ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

CLASSIFICATION OF PATHOLOGICAL VOICE INTO NORMAL/BENIGN/MALIGNANT STATE

Cheol-Woo Jo, Dae-Hyun Kim

Department of Control & Instrumentation Engineering
Changwon National University
Changwon, Kyeongnam, 641-773 KOREA
cwjo@sarim.changwon.ac.kr

10.21437/Eurospeech.1999-147

ABSTRACT

In this paper we propose a new method to classify the pathological voice into normal, benign and malignant cases. New parameter is proposed to discriminate each class. New parameter is based on cepstral analysis technique. Pathological speech signal is collected at the hospital. Normal speech signal is also contained at the same database and analyzed as well. Then the results are compared to find the differences between normal and pathological speech. Source components are separated using cepstrum after obtaining residual signal from speech. Then the ratios between harmonic components and noise components are obtained from the original signal and residual signal. Finally a neural network is used to train and classify normal, benign and malignant states of speech.

Keywords: pathological, cepstrum, classification

1. INTRODUCTION

These days the importance of human health is increasing more and more. And there is needs for distinguishing between normal and pathological speech signals for diagnostic purpose. Generally medical doctors are using special apparatus, similar to endoscope, to diagnose the mal-functioning vocal cords. Using such apparatus, doctors can see vocal cords directly. But there are needs to diagnose vocal cords by only speech signal. And there have been many researches about such applications.[1][2]

About source separation many methods has been suggested. Well known linear prediction method can provide basic idea of source separation.[3] In this paper, based on the fact that pathological speech can have more noise components than normal speech in their estimated source, we suggest a method to discriminate between normal and pathological speech using cepstral analysis. And it is compared with other conventional parameters.

2. CHARACTERISTICS OF PATHOLOGICAL SPEECH

There are many diseases which can be happened on vocal cord. Some kind of diseases do not affect the quality of speech signal. But also there are some which do affect. Human listeners, doctors or even plain peoples, can distinguish the differences between normal and abnormal speech. Generally some conceptual measures are used for such diagnosis. For example, husky, creaky, etc. As a numerical parameter, which can measure the variations of vocal sources, jitter and shimmer are most widely used ones. But these two parameters require pitch information to be computed. And extracting pitch is sometimes difficult. Especially for pathological voices, pitch on voice is often irregular. So this affects the performance of the pitch extractor and accordingly this affects the jitter and shimmer leading wrong decision. Generally pathological speech has more higher frequency noise and more irregular pitch.

In this paper, various vocal diseases including vocalfold cancer are used. That is because our main aim is an early finding of cancer and prevention of them.

3. CLASSIFICATION PARAMETERS

As parameters for classification, in addition to the conventional jitter and shimmer, we propose a new parameter.

Cepstral analysis is a well known technique to analyse speech signals and acoustic signals. We can use cepstrum to measure the harmonic components of rotating machines, and to measure the pitch and vocal tract characteristics of speech signal. Equation (1) is the definition of the cepstrum of signal x(t).

$$c(t) = idft(20 \cdot \log(|dft(x(t))|) \quad (1)$$

Here dft means discrete fourier transform, idft means inverse discrete fourier transform.

We computed cepstrum from the residual signal of pathological speech. From the result of the cepstrum, we separated pitch components and vocal tract components from the noise components. Actually residual signal has little vocal tract components but to reduce the remaining effect of vocal tract from the residual signal, both vocal tract component and pith components are removed together. Figure1 shows how the noise component is separated from the cepstrum.[3]

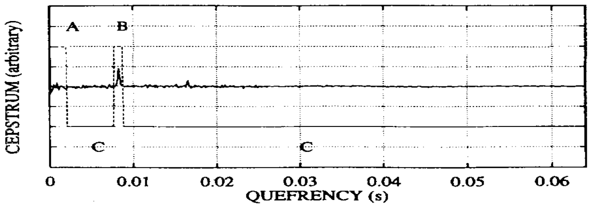
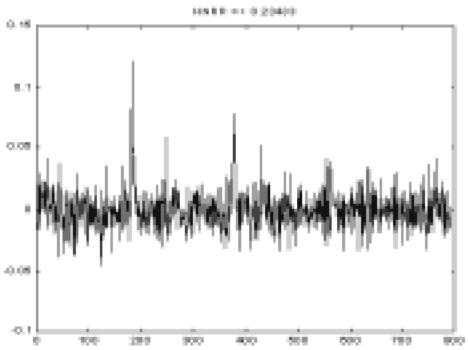
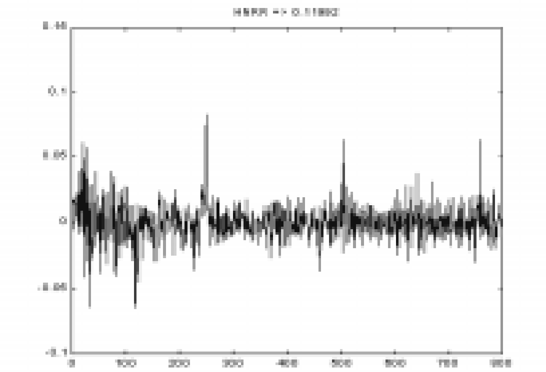


Figure 1. Extraction of noise components from cepstrum[3]

Figure 2 shows two results of cepstrual analysis from normal and abnormal cases.



(a) Cepstrum for normal speech



(b) Cepstrum for abnormal speech

Figure 2. Results from cepstral analysis

From the separated source signals we designed new parameters based on the harmonics-to-noise ratio.[4] We named it as HNRR(Harmonic to Noise Ratio from Residuals). Separate from this we applied the same process to the original speech , not residual speech. Then we obtained harmonics-to-noise ratio from the cepstrum. In this case cepstrum includes all the vocal tract components. We named it as HNRS(Harmonic to Noise Ratio from Speech).

$$HNRR = \frac{A+B}{C} \text{ for LPC residual signal} \quad (1)$$

$$HNRS = \frac{A+B}{C} \text{ for original signal} \quad (2)$$

From the process of computing these two parameters we estimate the characteristics of these two parameters as follows. HNRR represents the ratio of noise components and harmonic components from the voice source. HNRS represents the ratio of noise components and harmonic components. Generally pathological speech have more noisy components and more irregular pitch. So in case of pathological speech residual signal from the linear prediction can show much less pitch components than in case of normal speech. Accordingly cepstrum of the residual signal will show less pitch components (i.e. less harmonic components). From this fact , we can assume like this. HNRR will show smaller value in pathological speech and bigger value in normal speech. From the same reason we can assume the HNRS will show the similar tendency as in HNRR. But pitch component will be much bigger in normal speech compared to the pathological speech because less explicit pitch components are expected from the residual signal. From these points we can assume these parameters to be effective parameters which shows the variation of source characteristics.

Another parameter NHR is currently being used by some diagnostic tools,. This parameter measured the harmonic components at the bandwidth between 70 Hz and 4500Hz and noise components between 1500 Hz and 4500Hz and obtained the ratio from the result of FFT. So NHR is different from HNRR and HNRS in that these parameters use total bandwidth.

Summarizing the characteristics of suggested parameters, At normal speech HNRR will be bigger than HNRS because residual signals can emphasize the pitch components. But in pathological speech added noise component decreases harmonic components and it is expected that HNRR will be smaller than HNRS. But these two parameters show similar characteristics. So we used only HNRR for later experiment.

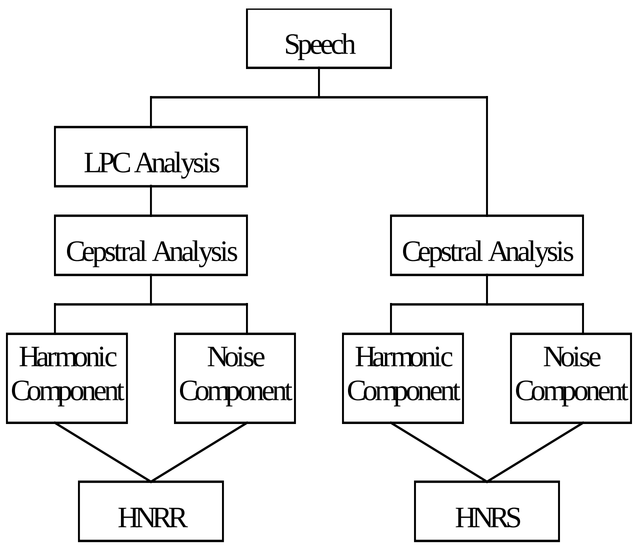


Figure 3. Computation of HNRR, HNRS

4. EXPERIMENTS AND DISCUSSIONS

We computed HNRR and HNRS using the pathological speech database which were collected at the ENT department of hospital. In this database total 66 normal and pathological speech are included. It consists of 33 normal voices, 17 benign cases and 16 malignant cases. Malignant cases are all cancer cases. Jitter, shimmer and other 31 parameters are computed from MDVP software from CSL of KAY. All data are /a/ sound sampled with the sampling rate 16KHz and with 16bit resolution. Sampled signals are stored in DAT and again converted into files using CSL.

To analyse pathological speech, we first compute jitter and shimmer using MDVP. The HNRR is computed separately.

To classify the speech 3 layered neural network is used. Due to the complex distrubution of parameters, classification between normal and abnormal cases are first performed. Then ,from the result ,abnormal cases are classified into benign and malignant case.

Figure 4 shows the basic structures of 3 layered neural network used. Figure 5 shows overall building blocks of the neural network. To find the optimal parameters, we grouped three parameters in 4 different combinations as follows, jitter-shimmer, shimmer-HNRR, jitter-HNRR, jitter-shimmer-HNRR. And experiments are performed for each parameter set per each stage of neural network. 2/3 of data set is used to train the neural networks. And the remaining 1/3 of data is used to test the neural network.

Table 1 shows the results from the experiments about first stage. For the data set , which was used to train the neural network, the correct classification ratio is the highest when all three parameters are used together. The ratio is 88.64%. Also the trained neural network showed 90.1% correct classification for the test data set. Accordingly the combination of three parameters can be considered as a best one.

From the second stage neural network, HNRR-Jitter combination showed best classification. At the table 2 100% ratio is not believable because of the small data set size. Excluding that the best combination is HNRR-Jitter.

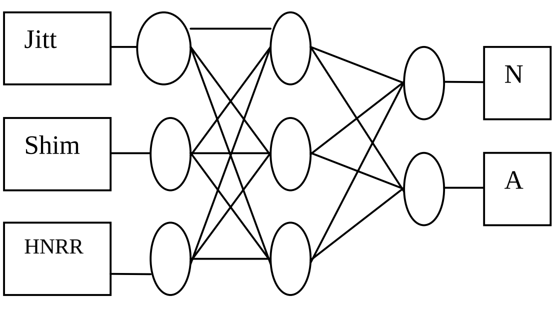


Figure 4. 3 –layered neural network for one stage

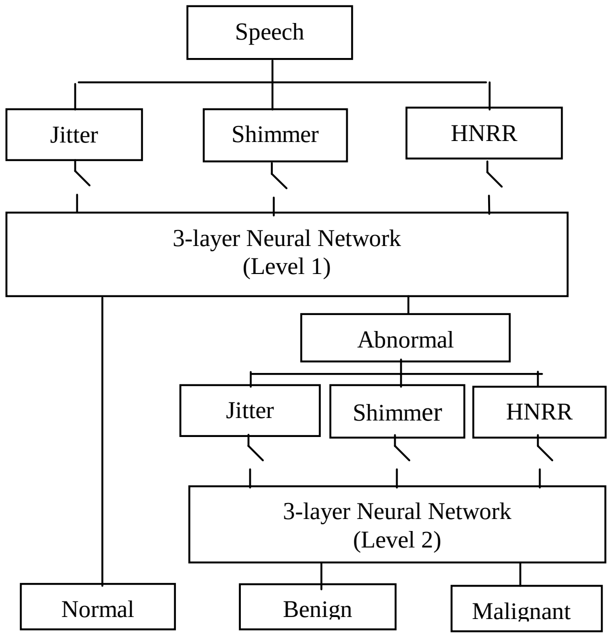


Figure 5. Overall neural network structure

Results from the neural networks show good classification between normal, benign and malignant state of pathological speech. Overal average correct classification rate is approximately 90% when jitter-shimmer-HNRR is chosen at the stage 1 and HNRR-jitter is chosen at the stage 2 at the same time. This rate is comparable to human expert's. At the preliminary experiments using single neural networks, it was not possible to distinguish the three cases at such high correctness. So using second level neural networks, it was possible to classify pathological speech into three classes with good correctness.

5. CONCLUSION

In this paper we proposed two parameters, which can discriminate pathological voices from unknown speech. Based on cepstral analysis on the residual signal and original speech new parameters HNRR and HNRS are suggested. Also using two 3 layered neural networks we classified pathological speech into normal, benign and malignant state. And it is shown that by using neural network, we can classify speech signal into normal, benign and malignant state. The result can be used usefully to pre-screening the possible vocal folds patients by only measuring speech signal. In that way this method can contribute to the diagnosis process of vocal folds.

Further research directions relating this research is as follows. Collection of more pathological speech is required to make the reliability of the result higher. Currently the size of database is still too small to extract general conclusion about a parameter. Next, more speech database from Korean patients and normal peoples should be collected to apply this clinically to the patients. The collection process is going on now.

Table 1. Correct classification rates at stage 1

		Jitt-Shim		Shim-HNRR		HNRR-Jitt		Jitt-Shim-HNRR	
		N	A	N	A	N	A	N	A
Train Data	N	21	1	19	3	18	4	21	1
	A	6	16	2	20	3	19	4	18
Ratio		84.09%(7/44)		88.64%(5/44)		84.09%(7/44)		88.64%(5/44)	
Test Data	N	11	.	9	2	9	2	10	1
	A	1	10	1	10	2	9	1	10
Ratio		95.45%(1/22)		86.36%(3/22)		81.81%(4/22)		90.9%(2/22)	

Table 2. Correct classification rates at stage 2

		Jitt-Shim		Shim-HNRR		HNRR-Jitt		Jitt-Shim-HNRR	
		N	A	N	A	N	A	N	A
Train Data	N	11	.	9	2	9	2	11	.
	A	.	11	0	11	.	11	.	11
Ratio		100%(0/22)		90.9%(2/22)		90.9%(2/22)		100%(0/22)	
Test Data	N	6	.	6	.	5	1	6	.
	A	3	3	3	3	.	6	6	.
Ratio		75%(3/12)		75%(3/12)		91.67%(1/12)		50%(6/12)	

ACKNOWLEDGEMENT

This paper is a partial result from the research funded by Korea Research Foundation. The authors wish to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997.

REFERENCES

[1] F.Plante, H.Kessler, B.Cheetham, J.Earis, "Speech Monitoring of Infective Laryngitis", Proceedings of ICSLP'96,pp.749-752, Philadelpia, 1966

[2] M.N.Vieira,F.R.McInnes,M.A.Jack,"Robust F0 and Jitter Estimation in Pathological Voices", pp.745-748, Proc. Of ICSLP'96, Philadelpia, 1996

[3] B.Yegnanarayana, C.d'Alessandro,V.Darsinos,"An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components", pp.1-10,IEEE Trans. On Speech and Audio Processing, Vol.6, No.1, Jan. 1998

[4] Cheol-Woo Jo, Dae-Hyun Kim,"On the Classification of the Pathological Speech", pp.388-391, Proceedings of the 15th Korean Speech Communication and Signal Processing Workshop, Aug. 1998