Classification of Pathological Voice Using Artigicial Neural Network with Normalized Parameters

Tao Li · Il-suh Bak · Cheolwoo Jo*

ABSTRACT

In this paper we examined the effect of normalization on discriminating the pathological voice into normal and abnormal classes using artificial neural network. Average values per each parameter were used to normalize each set of parameter values. Artificial neural networks were used as classifiers. And the effect of normalization was evaluated by comparing the discrimination results between original and normalized parameter sets.

Keywords : normalization, pathological, discrimination, neural network

I. Introduction

These days there are many attempts to analyze and discriminate the pathological and normal voice by the original parameters (Jitter, Shimmer, NHR, SPI, etc.). The major purpose of such researches is to obtain some good standards and methods to classify and diagnose the patients who have diseases on their vocal folds. [1] [2] [3] [4] [5] [6]

Even though there are some previous researches about discrimination of pathological voice, those only utilize original parameters' values as the data. Also artificial neural networks have been widely used as a classifier because of random and complex characteristics of the pathological voice parameters. But the differences of the ranges of values among these parameters is very large. When bigger values and relatively much smaller values are input into the network for training at the same time, the effect of the parameters with the different magnitudes is not checked yet.

In this paper we suggest a normalization method to scale each parameter group's

* SASPL, School of Mechatronics, Changwon National University, Korea

values and measure the effect of normalization using the classification rate from the artificial neural networks.

II. Data Collection

To collect original voice data, a collection system was installed in a room of the ENT department of hospital. The recording process was performed semi-automatically with the intervention of operator to control the quality and procedure. Also the voice materials from the same speaker were collected using DAT and CSL. [7] [8] The sampling rate was 50 KHz and the resolution 16 bits. The collection was conducted in a hospital soundproof room. All the subjects were asked to pronounce /a/. The patients' ages ranged between 23 and 75. Total voice data included 41 normal cases (33 males and 8 females)and 59 pathological cases (43 males and 16 females) after removing invalid data from the raw data sets. The vocal diseases considered consisted of Vocal Polyposis, Hyperadduction, Vocal Cord Palsy, Vocal Nodule and Glottic Cancer, etc. The parameters used were Jitter, Shimmer, NHR (Noise-to-Harmonic Ratio), SPI (Soft Phonation Index), APQ (Amplitude Perturbation Quotient) and RAP (Relative Average Perturbation). They were the 6 different kinds of parameters. [3]

III. Normalization

It is known that the units and magnitude ranges of the parameters Jitter, Shimmer, NHR, SPI, APQ, RAP, STD etc. are different. For example, Jitter is a percentage value but STD's unit is in Hz. And Shimmer's magnitude is much bigger than that of NHR. In the above, the measured parameter is 30.659 in one case of Shimmer, but 0.1296 in another case of NHR. As seen, there is great difference between these two parameters' magnitudes. When using an artificial neural network as a classifier with different input parameters, parameters with bigger value range may affect the classification rate more than those with smaller one.

Now in order to let these different parameters have the similar magnitude range, we normalized the 100 original measured values (41 are the normal data and 59 the pathological ones) for each parameter respectively (there are 6 kinds of parameters). Then we tried to observe the improvement of the classification rate with the normalized values comparing to that with the original values. By doing so, the effect of normalization can be measured and also we can measure how much each parameter affects the classification result under normalization.

Equation (1) and (2) shows how it is performed.

We then obtain

$$M_q = \frac{\sum_{i=1}^{K} P_i + \sum_{j=1}^{L} P_j}{K+L} \tag{1}$$

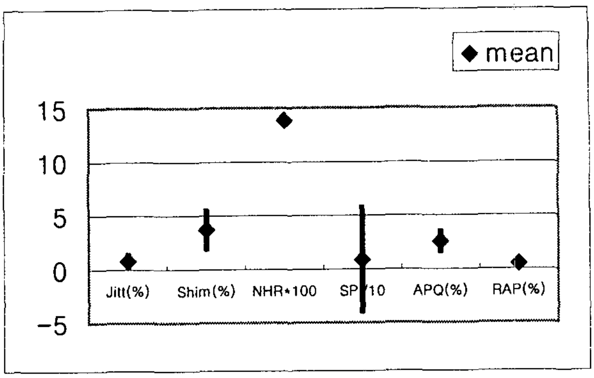
where P_i, $1 \le i \le K$, P_j, $1 \le j \le L$ are the original measured values of the normal and the pathological cases for parameter q respectively, K and $L$are the number of normal and pathological parameters respectively (K=41 and L=59 in this paper), and M_q is the mean value of the parameter q.

$$P_{nq} = \frac{P_o}{M_q} \tag{2}$$

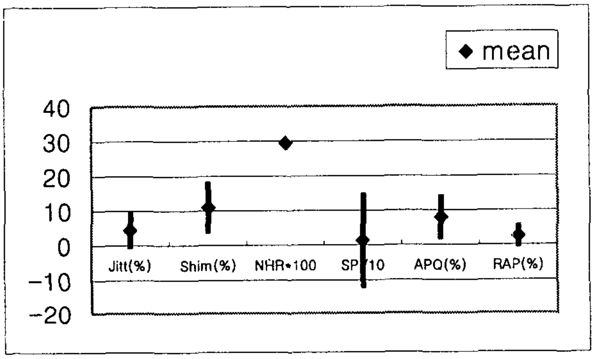
where P_o and P_{nq} are the original measured value and the corresponding normalized value for the specific parameter respectively. Then the normalized value for each parameter can be obtained by the equations (1) and (2).

After analyzing the collected voice materials using the analyzer and the above normalization method, we obtained the 6 different kinds of parameters (Jitter, Shimmer, NHR, SPI, APQ and RAP) which had the original measured values and especially the corresponding normalized values in this paper. And there were 100 original measured values and 100 corresponding normalized values for any kind of parameter. Also the 6 different kinds of parameters were divided into 3 categories according to their characteristics. They were pitch related (Jitter, RAP), amplitude related (Shimmer, APQ) and noise related (NHR, SPI). [6] [7]

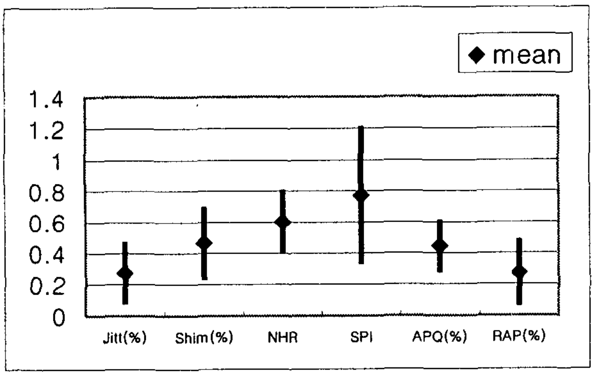
Fig.1 shows the comparisons of the original and normalized parameters from the normal and pathological voices. The graphs show the relative change of each parameter when DAT parameters are considered 1. [7]



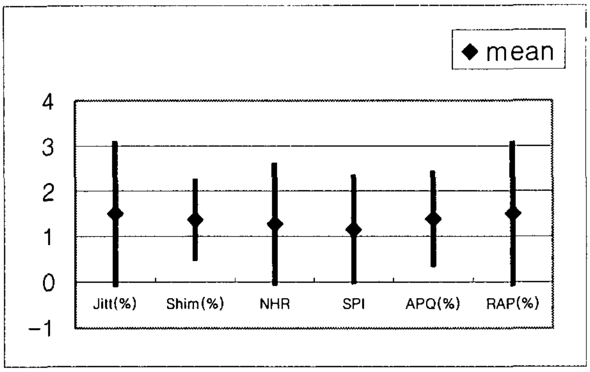
(a) Original normal parameter data



(b) Original pathological parameter data



(c) Normalized normal parameter data



(d) Normalized pathological parameter data

Fig.1 The comparisons of the original and normalized parameters

IV. Experiments and Results

As a classifier of this experiment, a three-layered artificial neural network was used. The number of input layers was varied from to 3 to 6 to find the optimal set of parameters. The number of output layers was fixed to 2. The number of hidden layers was set 3, 6, and 9 for the 3 inputs case, and 6, 9, and 12 for the 6 inputs case. [6] Because the total number of data was small, we tried to train and test the neural network by splitting total data sets into two parts. Two thirds of the data were used for training. The remaining one third was used for testing. In each training session, the neural network was trained and tested separately using different combinations of data sets. This was to compensate for the small size of the data sets.

The original and normalized parameters were used to train and test the same structure neural network. In order to accurately compare the difference between the classification results from the neural network using the original parameter inputs and the result using the normalized parameter inputs, we kept the same order when the two sets of different parameters were inputted into the different neural networks. In other words, the inputting order of normalized parameters was corresponded to the original order.

Table. 1 The classification rate (%)

Neural Network Structure	Times	Original Data		Normalized Data	
		Test Data	Training Data	Test Data	Training Data
3 Inputs 3 Hidden layers 2 Outputs 3I3H	1st Run	87.5000	91.1765	84.3750	97.0588
	2nd Run	84.3750	86.7647	84.3750	97.0588
	3rd Run	84.3750	86.7647	84.3750	86.7647
	4th Run	81.2500	98.5294	78.1250	97.0588
	5th Run	81.2500	94.1176	78.1250	95.5882
3 Inputs 6 Hidden layers 2 Outputs 3I6H	1st Run	81.2500	92.6471	84.3750	100.0000
	2nd Run	81.2500	91.1765	81.2500	98.5294
	3rd Run	78.1250	95.5882	78.1250	100.0000
	4th Run	75.0000	100.0000	78.1250	100.0000
	5th Run	75.0000	97.0588	78.1250	100.0000
3 Inputs 9 Hidden layers 2 Outputs 3I9H	1st Run	87.5000	98.5294	84.3750	100.0000
	2nd Run	84.3750	100.0000	81.2500	100.0000
	3rd Run	84.3750	94.1176	81.2500	100.0000
	4th Run	78.1250	95.5882	81.2500	97.0588
	5th Run	75.0000	100.0000	78.1250	100.0000
6 Inputs 6 Hidden layers 2 Outputs 6I6H	1st Run	81.2500	100.0000	84.3750	100.0000
	2nd Run	78.1250	100.0000	84.3750	100.0000
	3rd Run	75.0000	98.5294	81.2500	100.0000
	4th Run	71.8750	100.0000	81.2500	100.0000
	5th Run	71.8750	100.0000	78.1250	100.0000
6 Inputs 9 Hidden layers 2 Outputs 6I9H	1st Run	81.2500	100.0000	87.5000	100.0000
	2nd Run	81.2500	100.0000	84.3750	100.0000
	3rd Run	78.1250	100.0000	81.2500	100.0000
	4th Run	78.1250	100.0000	78.1250	100.0000
	5th Run	75.0000	100.0000	75.0000	100.0000
6 Inputs 12 Hidden layers 2 Outputs 6I12H	1st Run	84.3750	97.0588	87.5000	100.0000
	2nd Run	78.1250	100.0000	81.2500	100.0000
	3rd Run	78.1250	100.0000	81.2500	100.0000
	4th Run	78.1250	100.0000	81.2500	100.0000
	5th Run	75.0000	100.0000	78.1250	100.0000

Table. 1 shows the classification rate from neural network training and testing with 6 parameters. Experiments were performed using 3 and 6 parameters respectively to see the different effects of the original and normalized parameters on discriminating the pathological voice into normal and abnormal classes. In case of 3 parameters, Jitter, Shimmer and NHR were used. And additional 3 parameters (SPI, APQ and RAP) were used for 6 parameters. There were 24 sets of result data in total.

V. Discussion

From the experimental result we couldn't observe the significant differences between the corresponding classification rates when the original and normalized parameters were inputted into the different neural networks Table 1. The results looked very similar. In order to obtained the observation results directly, we chose the best classification rate from every 24 sets of parameters to plot the changing trend curve of the classification rate as shown in Fig.2. In the Fig.2, all corresponding curves were very close but slightly better results were obtained for some neural net configurations.

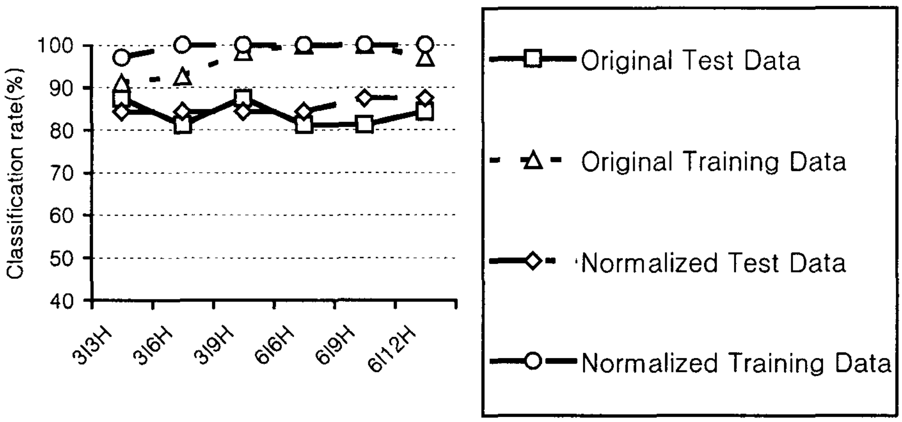


Fig. 2 The changing trend curve of the classification rate

In the original parameter set, the value range of NHR was about 100 times bigger than the range of other parameter's values. After normalization, the range became similar to those of the other's parameters. But the classification result didn't change much, so we conclude NHR didn't play a significant role in the classification. The relative change of values of other parameters before and after normalization was not great. The normalization process didn't affect the performance of the network.

VI. Conclusion

In this paper we collected pathological voice samples using DAT. The normalization method of the original parameters was introduced. Artificial neural networks were used to classify the voice into normal and abnormal states by original and normalized parameters.

We didn't observe a significant improvement or decrease in performance by normalizing parameters. We conclude that the normalization process is not necessary for the classification of pathological voice when using artificial neural networks.

The voice data sample was not large enough to generalize the performance and more data collection is required.

Acknowledgement

This study is supported by a grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (No. 02-PJ1-PG10-31401-0005).

References

[1] Godino-Llorente, Santiago Aguilera-Navarro*, Carlos Hernandez-Espinosa**, Mercedes Fernandez-Redondo**, Pedro Gomez-vilda+, 1999, "*On the selection of meaningful speech parameters used by a pathologic/non pathologic voice register classifier,*" *Eurospeech' 99,* Budapest, Hungary, vol. 1, pp. 563-566.

[2] Jo, C.W., Kim, D.H., Wang, S.G., 1999, "Classification of pathological voice into normal /benign/malignant state," Eurospeech' 99, Budapest, Hungary, pp. 399-402.

[3] Kim, D.H., Jo, C.W., 1998, "A study on the classification of pathological voice," 15th Workshop on Speech Communication and Signal Processing, pp. 388-391.

[4] Kim, K.I., Jo, C.W., Kim, D.Y., Wang, S.G., Jeon, G.O., Ahn, S.H., Kim, K.R., Kim, Y.J. 2000, "Collection, analysis and classification of pathological voice using ARS and neural network," 13rd Korean Conference on Signal Processing, pp. 955-958.

[5] Jo, C.W., Kim, D.H., Kim, K.I., Wang, S.G., Jeon, G.O., Ahn, S.H., Kim, K.Y., Kim, Y.J., 2000, "*Implementation of analysis tools for pathological voice,*" *17th Workshop on*

Speech Communication and Signal Processing, pp. 211−214.
[6] Jo, C.W., Kim, K.I., Kim, D.H., Wang, S.G., 2001, "*Screening of pathological voice from ARS using neural networks*," *International Workshop on Maveba*, Firenze, Italy, September 13−15.
[7] Operations Manual, 1993, "Multi−Dimensional Voice Program (MDVP)," Model 4305 Kay Elemetrics Corp.
[8] Jo, C.W., Kim, K.G., Kim, D.H., Wang, S.G., Jeon, G.R., Denmark, 2001, "*Comparisons of acoustical characteristics between ARS and DAT voice*," *Eurospeech' 2001.*

Received : February 10, 2004
Accepted : March 10, 2004

▲ Tao Li
SASPL, School of Mechatronics, Changwon National University
9 Sarim−dong, changwon, kyungnam, Korea (641−773)
Tel: +82−55−279−7559
Fax: +82−55−262−4064
e−mail: litao_com@hotmail.com

▲ Il−suh Bak
SASPL, School of Mechatronics, Changwon National University
9 Sarim−dong, changwon, kyungnam, Korea (641−773)
Tel: +82−55−279−7559
Fax: +82−55−262−4064
e−mail: ilsuh@korea.com

▲ Cheolwoo Jo
SASPL, School of Mechatronics, Changwon National University
9 Sarim−dong, changwon, kyungnam, Korea (641−773)
Tel: +82−55−279−7552
Fax: +82−55−262−4064
e−mail: cwjo@sarim.changwon.ac.kr