Class-Dependent Acoustic Features for Support Vector Machine Based Consonant Class Discrimination in Dysarthric Speech

Woo Kyeong Seong and Ji Hun Park

School of Information and Communications
Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea

{wkseong, jh_park }@gist.ac.kr

Abstract

In this paper, class-dependent acoustic features are investigated to discriminate consonant class in dysarthric speech. In dysarthric speech, imprecise articulation causes different distortion of consonants depending on each consonant class. For this reason, discrimination of consonant class can play an important role in dysarthric speech processing as a preprocessing step. Therefore, to discriminate each consonant into one of five different classes such as stop, fricative, affricate, nasal, and glide, the discrimination is performed based on a support vector machine (SVM) is employed, which is constructed by using the class-dependent acoustic features combined with mel-frequency cepstral coefficients (MFCCs). It is shown from the discrimination experiments that an SVM using the class-dependent acoustic features relatively reduces average discrimination error rate by 7.67%, compared to that using only MFCCs.

Keywords: *Consonant class discrimination, dysarthric speech, support vector machine, class-dependent acoustic features*

1. Introduction

Dysarthria involves speech impairment due to damaged control of the oral, pharyngeal, or laryngeal articulators [1]. Due to the impairment of speech, it is difficult to apply conventional speech processing techniques, such as noise suppression [2-4] or automatic speech recognition (ASR) [5-9] to dysarthric speech. To deal with dysarthric speech, for dysarthric ASR, error-correction techniques based on modeling pronunciation variations have been developed to increase dysarthric ASR performance [10-11]. However, the impairment of dysarthric speech distorts consonants more than vowels due to the imprecise movements of the articulators [12], and the degree of distortion varies depending on the consonant class [13], which substantially causes the pronunciation variation modeling in [10-11] to be under-estimated. Thus, it would be better to discriminate consonant class as a preprocessing step of pronunciation variation models in dysarthric ASR.

As an attempt to recognize a consonant of dysarthric speech, a method was proposed based on a support vector machine (SVM) [14], where an average frame error rate of 5.5% was achieved. The reason why the method in [14] achieved such a low error rate was in that only a subset of consonants was considered for the classification. Thus, it is challengeable to recognize individual consonant spoken by dysarthric speakers [15]. In particular, consonant class discrimination requires additional acoustic features to reflect the characteristics of each consonant class.

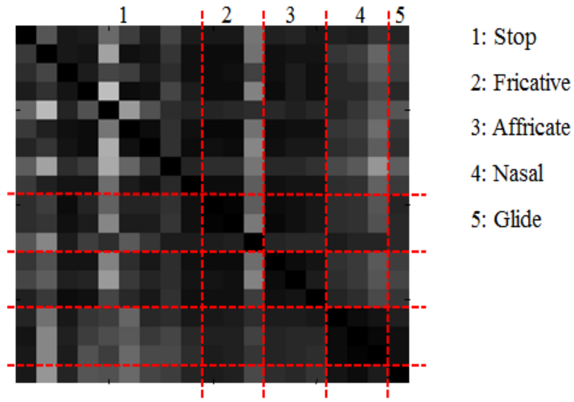

Figure 1. Illustration of a consonant confusion matrix of dysarthric speech

In this paper, class-dependent acoustic features are proposed for consonant class discrimination in dysarthric speech, where each consonant is classified into one of five different classes according to the manner of articulation such as stop, affricate, fricative, nasal and glide. In particular, each class-dependent acoustic feature is dedicated to a specific consonant class. After that, an SVM using the combination of mel-frequency cepstral coefficients (MFCCs) and class-dependent acoustic features is constructed to classify each consonant frame of dysarthric speech. Note that the reason why an SVM is used as a classifier is its reliable performance in classification problems [16] rather than other classifiers such as the Gaussian mixture model or a neural network [17-18].

The remainder of this paper is organized as follows. In Section 2, the effects of pronunciation variations on consonant class in dysarthric ASR are briefly described. In Section 3, the proposed class-dependent acoustic features are described in detail. Section 4 presents a consonant class discrimination system based on an SVM. In Section 5, the performance of the proposed method is analyzed. Finally, this paper is concluded in Section 6.

2. Effects of Pronunciation Variations on Consonant Class

In dysarthric ASR systems, error-correction methods have been applied to increase dysarthric ASR performance. Specifically, error-correction can be achieved by modeling pronunciation variations on a phoneme-level based on the Kullback-Leibler (KL) distance between acoustic models [11]. However, the pronunciation variation models may not accurately reflect the characteristics of dysarthric speech, where various pronunciation variations exist depending on consonant class. Figure 1 illustrates a confusion matrix for consonants of dysarthric speech, in which off-diagonal elements represent confusions of a phoneme where the gray-scale of each element represents the degree of confusability (a darker gray-scale indicates a lower degree of confusability). As shown in the figure, the degree of confusability is shown to differ according to consonant class. In this example, in stop, fricative, and affricate, confusability of stop, fricative, and affricate sounds is higher than that of nasal and glide sounds. It can be inferred that unvoiced sounds in stop, fricative, and affricate are more vulnerable to be mis-recognized because their characteristics are similar to that of noise. This demonstrates that the performance of error-correction based on pronunciation variation models can be improved by weighting pronunciation variation models according to the degree of errors on consonant class in the error-correction method. This implies that a study for discriminating consonant class is essential as a preliminary research for dysarthric ASR.

Table 1. Proposed class-dependent acoustic features for consonant class discrimination

Consonant class	Acoustic feature
Stop	Maximum normalized spectral slope [19]
	Spectral center of gravity with high-band [20]
	Spectral flatness [19]
Fricative	Zero crossing rate [21]
	Log root mean square energy [22]
	Spectral center of gravity [20]
Affricative	Burst degree [23]
	Bisector frequency [23]
	Band energy ratio [22]
Nasal	Energy ratio [24]
	Spectral peak frequency [24]
	Sum of magnitude difference [24]
Glide	Envelope variance measure [24]
	Sonorant measure [25]
	Short-time peak-to-peak [26]

3. Proposed Class-Dependent Acoustic Features

In this section, the proposed class-dependent acoustic features are described for discrimination of consonant classes. In particular, three acoustic features are defined for each class, which results 15-dimensional acoustic features. Each acoustic feature reflects distinctive characteristics of its corresponding consonant class.

3.1. Acoustic Features for Stops

In order to discriminate stops from other consonant classes, three different acoustic features are proposed, as shown in the first row of Table 1. First, maximum normalized spectral slope (MNSS) [19] is calculated as

$$MNSS(i) = \frac{\max_{n}\left\{ \left|X(n;i)\right|^2 - \left|X(n-1;i)\right|^2 \right\}}{\sum_{n=1}^{N} \left|X(n;i)\right|^2} \tag{1}$$

where $X(n;i)$ is a spectral component of the n-th frequency bin at the i-th frame, and N indicates the number of points of fast Fourier transform (FFT). In this paper, N is set to 256. In addition to MNSS, center of gravity with high-band (SCGH) [20] and spectral flatness (SF) [19] are used to discriminate stops from fricatives, spectral, which are defined as

$$SCGH(i) = \frac{\sum_{n=N_H}^{N-1} f(n) X(n;i)}{\sum_{n=N_H}^{N-1} X(n;i)}, \tag{2}$$

$$SF(i) = \frac{\exp(\frac{1}{N}\sum_{n=0}^{N-1}\ln\Gamma(n;i))}{\frac{1}{N}\sum_{n=0}^{N-1}\Gamma(n;i)} \tag{3}$$

where $f(n)$ is a center frequency of the *n*-th frequency bin and N_H is the minimum frequency of high-band. In (3), $\Gamma(n;i)$ indicates a power spectrum of the *n*-th frequency bin at the *i*-th frame.

3.2. Acoustic Features for Fricatives

For discriminating a fricative sound from the others, zero crossing rate (ZCR), log root mean square energy (LRMS), and spectral center of gravity (SCG) are chosen as shown in the second row of Table 1. These three acoustic features reflect voiced and unvoiced characteristics as well as fricative and affricative discrimination. First of all, ZCR is defined as [21]

$$ZCR(i) = \frac{1}{N-1}\sum_{n=1}^{N-1}\frac{|\mathrm{sgn}(x(n;i)) - \mathrm{sgn}(x(n-1;i))|}{2} \tag{4}$$

where $x(n;i)$ refers to the *n*-th speech sample in time domain at the *i*-th frame, N is a frame length, and $\mathrm{sgn}(x)$ is equal to 1 for $x \geq 0$ but $\mathrm{sgn}(x) = -1$ otherwise. Second, LRMS is given as [22]

$$LRMS(i) = \log\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x(n;i)^2} \tag{5}$$

which becomes smaller for an unvoiced sound. Finally, SCG which reflects unvoiced characteristics is defined as [20]

$$SCG(i) = \frac{\sum_{n=0}^{N-1} f(n)X(n;i)}{\sum_{n=0}^{N-1} X(n;i)} \tag{6}$$

which becomes larger for an unvoiced sound.

3.3. Acoustic Features for Affricates

When affricate sounds are pronounced, stop and fricative characteristics are involved due to the nature of affricate. Particularly, the burst sound during affricates is measured by burst degree (BD) [23], and it is calculated as

$$BD(i) = \frac{W_1 \frac{1}{avg(local\ max\ interval)} + W_2 \log E}{W_1 + W_2} \tag{7}$$

where W_1 and W_2 are the weighting factors and they are set 4 and 1 in this paper, respectively. In (7), *local max interval* refers to the distance between neighboring local maxima.

As a second feature for affricates, bisector frequency (BF) [23] is computed as

$$BF(i) = \frac{freqIndex}{N} F_s \tag{8}$$

where F_s is a sampling frequency and it is set to 16 kHz in this paper. In addition, $freqIndex$ indicates the frequency bin bisecting the energy of whole frequency bin, and it is defined as

$$freqIndex = \arg\min_{1<k<N} \left| \sum_{n=1}^{k} |X(n;i)| - \frac{1}{2} \sum_{n=1}^{N} |X(n;i)| \right|. \tag{9}$$

In order to discriminate affricates from fricatives, band energy ratio (BER) [22] is calculated as

$$BER(i) = 10 \log_{10} \frac{E_{B_H}}{E_{B_L}} \tag{10}$$

where E_{B_H} and E_{B_L} represent band energies between a high band, B_H, and a low band, B_L, that occupy 5~10 kHz and 0.5~3 kHz, respectively, in this paper.

3.4. Acoustic Features for Nasals

In order to discriminate nasal sounds from the others, voiced characteristics can be included in the acoustic features. To this end, energy ratio between low-band and high-band (ER) [24] is first defined as

$$ER(i) = \frac{\sum_{n=1}^{N_L} |X(n;i)|^2}{\sum_{n=N_{H_1}}^{N_{H_2}} |X(n;i)|^2} \tag{11}$$

where a numerator and a denominator represent the low-band and the high-band energy, respectively, and N_L, N_{H_1}, and N_{H_2} are set to 320 Hz, 320 Hz, and 5360 Hz, respectively, in this paper.

As another acoustic feature for discriminating nasals from glides, spectral peak frequency (SPK) [24] is defined as

$$SPK(i) = \arg\max_{0 \le n \le N_{L_2}} |X(n;i)| \tag{12}$$

where N_{L_2} is set to 800 Hz. In addition, sum of difference in amplitude (SDA) [24] is given as

$$SDA(i) = \left(\sum_{n=0}^{N_1-1} |X(n;i)| - \sum_{n=N_1}^{N_2-1} |X(n;i)| \right) + \left(\sum_{n=0}^{N_1-1} |X(n;i)| - \sum_{n=N_2}^{N_3-1} |X(n;i)| \right) + \left(\sum_{n=N_1}^{N_2-1} |X(n;i)| - \sum_{n=N_2}^{N_3-1} |X(n;i)| \right) \quad (13)$$

where N_1, N_2, and N_3, indicate 788 Hz, 2 kHz, and 3 kHz, respectively.

3.5. Acoustic Features for Glides

Three acoustic features for the glide class discrimination are chosen such as envelope variance measure (EVM), sonorant measure (SM), and short-time peak-to-peak (STPP). First, EVM is defined as [24]

$$EVM(i) = \sum_{n=0}^{N-1} \left(H(n;i) - \mu(i) \right) \quad (14)$$

where $H(n;i)$ refers to the n-th component of the Hilbert envelope at the i-th frame, and $\mu(i)$ is the average of $H(n;i)$ over n. $H(n;i)$ is represented as

$$H(n;i) = x(n;i) * \frac{1}{\pi n}. \quad (15)$$

Similarly to vowels, glides include sonorant sounds. Thus, SM is extracted from the band-limited energy as [25]

$$SM(i) = \sum_{n=N_{L_1}}^{N_{L_2}} |X(n;i)|^2 \quad (16)$$

where N_{L_1} and N_{L_2} are set to 100 Hz and 400 Hz, respectively. Furthermore, to discriminate glides from unvoiced sounds, STPP is computed by [26]

$$STPP(i) = \max_{0 \le n < N} x(n;i) - \min_{0 \le n < N} x(n;i). \quad (17)$$

4. SVM-Based Consonant Class Discrimination

In order to discriminate consonant class, an SVM is constructed using feature vectors composed of the proposed acoustic features as shown in Figure 2. First, each consonant speech signal is segmented into 20 msec frames with an overlap of 10 msec. Then, a 39-dimensional MFCC vector is extracted for each frame and concatenated once every five consecutive frames, which results in a 195-dimensional MFCC feature vector for each frame. Next, five consecutive frames are grouped to provide the class-dependent feature vector of the center frame, where duplicates of first or last feature vectors are used at the beginning and ending two frames, respectively. Then, the extracted acoustic feature vector is concatenated with a 195-dimensional MFCC vector for each frame. Consequently, the concatenated 210-dimensional feature vector is used as an input vector for SVM.

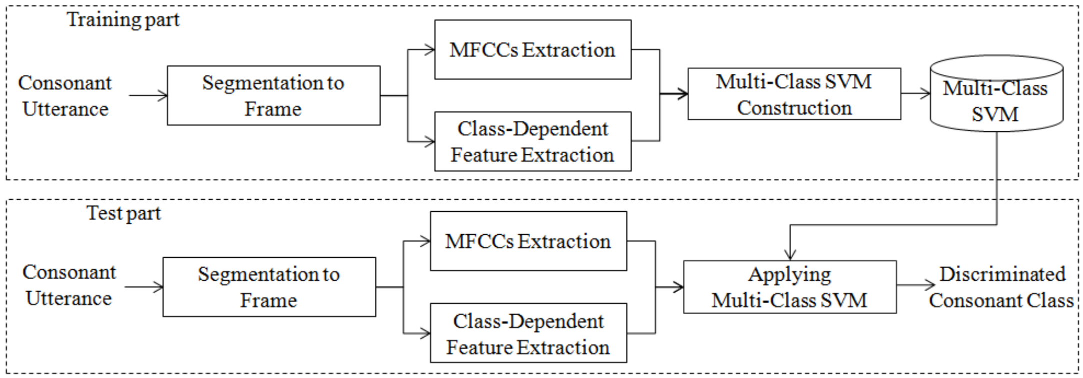

Figure 2. Block diagram of consonant class discrimination based on multi-class SVM using the proposed class-dependent acoustic features

With the input feature vectors, a multi-class kernel-based SVM [27] is trained to estimate prototypes for each class as a means of minimizing an empirical error that is defined as

$$\varepsilon(\mathbf{M}) = \frac{1}{N}\sum_{i=1}^{N}\left[H_M(\bar{x}_i \neq y_i)\right] \tag{18}$$

where $\mathbf{M}$ is a prototype $k \times n$ matrix, $\bar{x}_i$ and y_i are the i-th training data from a domain $\mathbf{X} \in \Re^n$ and corresponding class label from the set $Y = \{1, \cdots, k\}$. In (18), $[H]$ is 1 if the hypothesis, H, is true and 0 otherwise. Note that k and n are set to 5 and 210 in this paper. Then, each frame is discriminated into one of five consonant classes from the constructed SVM, $H_M(\bar{x})$, such as

$$H_M(\bar{x}) = \arg\max_{1 \le r \le k}\left\{\overline{M}_r \cdot \bar{x}\right\} \tag{19}$$

where $\bar{x}$ represents a feature vector of each frame and $\overline{M}_r$ indicates the r-th row of $\mathbf{M}$, which is the prototype for the r-th class.

5. Performance Evaluation

To evaluate the performance of discrimination of consonant class using the proposed class-dependent acoustic features, two different SVMs were constructed: one using 195-dimsional MFCCs, and the other using 210-dimensional features that are composed of 195-dimensional MFCCs and the proposed 15-dimensional class-dependent acoustic features. In this evaluation, 160 utterances of English consonants from eight dysarthric speakers were used [28], which resulted in 2,400 frames in total. However, the number of frames was insufficient to verify the effectiveness of the proposed consonant class discrimination method. Moreover, the data should be split into a training subset and a test subset. Thus, the 4-fold cross-validation method was used in this paper. In other words, four rounds of validations were performed with three folds for training SVMs and one fold for testing the consonant class discrimination methods.

Table 2. Comparison of average frame error rates (%) of the consonant class discrimination using MFCCs only and the combination of MFCCs and proposed class-dependent acoustic features for mild and severe dysarthric speech

Feature (Dimension)	Severity of dysarthria		
	Mild	Severe	Avg.
MFCC (195)	48.70	54.86	51.78
MFCC+Class-Dependent Feature (210)	43.91	51.70	47.81

Table 2 compares average frame error rates (FERs) of the two methods, which were constructed by different feature parameters. As shown in the table, the consonant class discrimination method using the combined features of MFCCs and class-dependent acoustic features relatively lowered the average FER by 7.67%, compared to that using MFCCs only.

6. Conclusion

In this paper, class-dependent acoustic features were proposed for consonant class discrimination in dysarthric speech, where an SVM was utilized as a classifier. The class-dependent acoustic features reflected characteristics of each consonant class, including features in a time or frequency domain. The acoustic features were extracted for five consecutive frames and combined with MFCCs for an input vector of SVM. The proposed method tried to classify each frame of consonant signals into one of five classes such as stop, affricate, fricative, nasal, and glide., It was shown from the experiments that the consonant class discrimination method using the combined features of MFCCs and proposed class-dependent acoustic features relatively lowered the average frame error rate by 7.67%, compared to that using MFCCs only. As a future work, the proposed consonant class discrimination will be integrated into an error-correction technique of dysarthric ASR in a way that pronunciation variation models are weighted according to consonant class.

Acknowledgements

This work was supported in part by the Technology Innovation program (10036461, Development of an embedded key-word spotting speech recognition system individually customized for disabled persons with dysarthria) by the Ministry of Trade, industry & Energy (MI, Korea).

References

[1] D. Haines, "Neuroanatomy: an Atlas of Structures, Sections, and Systems", Lippingcott Williams and Wilkins, Hagerstown, MD, **(2004)**.
[2] S. M. Kim, H. K. Kim, S. J. Lee and Y. Lee, "Multiple Likelihood Ratio Test-Based Voice Activity Detection Robust to Impact Noise in a Car Environment", Information: an International Interdisciplinary Journal, vol. 16, no. 3, **(2013)** March, pp. 3141-3151.
[3] S. M. Kim and H. K. Kim, "Probabilistic Spectral Gain Modification Applied to Beamformer Based Noise Reduction in a Car Environment", IEEE Transactions on Consumer Electronics, vol. 57, no. 2, **(2011)** May, pp. 866-872.

[4] H. K. Kim and R. C. Rose, "Cepstrum-Domain Model Combination Based on Decomposition of Speech and Noise Using MMSE-LSA for ASR in Noisy Environments", IEEE Transactions on Speech, Audio and Language Processing, vol. 17, no. 4, **(2009)** May, pp. 704-713.
[5] Y. R. Oh and H. K. Kim, "A Hybrid Acoustic and Pronunciation Model Adaptation Approach for Non-Native Speech Recognition", IEICE Transactions on Information and Systems, E93-D(9):2379-2387, **(2010)** September.
[6] Y. H. Lee and H. K. Kim, "Entropy Coding of Compressed Feature Parameters for Distributed Speech Recognition", Speech Communication, vol. 52, no. 5, **(2010)** May, pp. 405-412.
[7] Y. R. Oh, J. S. Yoon, H. K. Kim, M. B. Kim and S. R. Kim, "A Voiced-Driven Scene-Model Recommendation Service for Portable Digital Imaging Devices", IEEE Transactions on Consumer Electronics, vol. 55, no. 4, **(2009)** November, pp. 1739-1747.
[8] Y. R. Oh, Y. G. Kim, H. K. Kim, M. S. Lee and H. J. Bae, "Phonetically Balanced Text Corpus Design Using a Similarity Measure for a Stereo Super-Wideband Speech Database", IEICE Transactions on Information and Systems, vol. E94-E, no. 7, **(2011)** July, pp. 1459-1466.
[9] M. A. Ali, M. Hossain and M. N. Bhuiyan, "Automatic Speech Recognition Technique for Bangla Words", International Journal of Advanced Science and Technology, vol. 50, **(2013)** January, pp. 51-60.
[10] W. K. Seong, J. H. Park and H. K. Kim, "Dysarthric Speech Recognition Error Correction Using Weighted Finite State Transducers Based on Context-Dependent Pronunciation Variation", Lecture Notes in Computer Science, vol. 7383, **(2012)** July, pp. 475-482.
[11] W. K. Seong, J. H. Park and H. K. Kim, "Performance Improvement of Dysarthric Speech Recognition Using Context-Dependent Pronunciation Variation Modeling Based on Kullback-Leibler Distance", Advanced Science and Technology Letters, vol. 14, **(2012)** August, pp. 53-56.
[12] V. Young and A. Mihailidis, "Difficulties in Automatic Speech Recognition of Dysarthric Speakers and Implications for Speech-Based Applications Used by the Elderly: a Literature Review", Assistive Technology, vol. 22, no. 2, **(2010)**, pp. 99-112.
[13] W. K. Seong, J. H. Park and H. K. Kim, "Effects of noise suppression on consonant pronunciation variations of dysarthric speech", Proceedings of the 4th International Symposium on QoLT, Incheon, Korea, **(2012)** October 1-2.
[14] Z. M. Chen, W. X. Ling and J. H. Zhao, "Consonant Recognition of Dysarthria Based on Wavelet Transform and Fuzzy Support Vector Machine", Journal of Software, vol. 6, no. 5, **(2011)** May, pp. 887-893.
[15] D. L. Kendall, M. R. McNeil, S. Shaiman and M. A. Simonian, "Phonetic Variability in Flaccid Dysarthric Speech", International Journal of Speech-Language Pathology, vol. 1, no. 2, **(1999)**, pp. 107-111.
[16] K. Prasad, P. Lotia and M. R. Khan, "A Review on Text-Independent Speaker Identification Using Gaussian Supervector SVM", International Journal of u- and e- Service, Science and Technology, vol. 5, no. 1, **(2012)** March, pp. 71-82.
[17] C. M. Velu, P. Vivekanadan and K. R. Kashwan, "Indian Coin Recognition and Sum Counting System of Image Data Mining Using Artificial Neural Networks", International Journal of Advanced Science and Technology, vol. 31, **(2011)** June, pp. 67-80.
[18] A. Taherian and A. S. Mahdi, "Noise Resistant Identification of Human Iris Patterns Using Fuzzy ARTMAP Neural Network", International Journal of Security and Its Applications, vol. 7, no. 1, **(2013)** January, pp. 105-118.
[19] Y. Y. Kong and A. Mullangi, "On the Development of a Frequency-Lowering System that Enhances Place-of-Articulation Perception", Speech Communication, vol. 54, no. 1, **(2012)** January, pp. 147-160.
[20] J. S. Perkell, M. L. Matthies, M. Tiede, H. Lane, M. Zandipour, N. Marrone, E. Stockmann and F. H. Guenther, "The Distinctiveness of Speakers' /s/-/S/ Contrast is Related to Their Auditory Discrimination and Use of an Articulatory Saturation Effect", Journal of Speech, Language, and Hearing Research, vol. 47, no. 6, **(2004)** December, pp. 1259-1269.
[21] D. Ruinskiy, N. Dadush and Y. Lavner, "Spectral and Textural Feature-Based System for Automatic Detection of Fricatives and Affricates", Proceedings of 26th IEEE Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, **(2010)** November, pp. 771-775.
[22] L. F. Weiqelt, S. J. Sadoff and J. D. Miller, "Plosive/fricative Distinction: the Voiceless Case", Journal of the Acoustical Society of America, vol. 87, no. 6, **(1990)** June, pp. 2729-2737.
[23] C. Y. Lin, K. T. Chen and J. S. R. Jang, "A Hybrid Approach to Automatic Segmentation and Labeling for Mandarin Chinese Speech Corpus", Proceedings of Interspeech, Lisbon, Portugal, **(2005)** September, pp. 1553-1556.
[24] T. Pruthi and C. Y. Espy-Wilson, "Acoustic Parameters for Automatic Detection of Nasal Manner", Speech Communication, vol. 43, no. 3, **(2004)** August, pp. 225-239.
[25] C. Y. Espy-Wilson, "Acoustic Measures for Linguistic Features Distinguishing the Semivowels /wjrl/ in American English", Journal of the Acoustical Society of America, vol. 92, no. 2, **(1992)** August, pp. 736-757.

[26] A. Horner and J. Beauchamp, "A Genetic Algorithm-Based Method for Synthesis of Low Peak Amplitude Signals", Journal of the Acoustical Society of America, vol. 99, no. 1, **(1996)** January, pp. 433-443.
[27] K. Crammer and Y. Singer, "On the Algorithmic Implementation of Multiclass Kernel-Based Vector Machines", Journal of Machine Learning Research, vol. 2, **(2001)** March, pp. 265-292.
[28] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio and H. T. Bunnell, "The Nemours Database of Dysarthric Speech", Proceedings of the 4th International Conference on Spoken Language Processing, Philadelphia, PA, **(1996)** October, pp. 1962-1965.

Authors

Woo Kyeong Seong

Woo Kyeong Seong received a B.S. degree in Electronics Engineering from Inha University, Korea in 2010. He is currently pursuing a combined-MS-Ph.D degree at Gwangju Institute of Science and Technology (GIST). His current research interests include pronunciation variation modeling for speech recognition.

Ji Hun Park

Ji Hun Park received a B.S. degree in Electronics Engineering from Kwangwoon University, Korea in 2006, and an M.S. degree in Information and Communications Engineering from the Gwangju Institute of Science and Technology (GIST), Korea in 2008. He is currently pursuing a Ph.D. degree at GIST. His current research interests include signal processing for noise robust speech recognition.