International Journal of Scientific and Technical Advancements

ISSN: 2454-1532

Transformation of Alaryngeal Speech Using Line Spectral Frequencies

Manoj Kumar[1#], Parveen Lehana[2], Swapna Devi[3], Romilla Malla Bhat[4]
[1]Department of ECE, GCET, Jammu, J&K, India
[2,4]Department of Electronics, University of Jammu, Jammu, J&K, India
[3]Department of ECE, NITTTR, Chandigarh, India
[#]Email address: pklehana@gmail.com

***Abstract*—** In normal speech generation, air stream from the lungs acts as dc energy source. This dc flow is changed to ac flow by the vocal chords in the larynx or stoppage / constriction in the vocal tract. The vocal tract also provides the spectral shaping for the resulting speech. Persons who suffer from the diseases such as throat cancer or laryngeal impairment require the removal of vocal chords. Such persons use artificial larynx to generate the speech as an alternative to vocal folds. The quality of the alaryngeal speech so produced is very annoying and it is difficult to interpret the meaning. In this paper, three schemes to enhance the quality of alaryngeal speech have been investigated. For this study, we have used natural vowels and alaryngeal vowels uttered by using artificial larynx.

In the first method, the impulse response of the vocal tract is estimated from the LPC spectrum of the alaryngeal speech. The output is generated by convolving this impulse response with a natural glottal excitation. In second method, the source and the target vowels are analyzed using harmonic plus noise model (HNM). The transformation function is estimated by dividing the corresponding harmonic amplitudes of the source and target vowels. This transformation function is used to transform the harmonic amplitudes of the source to the harmonic amplitudes of the target. Then modified harmonic amplitudes estimated from the transformation function are interpolated at the desired harmonics of the target pitch before synthesizing the output. The third method is similar to the second method except that the transformation function is calculated by using the line spectral frequencies instead of the harmonic amplitudes.

***Keywords*—** Alaryngeal speech, glottis, transformation, excitation, larynx, line spectral frequencies.

I. Introduction

When the normal person speaks, he makes use of the vocal chords in the larynx. In normal speech synthesis, air stream from the lungs act as energy source, the vocal chords in the larynx act as the vibration source for the sound, and the vocal tract provides the spectral shaping for the resulting speech [1]. Persons who suffer from the diseases such as throat cancer or some time due to injury require that their larynx and vocal chord be removed by surgical operation. Such persons require external aids to communicate. These patients either use artificial larynx [2], [3] or esophagus to generate the excitation for the vocal tract.

An artificial larynx or electrolarynx is hand held electro-mechanical vibrator [2], [3]. This is pressed against the throat or cheeks for producing the excitation in the vocal tract. The speech is generated by changing the shape of the vocal tract similar to natural speech production mechanism. The artificial larynxes are broadly classified into three categories: external and internal pneumatic, intra-oral and implantable electronic, and external electronic (or transcervical) [2], [3], [4]. The pneumatic artificial larynxes use the air exhaled out from the lungs to produce the vibrations. It is further divided into external and internal. This device uses a tube fitted from stoma to the mouth. A vibrating reed is fixed into the tube. During exhalation, the air from the lungs moves out through the stoma and makes the reed to vibrate.

The internal pneumatic larynx [2], [3] uses a metal reed fitted inside the pharynx. To speak, the speaker closes the stoma with his finger and exhales through the reed to set it into vibrations. Since the pulmonary air is used for the production of the sound, this method of sound production is similar to natural method of sound production. On the other hand, the internal electronic larynx uses an electronic internal vibration generator. There are two types of internal electronic larynxes: implantable and intra-oral. In the implantable device, excitation source is placed below the pharynx as the natural sound source. Intra-oral type artificial larynx consisted of an Edison type phonograph cylinder, driven by an electromotor. The output of the phonograph is connected to a receiver, which is directly fitted to the patient's nose

Electronic larynx [3], [4] consists of an electronic vibration generator wherein the vibrations are produced using an electromagnet. Although the pneumatic artificial larynxes provide better output, electrolarynxes are the preferred choice because of their small size and convenience. However, they produce a strong background noise, which degrades the quality of speech output considerably.

The artificial larynx when coupled to the neck causes the vibrations to propagate through the neck tissue on to the vocal tract. The neck tissue is a highly non-uniform mass of muscle and membrane. When the sound propagates through such a medium, there is an amplitude variation and phase shift of various harmonics of the impressed sound wave [5]. Secondly, since the transmission loss is inversely proportional to frequency, the low frequency components in the signal are attenuated. Sometimes the vibrations may not propagate through the medium at all. Such is the case when the neck

Manoj Kumar, Parveen Lehana, Swapna Devi, and Romilla Malla Bhat, "Transformation of Alaryngeal Speech Using Line Spectral Frequencies," *International Journal of Scientific and Technical Advancements*, Volume 5, Issue 2, pp. 37-44, 2019.

muscles have thickened due to the radiation generally given after the laryngectomy operation [5].

In esophageal speech, the patient transports a small amount of air into the esophagus. Probably due to an increased thoracic pressure, the air is forced back past the pharyngo-esophageal (PE) segment and eddy currents are set-up. The esophagus acts as a resonator and shapes the eddy currents. The modified flow of eddy currents behaves like an excitation and is able to generate understandable speech [6], [7]. There are various techniques to transport air to the esophagus. One of the techniques involves swallowing of air into the stomach. The other is called injection technique in which air is forced to the pharynx and is further pushed to the esophagus by the back of the tongue. The synchronization of these two phases is of great importance for transporting the air into the esophagus. The patient creates a pressure in the esophagus that is lower compared to the atmospheric pressure and the air will flow from the mouth to the esophagus.

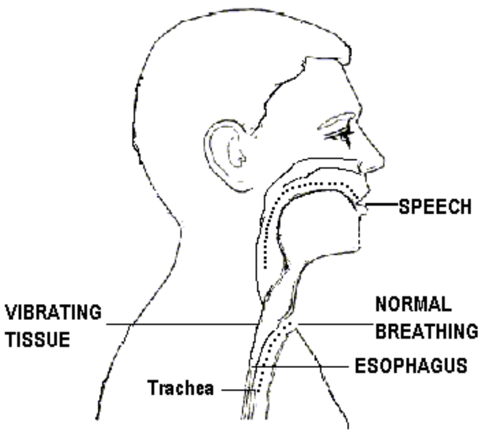


Fig. 1. Mechanism of esophagus speech **Error! Reference source not found.**].

Although the esophageal speech does not require expensive devices and prostheses, rigorous training is required to learn the art of generating this type of speech. The success rate of acquiring useful voice production is reported to be as low as 25%. Furthermore, esophageal speech results in low-pitched and low intensity speech which frequently results in poor intelligibility. On the other hand, the speech generated by artificial larynx is also associated with many shortcomings. The speech generated by this device is monotonous as the control of pitch contour is difficult to incorporate. Inefficient coupling of the device to the body results in the deficiency of low frequency. The other difficulties of the speech generated by artificial larynx are the presence of background noise and substitution of voiced segments instead of unvoiced segments. All these problems deteriorate the quality of the speech generated by this technique. In this paper three schemes for enhancing the quality of the alaryngeal speech are investigated. Various methods explored by other researchers for the enhancement of alaryngeal speech are discussed in the next section. The methodology and results are presented in the subsequent sections.

II. ENHANCEMENT OF ALARYNGEAL SPEECH

The alaryngeal speech generated by the artificial larynx is composed of two components. The first component is introduced by the filtering of the excitation through the vocal tract and the second component comes directly from the vibrator (Fig 2). The background noise produced due to the leakage of vibrations from the artificial larynx can be reduced by acoustic shielding, vibrator design, or other signal processing techniques. The mechanical improvements in the design of the artificial larynx do not provide a long lasting and efficient solution for reducing the background noise [9], [10]. The economical solution may be by employing digital signal processing for improving the quality and intelligibility of the speech. Much work is not carried out for the enhancement of alaryngeal speech. There are only few research papers available. The different algorithms can be classified into six categories. These are explained in the following sub sections.

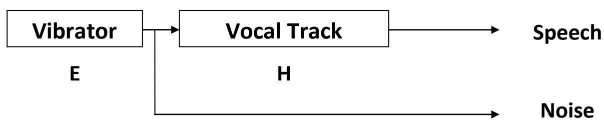


Fig 2. Signal processing model for alaryngeal speech.

SPECTRAL SUBTRACTION

In spectral subtraction [11], [12], [13] the clean speech and the noise are assumed uncorrelated, and therefore the magnitude spectrum of the noisy speech signal equals the sum of magnitude spectrum of noise and clean speech [14], [15]. In case of alaryngeal speech, speech signal and background interference are not uncorrelated. In this case, the noisy speech is given by

$$x(n) = e(n) * h_v(n) + e(n) * h_l(n) \tag{1}$$

where $e(n)$ be the excitation signal, $h_l(n)$ impulse responses of the vocal tract, and $h_l(n)$ is impulse response of the leakage path. Taking short-time Fourier transform on either side and imposing the condition that clean speech and the leakage signal are uncorrelated, we can write

$$\left|X_n(e^{j\omega})\right|^2 = \left|E_n(e^{j\omega})\right|^2 \left[\left|H_v(e^{j\omega})\right|^2 + \left|H_l(e^{j\omega})\right|^2\right] \tag{2}$$

Manoj Kumar, Parveen Lehana, Swapna Devi, and Romilla Malla Bhat, "Transformation of Alaryngeal Speech Using Line Spectral Frequencies," *International Journal of Scientific and Technical Advancements*, Volume 5, Issue 2, pp. 37-44, 2019.

The estimate of the squired magnitude of the clean speech was got by subtracting the short-time average value of the squared magnitude of the noise from the squired magnitude of the noisy speech. For synthesis, the phase of the original noisy signal can be retained. The other method may be by using minimum phase system for phase estimation. In real practice, the clean speech and the leakage signal were correlated and hence result in some of the frequency components becoming negative, causing narrow random spikes of value between zero and maximum during non-speech segment, known as residual noise. When converted back to the time domain, the residual noise sounds as sum of tone generators with random frequencies turned on and off. During speech period, this noise residual will be perceived at frequencies, which are not masked by the speech. This problem can be slightly reduced by using modified spectral subtraction method [15].

HARMONIC PLUS NOISE MODEL

Harmonic plus noise model has also been used for enhancing the alaryngeal speech [16], [17]. The alaryngeal speech and the leakage signal were analyzed using HNM and average harmonic spectrum of the leakage noise was subtracted from the harmonic magnitude spectrum of the noisy speech in each frame. HNM synthesis was carried out retaining the original phase spectra. The output was reported more natural and intelligible as compared to input speech.

AUDITORY MASKING

The auditory masking [18] approach takes into account the frequency domain masking properties of the human auditory system for a subtractive-type enhancement process. Subtractive-type algorithms can efficiently reduce the radiated noise of EL speech but not to reduce the additive noise from the environment due to the use of fixed subtraction parameters. Considering the particular characteristics of EL speech, a new computationally efficient algorithm based on the perceptual weighting technique was developed to adapt the subtraction parameters. The authors reported a significant reduction of the residual noise. Acoustic and perceptual experiments confirmed that the enhanced EL speech was more pleasant to human listeners and the proposed algorithm resulted in improved performance over classical subtractive-type algorithms.

VECTOR QUANTIZATION

In vector quantization (VQ) [19], the spectral space of an input talker was represented by discrete acoustic classes [20]. A mapping codebook that specifies the output vector of an input codeword was generated through a supervised learning procedure. Spectral conversion was accomplished by applying the mapping codebook to each input spectrum. VQ-based spectral conversion method has two major sources of error/distortion. First, the reduction of a continuous spectral space into a discrete codebook introduces quantization noise, which inevitably creates a difference between a given spectrum and its corresponding codeword in the codebook. Second, under the cepstral representation, the codewords created by the VQ process typically were the means of a set of spectral clusters, thus, have individual formant bandwidth larger than the original. In an effort to reduce quantization noise, Shikano et al. [21] proposed a fuzzy vector quantization method in which an input spectrum was coded as a weighted interpolation of a set of codewords. This weighted interpolation has the potential to reduce quantization noise because the spectral space was now approximated by many interconnected lines between codewords rather than by a point grid of codewords. The weighted interpolation, however, increases the bandwidth of the final coded spectrum.

A linear multivariate regression (LMR) approach for spectral conversion was used as an alternative to the VQ-based method [20], [22], [23]. In this approach, the spectral space of the input talker was partitioned by a few large clusters, and the spectra within each cluster were mapped linearly. The mapping matrix was obtained using procedures of least-square approximation. Because the mapping in a given region of the spectral space was continuous, the conversion distortions due to quantization and spectral averaging were minimized in a least square sense. The transitions between clusters in a connected speech, however, could be discontinuous resulting in audible clicks in the converted speech [23].

The enhancement achieved by the modified LMR-based approach was comparable to that of the modified VQ-based approach. Results of perceptual evaluations also revealed that speech conversion techniques were more effective on alaryngeal speech with articulatory deficits when comparing to enhancement achieved by voice source replacement alone.

ADAPTIVE FILTERING

An adaptive filter for noise removal can also be used for noise removal. It is based on the assumption that the clean speech and the additive noise are un-correlated. The filter has variable weights which were estimated by the difference of desired and processed noisy input. When this error is minimized, the weights were fixed and the filter becomes ready for processing the noisy input. This concept has been also used for the enhancement of alaryngeal speech [24] and appreciable improvement was reported.

PATTERN RECOGNITION

In pattern recognition [25] enhancement of the alaryngeal speech was achieved using two phases. In the first phase, the background noise was removed by using adaptive filtering, originally proposed in [26]. The preprocessed signal was filtered with a low pass filter with cutoff frequency of 900Hz for detecting the silence intervals before using second phase. If silence was detected, the segment was concatenated with the previous one to produce the output signal. If voicing was detected in the segments (30 ms), the segments were analyzed for estimating the pitch. Based on the value of the pitch, the segment was declared as voiced or unvoiced. If the segment was voiced, it was replaced with a normal segment using pattern matching based on linear prediction coefficients (LPC) [27], else, it was left unchanged.

Manoj Kumar, Parveen Lehana, Swapna Devi, and Romilla Malla Bhat, "Transformation of Alaryngeal Speech Using Line Spectral Frequencies," *International Journal of Scientific and Technical Advancements*, Volume 5, Issue 2, pp. 37-44, 2019.

For pattern identification in the alaryngeal speech, ANN with structure 12-25-8 was used to find vowel-consonant combinations. Neural networks were trained using the back propagation algorithm with 650 different voiced segments with a convergence factor equal to 0.009, achieving a mean square error of 0.1 after 400,000 iterations [28]. Once the voiced segment was identified; it was replaced by its equivalent voiced segment of normal speech stored in a codebook and concatenated with the previous segments. Spectrograms of enhanced speech signals were reported to be similar to the spectrograms of normal speech signals. Using objective and subjective evaluation, it was further reported that the proposed system provided improvement in the intelligibility as well.

III. METHODOLOGY

The investigations of the transformation functions for the enhancement of the alaryngeal speech were carried out on the cardinal vowels only. The recording was done in an acoustic room at a sampling rate of 10 k samples/s. Speech was recorded from 5 males, 5 females, and one speaker who used artificial larynx for speaking. For normal speakers, speech was recorded using both natural method of speech production and the method involving artificial larynx. The same vowels from these three recordings were aligned and cut for all the speakers.

Three methods were investigated for improving the quality of the alaryngeal speech. In the first method, the impulse response of the vocal tract is estimated from the LPC spectrum of the alaryngeal speech. The average LPC spectrum of the vowels is found and impulse response is derived by taking inverse Fourier transform. In case of alaryngeal speech, the glottal excitation is very noisy; hence, the output generated by the vocal tract is unintelligible and unnatural. It means if the noisy excitation provided by the artificial larynx is replaced by the natural excitation already recorded from a person having normal speech production system then the synthesized speech may be expected to be close to the normal one. This can be achieved by convolving the impulse response derived from alaryngeal speech with a natural glottal excitation.

In second method, the source and the target vowels are analyzed using harmonic plus noise model (HNM). These harmonic amplitudes are interpolated at the integral multiples of a constant pitch frequency (100 Hz). This provides us same number of harmonics for the source and the target in each frame. Now the average harmonic amplitudes are estimated across all the frames of the given vowel. The transformation function is estimated by dividing the corresponding harmonic amplitudes of the source and target vowels. This transformation function is used to transform the harmonic amplitudes of the source to the harmonic amplitudes of the target. The modified harmonic amplitudes are interpolated the desired harmonics of the target pitch and output is synthesized.

The third method is similar to the second method except that the transformation function is calculated by using the line spectral frequencies instead of the harmonic amplitudes. This scheme is shown in Fig 3. The source and the target vowels are analyzed using HNM and harmonic amplitudes so obtained are converted to line spectral frequencies (LSFs) of the order of 30. The relation between the source and the target in LSF domain is modeled using a polynomial of fifth degree. This polynomial represents the transformation function between the source and the target for the given vowel. After estimating the transformation function, the alaryngeal vowel is analyzed using HNM and the harmonic amplitudes are converted to LSFs and the target LSFs are estimated using the transformation function already computed. These transformed LSFs are converted back to the harmonic amplitudes. These harmonic amplitudes are interpolated at the desired harmonics of the target pitch frequency and output is generated using HNM synthesis (Fig 4). Informal listening tests were carried out for the assessment of the quality of the synthesized speech.

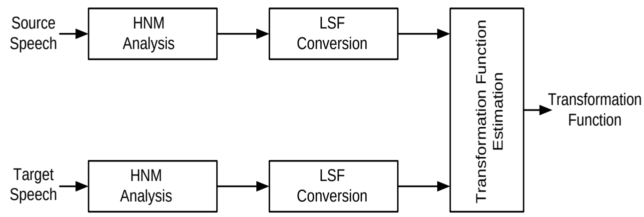


Fig. 3. Estimation of the transformation function.

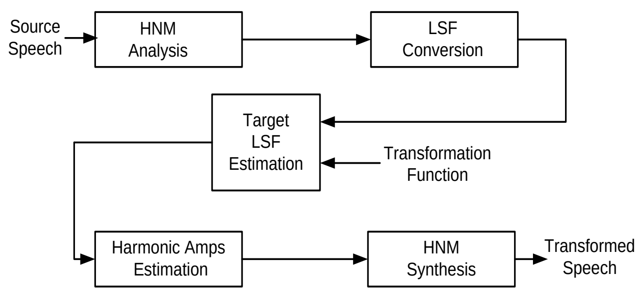


Fig. 4. Transformation of the alaryngeal speech.

IV. RESULTS

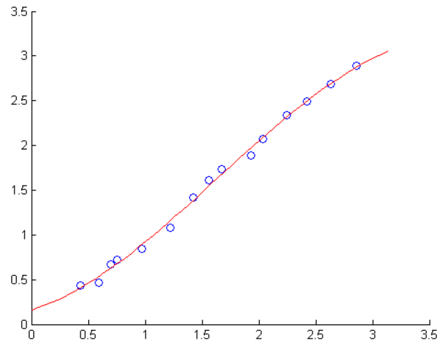
Investigations carried out by changing the shape of glottal excitation for improving the quality of the alaryngeal speech using impulse response derived from LPC spectrum of the alaryngeal vowel show that there is some improvement in the synthesized speech although the modified alaryngeal speech is slightly mechanical and buzzy. This may be due to the inaccuracy of the impulse response estimated from the corrupted alaryngeal speech due to leakage of the acoustical vibrations from the shield of the artificial larynx.

Second and third method use the mapping from the source

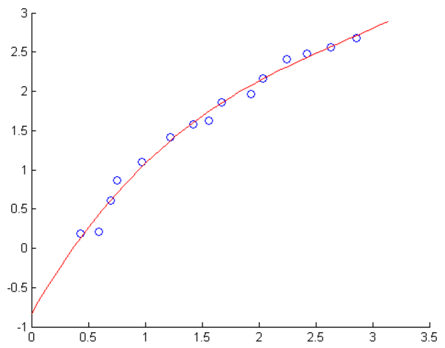
Manoj Kumar, Parveen Lehana, Swapna Devi, and Romilla Malla Bhat, "Transformation of Alaryngeal Speech Using Line Spectral Frequencies," *International Journal of Scientific and Technical Advancements*, Volume 5, Issue 2, pp. 37-44, 2019.

ISSN: 2454-1532

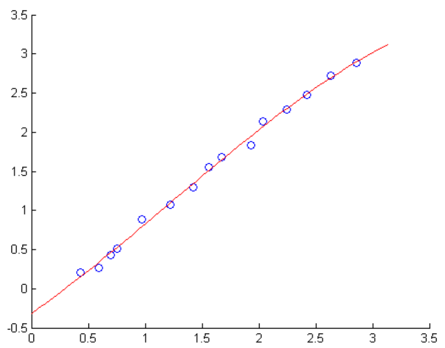
acoustic space to the target acoustic space. In the second method, the transformation function from the alaryngeal vowel to the natural vowel was estimated using harmonic amplitudes.



(a)



(b)



(c)

Fig 5. LSF based transformation functions from the alaryngeal /*a*/ to natural /*a*/ (a), alaryngeal /a/ to natural /*i*/ (b), alaryngeal /*a*/ to natural /*u*/ (c).

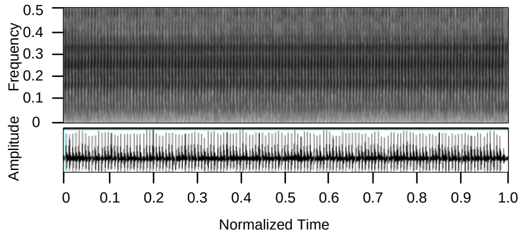
This transformation was further used to transform the alaryngeal vowels for enhancing the quality. It was observed that the quality of the synthesized output was not satisfactory. Analysis of the method showed that there were many problems associated with this scheme. The interpolation of the harmonic amplitudes does not provide satisfactory spectrum envelop at the desired target pitch frequency and hence the quality is affected. Second, the range of the amplitudes is very large. Third, the ratio of the harmonic amplitudes becomes difficult to calculate for very small amplitudes of the source speech.

It is also observed that the transformation functions for the mapping from natural to natural is very complex as compared to the mapping from the alaryngeal to natural. In other words, the mapping from the alaryngeal speech to natural is smooth. Hence, the distinct mappings for each cardinal vowel shows that the alaryngeal sound can be transformed towards natural speech using these transformation functions. Although, this enhances the naturalness and intelligibility of the alaryngeal speech but the quality is not that appealing.

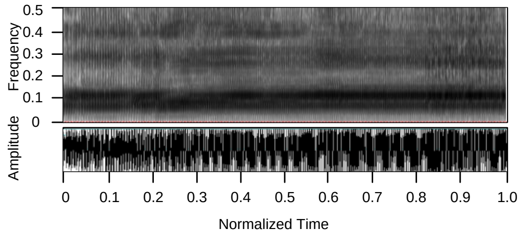
Analysis of the second method showed that there were many problems associated with this scheme. The interpolation of the harmonic amplitudes does not provide satisfactory spectrum envelop at the desired target pitch frequency and hence the quality is affected. Second, the range of the amplitudes is very large. Third, the ratio of the harmonic amplitudes becomes difficult to calculate for very small amplitudes of the source speech. To overcome these problems, the transformation function was estimated using LSFs. The harmonic amplitudes were converted to LSFs and the relation between these was approximated using fifth order polynomial. There are many advantages of the LSFs. First, the range of these is limited to 0 to pi only. Second, these can be interpolated very efficiently. Also, these do not pose any problem for division. The LSFs of the alaryngeal vowels and the corresponding natural target vowels were plotted. A linear relation was established to estimate the mapping function from alaryngeal to natural. Some of the mapping functions are shown in Fig 5. These transformation functions were used to modify the alaryngeal speech in order to improve the quality. Few spectrograms of the recorded and modified vowels are shown in Fig 6. The results obtained from the analysis of the output can be summarized as follows.

1. The transformation function is different across speakers and vowels.
2. Since the range of the LSFs is very limited hence the variation of the transformation functions is minute yet discernible.
3. The quality of the transformed vowel is best when the transformation involves the same vowel.
4. The quality of the transformed speech is independent of the source and target speaker.

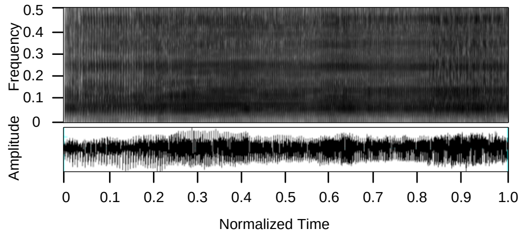
It should be noted that the above results were derived for cardinal vowels only. We are investigating LSF scheme to other speech units as well, wherein any alaryngeal sound can be transformed towards natural speech using these transformation functions. transformation function estimated from the LSFs provided best results and the quality of speech generated was almost very near to the natural speech.



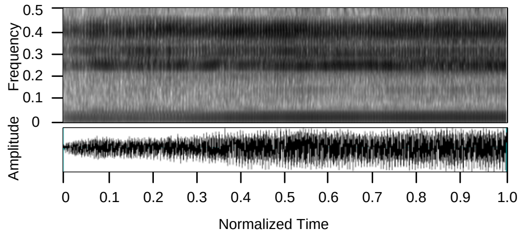
(a) Alaryngeal /*i*/ (Speaker code: pkl)



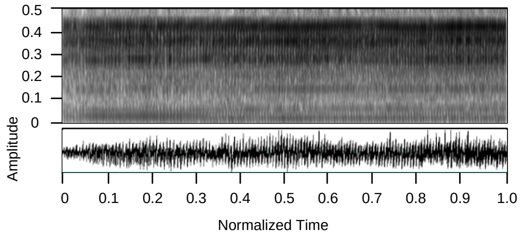
(b) Natural /*a*/ (pkl)



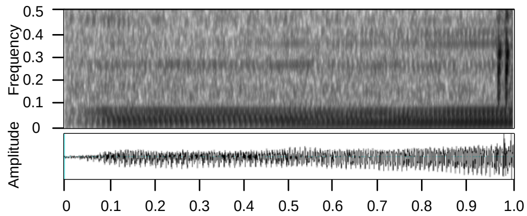
(e) Transformed alaryngeal /*i*/ to natural /*a*/



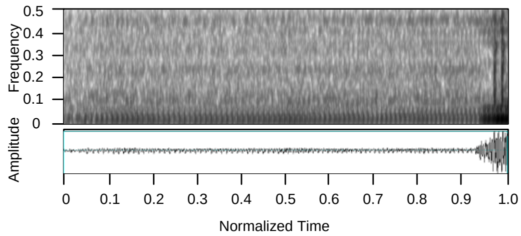
(c) Natural /*i*/ (pkl)



(f) Transformed alaryngeal /*i*/ to natural /*i*/



(d) Natural /*u*/ (pkl)



(g) Transformed alaryngeal /*i*/ to natural /*u*/

Fig 6. Spectrograms of the recorded and transformed alaryngeal vowel using LSF based transformation function.

V. CONCLUSION

Three methods were investigated for transforming the alaryngeal speech to natural speech. In the first method, the impulse response of the vocal tract was estimated from the LPC spectrum and convolved with the natural glottal excitation to generate the output. But this method did not provide satisfactory output. Second and the third method used transformation function. One method used harmonic amplitude and the other one employed LSFs. The

REFERENCES

[1] [1] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Englewood Cliffs,New Jersey: Prentice Hall, 1978.

[2] [2] Y. Lebrun, “History and development of laryngeal prosthetic devices,” The Artificial Larynx, Amsterdam: Swets and Zeitlinger, pp. 19-76, 1973.

Manoj Kumar, Parveen Lehana, Swapna Devi, and Romilla Malla Bhat, “Transformation of Alaryngeal Speech Using Line Spectral Frequencies,” *International Journal of Scientific and Technical Advancements*, Volume 5, Issue 2, pp. 37-44, 2019.

[3] [3] L. P. Goldstein, “History and development of laryngeal prosthetic devices,” Electrostatic Analysis and Enhancement of Alaryngeal Speech, pp. 137-165.
[4] [4] "Speech aids", http://jkemp.larynxlink.com/speechaids.htm, Jan 2002
[5] [5] Q. Yingyong and B. Weinberg, “Low-frequency energy deficit in electro laryngeal speech,” J. Speech and Hearing Research, vol. 34, pp.1250- 1256, 1991.
[6] [6] T. Sato, “Esophageal speech and rehabilitation of laryngeetomized,” Kanehara & Co., Ltd., Toky, 1993.
[7] [7] E. Naguchi and K. Matsuri, “An evaluation of esophagi , ,al speech enhancement,” Acoust. Soc. Jp., pp. 421-422, 1996.
[8] [8] P. K. Lehana, R. K Gupta, and S. Kumari, “Enhancement of esophagus speech using Harmonic plus noise model,” IEEE Tencon-2004, Thailand, 24-25 November, 2004.
[9] [9] C. Y. Espy-Wilson, V. R. Chari, and C.B. Huang, “Enhancement of alaryngeal speech by adaptive filtering,” in Proc. ICSLP 96, pp. 764-771, 1996.
[10] [10] "Artificial larynx with PZT ceramics," http://www. nagoya_u.ac.jp/ activity / 1999-e / VOICE_99E.html, Jan 2002.
[11] [11] P. C. Pandey, S. M. Bhatnagar, G. K. Bachher, and P. K. Lehana, “Enhancement of alaryngeal speech using spectral subtraction,” in Proc. DSP2002 (1-3 July 2002), Santorini, Greece, 591-594.
[12] [12] Hanjan Liu, Qin zhao, Mingxi Wan and Supin Wang. “Application of spectral subtraction method on enhancement of electro larynx speech” J. Acoust. Soc. Am., vol. 120, no. 1, pp. 398-406, 2006.
[13] [13] S. S. Pratapwar, P. C. Pandey, and P. K. Lehana, “Reduction of background noise in alaryngeal speech using spectral subtraction with quantile based noise estimation,” in Proc. of 7th World Multiconference on Systemics, Cybernetics and Informatics SCI, (Orlando, USA, 2003).
[14] [14] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE ASSP-27, pp. 113-120, 1979.
[15] [15] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” Proc. ICASSP, pp. 208-211, 1979.
[16] [16] P. K. Lehana and P. C. Pandey, “Speech enhancement during analysis-synthesis by harmonic plus noise model,” J. Acoust. Soc. Am., vol. 120, pp. 3039, 2006
[17] [17] Gidda Reddy, Prem C. Pandey, and Parveen, “Application of harmonic plus noise model for enhancement of speaker recognition,” J. Acoust. Soc. Am., vol. 120, pp. 3040, 2006.
[18] [18] Hanjan Liu, Qin zhao, Mingxi Wan,Member IEEE and Supin Wang. “Enhancement of electrolarynx speech based on auditory masking,” IEEE Tran. Bio. Eng., vol. 53, no. 5, pp. 865-874, 2006.
[19] [19] B. Ning and Q. Yingyong, “Application of speech conversion to alaryngeal speech enhancement,” IEEE Tran. Speech and audio Proc., vol. 5, no. 2, pp. 97-105, 1997.
[20] [20] Y. Linde, A. Buzo, and R. Gray, “An algorithm for vector quantizer design,” IEEE Trans. Commun., vol. 28, pp. 84–95, 1980.
[21] [21] K. Shikano, S. Nnakamura, and M. Abe,”speaker adaption and voice conversion by codebook mapping,” in Proc IEEE Int. Symp. Circuits and Systems, 1991, vol. 1, pp. 594-597.
[22] [22] L. Rabiner, S. Levinson, and M. Sondhi, “On the application of vector quantization and hidden Markov models to speaker-independent, isolated word recognition,” Bell System Tech. J., vol. 62, pp. 1075–1105,1983.
[23] [23] H. Valbret, E. Moulines, and J. Tubach, “Voice transformation using PSOLA technique,” speech Commun., vol. 11, pp. 175–187, 1992.
[24] [24] C. Y. Espy-wilson, V. R Chari, C. B Huang, “Enhancement of alaryngal speech by adaptive filtering,” ECS Engineering Department, Borton University, Borton, MA02215.
[25] [25] A. T. Gualberto, N. M. Marico, and P. M. Hector “Enhancement and restoration of alaryngeal speech signals,” in Proc. 16th IEEE Conference of Electronics, Communication and Computer, vol. 9, no. 6, pp. 207695-2505-9/06, 2006.
[26] [26] G. Aguilar-Torres, M. Nakano-Miyatake, and H. Perez-Meana, “Alaryngeal speech enhancement using pattern recognition techniques,” IEICE Transactions on Information and Systems, Vol. E88-D, No. 7, 2005.
[27] [27] J. Markel and A. H. Gray Jr., "Linear prediction of speech," Springer-Verlag, Berlin Heidelberg, New-York,1976.
[28] S. Haykin, “Neural networks: A comprehensive foundation,” Prentice Hall, Englewood Cliffs NJ, 1994.

Manoj Kumar, Parveen Lehana, Swapna Devi, and Romilla Malla Bhat, “Transformation of Alaryngeal Speech Using Line Spectral Frequencies,” *International Journal of Scientific and Technical Advancements*, Volume 5, Issue 2, pp. 37-44, 2019.