INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Automatic Severity Classification of Korean Dysarthric Speech Using Phoneme-Level Pronunciation Features

Eun Jung Yeo[1], *Sunhee Kim*[2], *Minhwa Chung*[1]

[1]Department of Linguistics, Seoul National University, Republic of Korea
[2]Department of French Language Education, Seoul National University, Republic of Korea

ej.yeo@snu.ac.kr, sunhkim@snu.ac.kr, mchung@snu.ac.kr

Abstract

This paper proposes an automatic severity classification method for Korean dysarthric speech by using two types of phoneme-level pronunciation features. The first type is the percentage of correct phonemes, which consists of percentage of correct consonants, percentage of correct vowels, and percentage of total correct phonemes. The second type is related to the degree of vowel distortion, such as vowel space area, formant centralized ratio, vowel articulatory index, and F2-ratio. The baseline experiments use features from our previous study, consisting of MFCCs, voice quality features, and prosody features. Compared to the baseline, experiments including phoneme-level pronunciation features achieve a relative percentage increase of 32.55% and 33.84% in F1-score for support vector machine and feed-forward neural network classifiers, respectively. Our best performance reaches an F1-score of 77.38%, which is a relative percentage increase of 10.39% compared to the best previous results conducted on the Korean dysarthric QoLT corpus. Furthermore, with feature selection applied, all seven phoneme-level pronunciation features are chosen, accounting for the highest percentage of the selected feature set by both recursive feature elimination and extra trees classifier feature selection algorithms. Results indicate that phoneme-level pronunciation features are useful in enhancing the performance for automatic severity classification of dysarthric speech.

Index Terms: dysarthria, automatic severity classification, intelligibility assessment, machine learning, feature selection

1. Introduction

Dysarthria is a motor speech disorder, which involves impairments in various speech dimensions such as respiration, phonation, resonance, prosody, and articulation [1]. Speech pathologists conduct diagnosis on the severity levels of dysarthria, which is crucial in understanding the status of the patients and the progress of treatments. Perceptual evaluation of speech intelligibility by trained speech pathologists is the gold standard for such diagnosis. However, not only is the number of speech pathologists limited compared to that of dysarthric speakers, but the assessment itself is a highly time- and labor-consuming task. Therefore, automatic assessment that strongly correlates with the expert's score could be practical in a diagnostic field [2].

There are two main approaches for automatic assessment of dysarthria. The first approach is to propose a novel complex neural network that takes raw speech signals as input [3][4]. However, this approach often suffers from multi-classification, because most dysarthric speech corpus does not contain enough speakers for each severity level. The second approach is to extract features that show distinctive characteristics between severity levels. It has been reported that features such as spectral features [5][6], glottal features [7], articulation features [8][9], and prosody features [10][11][12] are useful in automatic severity classification for dysarthria.

Misarticulation is one of the most salient cues of dysarthria [13]. The most common features related to articulation used for automatic severity classification are frame-level pronunciation features such as MFCCs or PLPs [5][6][10]. These features reflect the overall characteristics of speech but are insufficient to capture the characteristics of dysarthria in two ways [14]. First, acoustic features like MFCCs or PLPs incorporate speaker properties, and may fail to capture speaker-independent characteristics of dysarthria. Second, the frame length often does not align with the length of each phoneme. Hence, characteristics of dysarthric speech that stand out at phoneme-level may not be amply captured by frame-level features. On the other side, studies that propose to use phoneme-level pronunciation features for automatic severity classification often do not consider features related to other speech dimensions such as voice quality, prosody, or frame-level pronunciation features [8][9].

Phoneme-level pronunciation features have been widely used for severity classification of dysarthria in clinical settings. Studies have reported strong correlations between these measures and the severity levels of dysarthria in English [13][15]. Analysis of Korean dysarthric speakers also finds significant correlations between such measures and severity levels [16][17]. Further, Percentage of Correct Consonants, a phoneme-level pronunciation feature calculated by the ratio of the number of correct consonants out of the total number of target consonants, is commonly used by speech pathologists in Korea to diagnose the severity levels of dysarthria [18].

Phoneme-level pronunciation features are generally used in articulation assessments in a clinical setting but have been overlooked in the field of automatic severity classification. Studies that use phoneme-level pronunciation features do not consider the effect of other speech dimension features that are reported to be helpful in previous studies. This paper proposes an automatic severity classification method for Korean dysarthric speech by using phoneme-level pronunciation features, along with features from different speech dimensions.

The paper is organized as follows: Section 2 describes the overall methods, including feature extraction, feature selection, and classification. Section 3 presents the experimental results, which is followed by the conclusion in Section 4.

http://dx.doi.org/10.21437/Interspeech.2021-1353

2. Methods

2.1. System overview

Figure 1 shows the system overview including phoneme-level pronunciation features, indicated in bold. The first step is feature extraction. Extracted features consist of spectral features (MFCCs), voice quality features, prosody features, and phoneme-level pronunciation features. Next, the optimal feature set is chosen by a feature selection algorithm. Then the selected features are used as inputs to the classifier, which implements classification into five severity levels: healthy, mild, mild-to-moderate, moderate-to-severe, and severe. The performance of the classifier is evaluated using two metrics: classification accuracy and F1-score.

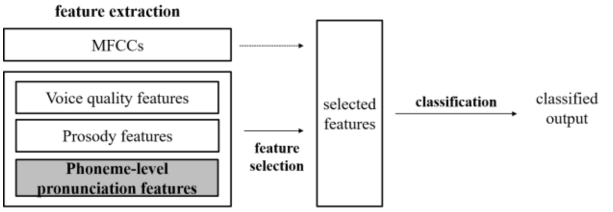

Figure 1: *System overview*

2.2. Feature extraction

Table 1 presents the full feature list used in this study. We separate the features into 4 categories: spectral features (MFCCs), voice quality features, prosody features, and phoneme-level pronunciation features. In this paper, the same feature set from our previous study [10] is used for spectral features, voice quality features, and prosody features.

Table 1: *Feature list*

Category		Features
spectral features		MFCCs
voice quality features		jitter, shimmer, HNR # of voice breaks, % of voice breaks
prosody features	speech rate	speech duration, total duration, speaking rate, articulation rate
	pitch	F0 mean/std/min/max/med/25quartile/75quartile
	rhythm	%V, deltas, Varcos, rPVIs, nPVIs
phoneme-level pronunciation features	percentage of correct phonemes	PCC, PCV, PCT
	degree of vowel distortion	VSA, FCR, VAI, F2-Ratio

2.2.1. MFCCs

MFCCs have been commonly used for automatic assessment of impaired speech [5][6][10]. We extract 12-dim MFCCs and log energy from all utterances by using librosa [19].

2.2.2. Voice quality features

Speakers with dysarthria have impaired laryngeal regulation and are negatively affected in terms of voice quality [20]. Five voice quality features are used: jitter, shimmer, Harmonic to Noise Ratio (HNR), number of voice breaks, and degree of voice breaks.

Jitter evaluates the regular period of F0 change, and shimmer assesses the regular period of intensity. HNR is the ratio of harmonic energy to noise energy. Jitter, shimmer, and HNR are indicators of voice disorders, and are often used by speech therapists. The number of voice breaks and the degree of voice breaks reflect the vocal capability to sustain phonation. All voice quality features are extracted by using Praat [21], which is an open-source speech analysis software widely used in speech pathology. For jitter, shimmer, and HNR, minimum pitch and maximum pitch is each set to 70 Hz and 625 Hz [22]. For features related to voice beaks, voice breaks are considered as all inter-pulse intervals longer than 17.86ms, calculated by the distance of successive glottal pulses set to 1.25 divided by the pitch floor of 70 Hz. The degree of voice breaks is calculated by dividing the total duration of voice breaks by the total duration of speech.

2.2.3. Prosody features

Most dysarthric speakers have difficulties in various prosodic aspects [23]. To effectively capture the characteristics of dysarthria, three types of prosody features are used: speech rate, pitch, and rhythm.

For speech rate, we include speech duration, total duration, speaking rate, and articulation rate. Speech duration refers to the duration excluding the pause sections from the total duration. Speaking rate refers to the number of produced syllables divided by total duration, while articulation rate means the number of produced syllables divided by speech duration. Speech rate features are extracted by using Parselmouth [24].

As for pitch features, the mean, standard deviation, minimum, maximum, median, 25 quartile, and 75 quartile of F0 are used. All features are extracted by using Praat. Minimum and maximum pitch is set to 70 Hz and 625 Hz, respectively.

For rhythm, %V, deltas, Varcos, rPVIs, nPVIs are used. %V refers to the proportion of vocalic intervals of an utterance. Deltas are the standard deviation of vocalic or consonantal intervals, and Varcos are the normalized values for each delta. rPVIs are the temporal succession of the vocalic or consonantal intervals, and nPVIs are the normalized rPVIs. Normalized values are proposed to reduce the influence of the speech rate on raw values. Rhythm features are extracted by using Correlatore 2.3.4 [25].

2.2.4. Phoneme-level pronunciation features

Phoneme-level pronunciation features can be divided into two types: the percentage of correct phonemes and the degree of vowel distortion. The percentage of correct phonemes features are adapted from the Urimal Test of Articulation and Phonology (U-TAP), an articulation assessment toolkit used in Korean clinical settings [26]. For the degree of vowel distortion, we modified the features from [27][28], which evaluated the vowel characteristics of Korean dysarthric speakers.

Features of the percentage of correct phonemes are as follows: percentage of correct consonants (PCC), percentage of correct vowels (PCV), and percentage of total correct phonemes (PCT). PCC/PCV/PCT refers to the ratio of the number of correct consonants/vowels/phonemes out of the total number of target consonants/vowels/phonemes.

Features are extracted by using a speech recognizer trained with healthy speakers. The alignments between the phoneme sequence from an automatic speech recognition (ASR) model and the canonical pronunciation sequence are used. PCC, PCV, PCT are calculated based on the number of matches between the two sequences. Kaldi toolkit [29] is used for ASR training. The acoustic model is trained on AI Hub corpus [30], an open Korean spontaneous speech database. The amount of training data is about 2,700 hours with 3-way speed perturbation. The acoustic model consists of time-delay neural network (TDNN) layers with ReLU activation function. All hidden layers have 1280 nodes with batch normalization applied. Root Mean Square Error between human-calculated and auto-calculated measurements are 38.01, 37.54, 36.98 for PCC, PCV, PCP, respectively. Both measurements show a tendency for lower PCC, PCV, and PCP as dysarthria becomes more severe, which is also consistent with the results from previous studies [17][18].

Features related to the degree of distorted vowels are extracted as follows: Vowel Space Area (VSA), Formant Centralized Ratio (FCR), Vowel Articulatory Index (VAI), and F2-ratio. Table 2 presents the formulas for each feature. VSA refers to the area of the F1-F2 vowel diagram. Among the various types of vowel diagrams, vowel rectangle is used because significant differences are seen between healthy controls and dysarthric speakers with vowel rectangles but not with vowel triangles [27]. FCR and VAI are the indicators for vowel centralization. The two being in an inverse relationship, the value of FCR increases as the value of VAI decreases. F2-ratio refers to the ratio of the F2 value of /i/ and /o/; each vowel represents an unrounded vowel and a rounded vowel.

Vowel distortion features are extracted by using the speech recognizer described above and Praat. Forced alignments are generated by the speech recognizer. Then formants from the center of each vowel are extracted by Praat. In formant extraction, the maximum formant frequency is set depending on gender, 5500 Hz for females and 5000 Hz for males. Dysarthric speakers have smaller VSA, lower VAI and F2-ratio, and higher FCR compared to healthy speakers, which follows the findings of previous studies [26][27].

Table 2: *Formulas for the degree of vowel distortion features*

Category	Features
VSA	$\frac{1}{2}\lvert[(F2/i/*F1/\varepsilon/ + F2/\varepsilon/*F1/a/ + F2/a/*F1/o/ + F2/o/*F1/i/) - (F1/i/*F2/\varepsilon/ + F1/\varepsilon/*F2/a/ + F1/a/*F2/o/ + F1/o/*F2/i/)]\rvert$
FCR	$\frac{F2/o/ + F2/a/ + F1/i/ + F1/o/}{F2/i/ + F1/a/}$
VAI	$\frac{F2/i/ + F1/a/}{F1/i/ + F1/o/ + F2/o/ + F2/a/}$
F2-ratio	$\frac{F2/i/}{F2/o/}$

2.3. Feature Selection

Feature selection is applied to choose the optimal set of features for classification. We use Recursive Feature Elimination (RFE) and Extra Trees Classifier (ETC) feature selection algorithms provided by the scikit-learn library [31]. These two feature selection algorithms are used because they showed the best performances in our preliminary experiments.

2.4. Classification

We use support vector machine (SVM) and feed-forward neural network (MLP) classifiers. SVM has been often used for impaired speech classifications [6][7][10]. In this paper, C and gamma, the hyperparameters of SVM, are optimized through grid search between 10^{-4} and 10^{4}. The hyperparameters of MLP, such as the number of hidden layers (1 to 5), learning rate (0.0001 to 0.1), optimizer (SGD, lbfgs, adam), and activation function (logistic, tanh, ReLU) are also optimized through grid search.

3. Experiment

3.1. Corpus

Quality of Life Technology (QoLT) corpus is a Korean dysarthric speech corpus, constructed to develop ASR systems for people with dysarthria [32]. The corpus consists of isolated words and restricted sentences. Isolated words may be enough in detecting speakers with severe or moderate dysarthria but may not be sufficient for speakers with mild dysarthria [10]. Furthermore, voice quality and prosody are better represented by sentences than words. Hence only restricted sentences are used in this paper.

The corpus has 10 healthy speakers (5 male, 5 female) and 70 dysarthric speakers (45 male, 25 female). For dysarthric speakers, 24 subjects are graded as mild, 24 subjects as mild-to-moderate, 15 subjects as moderate-to-severe, and 7 subjects as severe. Since Korean has a variety of dialects, it would be ideal to include speakers from the same region. However, due to a limited number of dysarthric speakers, we include all speakers in the corpus. All 10 healthy speakers and 60 dysarthric speakers are from Seoul or Gyeonggi province, while 10 dysarthric speakers are from other provinces. Among 10 speakers, 9 speakers have mild dysarthria, and 1 speaker has mild-to-moderate dysarthria.

The corpus contains 5 phonetically balanced sentences repeated twice by each speaker. However, one sentence is excluded in our experiments because it does not contain the vowel /o/. Accordingly, 80 utterances from healthy speakers and 560 utterances from dysarthric speakers are used. The experiments are conducted for a speaker-independent scenario, where speakers for training and test are separated. The ratio of training and test utterances is set to 7:3.

3.2. Severity classification

As shown in Table 3, when phoneme-level pronunciation features are included, the performance of severity classification is improved. Compared to the baseline experiments, the proposed experiments achieve a relative increase of up to 32.55% and 33.84% on F1-score for SVM and MLP classifiers, respectively. Our best performance is the ETC-SVM experiment, with an F1-score of 77.38%.

Table 3: *F1-score for severity classification (%)*

Feature selection	Classifier	Baseline	Proposed	Relative increase
X	SVM	60.09	69.41	15.51
	MLP	59.96	69.13	15.29
RFE	SVM	58.97	74.4	26.17
	MLP	**53.60**	**71.74**	**33.84**
ETC	**SVM**	**58.44**	**77.38**	**32.55**
	MLP	60.7	76.76	26.46

Further analysis is conducted for the ETC-SVM experiment by using the two different types of pronunciation features: the percentage of correct phonemes and the degree of vowel distortion. Including the either shows better performances than the baseline experiment, achieving 24.20% and 28.99% relative increase. However, the best performance is attained when both types of pronunciation features are used, with 32.41% relative increase. The results are presented in Table 4.

Table 4: *F1-score by using different types of phoneme-level pronunciation features (%)*

Features	F1-score	Relative increase
Baseline	58.44	—
+Percentage of correct phonemes	72.58	24.20
+Degree of vowel distortion	75.38	28.99
+Both	**77.38**	**32.41**

ETC-SVM experiment is also examined by severity levels of dysarthria. As shown in Table 5, classification accuracy improves for healthy, mild, mild-to-moderate speakers when phoneme-level pronunciation features are included. Classification accuracy for each severity level increases from 25%, 59.38%, 60.94% to 100%, 85.94%, 64.06%, respectively. However, no improvements are found for moderate-to-severe and severe speakers. This implies that moderate-to-severe and severe speakers can be sufficiently classified using voice quality and prosody features.

Table 5: *Classification accuracy by severity levels (%)*

Severity Level	Baseline	Proposed
Healthy	**25.00**	**100.00**
Mild	**59.38**	**85.94**
Mild-to-Moderate	**60.94**	**64.06**
Moderate-to-Severe	71.88	71.88
Severe	75.00	75.00

3.3. Analysis of selected features

Figure 3 presents the number of selected features in each experiment. In the baseline experiment, RFE selects 4 voice quality features and 9 prosody features, and ETC selects 2 voice quality features and 9 prosody features. However, after including the phoneme-level pronunciation features, 1 voice quality feature and 3 prosody features are excluded from the selected feature set with RFE, and 1 voice quality feature and 8 prosody features are excluded through ETC. In contrast, all 7 phoneme-level pronunciation features are selected by both RFE and ETC algorithms, accounting for the highest percentage of the selected feature set. Details of the selected features are provided in Table 6. Features selected by both algorithms are indicated in bold.

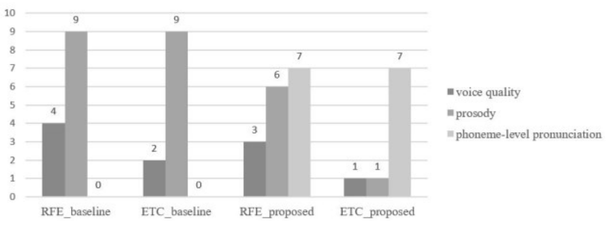

Figure 3: *Number of selected features by experiments*

Table 6: *Selected features by RFE and ETC algorithm*

Feature set		RFE	ETC
voice quality features		jitter, number of voice breaks, **degree of voice breaks**	**degree of voice breaks**
prosody features	speech rate	articulation rate, **speech duration**, total duration	**speech duration**
	pitch	F0 Mean, F0 Med	-
	rhythm	Varco-C	-
phoneme-level pronunciation features	percentage of correct phonemes	**PCC, PCV, PCT**	**PCC, PCV, PCT**
	degree of vowel distortion	**VSA, FCR, VAI, F2-Ratio**	**VSA, FCR, VAI F2-Ratio**

4. Discussion and Conclusion

We have proposed an automatic severity classification method of dysarthric speech by using two types of phoneme-level pronunciation features: the percentage of correct phonemes and the degree of vowel distortion. Compared to the baseline experiments using MFCCs, voice quality features, and prosody features, our proposed method achieves a relative increase of 32.55% and 33.84% in F1-score for SVM and MLP classifiers, respectively. Our best performance reaches 77.38% F1-score, with a relative percentage increase of 10.39% compared to the previous best results conducted on QoLT corpus [10].

The selected features resulting from the baseline experiments are in line with the previous study [10], which reports that all three kinds of prosody features are helpful in automatic severity classification of dysarthric speech. However, when the phoneme-level pronunciation features are added to the baseline features, many prosody features, especially pitch and rhythm features are excluded. On the other hand, all 7 phoneme-level pronunciation features are selected, occupying the highest proportion of the selected feature set for both RFE and ETC algorithms. This may imply that for Korean, pitch and rhythm may be less important in classifying the severity levels of dysarthria compared to the proposed phoneme-level pronunciation features.

In conclusion, this study demonstrates the importance of phoneme-level pronunciation features in automatic severity classification of dysarthric speech, which have been overlooked in the field of speech engineering. Future work includes applying the proposed method on different languages such as English, to generalize the usefulness of phoneme-level pronunciation features towards severity classification. Moreover, cross-linguistic analysis can provide interesting insights into understanding the characteristics of dysarthric speech for each language. Lastly, further experiments with more dysarthric speakers should be conducted.

5. Acknowledgements

This research is supported by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: R2019080018)

6. References

[1] Y. Kim, R. D. Kent, and G. Weismer, "An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria," *American Speech-Language-Hearing Assiciation,* 2011.

[2] M. J. Kim, and H. Kim, "Automatic assessment of dysarthric speech intelligibility based on selected phonetic quality features," in *Proceedings of International Conference on Computers for Handicapped Persons,* pp. 447-450, Berlin: Springer, 2012.

[3] C. Bhat and H. Strik, "Automatic assessment of sentence-level dysarthria intelligibility using BLSTM," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 322-330, 2020.

[4] J. Millet and N. Zeghidour, "Learning to detect dysarthria from raw speech," *IEEE International Conference on Acoustics, Speech and Signal Processing,* pp. 5831-5835, May. 2019.

[5] B. A. Al-Qatab, and M. B. Mustafa, "Classification of Dysarthric Speech According to the Severity of Impairment: an Analysis of Acoustic Features," *IEEE* Access, 9, pp. 18183-18194., 2021.

[6] A. Benba, A. Jilbab, A. Hammouch, and S. Sandabad, "Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease," *IEEE on International conference on electrical and information technologies*, pp. 300-304, Mar. 2015.

[7] N. P. Narendra, and P. Alku, "Dysarthric speech classification from coded telephone speech using glottal features," *Speech Communication*, vol. 110, pp. 47-55, 2019.

[8] M. J., Kim, Y. Kim, and H. Kim, "Automatic intelligibility assessment of dysarthric speech using phonologically-structured sparse linear model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 4, pp. 694-704, 2015.

[9] K. L. Lansford, & J. M. Liss, "Vowel acoustics in dysarthria: Speech disorder diagnosis and classification", *American Speech-Language-Hearing Assication,* vol. 57, no. 1, 2014.

[10] A. Hernandez, S. Kim, and M. Chung, "Prosody-based measures for automatic severity assessment of dysarthric speech," *Applied Sciences*, vol. 10, no. 19, pp. 6999, 2020.

[11] A. Hernandez, E. J. Yeo, S. Kim, and M. Chung, "Dysarthria Detection and Severity Assessment using Rhythm-Based Metrics," in *Proceedings of INTERSPEECH*, Shanghai, China, Oct. 2020, pp. 25-29.

[12] K. L. Kadi, S. A. Selouani, B. Boudraa, and M. Boudraa, "Discriminative prosodic features to assess the dysarthria severity levels," *Proceedings of the World Congress on Engineering*, vol. 3, Jul. 2013.

[13] L. J. Platt, G. Andrews, M. Young, and P. T. Quinn, "Dysarthria of adult cerebral palsy: I. Intelligibility and articulatory impairment," *Journal of Speech, Language, and Hearing Research*, vol. 23, no. 1, pp. 28-40. 1980.

[14] A. Tripathi, S. Bhosale, and S. K. Kopparapu, "Improved Speaker Independent Dysarthria Intelligibility Classification Using Deepspeech Posteriors," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6114-6118, May. 2020.

[15] H. Kim, K. Martin, M. Hasegawa-Johnson, and A. Perlman, "Frequency of consonant articulation errors in dysarthric speech," *Clinical linguistics & phonetics*, vol. 24, no. 10, pp. 759-770, 2010.

[16] M. J. Kim, "Phonological error patterns of preschool children in the 'Korean Test of Articulation for Children's phonetics study," *Communication Sciences & Disorders*, vol. 11, no. 2, pp. 17-31, 2006.

[17] S. M. Hong, P. Y. Jeong, H. S. Sim, S. M. Hong, P. Y. Jeong, P. and H. S. Sim, "Comparison of Perceptual Assessment for Dysarthric Speech: The Detailed and General Assessments," *Communication Sciences & Disorders*, vol. 23, no. 1, pp. 242-253, 2018.

[18] Y. M. Lee, J. E. Seong, H. S. Shim, J. H. Han, and H. N. Song "Analysis of Articulation Error Patterns Depending on the Level of Speech Intelligibility in Adults with Dysarthria," *The Korean Academy of Speech-Language Pathology and Audiology*, vol. 17, no. 1, pp. 130-142, 2012.

[19] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference,* vol. 8, pp. 18-25, Jul. 2015.

[20] I. H. Seo and C. J. Seong, "Voice quality of dysarthric speakers in connected speech," *Journal of the Korean Society of Speech Science*, vol. 5, no. 4, pp. 33-41, 2013.

[21] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International* vol. 5, no. 9/10, pp. 341-345, 2001.

[22] S., Shim, H., Kim, J., Kim, & J. Shin, "Difference in Voice Parameters of MDVP and Praat Programs according to Severity of Voice disorders in Vocal Nodule," *Phonetics and Speech Sciences*, vol. 6, no. 2, pp. 107-114, 2014

[23] K. J., Schlenck, R., Bettrich, & K. Willmes, "Aspects of disturbed prosody in dysarthria," *Clinical linguistics & phonetics,* vol. 7, no. 2, pp. 119-128, 1993

[24] Y., Jadoul, B. Thompson, & B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics, 71*, 1-15, 2018. https://doi.org/10.1016/j.wocn.2018.07.001

[25] P. Mairano, and A. Romano, "Un confronto tra diverse metriche ritmiche usando Correlatore," In: Schmid, S., Schwarzenbach, M. & Studer, D. (eds.) La dimensione temporale del parlato, (Proc. of the V National AISV Congress, University of Zurich, Collegiengebaude, 4-6 February 2009), Torriana, pp. 79-100, 2010.

[26] Y. T. Kim, and M. J. Shin, "Urimal test of articulation and phonology". *Hakjisa:* Seoul 2004,

[27] Y. Kang, K. Yoon, H. Lee, and C. Seong, "A comparison of parameters of acoustic vowel space in patients with Parkinson's disease," *Journal of the Korean Society of Speech Sciences*, vol 2, pp. 185–192, 2010.

[28] S.,Kim, J. H., Kim, & D, Ko, "Characteristics of Vowel Space and Speech Intelligibility in Patients with Spastic Dysarthria," *Communication Sciences & Disorders,* vol. 19, no. 3, pp. 352-360, 2014.

[29] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, and K. Vesely, "The Kaldi speech recognition toolkit," *In IEEE 2011 workshop on automatic speech recognition and understanding*, 2011

[30] AIHub Homepage. Available online: http://www.aihub.or.kr/aidata/105 (accessed on 5 August 2020).

[31] O. Kramer, "Scikit-learn. In Machine learning for evolution strategies," pp. 45-53, Cham : Springer, 2016.

[32] D. L. Choi, B. W. Kim, Y. W. Kim, Y. J. Lee, Y. Um, and M. Chung, "Dysarthric Speech Database for Development of QoLT Software Technology," in *Proceedings of International Conference on Language Resources and Evaluation*, pp. 3378-3381, 2012.