were carried out at 12 kHz to 40 kHz frequencies. The investigations showed that non-linearly generated excitation may be used for reducing the background noise in alaryngeal speech for laryngeotomee. The speech generated was relatively intelligible and natural as compared to other the speech generated by other types of excitation. The subjective test using MOS showed that the quality of the synthesized speech using non-linear excitation is 2.1, 4.0 and 3.7 with laryngeal vocal tract parameters, for the three cardinal vowels as compared to the MOS of 1.4, 1.0 and 2.3 with alaryngeal vocal tract parameters. The quality of the speech may further be improved by taking more number of ultrasonic waves in the form of parametric arrays. These investigations are on the future plan of our research.

6. Ackowledgement

The authors would like to thank Prof. P. C. Pandey from IIT Bombay for his rare insight and invaluable guidance that made this work possible.

7. References

1. Fant G. Acoustic theory of speech production. The Hague, Netherlands: Mouton and Company; 1970.
2. Rabiner LR, Schafer RW. Digital processing of speech signals. Prentice Hall: The University of Michigan; 1978.
3. Manjari M, Popli R, Paul S, Gupta VP, Kaholon SK. Prevalence of oral cavity, pharynx, larynx and nasal cavity malignancies in Amritsar, Punjab. Indian Journal of Otolaryngology and Head and Neck Surgery. 1996 Jul; 48(3):191–5.
4. Bhattacharjee A, Chakraborty A, Purkaystha P. Prevalence of head and neck cancers in the north east - an institutional study. Indian Journal of Otolaryngology and Head and Neck Surgery. 2006 Jan; 58(1):15–9. PMid: 23120228. PMCid: PMC3450618.
5. Bakshi J, Panda NK, Sharma SC, Gupta A, Mann SB. Survival patterns in treated cases of carcinoma larynx in North India: A 10-year follow-up study. Indian Journal of Otolaryngology and Head and Neck Surgery. 2005 Apr; 57(2):103–7. PMid: 23120142. PMCid: PMC3450955.
6. Boone DR. The voice and voice therapy. New Jersey: Prentice-Hall; 1971.
7. Perkins WH. Speech pathology: An applied behavioral science. CV Mosby Company: The University of Michigan; 1977.
8. Lebrun Y. History and development of laryngeal prosthetic devices. The Artificial Larynx; 1973. p. 19–76.
9. Goldstein LP. History and development of laryngeal prosthetic devices. Electrostatic Analysis and Enhancement of Alaryngeal Speech; 1982. p. 137–65. PMid: 6283108. PMCid: PMC256734.
10. Green G, Hults M. Preferences for three types of alaryngeal speech. Journal of Speech and Hearing Disorders. 1982 May; 47(2):141–5. PMid: 7176589. Crossref
11. Diedrich WM, Youngstrom KA. Alaryngeal speech. Washington, D.C.: Charles C. Thomas Publisher; 1966.
12. Singer MI, Blom ED. An endoscopic technique for restoration of voice after laryngectomy. Annals of Otology, Rhinology and Laryngology. 1980 Nov; 89(6):529–33. PMid: 7458140. Crossref
13. Liu H, Zhao Q, Wan M, Wang S. Enhancement of electrolarynx speech based on auditory masking. IEEE Transactions on Biomedical Engineering. 2006 May; 53(5):865–74. PMid: 16686409. Crossref
14. Uemi N, Ifukube T, Takahashi M, Matsushima JI. Design of a new electrolarynx having a pitch control function. Proceedings of 3rd IEEE International Workshop. RO-MAN'94 Nagoya; 1994. p. 198–203.a
15. Nakamura K, Toda T, Saruwatari H, Shikano K. Speaking-aid systems using GMM-based voice conversion for electrolaryngeal speech. Speech Communication. 2012 Jan; 54(1):134–46. Crossref
16. 16. Espy-Wilson CY, Chari VR, Huang CB. Enhancement of alaryngeal speech by adaptive filtering. Proceedings of Fourth International Conference in Spoken Language; 1996. p. 764–7. Crossref
17. Bi N, Qi Y. Application of speech conversion to alaryngeal speech enhancement. IEEE Transactions on Speech and Audio Processing. 1997 Mar; 5(2):97–105. Crossref
18. Pandey PC, Bhandarkar SM, Bachher GK, Lehana PK. Enhancement of alaryngeal speech using spectral subtraction. Proceedings of 14th International Conference of Digital Signal Processing; 2002. p. 591–4. Crossref
19. Murakami K, Araki K, Hiroshige M, Tochinai K. A method for speech transform from electrolaryngeal speech to normal speech. IEICE Transactions. 2004; 87:1030–40.
20. Liu H, Ng ML. Electrolarynx in voice rehabilitation. Auris Nasus Larynx. 2007 Sep; 34(3):327–32. PMid: 17239553. Crossref
21. Saikachi Y, Stevens KN, Hillman RE. Development and perceptual evaluation of amplitude-based F0 control in electrolarynx speech. Journal of Speech, Language and Hearing Research. 2009 Oct; 52(5):1360–9. Crossref
22. Goldstein EA, Heaton JT, Kobler JB, Stanley GB, Hillman RE. Design and implementation of a hands-free electrolarynx device controlled by neck strap muscle electromyographic activity. IEEE Transactions on Biomedical Engineering. 2004 Feb; 51(2):325–32. PMid: 14765705. Crossref
23. Ahmadi F, McLoughlin I. The use of low-frequency ultrasonics in speech processing. Signal Processing. Sebastian Miron, ed. INTECH Open Access Publisher: 2010. p. 503–28. Crossref
24. Pompei FJ. The use of airborne ultrasonics for generating audible sound beams. Audio Engineering Society Convention 105 Audio Engineering Society; 1998 Sept.
25. Ji P, Yang J, Gan WS. The investigation of localized sound generation using two ultrasound beams. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control. 2009 Jun; 6(6):1282–7. PMid: 19574137. Crossref
26. Bhat RM, Lehana PK. Ultrasonic larynx. IP India 2595/DEL/2013; 2013 Sept.
27. Bhat RM, Lehana P. Investigations of the synthesis and evaluation of alaryngeal speech generated using ultrasonic excitations. Paper presentation in INTERSPEECH; Dresden. 2015.
28. Bhat RM, Lehana PK. Efficiency of synthetic excitation obtained by interference of ultrasonic waveforms for reducing background noise for laryngeotomee. The Journal of the Acoustical Society of America. 2016 Apr; 139(4):2222. Crossref
29. Bhat RM, Lehana P. Exploring non linearity in acoustics for generating glottal excitation for laryngeotomee. International Journal of Scientific and Technical Advancements. 2016 Sept; 2(4):307–10.

30. Mills P, Zara J. 3D simulation of an audible ultrasonic electrolarynx using difference waves. PloS One. 2014 Nov; 9(11):e113339. PMid: 25401965 PMCid: PMC4234661. Crossref
31. Merriam-Webster. Webster's Ninth New Collegiate Dictionary. Merriam-Webster Incorporation: 1990.
32. Buhler O. A Brief introduction to classical, statistical and quantum mechanics. New York University, New York: American Mathematical Society; 2006. Crossref
33. Benenson W, Harris JW, Stocker H, Lutz H, editors. Handbook of physics. 4th ed. New York: Springer-Verlag Science and Business Media; 2000.
34. Avallone EA, Baumeister T, Sadegh A, Marks LS. Marks standard handbook for mechanical engineers. 11th ed. McGraw-Hill Professional; 2006.
35. Blitz J. Fundamentals of ultrasonics. 2nd ed. London: Butterworth and Company; 1967.
36. Karal FC Jr, Keller JB. Elastic wave propagation in homogeneous and inhomogeneous media. The Journal of the Acoustical Society of America. 1959 Jun; 31(6):694–705. Crossref
37. Keller JB, Karal FC Jr. Geometrical theory of elastic surface-wave excitation and propagation. The Journal of the Acoustical Society of America. 1964 Jan; 36(1):32–40. Crossref
38. Crocker MJ. Handbook of Acoustics. New York: A Wiley-Interscience Publication; 1998.
39. Watkinson J. The art of sound reproduction. CRC Press: 2012. PMCid: PMC3411993.
40. Cheeke JD. Fundamental and applications of ultrasonic waves. 2nd ed. Boca Raton, FL: CRC Press; 2002. PMid: 12216789.
41. Enflo BO, Hedberg CM. Theory of nonlinear acoustics in fluids. Dordrecht, Netherlands: Kluwer Academic Publishers; 2002. PMid: 12400955.
42. Castor K, Gerstoft P, Roux P, Kuperman WA, McDonald BE. Long-range propagation of finite-amplitude acoustic waves in an ocean waveguide. The Journal of the Acoustical Society of America. 2004 Oct; 116(4):2004–10. Crossref
43. Gagen MJ. Novel acoustic sources from squeezed cavities in car tires. The Journal of the Acoustical Society of America. 1999 Aug; 106(2):794–801. Crossref
44. Hamilton MF, Blackstock DT, editors. Nonlinear acoustics. San Diego: Academic press; 1998.
45. Hoffelner J, Landes H, Kaltenbacher M, Lerch R. Finite element simulation of nonlinear wave propagation in thermoviscous fluids including dissipation. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control. 2001 May; 48(3):779–86. PMid: 11381703. Crossref
46. Morse PM, Ingard KU. Theoretical acoustics. International Series in Pure and Applied Physics. New York: McGraw Hill; 1968.
47. Kuznetsov VP. Equations of nonlinear acoustics. Soviet Physics Acoustics-USSR. 1971 Jan; 16(4):467–70.
48. Makarov S, Ochmann M. Nonlinear and thermoviscous phenomena in acoustics, Part II. Acta Acustica United with Acustica. 1997 Mar; 83(2):197–222.
49. Naugolnykh KA, Ostrovsky LA. Nonlinear wave processes in acoustics. New York: Cambridge University Press; 1998.
50. Pierce AD. Acoustics. New York: McGraw-Hill; 1981. p. 481–94.
51. Soderholm LH. On the Kuznetsov equation and higher order nonlinear acoustics equations. Proceedings of AIP Conference; 2000. p. 133–6. Crossref
52. Wilson EL, Khalvati M. Finite elements for the dynamic analysis of fluid-solid systems. International Journal for Numerical Methods in Engineering. 1983 Nov; 19(11):1657–68. Crossref
53. Shaughnessy DO. Speech analysis. Speech Communications. 2nd ed. Hyderabad: Universities Press Private Limited; 2001. p. 192–209.
54. Pham TD, Wagner M. A geostatistical model for linear prediction analysis of speech. Pattern Recognition. 1998 Dec; 31(12):1981–91. Crossref
55. Kura VB. Novel pitch detection algorithm with application to speech coding. [Doctoral dissertation]. University of New Orleans.
56. Madane AR, Shah Z, Shah R, Thakur S. Speech compression using linear predictive coding. Proceeding of the International Workshop on Machine Intelligence Research; 2009. p. 119–21.