

Esophageal Cancer Recognition using an Artificial Neural Networks

Anichur Rahman
Department of Computer
Science and Engineering
Mawlana Bhashani
Science and Technology
University, Santosh,
Tangail, Bangladesh

Saibur Rahman
Department of Computer
Science and Engineering
Mawlana Bhashani
Science and Technology
University, Santosh,
Tangail, Bangladesh

Md. Tarequl Islam
Senior Lecturer
Department of Computer
Science and Engineering
Khwaja Yunus Ali
University, Enayetpur,
Sirajganj, Bangladesh

Mohammad Motiur
Rahman
Professor, Department
of Computer Science
and Engineering
Mawlana Bhashani
Science and Technology
University, Santosh,
Tangail, Bangladesh

ABSTRACT

Esophageal cancer patients don't have ability to speak usually because of absence of their vocal chords which have been detached. While talking every individual patient make different vocal sound from each other. Esophageal speech recognition is one of the major applications that can be incorporated in supermarkets in case of identifying esophageal speech from the vocal sounds which have been imported from different persons. Meanwhile, the speech recognition technology has been improving rapidly. However, to date, the esophageal speech recognition technology has been developing to identify esophageal abnormality appropriately. This research work describes a system for esophageal speech recognition. The key part of the speech recognition system is speech extraction, feature extraction and recognition of esophageal speech. The Artificial Neural Network (ANN) has been used to extract the feature and characteristic of esophageal speech. The proposed system can recognize the esophageal speech nearly up to 97 % acceptably. From the experimental results, it can be concluded that the proposed system is better than any other recent proposed methods.

Keywords

Esophageal speech, Esophageal Cancer, Speech Recognition, Artificial Neural Network (ANN), Feature Extraction, Classification.

1. INTRODUCTION

Esophageal individual person produces esophageal vocal sound which is also known as esophageal voice, is a process of speech generation that involves oscillation of the esophagus. This contrasts with traditional laryngeal speech which involves oscillation of the vocal folds. Instead, air is injected into the upper esophagus and then released in a controlled manner to produce sound used to generate speech. The Esophageal cancer patient does not have ability to speak normally because of the nonexistence of their vocal chords. The simplest way for the cancer patient to produce speech is esophageal speech. Although, the procedure of recognizing speech has been updated gradually. On the other hand, the voice recognition technology also has been cast-off to identify esophageal speech adequately, whatsoever, the esophageal speech recognition technology must be developed.

Nowadays research on the focus of esophageal speech recognition is growing swiftly. A lot of applications of

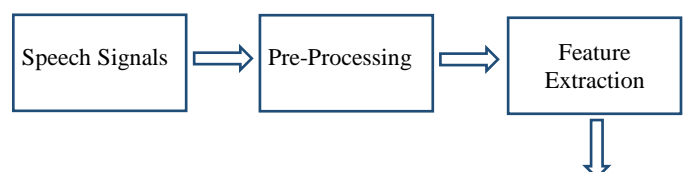
esophageal speech recognition are introduced. We Believe that this research technology will be more appropriate to identify esophageal speech.

There are several methodologies to improve the quality of alaryngeal speech, esophageal speech and ALT speech. Daub 4 Wavelet, Haar Wavelet, Mexican hat Wavelet, Morlet Wavelet, HMM are used as such kinds of approaches. To reduce the background noise of ALT speech, several adaptive filter based speech enhancement algorithms have been proposed [1], [2]. These algorithms improve the intelligibility, considering less the quality besides reducing background noise of ALT produced speech. An analysis-synthesis-based method is used to improve the alaryngeal speech quality [3].

This paper proposes an esophageal speech enhancement system based on feature extraction method using Artificial Neural Network (ANN). The core part of this system of esophageal speech recognitions are extract voice first, then extract feature and lastly classify and recognize speech. The Artificial Neural Network –ANN is used to extract the feature and characteristic of esophageal speech. In this system, firstly extract the normal voice signals and the esophageal voice signals to reduce the background noise. Later on, the signal is per-processed and then extract the statistical feature. Finally, the system matches the statistical values for usual and esophageal speech with target values. Depends on the statistical values of that particular usual speech and the esophageal speech, system can identify the nature of speech by matching corresponding statistical values with system target values.

2. METHOD

The core parts of our system for esophageal speech recognition are speech extraction, feature extraction and recognition of esophageal speech. The features used in the esophageal speech recognition can be categorized by statistical features. The steps involved in the esophageal speech recognition are explained below and also shown diagrammatically in figure 1.



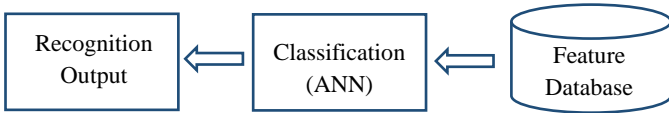


Figure 1: Framework for the Proposed Method

2.1 SPEECH DATA COLLECTION

In this step the speech of a person is received in a waveform.

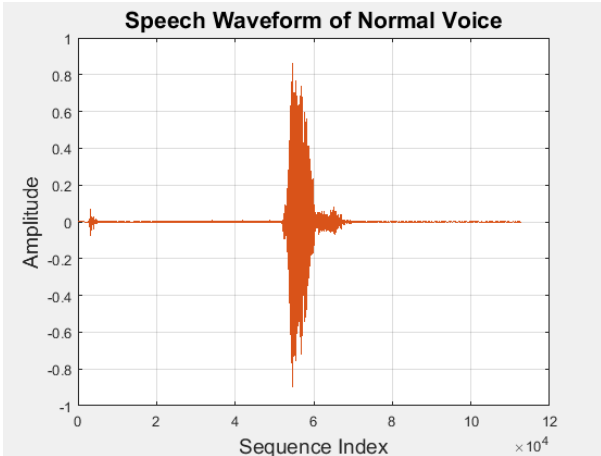


Figure 2 (a): Normal speech (Sample - 1)

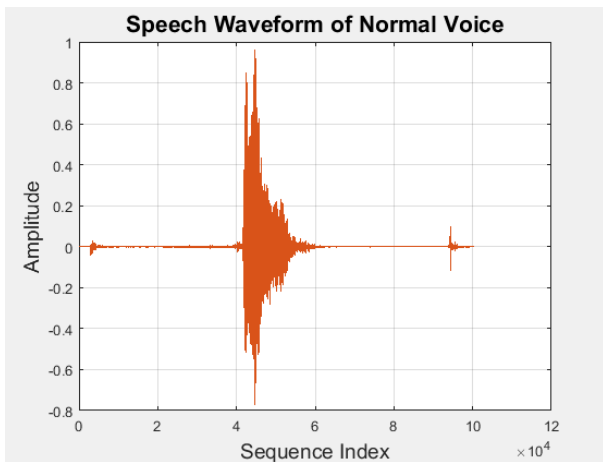


Figure 2 (b): Normal speech (Sample - 2)

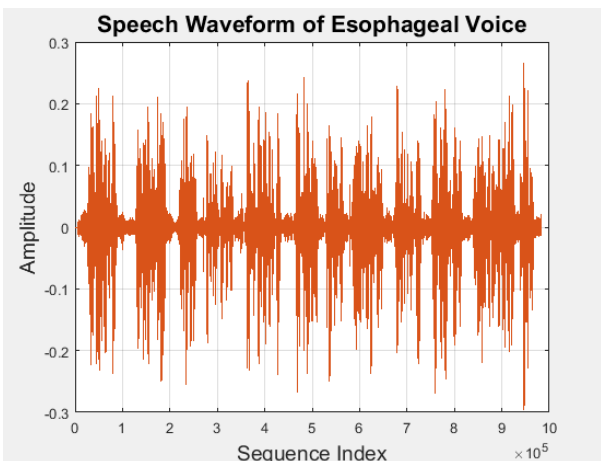


Figure 3 (a): Esophageal Speech (Sample - 1)

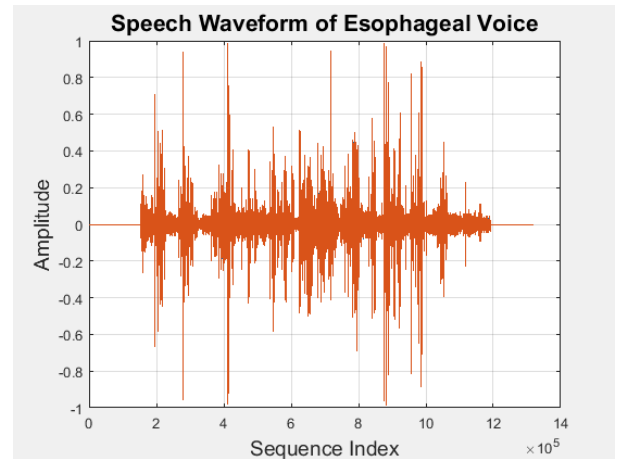


Figure 3 (b): Esophageal Speech (Sample -2)

The first stage consists of recording a speech file, from an esophageal speaker and general speaker. The audio data is kept as a WAV file. The speeches are normal speech and esophageal speech. We used 100 normal speeches and 110 esophageal speeches for sample data speech. The plot diagram for normal speech and esophageal speech have been presented in figure 2 and figure 3 respectively. Many types of software are available by which the speech of a person can be recorded [8].

2.2 SPEECH DATA PRE-PROCESSING

Pre-processing plays a very important role to process of speech signal, especially when the background noise is completely undesirable. This step is necessary to make sure that the signal is less vulnerable to noise. Pre-processing of speech signal serves various purposes in any speech processing application. It improves the readability and accuracy of automatic speech recognition. Pre-processing generally involves endpoint detection, pre-emphasis filtering, noise filtering, smoothing, framing, windowing, echo cancelling and channel etc.

2.3 FEATURE EXTRACTION

The speech produced of every individual person differs from each other. This is because the speech of every individual has different individual characteristics embedded in utterances. Therefore, we use feature extraction process to separate one speech from the other person. Speech recognition or front end analysis are the other names of feature extraction.

$$Mean = \frac{\sum_{i=1}^n X_i}{n}$$

$$Standard\ Deviation\ (SD) = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$Variance = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

$$\text{Average Power, } P = \frac{1}{N} \sum_{i=1}^n X_i^2$$

$$\text{Avg. Magnitude } X_{pwr}(m) = |X(m)|^2$$

$$\text{Zero - crossing, } zcr = \frac{1}{N-1} \sum_{i=1}^{N-1} I \{S_i S_{i-1} < 0\}$$

Feature	Normal Speech			
	Sample-1	Sample-2	Sample-3	Sample-4
Mean	-0.0001	-0.0001	-0.00011	-0.0001
SD	0.060641	0.066542	0.050693	0.108929
Variance	0.003677	0.004428	0.002577	0.011865
Avg. PW.	0.003677	0.004428	0.002577	0.011865
Avg. Mg.	0.013215	0.01777	0.011815	0.036661
Zero Crs.	5443	5893	6586	4862

Table-1: Experimental speech data set for normal voice

Feature	Esophageal Speech			
	Sample-1	Sample-2	Sample-3	Sample-4
Mean	-0.00785	0.00024	-0.00125	0.00019
SD	0.096059	0.050354	0.030934	0.033809
Variance	0.009227	0.002535	0.000957	0.001143
Avg. PW.	0.009289	0.002535	0.000957	0.001143
Avg. Mg.	0.052983	0.025779	0.018419	0.017543
Zero Crs.	5642	90971	93401	113482

Table-2: Experimental speech data set for esophageal voice

In this system we extracted feature by using Mean, Standard Deviation, Variance, Average power, Zero-crossing, Average Magnitude. Table-1 and Table-2 depicts the experimental data set for both normal and esophageal speech. By using this data set we learnt the system. This extraction procedure can be used to extract different individual characteristics in speech data. Applications like Speech and Speaker Recognition needs efficient feature extraction techniques from speech signal where most of the voiced part contains speech or speaker specific attributes.

2.4 SPEECH CLASSIFICATION

Using Artificial Neural Network in our method the system can classify the normal speech and esophageal speech. In this method the system tested 100 normal speeches and 110 different esophageal voice segments. We identified the target factor based on the test data set used while learning the system. Based on that particular target factor normal speech and esophageal speech are classified.

3. RESULT

The feature value which is generated from the statistical process will be compared to identify the speech type.

In our proposed method, we can simply distinguish the esophageal speech from normal speech by using the target factor value used in the system. By comparing the target value with the statistical feature value generated we can recognize the types of speech whether it is esophageal or normal.

In this method, we have taken 100 samples of normal speech and 110 esophageal speeches to learn system. After learning, our proposed system exhibits that it can be able to recognize speech about 97% accurately. Compared to the other methods we can say that our system is much better to recognize the esophageal speeches. In table 1, the observation for Classification and accuracy for esophageal speech recognition with other methods have been represented.

Method Name	No. of Sample	Input Type	Sample Type	Accuracy
Daub4 wavelet	160	Audio voice	Normal & Esophageal	75%
Morlet wavelet	450	Audio voice	Normal & Esophageal	89%
HMM	250	Audio voice	Normal & Esophageal	93%
ANN	200	Audio voice	Normal & Esophageal	95%
Proposed Method	210	Audio voice	Normal & Esophageal	97%

Table 3: Classification and accuracy for esophageal speech recognition with other methods.

4. FUTURE WORK

The number of input speech was not too high therefore accuracy may differ. Even the best speech recognition systems sometimes make errors. If there is noise or some other sound in the room, the number of errors will increase. Speech Recognition works best if the microphone is close to the user. In future we wish to dream up to improve our system so that it can work in very complex situation by using large number of data set of speech which can improve the accuracy more.

5. CONCLUSION

Esophageal Speech recognition is one of the greatest significance to many of the applications. In this research for esophageal speech recognition, Artificial Neural Network has been used. Two types of speech have been taken as sample data such as normal speech and esophageal speech. The proposed method is composed of three steps. Firstly, collected sample data has been per-processed and statistical features are extracted next from the per-processed segmented speech. Finally, speeches are classified using Artificial Neural network (ANN) whether to identify the patient is having esophageal cancer or not. The proposed system can recognize speech correctly up to 97% and recognize the esophageal speech. The result of testing shows that the proposed esophageal speech recognition system is superior than recent other methods published. This research allows rewarding with great experience, arise huge confident in our mind, meet with brief ocean of knowledge. We believe that this prototype might help the people in health care purpose.

6. REFERENCES

- [1] Espy-Wilson, C., Chari V. & Huang C., Enhancement of alaryngeal speech by adaptive filtering, Technical report, Boston University, Boston, MA, 2000.
- [2] Becerril H., Nakano-Miyatake M. & Perez-Meana H., Development of an adaptive system for voice enhancement in persons with artificial larynx using DSP, *Cientifica*, Vol. 8, No. 2, April 2004, pp. 12-20.
- [3] K. Matsui and N. Hara, Enhancement of esophageal speech using format synthesis, *IEEE International Conference on Acoustic, Speech and Signal processing*, Vol. 1, 1999, pp. 81-84.
- [4] Magdalena Marlin Amanda, application of voice recognition in cryptograph of public key, Tugas AKhir Petr
- [5] Achmad Basuki, Miftahul Huda, Tria Silvie Amalin, application of voice recognition in musical request, Jurusan Teknik Telekomunikasi Politeknik Elektronika Negeri Surabaya.
- [6] Ajub Ajulian Z., Achmad Hidayatno, Muhammad Widyanto Tri Saksono, application of voice recognition in car controller.
- [7] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, *Senior Member, IEEE*, "Neural Networks used for Speech Recognition" *JOURNAL OF AUTOMATIC CONTROL, UNIVERSITY OF BELGRADE, VOL. 20:1-7, 2010*©