

Research Article

Automatic Assessment of Pathological Voice Quality Using Higher-Order Statistics in the LPC Residual Domain

JiYeoun Lee^{1,2} and Minsoo Hahn²

¹Division of Head and Neck Surgery, School of Medicine, University of California at Los Angeles, Los Angeles, CA 90024, USA

²Division of Multimedia Communications and Processing, School of Engineering, Korea Advanced Institute of Science and Technology, 335 Gwahangno, Yuseong-gu, Daejeon 305-701, South Korea

Correspondence should be addressed to JiYeoun Lee, jiyeeounlee09@gmail.com

Received 4 May 2009; Revised 15 September 2009; Accepted 2 December 2009

Recommended by Gregor Rozinaj

A preprocessing scheme based on linear prediction coefficient (LPC) residual is applied to higher-order statistics (HOSs) for automatic assessment of an overall pathological voice quality. The normalized skewness and kurtosis are estimated from the LPC residual and show statistically meaningful distributions to characterize the pathological voice quality. 83 voice samples of the sustained vowel /a/ phonation are used in this study and are independently assessed by a speech and language therapist (SALT) according to the grade of the severity of dysphonia of GRBAS scale. These are used to train and test classification and regression tree (CART). The best result is obtained using an optimal decision tree implemented by a combination of the normalized skewness and kurtosis, with an accuracy of 92.9%. It is concluded that the method can be used as an assessment tool, providing a valuable aid to the SALT during clinical evaluation of an overall pathological voice quality.

Copyright © 2009 J. Lee and M. Hahn. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Pathological voice quality assessment has attracted attention for many years, inducing a large amount of research based on acoustical, aerodynamic, and physiological measurements [1–6]. Our goal is to assess an overall pathological voice quality which is scored on a four-point scale: 0 = normal, 1 = mild deviance, 2 = moderate deviance, and 3 = severe deviance.

Gu et al. suggested three objective quality assessment measures such as Itakura-Saito (IS) distortion, log-likelihood ratio (LLR), and log-area-ratio (LAR). By evaluating speech database of thirteen sentences, the IS measure showed a strong correlation with mean opinion score (MOS). Therefore, the IS measure was suggested to be more suitable than LLR and LAR for use as a reliable tool to evaluate an overall quality of disordered speech [1].

An artificial neural network (NN) was investigated using various combinations of short-term and long-term time-domain and frequency-domain parameters extracted from electrical impedance signals by R.T. Ritchings et al. In 77 abnormal speech signals, the voice quality was independently

assessed by a speech and language therapist (SALT) according to their seven-point ranking of subjective voice quality. The best result was obtained using 21 input parameters, for which an accuracy of 92% was achieved [4].

Muzeyyen Dogan et al. performed an acoustic analysis using the multidimensional voice program (MDVP; Kay Elemetrics Corporation, Lincoln Park, NJ). Voice handicap index (VHI), grade of the severity of dysphonia (G), roughness (R), and breathiness (B) scales were used for subjective evaluations. They found that maximum phonation time, frequency, and amplitude perturbation parameters were impaired in 40 asthmatic patients [6].

Although all the achievements and conclusions are not easily comparable due to a lack of uniformity when computing and presenting the results, these works represent novel contributions to pathological voice quality assessment [1–6]. However, improving its performance is still clinically important since the tool enables to accurately detect the pathological voice quality in a quantitative way without professional doctors and medical instruments.

In this paper, we propose a novel scheme of pathological voice quality measurement, higher-order statistics (HOSs)

analysis based on the linear prediction coefficient (LPC) residual, to assess an overall quality of pathological voices. HOSs have been recently applied in the automatic detection of voice pathologies [7]. Although the HOSs hold promise as one possible marker of classification between normal and pathological voices, there have not yet been studies applying HOSs in connection with an overall voice quality under pathological voice quality assessments. And the LPC residual of the speech signal corresponds to an estimate of the excitation signal from a mathematical model of the vocal tract. Some parameters extracted from the LPC residual turn out to distinguish between normal speakers and dysphonic patients [2, 5, 8]. These facts suggest that the combination of the HOSs analysis and the LPC residual may help to effectively construct some important information for distinguishing pathological voice quality [8].

The rest of the paper is organized as follows. In Section 2, a brief review of LPC residual, HOSs analysis, and classifier is presented. The experiments which utilize HOSs of the LPC residual and block diagram as a classification tool are shown in Section 3, and in Section 4, the classification results to classify an overall quality of pathological voice are discussed. In Section 5, our result is compared with one of another approach. Finally, the conclusions are presented in Section 6.

2. Methodology

2.1. LPC Residual. The LPC residual corresponds to the excitation signal of the vocal tract model [9]. The LPC residual of normal voice may be modeled as a deterministic signal, consisting of the sinusoids with equal amplitudes. The frequencies of these sinusoids may be harmonically related. It has rather periodic and stable structure. However, the LPC residual of pathological voice may be modeled as a sum of incoherent sine waves whose phases are rather random. It is characterized by the large variation in the pitch period because the movement of the vocal folds is not balanced and an incomplete closure may appear in glottal cycles [10]. These have effect on an overall degree of the severity of dysphonia. So, as the degree is higher, the voices tend to have more and more irregular, aperiodic, and unstable. That is, it depends on whether the voice quality is a slight or serious according to the voice disorder [8]. Thus, the use of the LPC residual may bring some information like abnormal movement of vocal folds and turbulence noise for pathological voice quality assessment [2, 5].

In researches studying the relationship between LPC residual and disordered voices, the authors aimed to perform the LPC characteristics as an acoustic parameter for classification between normal and pathological voices. Marcelo de Oliveira Rosa et al. suggested an adaptive estimation method of LPC residual for voice pathology diagnosis. Through inverse filtering (Kalman and Wiener filters) of the voice signal, the LPC residual was estimated and seven acoustic features were extracted from it to evaluate the laryngeal diseases. The presented techniques showed that it is possible to evaluate the extent of the larynx diseases and identify them using adaptive filtering and acoustic measurements

[8]. Therefore, we conclude that they show the possibility of being able to automatically detect an overall quality of pathological voice related to the larynx diseases.

2.2. Higher-Order Statistics. HOS analysis has shown promising results as a classification index of pathological voice and also has the advantage of not requiring a periodic or quasiperiodic voice signal to permit reliable analysis. So, it is of particular value when dealing with a mixture of Gaussian and non-Gaussian processes and system nonlinearity. The application of HOS to speech processing has been primarily motivated by their inherent Gaussian suppression and phase preservation properties [7, 11]. Researchers in this area have been based on the assumption that speech has nonzero HOS characteristic that is distinct from those of Gaussian noise [7]. It has been also sufficiently used as a basis for voiced and unvoiced speech detection. Thus it may be more valuable for solving the noisy component generated in pathological voices which are similar to characteristics of unvoiced sound [10]. It may also be one possible acoustic marker that is sensitive to voice impairment.

It is well known that a speech signal, $x(n)$, which may be normal or pathological, can be expressed as in (1) [11]:

$$x(n) = s(n) + w(n), \quad (1)$$

where $s(n)$ is a non-Gaussian signal generated by the oscillation of the vocal folds and $w(n)$ is Gaussian noise which can be assumed to be zero in normal voices and not to be zero in pathological voices.

Pathological voices are corrupted by noise, $w(n)$, which is directly related to the perceived roughness of the voice [10]. If $s(n)$ and $w(n)$ are statistically independent, then the energy of $x(n)$ is the sum of speech and noise energies: $E_x = E_s + E_w$. Second-order statistics (SOSs) are directly affected in an additive way by the presence of noise [11]. Hence, when HOS analysis is applied to pathological voices, unstable and discontinuous statistics of $x(n)$ may be estimated because HOS analysis is blind to Gaussian processes. On the other hand, in normal voices, the HOS of only non-Gaussian measurements may be extracted because a Gaussian noise can be assumed to be zero. After all, the variation of a non-Gaussian signal which is produced by vibration of the vocal folds can be an important clue for quality assessment of the pathological and normal voices.

Research implemented by Alonso et al. proposed new seven HOS-based parameters that were obtained directly or indirectly in the frequency and time domain for the classification between normal and pathological voices. A success rate of 98.3% was obtained by using both the conventional and the HOS-based parameters with the NN classifier [7]. This paper is a little short on the detailed analyses why they propose new HOS parameters, that is, a general trend of HOS, experimental objective, and how it can be applied to pathological and normal voices. However, they show the possibility of being able to automatically classify the pathological voice with good performance using the HOS parameters. This paper is the only one that applies the HOS

analysis to pathological voice detection in time domain. Therefore, it is additionally necessary to study how well each HOS parameter can assess an overall quality of pathological voices and to confirm the experiments with an authorized database.

Among various HOSs, the normalized skewness, γ_3 , and the normalized kurtosis, γ_4 , are widely used as characteristic parameters. They are defined as in (2) [7, 11, 12]:

$$\gamma_3 = \frac{\sum_{n=1}^N (x_n - \mu)^3}{(N-1)\sigma^3}, \quad \gamma_4 = \frac{\sum_{n=1}^N (x_n - \mu)^4}{(N-1)\sigma^4}, \quad (2)$$

where x_n is the n th sample value and N is the number of the samples while μ and σ represent the mean and the standard derivation of x_n , respectively.

One way to quantify the higher-order cumulants is to compare it with a Gaussian bell curve (a normal random probability distribution) and then characterize the HOS distribution using skewness and kurtosis. Skewness measures symmetry. That is, the skewness for a normal distribution is zero, and any symmetric data should have skewness near zero. Positively skewed distributions appear to have a tail extended to the right of the bell curve. Similarly, negatively skewed distributions mean that the left tail is long relative to the right tail. Kurtosis represents the “peakedness” of the distribution, with steep distributions producing kurtosis values larger than three (leptokurtic) and flat distributions producing kurtosis values lower than three (platykurtic) [12].

2.3. Classifier. Classification and regression tree (CART) analysis is a common method to build statistical models founded on tree-based techniques [13]. One of the most important characteristics of the CART is that the optimal decision tree contains the rules which are easily readable by humans compared to other classification and regression methods such as vector quantization (VQ) and NN. Decision tree contains a binary question about some feature at each node. The leaves of the tree contain the best prediction based on the training data [13, 14]. To improve the performance of pathological voice quality measurement, there have been many studies on parameter extraction [2, 7, 8, 15, 16]. However, each parameter does not always guarantee the reliable performance in various kinds of conditions. Therefore, it may be necessary to use these parameters together to ensure the robustness in various conditions. Statistical approach can be considered as a solution to effectively combine the multiple parameters. We use the CART algorithm to evaluate the performance for the assessment of pathological voice quality using multiple parameters.

3. Experiments

3.1. Database. In 1981, the Japan Society of Logopedics and Phoniatics distributed a DVD-ROM database of a total of approximately 65 utterances scored with the GRBAS scale. The GRBAS scale is the authorized perceptual evaluation method proposed by above institute. For the clinician’s

assessment, it consists of the voice properties such as the grade of the severity of dysphonia (G), roughness (R), breathiness (B), asthenicity (A), and strain (S). It is also scored on a four-point grading scale: a normal is 0, a slight, 1, a moderate, 2, and finally, a severe, 3, for five different parameters to rate the degree of vocal quality [4, 6]. Only G parameter is taken into account for this study. It is marked as G0, G1, G2, and G3 voices shown in Table 1. These perceptual grades are determined by the juries composed of Japanese SALT.

Pathological voices used in this paper are composed of 63 male and female voices of aged from 7 to 78 (mean: 45.7). We add 20 normal Korean voice data to this pathological data, and finally, we have the 83 pathological and normal voice data. Among the 83 voices, 20 voices are Korean normal; 17 is associated with a voice of grade 1, 26 with a voice of grade 2, and 20 with a voice of grade 3. Subject information used in this paper is shown in Table 1. Since we are interested only in pathologies which affect the vocal folds, the experiment is carried out for the sustained vowel /a/ phonation (1–3 sec.). All voice data are down-sampled to 16 kHz with 16 bits. 70% and 30% of the data are used for the training and the testing set to build each set for a 5-fold cross-validation scheme [17].

3.2. HOS Analysis Based on LPC Residual. Figure 1 shows the distributions of γ_3 and γ_4 extracted from the LPC residual of G0, G1, G2, and G3 voices. The preprocessing involves taking 20 milliseconds frames with 10 milliseconds frames overlapping from the signals. Each frame is then preemphasized by forward differencing to reduce the effects of drifting signal amplitude, and it is multiplied by a hamming window, prior to LPC analysis. Then, the Levinson-Durbin algorithm is applied to derive the 10th-order all-pole LPC model. Finally, the normalized third- and fourth-order cumulants of the LPC residual are computed in each sentence. In the γ_3 distributions of Figure 1(a), the distributions of the G-scaled voices are skewed to the right, with the means of 1.06 ± 0.46 , 0.81 ± 0.41 , 0.51 ± 0.46 , and 0.32 ± 0.36 , respectively. Specifically, γ_3 distribution of normal voices is more skewed right and has a broad distribution with ratings distributed from 0.18 to 2.10 than that of other ones. In Figure 1(b), γ_4 distributions have a leptokurtic one ($\gamma_4 > 3$), overall (means of 10.27 ± 2.79 , 8.48 ± 2.26 , 6.97 ± 1.30 , and 6.81 ± 1.51 , resp.). As the voices are a normal, γ_4 spreads out rather widely with a large range and has more leptokurtic distribution. The distributions result in the distinct characteristics for these cumulants in terms of phase, periodicity, and harmonic components of G0, G1, G2, and G3 voices. The results of statistical analysis of the HOSs measurements are summarized in Table 2. The HOSs of the LPC residual show a clear distribution between G-scaled voices.

As discussed in Section 2.2, we can confirm that normal signal that has little Gaussian noise tends to be zero shifting when HOS is applied to it. On the other hand, pathological voice that has mixed noises including a lot of Gaussian noise tends to be distinguished from normal one according to the grade of voice quality.

TABLE 1: Summary of subject information. “F” and “M” in the second column stand for females and males, respectively.

Subject no.	Sex	Age (y)	Diagnosis	Voice quality	Subject No.	Sex	Age (y)	Diagnosis	Voice quality
(1)	M	28	Normal	G0	(43)	M	44	Polyps	G2
(2)	M	35	Normal	G0	(44)	F	30	Polyps	G2
(3)	F	25	Normal	G0	(45)	F	62	Polyps	G2
(4)	F	40	Normal	G0	(46)	F	32	Polyps	G2
(5)	F	21	Normal	G0	(47)	M	60	Polyps	G2
(6)	M	23	Normal	G0	(48)	M	56	Leukoplakia	G2
(7)	M	36	Normal	G0	(49)	M	26	Spasmodic dysphonia	G2
(8)	M	28	Normal	G0	(50)	F	52	Polyps	G2
(9)	F	33	Normal	G0	(51)	F	54	Polyps	G2
(10)	F	28	Normal	G0	(52)	M	55	Glottic cancer	G2
(11)	M	34	Normal	G0	(53)	F	21	Virilization	G2
(12)	F	29	Normal	G0	(54)	M	37	Cysts	G2
(13)	M	24	Normal	G0	(55)	F	54	Papilloma	G2
(14)	M	30	Normal	G0	(56)	M	62	Paralysis	G2
(15)	F	26	Normal	G0	(57)	F	78	Paralysis	G2
(16)	F	43	Normal	G0	(58)	M	23	Paralysis	G2
(17)	F	39	Normal	G0	(59)	M	63	Sulcus vocalis	G2
(18)	M	45	Normal	G0	(60)	M	58	Sulcus vocalis	G2
(19)	M	37	Normal	G0	(61)	F	61	Scarring	G2
(20)	M	22	Normal	G0	(62)	F	55	Scarring	G2
(21)	F	29	Nodules	G1	(63)	F	22	Scarring	G2
(22)	F	22	Nodules	G1	(64)	F	34	Laryngitis	G3
(23)	F	30	Nodules	G1	(65)	F	59	Laryngeal web	G3
(24)	F	37	Nodules	G1	(66)	M	70	Hemorrhage	G3
(25)	F	20	Nodules	G1	(67)	F	77	Polyps	G3
(26)	F	39	Nodules	G1	(68)	M	58	Glottic cancer	G3
(27)	F	22	Nodules	G1	(69)	M	64	Leukoplakia	G3
(28)	F	30	Nodules	G1	(70)	M	47	Cysts	G3
(29)	M	34	Laryngitis	G1	(71)	M	21	Papilloma	G3
(30)	F	49	Laryngeal web	G1	(72)	F	46	Paralysis	G3
(31)	M	7	Nodules	G1	(73)	M	57	Paralysis	G3
(32)	M	73	Polyps	G1	(74)	M	35	Paralysis	G3
(33)	F	53	Polyps	G1	(75)	M	79	Paralysis	G3
(34)	F	62	Virilization	G1	(76)	M	67	Paralysis	G3
(35)	F	19	Cysts	G1	(77)	M	69	Granuloma	G3
(36)	F	61	Cysts	G1	(78)	F	68	Polyps	G3
(37)	M	45	Granuloma	G1	(79)	M	39	Polyps	G3
(38)	F	23	Nodules	G2	(80)	F	50	Polyps	G3
(39)	M	38	Laryngitis	G2	(81)	F	53	Polyps	G3
(40)	M	50	Laryngitis	G2	(82)	F	43	Spasmodic dysphonia	G3
(41)	F	8	Nodules	G2	(83)	M	64	Glottic cancer	G3
(42)	M	49	Chronic laryngitis	G2					

3.3. *Classifier.* The CART software to be used in this paper is a commercial product manufactured and sold by Salford Systems (<http://www.salford-systems.com/>). The important feature is that all data management functions to note about CART are taken care of by the software. Once we have the data file, we can conduct the CART analysis with various commands.

A splitting rule is a method and strategy for growing a tree [13, 14]. For classification trees, the default rule is the Gini that generally works well across a broad range of problems and we utilized this method.

The largest tree is the starting point for CART analysis. CART first splits the root node, then splits the resulting children, then splits the grandchildren, and so on. However,

TABLE 2: Comparisons of G-scaled voices on the normalized skewness and kurtosis based on LPC residual.

	skewness				kurtosis			
	G0	G1	G2	G3	G0	G1	G2	G3
Mean	1.06	0.81	0.51	0.32	10.27	8.48	6.97	6.81
Median	0.99	0.74	0.56	0.28	9.84	7.57	6.92	6.45
Range	0.18–2.10	0.08–1.51	−0.19–1.34	−0.41–1.07	5.71–16.34	5.25–13.41	4.90–10.05	4.83–10.08
25%–75% range	0.72–1.41	0.59–1.11	0.00–0.71	0.09–0.61	7.97–11.96	6.91–9.78	5.75–7.68	5.34–8.07
Std. Dev	0.46	0.41	0.46	0.36	2.79	2.26	1.30	1.51
N	30	17	26	20	30	17	26	20

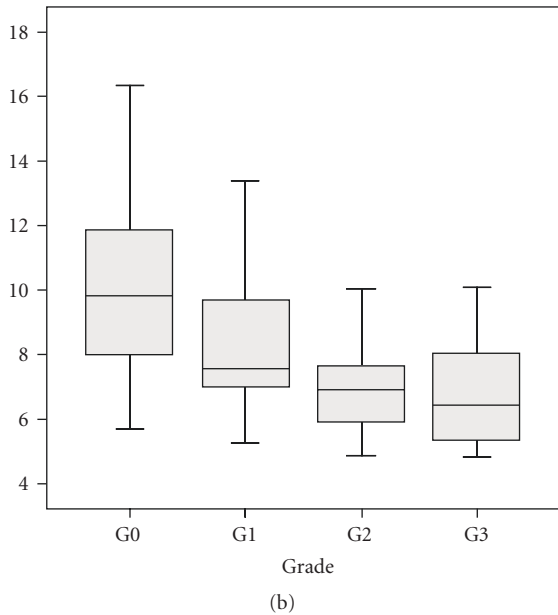
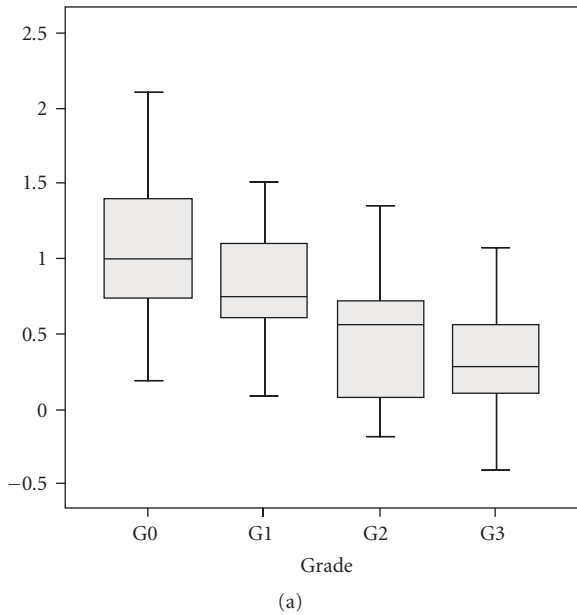


FIGURE 1: Higher-order statistics. (a) Distributions of the estimated normalized skewness. (b) Distributions of the estimated normalized kurtosis.

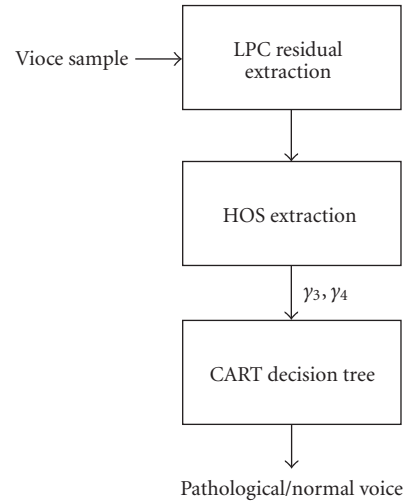


FIGURE 2: Overall procedure of proposed algorithm. (γ_3 : the normalized skewness, γ_4 : the normalized kurtosis).

this largest tree is quite a bit worse than the simple tree, indicating that the large tree is seriously “overfit” [13, 14]. In this software, several dimensions of the size of a tree-growing problem can be controlled with the LIMIT command. We prevented tree growth beyond a depth of 7 levels and did not prune the tree. Once we have the largest tree constructed, we can see the relative error profile. Then we chose the optimal tree that has the minimum cost regardless of size.

3.4. Block Diagram. Figure 2 presents the overall procedure of the proposed method for the pathological voice quality assessment, that is, G0, G1, G2, and G3 voices. The 10th-order LPC analysis is performed on speech sampled at 16 kHz. Then, the normalized skewness and kurtosis are calculated in the total LPC residuals. The final decision is processed according to decision tree of the normalized skewness and kurtosis through the CART analysis. The experiment is repeated 5 times and the dataset is randomly split in 5 different training and test sets, using each time a different subset for testing the performance. After the cross-validation, the final results are averaged across these repetitions, and confidence intervals can be computed using the standard deviation of the measures.

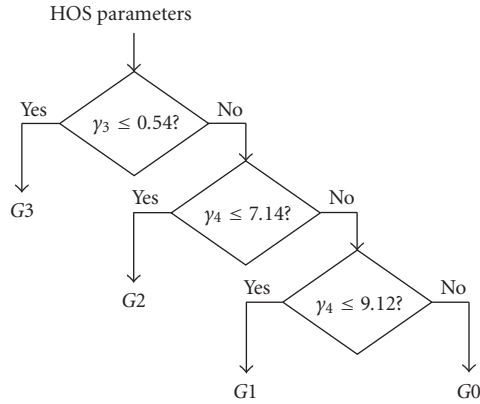


FIGURE 3: Optimal decision tree formed by HOS parameters.

Equation (3) represents a statistic used to measure the generalization error [17, 18]. Testing with N patterns and obtaining an accuracy p , the confidence interval (CI) for this measure is

$$CI = \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{N}}, \quad (3)$$

where the value z is obtained from a standard normal distribution as a function of the required confidence level α . Using $\alpha = 0.05$ (95% of confidence), z is 1.96.

4. Results

In this paper, CART analysis is used to combine the HOSs parameters and evaluate the performance. Since some parameters make good decisions for the classification among G0, G1, G2, and G3 voices and some do not, it is necessary to design a rule to make the final decision regarding the use of multiple inputs in the classifiers at the same time. By using the information of the multiple parameters extracted from pathological and normal voice, the CART makes a final decision whether the current phonation is normal, a slight, a moderate, or a severe pathological voice. The optimal decision tree generated by the normalized skewness and kurtosis as its inputs is shown in Figure 3. It excellently reflects the distribution characteristics shown in Figure 1.

Table 3 shows the confusion matrix with the CI obtained by averaging the results for each individual experiment. It is based on the decision tree shown in Figure 3. Each matrix cell indicates how many instances with the corresponding actual class label are predicted by the model to have the corresponding predicted class label. The diagonal numbers indicate the performance of correctly classified signals. The off-diagonal elements are associated with the performance of misclassifications. In this table, the final, that is, G0 versus G1 versus G2 versus G3, classification performance is averagely 92.9%. A small part of the voices are not classified as any of the defined classes and are designated as unclassified. Actually, misclassification is inevitable because we are not considered for the “roughness” and “breathiness” factors which may affect the performance in “GRBAS” scale.

5. Comparison with Another Approach

We utilized an algorithm developed by Lingyun Gu et al. to compare the results of our algorithm with ones of another approach. It is recently published as automatic algorithm based on pathological voice quality assessment [1]. They tested several well-known speech processing parameters, that is, IS, LLR, and LAR measures which evaluate the spectral envelope of the given voice data. And the voice produced by healthy people as the gold standard is used to compare with pathological voice. Finally, dynamic time warping (DTW) is used to align the two different speech segments and to classify the system accuracy. 70% and 30% of our voice data are used for the training and the testing set to build each set for a 5-fold cross-validation scheme like our experiments. Lingyun Gu et al. suggested that the IS measure showed a string correlation with the subjective tests. It is concluded that the IS measure is suggested to be more suitable than LLR and LAR for use as a reliable tool to evaluate the overall quality of disordered voice [1].

Table 4 shows the confusion matrix with the CI obtained by averaging the results for IS, LLR, and LAR experiments. In this table, the final, that is, G0 versus G1 versus G2 versus G3, classification performance is averagely 75.7%, 71.4%, and 67.2% for IS, LLR, and LAR, respectively. However, our performance is averagely 92.9%. It is concluded that our method using the HOSs parameters based on LPC residual is more effective than the method implemented by Lingyun Gu et al. It is also the best performance among published paper, higher than the performance measured by other authors, for automatic assessment of pathological voice quality.

6. Conclusion

The accurate assessment of pathological voice quality is a major research that has attracted attention in the field of biomedical engineering and voice disorder for many years. A meaningful quality assessment should be consistent with human responses and perception. Performance decisions using subjective measures are based on a group of listeners’ opinion of the quality of an utterance. Although it is not suggested to use objective quality methods to completely replace subjective ones, the objective quality evaluations have shown the strong ability to predict disordered voices in many researches and the results correlate very well with those evaluated by subjective quality measures. However, no studies have investigated the HOSs characteristics of LPC residual by means of the objective quality measure.

In this paper, we have applied HOS analysis to LPC residual for an overall pathological voice quality assessment. That is, this study represents a novel way to combine the LPC residual signal and higher-order cumulants. The normalized skewness and kurtosis of the LPC residual signals show statistically significant distributions to characterize an overall quality of pathological voice. A close correlation between the HOSs parameters and pathological voice quality is also demonstrated by means of the descriptive statistics. For the performance measurements, the CART algorithm is implemented. Especially, the optimal decision tree based

TABLE 3: Confusion matrix (accuracy \pm confidence interval%).

		True Classification			
		G0	G1	G2	G3
Predicted classification	G0	95.0 \pm 4.08	5.0 \pm 4.08	0.0	0.0
	G1	3.9 \pm 2.78	92.1 \pm 2.52	3.9 \pm 2.78	0.0
	G2	0.0	1.3 \pm 1.79	94.3 \pm 1.93	3.8 \pm 3.10
	G3	0.0	3.3 \pm 2.35	6.7 \pm 2.35	90.0 \pm 4.08

TABLE 4: Confusion matrix (accuracy \pm confidence interval%).

		True Classification				
		G0	G1	G2	G3	
Predicted classification	IS	75.0 \pm 0.38	25.0 \pm 0.38	0.0	0.0	
	LLR	G0	75.0 \pm 0.38	12.5 \pm 0.29	12.5 \pm 0.29	0.0
	LAR		66.7 \pm 0.41	26.7 \pm 0.39	6.6 \pm 0.22	0.0
	IS		16.7 \pm 0.33	75.0 \pm 0.38	8.3 \pm 0.24	0.0
	LLR	G1	12.5 \pm 0.29	75.0 \pm 0.38	12.5 \pm 0.29	0.0
	LAR		13.8 \pm 0.30	66.7 \pm 0.41	19.5 \pm 0.35	0.0
	IS		0.0	10.0 \pm 0.26	77.6 \pm 0.37	12.5 \pm 0.29
	LLR	G2	0.0	13.8 \pm 0.30	68.8 \pm 0.41	17.5 \pm 0.33
	LAR		0.0	13.8 \pm 0.30	68.8 \pm 0.41	17.5 \pm 0.33
	IS		0.0	10.2 \pm 0.27	14.8 \pm 0.31	75.0 \pm 0.38
	LLR	G3	0.0	16.7 \pm 0.33	16.7 \pm 0.33	66.7 \pm 0.41
	LAR		0.0	6.6 \pm 0.22	26.7 \pm 0.39	66.7 \pm 0.41

on the normalized skewness and kurtosis is proposed to effective combination method of the multiple parameters. The experiment demonstrates that the CART algorithm which uses HOS parameters together can provide an average classification performance, at 92.9%. Although many studies give a good idea of the variety of approaches, all the results are not easily comparable due to a lack of uniformity. However, we apply an algorithm proposed by Lingyun Gu et al. to our database, which is recently published as automatic algorithm for pathological voice quality assessment. While the best performance is averagely 75.7% in by Lingyun Gu et al. method, it shows averagely 92.9% in our method. Therefore, we can verify that our works provide the highest accuracy by utilizing combination of HOS analysis and LPC residual and suggesting the optimal decision tree. This is important, since the method is beneficial in improving the performance for diagnosing an overall quality of disordered voice.

In the future, our proposed method should be tested with large voice samples and rather long sentence. We will be investigated to complete an objective assessment of the voice quality according to the GRBAS scale. Finally, in actual clinical circumstances, it will be tested for the application of a monitoring system for patients.

Acknowledgment

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-357-D00202).

References

- [1] L. Gu, J. G. Harris, R. Shrivastav, and C. Sapienza, "Disordered speech assessment using automatic methods based on quantitative measures," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 9, pp. 1400–1409, 2005.
- [2] A. Kacha, F. Grenez, and J. Schoentgen, "Estimation of dysperiodicities in disordered speech," *Speech Communication*, vol. 48, no. 10, pp. 1365–1378, 2006.
- [3] P. Yu, M. Ouaknine, J. Revis, and A. Giovanni, "Objective voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements," *Journal of Voice*, vol. 15, no. 4, pp. 529–542, 2001.
- [4] R. T. Ritchings, M. McGillion, and C. J. Moore, "Pathological voice quality assessment using artificial neural networks," *Medical Engineering and Physics*, vol. 24, no. 7-8, pp. 561–564, 2002.
- [5] M. Wester, "Automatic classification of voice quality: comparing regression models and hidden markov models," in

- Proceedings of Symposium on Databases in Voice Quality Research and Education (VOICEDATA '98)*, pp. 92–97, Utrecht, The Netherlands, 1998.
- [6] M. Dogan, E. Eryuksel, I. Kocak, T. Celikel, and M. A. Sehitoglu, “Subjective and objective evaluation of voice quality in patients with asthma,” *Journal of Voice*, vol. 21, no. 2, pp. 224–230, 2007.
 - [7] J. B. Alonso, J. De Leon, I. Alonso, and M. A. Ferrer, “Automatic detection of pathologies in the voice by HOS based parameters,” *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 4, pp. 275–284, 2001.
 - [8] M. De Oliveira Rosa, J. C. Pereira, and M. Grellet, “Adaptive estimation of residue signal for voice pathology diagnosis,” *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 1, pp. 96–104, 2000.
 - [9] J. R. Deller Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, MacMillan, New York, NY, USA, 1993.
 - [10] R. D. Kent and M. J. Ball, *Voice Quality Measurement*, Thomson Learning, New York, NY, USA, 1st edition, 2000.
 - [11] E. Nemer, R. Goubran, and S. Mahmoud, “Robust voice activity detection using higher-order statistics in the LPC residual domain,” *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 3, pp. 217–231, 2001.
 - [12] J. M. Mendel, “Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications,” *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278–305, 1991.
 - [13] S. Gey and E. Nedelec, “Model selection for CART regression trees,” *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 658–670, 2005.
 - [14] M. M. Tanabian, P. Tierney, and B. Z. Azami, “Automatic speaker recognition with formant trajectory tracking using CART and neural networks,” in *Proceedings of Canadian Conference on Electrical and Computer Engineering*, pp. 1225–1228, Saskatoon, Canada, May 2005.
 - [15] D. Michaelis, M. Fröhlich, and H. W. Strube, “Selection and combination of acoustic features for description of pathologic voices,” *Journal of the Acoustical Society of America*, vol. 103, no. 3, pp. 1628–1639, 1998.
 - [16] K. Shama, A. Krishna, and N. U. Cholayya, “Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology,” *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 85286, 9 pages, 2007.
 - [17] N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, “Methodological issues in the development of automatic systems for voice pathology detection,” *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 120–128, 2006.
 - [18] J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, S. Aguilera-Navarro, and P. Gómez-Vilda, “An integrated tool for the diagnosis of voice disorders,” *Medical Engineering and Physics*, vol. 28, no. 3, pp. 276–289, 2006.