

# Parameter Estimations for Signal Type Classification of Korean Disordered Voices

JiYeoun Lee

Department of Biomedical Engineering, Jungwon University  
Donbu-ri #5, Goesan-eup, Goesan-gun, Chungcheongbuk-do 367-805, South Korea  
jylee@jwu.ac.kr

**Abstract**— Although many signal-typing studies have been published, they are primarily based on manual inspection and experts' judgments of voice samples' acoustic content. Software may be required to automatically and objectively classify pathological voices into the four signal types and to facilitate experts' opinion formation by providing specific signal type determination criteria. This paper suggests the coefficient of normalized skewness variation (CSV), coefficient of normalized kurtosis variation (CKV), and bicoherence value (BV) based on the linear predictive coding (LPC) residual to categorize voice signals. Its objective is to improve the performances of acoustic parameters such as jitter, shimmer, and the signal-to-noise ratio (SNR) in signal type classification. In this study, the classification and regression tree (CART) was used to estimate the performances of the acoustic, CSV, CKV, and BV parameters by using the LPC residual. In the investigation of acoustic parameters such as jitter, shimmer, and the SNR, the optimal tree generated by jitter alone yielded an average accuracy of 78.6%. When the acoustic, CSV, CKV, and BV parameters together were used to generate the decision tree, the average accuracy was 82.1%. In this case, the optimal tree formed by jitter and the BV effectively discriminated between the signal types. To perform accurate acoustic pathological voice analysis, signal type quantification is of great interest. Automatic pathological voice classification can be an important objective tool as the signal type can be numerically measured. Future investigations will incorporate multiple pathological data in classification methods to improve their performance and implement more reliable detectors.

**Keyword** - Signal typing, Disordered voice, Acoustic analysis, higher-order statistics, Linear predictive coding

## I. INTRODUCTION

In a 1994 summary statement for a voice and speech workshop, Titze proposed that signals should be categorized as type 1, 2, or 3 to identify the nature of the bifurcations in voice signals, and this classification is central to all other considerations in acoustic voice analysis [1]. In the system, type 1 signals are nearly periodic and display no qualitative changes in the analysed segments. Type 2 signals contain qualitative changes (bifurcations), subharmonic frequencies, or modulating frequencies in the analysed segments. Type 3 signals have no apparent periodic structure. The addition of a fourth type of voice to Titze's classification scheme was recently proposed [2]. Type 4 signals are primarily stochastic in behavior.

Sprecher et al. presented an overview of signal typing schemes by using 150 sustained phonations collected from the Kay Elemetrics database [2]. They manually analysed a narrowband spectrogram generated by Praat software to classify each signal type as 1, 2, 3, or 4. A window length of 50 ms, time step of 0.002 s, frequency step of 5 Hz, and dynamic range of 40 dB were employed. A Hamming window shape was used to create the spectrogram. The classification was based on the recommendations outlined by Titze with some modifications [1]. Acoustic, perceptual, and correlation dimension analyses were conducted for all voice types. Various interpretations of the voice types were also introduced.

Lee et al. manually classified Korean pathological voices into types 1, 2, 3, and 4 according to the recommendation of Sprecher et al. [2]-[5]. They confirmed that the differences between the perturbation parameters tended to increase from type 1 to type 4 signals. A sample selection method with a moving window and minimum perturbation scheme were proposed to analyse the typed signals [3].

Thus, although many studies related to signal typing have been published, they are based on visual inspection and the consensus judgment of many experts about the acoustic content of the voice samples [2]-[6]. The advantages of narrowband spectrograms include that they can be obtained reliably for each voice sample and that they can indicate whether further acoustic analyses are appropriate. However, the disadvantages of visual evaluations include that experts differ in their spectrogram interpretations and that it is time-consuming to acquire these judgments because many raters are needed to obtain sufficient inter- and intra-rater reliability. Therefore, software may be required to automatically and objectively classify pathological voices into the four signal types and to help suggest experts' opinions by providing specific criteria by which they can determine the signal type.

Novel parameters for automatic signal typing of pathological voices are proposed in this paper. Many studies have applied higher-order statistics (HOS) to pathological voices since Alonso et al.'s publication on automatic detection of voice pathologies by HOS-based parameters [7]-[9]. Further, the combination of HOS analysis and the linear predictive coding (LPC) residual may help to effectively construct important information to distinguish the signal types of pathological voices [10]-[12]. Although HOS analysis holds promise as one possible method of distinguishing between normal and pathological voices [13], no studies have applied HOS analysis to classify voices into signal types 1, 2, 3, and 4.

The objective of this study was to investigate a technique to enable automatic classification of pathological voices by signal type. Therefore, this investigation examined and assessed the different parameters used in the current literature for analysis of pathological voices by signal processing. Furthermore, new HOS-based parameters in the time and frequency domains are presented in this paper in order to categorize the signal types of pathological voices and to improve the classification performance. It was the first research in automatic signal type classification.

## II. RESEARCH METHODS

### A. Database

The voice samples used in this research were collected in the speech clinic, Otorhinolaryngology of Gangnam Severance Hospital. The database includes the pathological voices of 140 Korean subjects (74 female and 66 male) ranging in age from 19 to 80 years. Sustained vowel /a/ phonations (3.0–4.0 s in length) were used, and all voice data were sampled at 44.1 kHz. The information about the selected data, separated into the four types, is detailed in Table I. The types of diseases included are Reinke's edema, glottic cancer, intra-codal cysts, granuloma, laryngitis, papilloma, SD, vocal fold palsy, vocal nodules, vocal polyps, and sulcus vocalis.

### B. Signal typing

Narrowband spectrograms were generated by using the Praat software. They were created with a window length of 0.05 s, time step of 0.002 s, frequency step of 5 Hz, and dynamic range of 40 dB. A hamming window shape was used to generate the spectrograms [2]. The signal typing was conducted by three trained speech-language pathologists (SLPs) and speech specialists in advance. These specialists are well known in the field of speech disorder and have many years of speech therapy experience. At weekly meetings, the speech samples from each patient were reviewed and signal typing was performed, with the results shown in Table I.

The typical waveform and spectrogram characteristics for each signal type are shown in Figs. 1 and 2. Figure 1 shows the classic waveforms that are characteristic of signal types 1 through 4. The waveforms become increasingly complex for voices of signal types 3 and 4.

In Fig. 2(a), the spectrogram of signal type 1 is nearly periodic and displays no qualitative changes in the analysed segment. The spectrogram of signal type 2 in Fig. 2(b) shows subharmonic or modulating frequencies in the analysed segments. The spectrogram of signal type 3 displays qualitative changes in the analysed segment, as shown in Fig. 2(c). It also shows a smearing of energy across multiple harmonics. Although the fundamental frequency is often apparent, higher harmonics cannot be distinguished. Signal type 4 has no apparent periodic structure, as shown in Fig. 2(d). It also displays destroyed frequency bands with no apparent evidence of subharmonics, modulations, or bifurcations.

TABLE I. Subject characteristics for each signal type group. Displayed are number, age range, and gender ratio.

| Signal type | Number | Age   | Gender (F : M) |
|-------------|--------|-------|----------------|
| Type 1      | 35     | 21–45 | 27 : 8         |
| Type 2      | 35     | 22–70 | 17 : 18        |
| Type 3      | 35     | 17–80 | 17 : 18        |
| Type 4      | 35     | 38–86 | 13 : 22        |

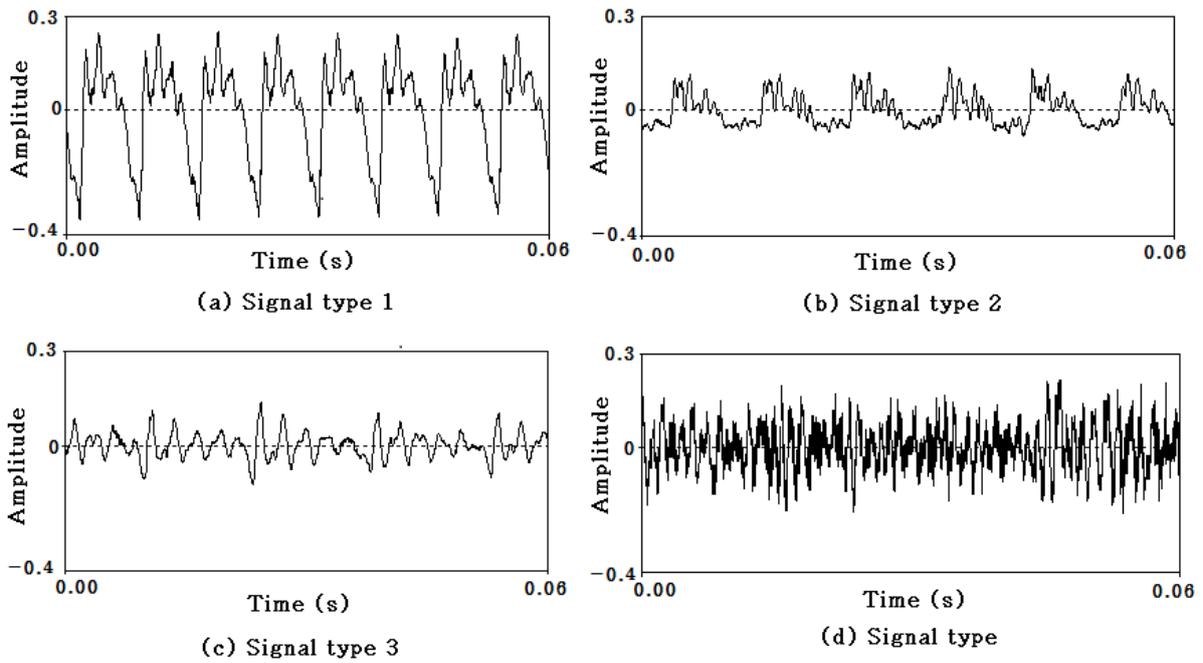


Fig. 1. Four waveforms of voice samples. (a), (b), (c), and (d) are samples of types 1, 2, 3, and 4 voices, respectively.

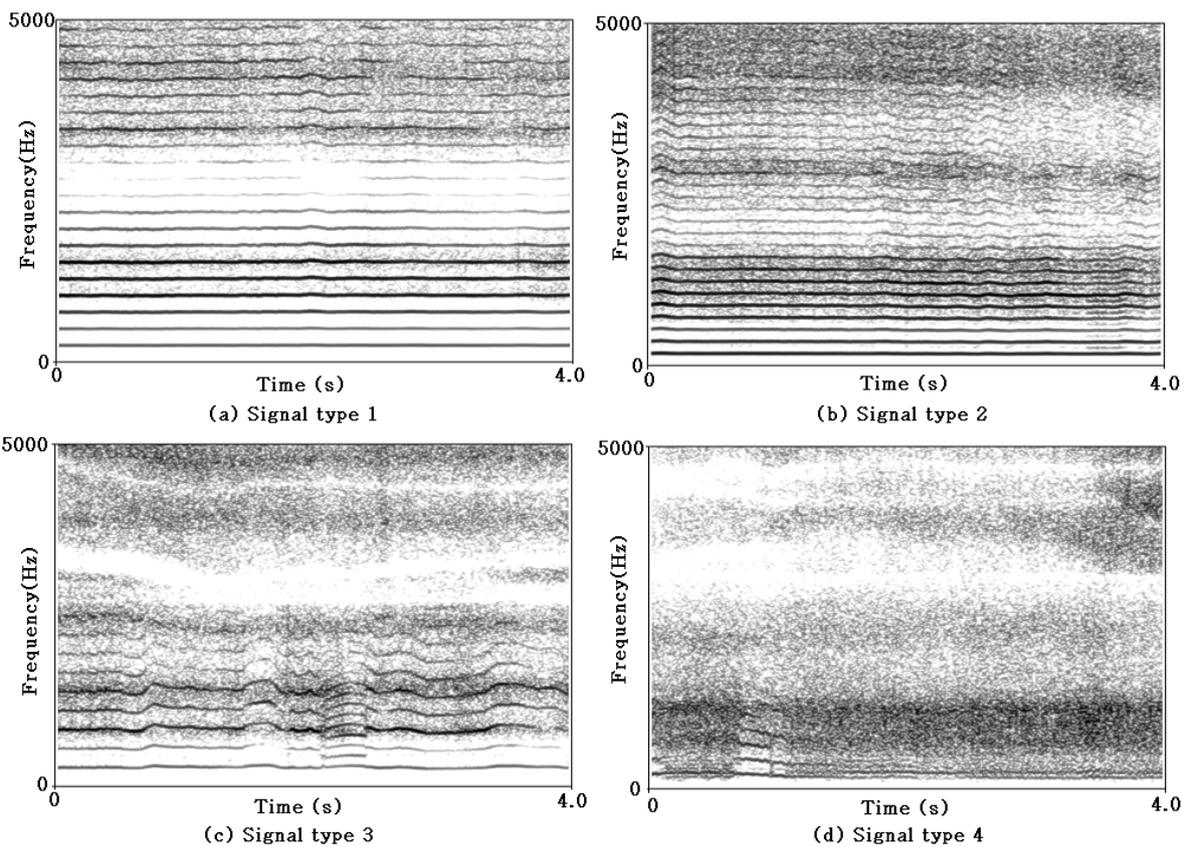


Fig. 2. Spectrograms generated from pathological voice signals. (a), (b), (c), and (d) are classified as types 1, 2, 3, and 4, respectively.

C. Perturbation analysis

Jitter (%), shimmer (%), and the signal-to-noise ratio (SNR) were obtained from TF32 software to implement perturbation analysis [14]-[15]. The perturbation parameters were estimated from an entire portion except for onset and offset. The numbers notated by “Trk” and “Err” are measures of the pitch-tracking reliability provided by TF32. A high “Trk” count signifies dramatic fluctuations in F0, while a high “Err” count indicates a large

discrepancy in the calculated fundamental frequency due to voice breaks in the sample. Therefore, a large “Err” count may exaggerate the jitter and shimmer values and diminish the SNR in what is already unsteady voicing.

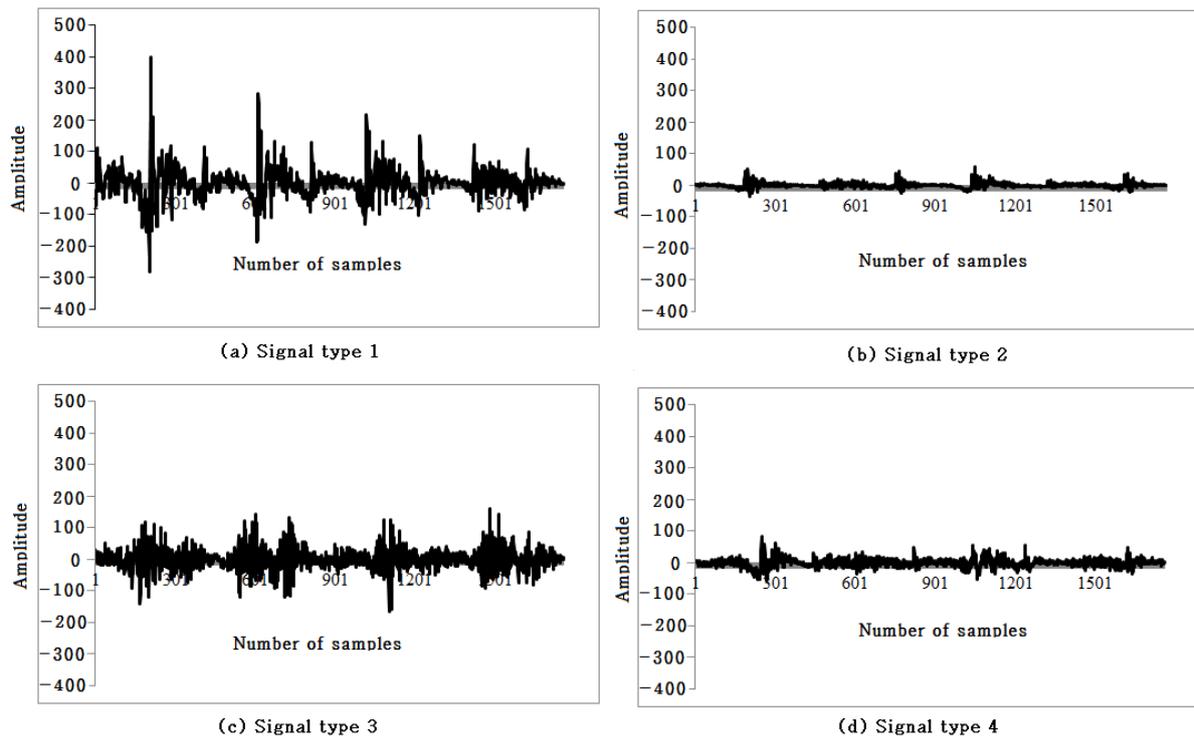


Fig. 3. LPC residuals estimated for signal types 1, 2, 3, and 4.

#### D. Linear predictive coding analysis

Among the researchers studying the relationship between the LPC residual and disordered voices, Lee et al. introduced the LPC characteristics as acoustic parameters to classify normal and pathological voices. They suggested HOS-based parameters obtained from the LPC residual for voice-quality assessment. The presented techniques demonstrated that it is possible to evaluate the extent of larynx diseases [16]-[17]. Therefore, LPC analysis was introduced in this study to investigate its ability to automatically classify the signal type of a pathological voice.

Figure 3 shows the LPC residuals estimated for signal types 1, 2, 3, and 4. In Fig. 3(a), the LPC residuals of signal type 1 seem to be harmonically related. Only some of the residuals are harmonically related in Figs. 3(b) and 3(c). It can be seen that the LPC residuals tend to contain no definite pulses and be unstructured in Fig. 3(d). As the degree is higher, the signals tend to be more irregular, aperiodic, and unstable. Generally the LPC residuals of pathological voices show differences in pulse regularity and noise between the signal types. Therefore, the use of LPC residuals may yield information regarding the abnormal movement of vocal folds and turbulence noise, which is useful for pathological voice type classification.

#### E. Higher-order statistics analysis

HOS analysis in the time domain has shown promise as a classification index for pathological voices and also has the advantage of not requiring a periodic or quasiperiodic voice signal to enable reliable analysis [7]-[9]. Therefore, the application of HOS analysis in speech processing has been primarily motivated by its inherent Gaussian suppression and phase preservation properties. It may also be one possible acoustic marker that is sensitive to voice impairment [13][16]-[17].

In this study, the normalized skewness ( $\gamma_{3t}$ ) and kurtosis ( $\gamma_{4t}$ ) were extracted in-frame, as shown in Eq. (1):

$$\gamma_{3t} = \frac{\sum_{n=1}^N x_r^3(n)}{[\sum_{n=1}^N x_r^2(n)]^{1.5}}, \quad \gamma_{4t} = \frac{\sum_{n=1}^N x_r^4(n)}{[\sum_{n=1}^N x_r^2(n)]^2}, \quad (1)$$

where  $x(n)$  is the speech sample value of the  $t^{\text{th}}$  frame, and  $N$  is the number of samples.

The proposed parameters are the coefficients of the normalized skewness variation (CSV) and of the normalized kurtosis variation (CKV), as indicated in Eq. (2):

$$CSV = \frac{\sqrt{\frac{1}{T-1} \sum_{t=1}^T (\gamma_{3t} - \bar{\gamma}_3)^2}}{\bar{\gamma}_3}, \quad CKV = \frac{\sqrt{\frac{1}{T-1} \sum_{t=1}^T (\gamma_{4t} - \bar{\gamma}_4)^2}}{\bar{\gamma}_4}, \quad (2)$$

where  $\gamma_{3t}$  and  $\gamma_{4t}$  are extracted in the  $t^{th}$  frame, and  $T$  is the number of the frames. Here,  $\bar{\gamma}_3$  and  $\bar{\gamma}_4$  are the means of the normalized skewness and kurtosis estimated from the analysed frames.

HOS analysis in the frequency domain is mainly applied by using a bispectrum or trispectrum. A bispectrum is a function of two frequencies, whereas a trispectrum is a function of three frequencies. Both spectrums contain complex values. This investigation utilized the bispectrum shown in Eq. (3):

$$B(f_1, f_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} C_{3x}(\tau_1, \tau_2) e^{-j \sum_{i=1}^2 f_i \tau_i}, \quad (3)$$

where  $C_{3x}(\tau_1, \tau_2)$  are the third-order cumulants and are defined by Eq. (4):

$$C_{3x}(\tau_1, \tau_2) = E\{x(t)x(t+\tau_1)x(t+\tau_2)\}. \quad (4)$$

Given estimates of the power spectra and bispectrum, the bicoherence can be estimated as indicated in Eq. (5):

$$bic_{auto}(f_1, f_2) = \frac{S_{xxx}(f_1, f_2)}{\sqrt{S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f_1+f_2)}}, \quad (5)$$

where  $S_{xx}$  and  $S_{xxx}$  are the power spectrum and bispectrum, respectively.

The proposed bicoherence value (BV) parameter results from the bicoherences. Thus the BV parameter is defined by Eq. (6):

$$BV = \left(\frac{1}{N-1}\right) \sum_{f_1=0}^{\dim_x-1} \sum_{f_2=0}^{\dim_y-1} bic(f_1, f_2), \quad (6)$$

where the dimensions of the bicoherence value are  $\dim_x$  and  $\dim_y$ , and  $N$  is the product of  $\dim_x$  and  $\dim_y$ .

#### F. Statistical analysis

Statistical analysis was conducted by using Sigma Stat 3.0 (Jandel Scientific, San Rafael, CA). One-way repeated-measure ANOVAs on ranks were performed to test the differences between the type 1, 2, 3, and 4 signals for each parameter of interest. An alpha of 0.05 was employed for all comparisons. Multiple pairwise comparisons were conducted with the Turkey method.

#### G. Classification and regression tree analysis

The classification and regression tree (CART) algorithm is commonly used to build statistical models through a tree-based technique. Most importantly, the optimal decision tree contains rules that are more easily readable by humans than are those of other classification and regression methods, such as vector quantization (VQ), neural networks (NNs), the Gaussian mixture model (GMM), and so on. The CART algorithm is also considered to be a solution that can effectively combine multiple parameters [18]-[20]. Since each parameter has separate characteristics and independently does not always guarantee reliable performance in various kinds of environments, it may be necessary to combine these parameters to ensure robustness in various conditions. Therefore, this investigation used the CART algorithm to classify the signal types by using multiple parameters.

### III. RESULTS AND ANALYSIS

#### A. Distribution of acoustic parameters

Figures 4(a), 4(b), 4(c), and 4(d) show the distributions of the "Err" count, jitter (%), shimmer (%), and SNR (dB) extracted from signal types 1, 2, 3, and 4, respectively. These figures were obtained by calculating the minimum, first quartile, median, third quartile, maximum values, and outliers of the acoustic parameters. The "Err" values increase significantly from type 1 to type 4 signals. It can be unreliable to analyse jitter, shimmer, and the SNR in a severely chaotic pathological voice, such as one exhibiting signal type 4.

As can be seen from Figs. 4(b) and 4(c), there is little difference in the distributions of acoustic parameters such as jitter and shimmer between type 1 and type 2 signals. However, for type 3 and 4 signals, these parameters tend towards higher values and slightly broader distributions than those for type 1 and 2 signals. Similarly, the SNR decreases from type 1 through type 4 voices, indicating that the evidence of harmonics decreases as the signal type increases.

**B. Distributions of CSV, CKV, and BV parameters**

Figure 5 shows the characteristics of the CSV and CKV estimated from the LPC residuals in type 1, 2, 3, and 4 signals. To extract an LPC residual, the pathological signal was divided into frames of 20 ms with an overlap of 10 ms. Each frame was then windowed by using a hamming window, the 12<sup>th</sup>-order autocorrelation coefficients were found, and then the reflection coefficients (12<sup>th</sup> order) were calculated from the autocorrelation coefficients by using the Levinson-Durbin algorithm. The original signal was passed through an all-zero analysis filter with the coefficients being the reflection coefficients obtained earlier. The output of the filter was the residual signal. In the case of a pathological voiced segment, not all of the harmonics may be related, and as a result, the CSV values may be zero or greater than zero for any practical value of the pitch.

Figures 5(a) and 5(b) present the CSV and CKV parameters in the form of box plots to provide better visualizations of signal types 1, 2, 3, and 4. As may be seen from Fig. 5(a), the CSV distribution is greater than zero for the type 1 signal, which is more harmonic than those of the other signal types, as expected. There is little difference between the CSV distributions of type 1, 2, 3, and 4 signals because each voice contains independent dynamic variation. However, there is a clear distinction between the CSV distribution of the type 4 signal and those of the other signal types. In Fig. 5(b), there are clear differences between the CKV parameter distributions of the different typed signals. In particular, in type 3 and 4 signals, the distributions tend towards higher values and are slightly broader than those of type 1 and 2 signals, indicating that the evidence of harmonics decreases as the signal type increases. Overall, it is evident that the CKV is sufficiently distinct to discriminate between the typed signals and can be used as a basis for automatic classification of signal types.

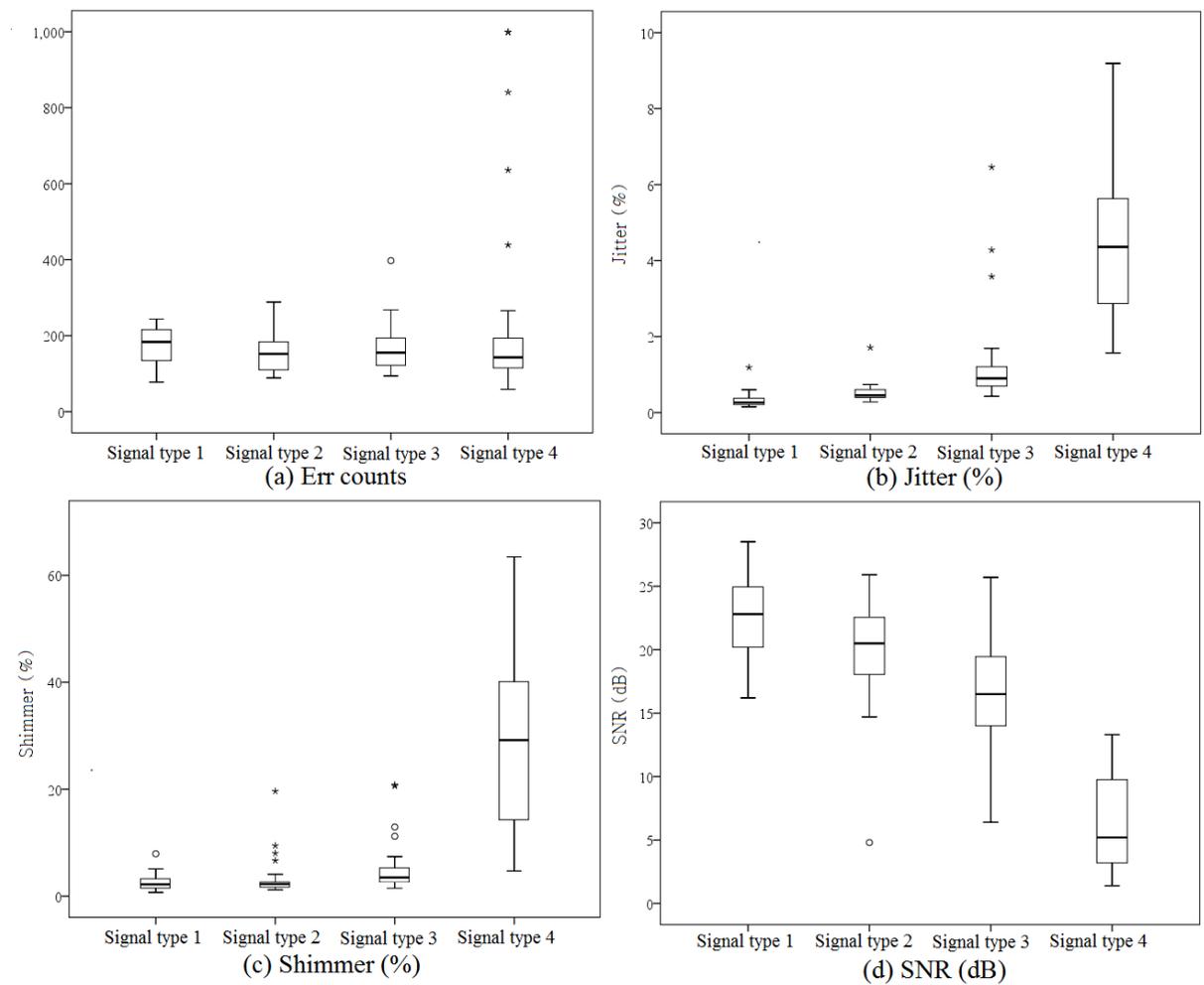


Fig. 4. Distributions of acoustic parameters estimated from each signal type. (a) "Err" counts, (b) jitter (%), (c) shimmer (%), and (d) SNR (dB).

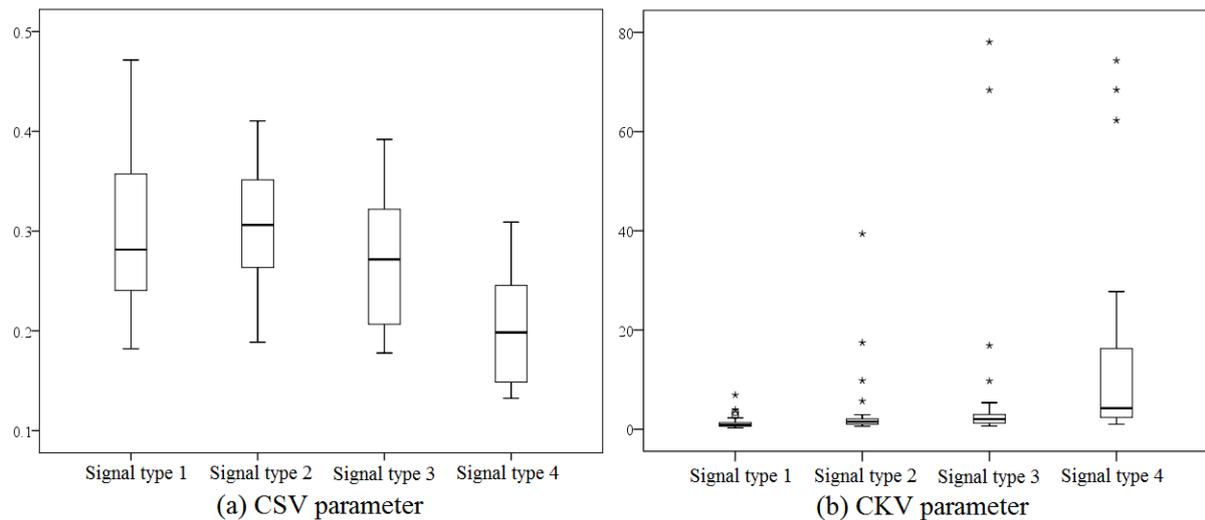


Fig. 5. Distributions of CSV and CKV parameters of different signal types. (a) and (b) are CSV and CKV distributions extracted from signal types 1 and 2 and from signal types 3 and 4, respectively.

Figures 6(a), 6(b), 6(c), and 6(d) present clear distinctions between the bicoherence distributions of the different signal types. In the bicoherence distribution of signal type 1, the greatest amplitude is observed, and the distribution tends towards higher values and a more assembled structure do than those of the other signal types. Overall, the distributions become broader and more dispersed appearance as signal type increases, while the amplitudes tend to decrease significantly from type 1 to type 4 signals.

Figure 7 shows the characteristics of the BV parameters estimated for signal types 1, 2, 3, and 4. The BV values decrease significantly as signal type increases, similarly to the behavior shown in Fig. 6. These distributions also presents clear distinctions between the signal types.

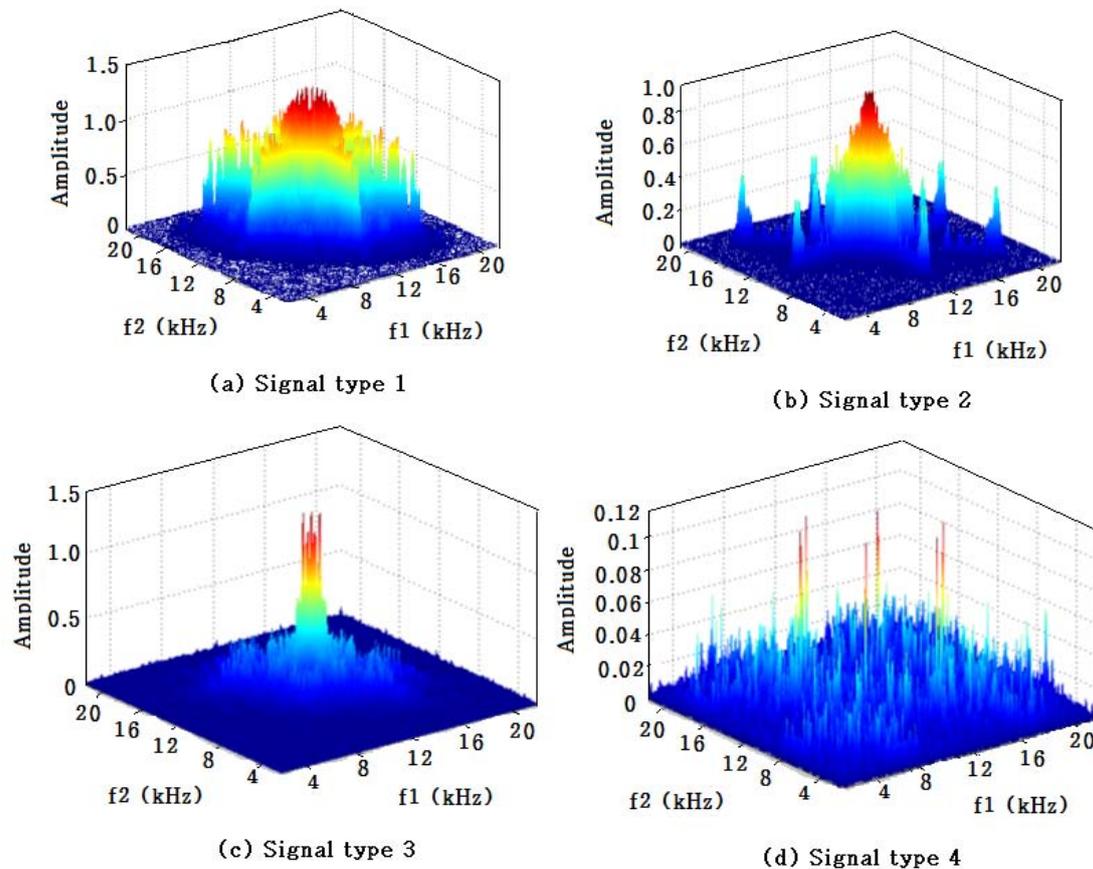


Fig. 6. Bicoherence distributions of different signal types. (a), (b), (c), and (d) are bicoherence aspects estimated from samples of signal types 1, 2, 3, and 4, respectively.

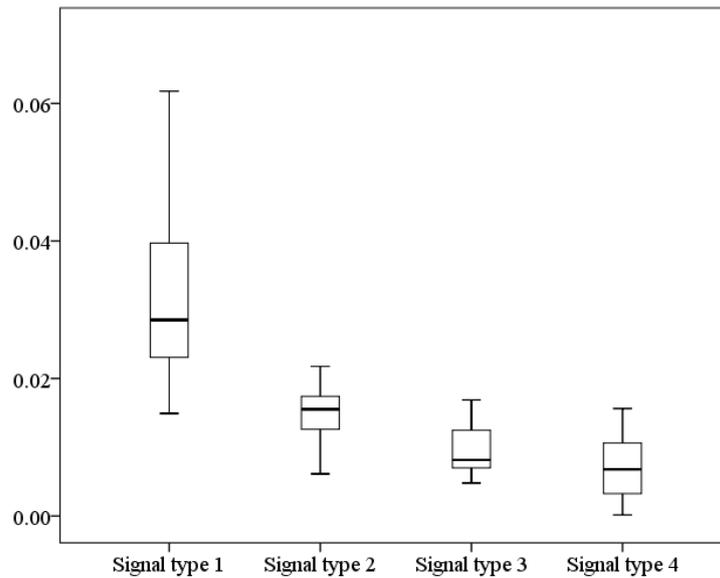


Fig. 7. Distributions of BV parameter among signal types.

TABLE II. Statistical analysis of signal types.

|                                 | Signal type 1 | Signal type 2 | Signal type 3 | Signal type 4 |
|---------------------------------|---------------|---------------|---------------|---------------|
| <b>Jitter (%) (p = 0.000*)</b>  |               |               |               |               |
| Signal type 1                   | -             | 0.990         | 0.434         | 0.000*        |
| Signal type 2                   | -             | -             | 0.624         | 0.000*        |
| Signal type 3                   | -             | -             | -             | 0.000*        |
| Signal type 4                   | -             | -             | -             | -             |
| <b>Shimmer (%) (p = 0.000*)</b> |               |               |               |               |
| Signal type 1                   | -             | 0.989         | 0.532         | 0.000*        |
| Signal type 2                   | -             | -             | 0.729         | 0.000*        |
| Signal type 3                   | -             | -             | -             | 0.000*        |
| Signal type 4                   | -             | -             | -             | -             |
| <b>SNR (dB) (p = 0.000*)</b>    |               |               |               |               |
| Signal type 1                   | -             | 0.033*        | 0.000*        | 0.000*        |
| Signal type 2                   | -             | -             | 0.000*        | 0.000*        |
| Signal type 3                   | -             | -             | -             | 0.000*        |
| Signal type 4                   | -             | -             | -             | -             |
| <b>CSV (p = 0.000*)</b>         |               |               |               |               |
| Signal type 1                   | -             | 0.960         | 0.291         | 0.000*        |
| Signal type 2                   | -             | -             | 0.110         | 0.000*        |
| Signal type 3                   | -             | -             | -             | 0.000*        |
| Signal type 4                   | -             | -             | -             | -             |
| <b>CKV (p = 0.002*)</b>         |               |               |               |               |
| Signal type 1                   | -             | 0.913         | 0.301         | 0.002*        |
| Signal type 2                   | -             | -             | 0.693         | 0.013*        |
| Signal type 3                   | -             | -             | -             | 0.199         |
| Signal type 4                   | -             | -             | -             | -             |
| <b>BV (p = 0.000*)</b>          |               |               |               |               |
| Signal type 1                   | -             | 0.000*        | 0.000*        | 0.000*        |
| Signal type 2                   | -             | -             | 0.681         | 0.003*        |
| Signal type 3                   | -             | -             | -             | 0.071         |
| Signal type 4                   | -             | -             | -             | -             |

### C. Statistical analysis

Table II shows the statistical analysis results from the different signal types. The purpose of a repeated-measure ANOVA is to determine if significant differences exist in the dependent variables such as jitter, shimmer, SNR, CSV, CKV, and BV that are used for automatic signal type classification. In Table II, since all the parameters still show p-values  $<.05$ , it is said that they are useful and meaningful for signal type classification. In particular, SNR and BV perform better statistically than do the other parameters.

### D. Classifier

The CART software used in this study is a commercial product manufactured and sold by Salford Systems, San Diego, CA. Importantly, all of the CART data management functions are performed by the software. For classification trees, the default rule is the Gini, which generally works well across a broad range of problems. In this experiment, tree growth was prevented beyond a depth of 7 levels, and trees were pruned back by using cross-validation to select the optimal tree size. Once the largest tree was constructed, the relative error was shown in the profile section. Then, the optimal tree was chosen to be that which minimized cost regardless of size. The performance was measured by a 5-fold cross-validation scheme.

The first experiment was conducted to measure the performance of acoustic parameters such as jitter, shimmer, and the SNR. Although the error counts were high for signal types 3 and 4, these signal types were included in this experiment because their distributions are clearly distinguishable, as shown in Fig. 4. Figure 8 shows the optimal decision tree generated by using jitter as the only acoustic parameter. It is believed that jitter is very useful for signal type classification. The average classification performance among signal types is also 78.6%. Secondly, when the acoustic, CSV, CKV, and BV parameters were used to generate the decision tree, the average performance was 82.1%. Figure 9 presents the optimal decision tree formed by the acoustic, CSV, CKV, and BV parameters. In this case, it is believed that jitter and the BV effectively discriminate between the signal types. The characteristic distributions of the various parameters shown in Figs. 4, 5, 6 and 7 are also reflected at each of the tree nodes.

The performance matrix in Table III was formed from the decision tree shown in Figs. 8 and 9. Table III compares the results obtained from the acoustic parameters and from combinations of the acoustic, CSV, CKV, and BV parameters. The diagonal numbers indicate the performances of correctly classified signals. The off-diagonal elements are associated with the performances of misclassifications. The best average performance is 82.1% when the combination of the acoustic and HOS-based parameters was utilized to classify the signal type as 1, 2, 3, and 4.

TABLE III. Classification performance (accuracy, %)

|   |               | Predicted classification |                      |                   |                      |
|---|---------------|--------------------------|----------------------|-------------------|----------------------|
|   |               | Signal type 1            | Signal type 2        | Signal type 3     | Signal type 4        |
| Actual classification                     | Signal type 1 | ** <b>62.86%</b>         | 31.42%               | 5.71%             | 0%                   |
|   |               | *** <b><u>85.71%</u></b> | 11.42%               | 2.86%             | 0%                   |
|   | Signal type 2 | ** 2.86%                 | <b>74.28%</b>        | 20.00%            | 2.86%                |
|   |               | *** 11.42%               | <b><u>85.71%</u></b> | 0%                | 5.71%                |
|   | Signal type 3 | ** 0%                    | 8.57%                | <b><u>80%</u></b> | 11.43%               |
|   |               | *** 8.57%                | 17.14%               | <b>62.86%</b>     | 11.42%               |
|   | Signal type 4 | ** 0%                    | 0%                   | 2.86%             | <b><u>97.14%</u></b> |
|   |               | *** 0%                   | 0%                   | 5.71%             | <b>94.29%</b>        |
| ** Acoustic parameters                    |               |                          |                      |                   |                      |
| *** Acoustic, CSV, CKV, and BV parameters |               |                          |                      |                   |                      |

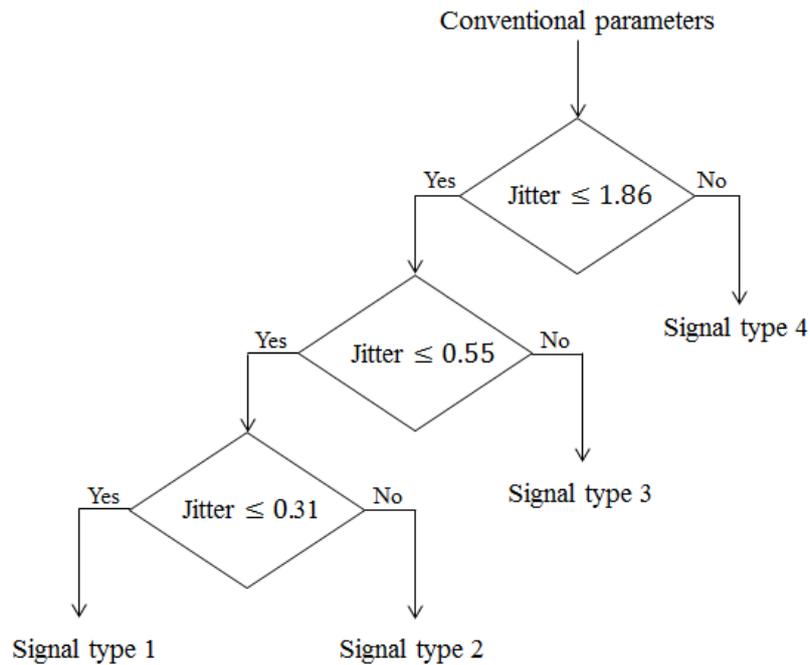


Fig. 8. Optimal decision tree generated by acoustic parameters.

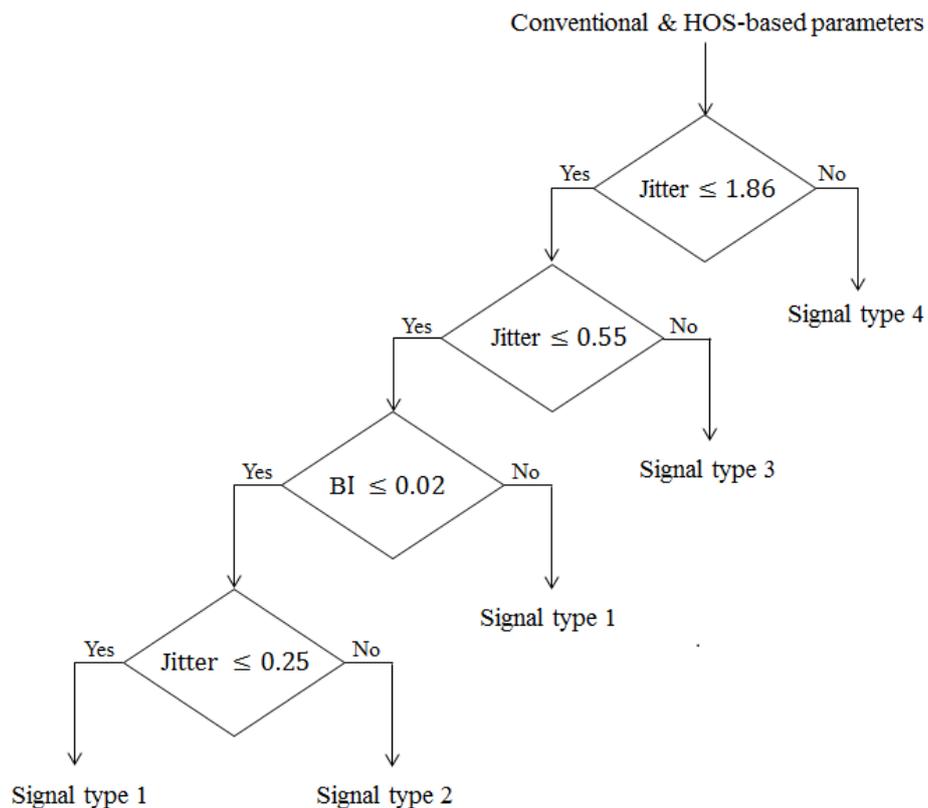


Fig. 9. Optimal decision tree generated by acoustic, CSV, CKV, and BV parameters.

**IV. CONCLUSION**

Since the voice and speech workshop in 1994, acoustic signal analysis has consisted of manual categorization as signal type 1, 2, 3, or 4. However, manual evaluation has the disadvantage that experts differ in their spectrogram interpretations. This process is also time consuming because many raters are needed to obtain sufficient inter- and intra-rater reliability. Therefore, it is necessary to implement a software-based method to automatically and objectively classify signals as type 1, 2, 3, or 4.

As shown in previous works which have attempted to exploit the skewness and kurtosis of speech signals, HOS analysis has shown promising results in pathological voice signal processing applications [7]-[9][13][16]-[17]. The LPC residual is also widely known as a parameter to that can distinguish between normal and pathological voices [16]-[17]. Therefore, in this study, a new parameter to describe the characteristics of signal types 1, 2, 3, and 4 was introduced by using HOS analysis and the LPC residual.

In this paper, the normalized skewness and kurtosis were examined for signal types 1, 2, 3, and 4, and strong correlations between the HOS coefficients and signal type were demonstrated. Therefore, it is suggested to use the coefficient of the normalized skewness variation (CSV), coefficient of the normalized kurtosis variation (CKV), and bicoherence value (BV) based on the LPC residual to categorize voice signals in order to improve the performance of the acoustic parameters used to determine the signal types.

The CART algorithm was used to estimate the performances of acoustic parameters such as jitter, shimmer, the SNR, the CSV, the CKV, and the BV by using the LPC residual. Since some parameters make accurate signal type classifications and some do not, it is necessary to design a rule to make the final decision by simultaneously incorporating multiple parameters in the classifiers.

In the experiment that incorporated the acoustic parameters such as jitter, shimmer, and the SNR, the optimal tree generated by only jitter yielded an average accuracy of 78.6%. It is believed that jitter is very useful for the automatic signal type classification. Next, when the acoustic, CSV, CKV, and BV parameters together were used to generate the decision tree, the average accuracy was 82.1%. In this case, the optimal decision tree formed only by jitter and the BV effectively discriminated between the signal types, confirming that the BV parameter using the LPC residual can be successfully used to classify signal types.

To acoustically analyse pathological voices, quantifying the signal type is of great interest. Applying automatic classification techniques to pathological voices can be an important objective method, as numerical values can be evaluated to determine the signal type. This method is attractive for its potential ability to detect illnesses early and for its applicability in telemedicine. Future investigations will develop classification methods to improve performance and implement more reliable detectors.

#### ACKNOWLEDGMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No. 2014-00540001).

#### REFERENCES

- [1] I. R. Titze, "Workshop on acoustic voice analysis: summary statement," National Center for Voice and Speech, pp. 1-36, 1995.
- [2] A. Sprecher, Y. Zhang, and A. Olszewski, "Updating signal typing in voice: addition of type 4 signals," *J. Acoust. Soc. Am.*, vol. 127, pp. 3710-3716, 2010.
- [3] J. Y. Lee and S. H. Choi, "Perturbation analysis using a moving window for disordered voices," *International Journal of Engineering, Science, and Innovative Technology*, vol. 3(1), pp. 1-10, 2012.
- [4] J. Y. Lee, S. H. Choi, J. J. Jiang, M. S. Hahn, and H. S. Choi, "Perturbation and Perceptual Analysis of Pathological Sustained Vowels According to Signal Typing," *Malsori and Speech Science*, vol. 2(2), pp. 109-115, 2010.
- [5] J. Y. Lee, S. H. Choi, J. J. Jiang, M. S. Hahn, and H. S. Choi, "Perturbation and Nonlinear Dynamical Analysis of Sustained Vowels in Normal and Pathological voices with Signal Typing," *Malsori and Speech Science*, vol. 2(1), pp. 113-120, 2010.
- [6] A. Behnman, C. Agresti, E. Blumstein, and N. Lee, "Microphone and electroglottographic data from dysphonic patients: Type 1, 2 and 3 signals," *J. Voice*, vol. 12, pp. 249-260, 1998.
- [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. 4, pp. 275-284, 2001.
- [8] J. Y. Lee, S. Jeong, and M. S. Hahn, "Pathological Voice Detection Using Efficient Combination of Heterogeneous Features," *IEICE Transactions on Information and Systems*, vol. E91-D(2), pp. 367-370, 2008.
- [9] J. Y. Lee, S. Jeong, H. S. Choi, and M. S. Hahn, "Objective pathological voice quality assessment based on HOS features," *IEICE Transactions on Information and Systems*, vol. E91-D(12), pp. 2888-2891, 2008.
- [10] E. Nemer, R. Goubran, and S. Goubran, "Robust voice activity detection using higher-order statistics in the LPC residual domain," *IEEE Transaction Speech and Audio Processing*, vol. 9(3), pp. 217-231, 2001.
- [11] M de Oliveira Rosa, J. C. Pereira, and M. Grellet, "Adaptive estimation of residual signal for voice pathology diagnosis," *IEEE Transaction on Biomedical Engineering*, vol. 47(1), pp. 96-104, 2000.
- [12] 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. 9, pp. 1400-1409, 2005.
- [13] J. Y. Lee, "A two-stage approach using Gaussian mixture models and higher-order statistics for a classification of normal and pathological voices," *Advances in Signal Processing on Euraship*, vol. 252, <http://asp.eurasipjournals.com/content/2012/1/252>, 2012.
- [14] P. Milenkovic, "TF32 User's Manual". Madison, WI, 2001.
- [15] D. K. Raymond and J. B. Martin, "Voice Quality Measurement," Thomson Learning, pp. 134-159, 2007.
- [16] J. Y. Lee and M. S. Hahn, "Automatic Assessment of Pathological Voice Quality Using Higher-Order Statistics in the LPC Residual Domain," *EURASIP Journal on Advances in Signal Processing*, Article ID 748207, 2009.
- [17] J. Y. Lee, S. B. Jeong, M. S. Hahn, A. Sprecher, and J. J. Jiang, "An efficient approach using HOS-based parameters in the LPC residual domain to classify breathy and rough voices," *Biomedical Signal Processing and Control*, vol. 6(2), pp. 186-196, 2011.
- [18] L. Breiman, J. H. Friedman, and R. A. Olshen, "Classification and Regression Trees," Wadsworth, Monterey, CA, USA, 1984.
- [19] S. Gey and E. Nedelec, "Model selection for CART regression trees," *IEEE Transactions on Information Theory*, vol. 51(2), pp. 658-670, 2005.
- [20] F. Questiera, R. Puta, D. Coomansa, B. Walczaka, and Y. Vander Heydena, "The use of CART and multivariate regression trees for supervised and unsupervised feature selection," *Chemometrics and Intelligent Laboratory Systems*, vol. 76(1), pp. 45-54, 2005.

**AUTHOR PROFILE**

JiYeoun Lee obtained B.S. in division of Electronics and Computer Engineering from Hanyang University in the year of 2001, M.S. and Ph.D in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST) in the year of 2003 and 2008. Dr. JiYeoun Lee is now an assistant professor at the faculty of Biomedical Engineering, Jungwon University. Her research interest fields are Biomedical Signal Processing, Noise Reduction, Sound Localization, and Audio Signal Processing.