Dysarthria Detection and Severity Assessment using Rhythm-Based Metrics

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

[1]Department of Linguistics, Seoul National University, Republic of Korea
[2]Departement of French Language Education, Seoul National University, Republic of Korea
abner@snu.ac.kr, ej.yeo@snu.ac.kr, sunkim@snu.ac.kr, mchung@snu.ac.kr

Abstract

Dysarthria refers to a range of speech disorders mainly affecting articulation. However, impairments are also seen in suprasegmental elements of speech such as prosody. In this study, we examine the effect of using rhythm metrics on detecting dysarthria, and for assessing severity level. Previous studies investigating prosodic irregularities in dysarthria tend to focus on pitch or voice quality measurements. Rhythm is another aspect of prosody which refers to the rhythmic division of speech units into relatively equal time. Speakers with dysarthria tend to have irregular rhythmic patterns that could be useful for detecting dysarthria. We compare the classification accuracy between solely using standard prosodic features against using both standard prosodic features and rhythm-based features, using random forest, support vector machine, and feed-forward neural network. Our best performing classifiers achieved a relative percentage increase of 7.5% and 15% in detection and severity assessment respectively for the QoLT Korean dataset, while the TORGO English dataset had an increase of 4.1% and 3.2%. Results indicate that including rhythmic information can increase accuracy performance regardless of the classifier. Furthermore, we show that rhythm metrics are useful in both Korean and English.

Index Terms: dysarthric speech, prosody, rhythm, dysarthria detection, severity assessment

1. Introduction

Speech impairments are one of the first indicators of degenerative or neurological disorders such as Parkinson or cerebral palsy. Typically, dysarthria is diagnosed by a trained speech therapist who uses perceptual tasks and/or an acoustic analysis to determine what aspects of speech are affected. While a professional diagnosis is important, their assessment can be biased and subjective. Therefore, an objective approach could aid therapist to provide a more robust diagnosis.

Although the most common cue of dysarthria is misarticulation, several studies have found impairments in the suprasegmental domain [1, 2]. Acoustic studies on Korean speakers with dysarthria also found significant differences in pitch range, speaking rate, and duration of utterances and pauses [3,4,5].

While there has been many studies looking at pitch, a less studied prosodic element is rhythm. Rhythm refers to the duration-based division of speech units into equal pieces. Early studies found significant differences in rhythm metrics between healthy and dysarthric speakers [6]. However, few studies have been conducted using rhythm metrics for automatic dysarthria detection. In [7], mel-frequency cepstral coefficients (MFCCs) are combined with seven rhythm metrics as input to GMM and MLP classifiers. Their results showed that depending on the classifier a 3% to 6% improvement in accuracy can be achieved compared to only using MFCCs. Similarly, a Gaussian Bayes-based classifier was used in [8] to determine which pair of rhythm metrics out of 10 measures was most useful for discriminating dysarthric speech from healthy speech. They found that measurements related to vowel segments were the most useful for discrimination. Lastly, a variety of prosodic measurements were taken in [9], including rhythm-based metrics. Results from this study showed that these prosodic features can be used in SVM and GMM classifiers to accurately assess dysarthria into four levels (healthy, mild L1, severe L2 and severe L3).

The results from previous studies are promising, but several limitations arise with the use of the Neymours database [10] which was used in all the above-mentioned studies. First, the speakers in the Neymours database are composed of 12 males, 11 with dysarthria and only one healthy control. The lack of both healthy speakers and female speakers may limit the generalizability. Another issue relates to the limited sentences structure. The database is mostly composed of simple carrier sentences where the format is always: ‘the X is Y-ing the Z”. X and Z coming from a set of 74 monosyllabic nouns, while Y was selected from a set of 37 disyllabic verbs. Using carrier sentences can alter the natural rhythm of language leading to misleading results.

The main contributions of this paper are as follows: First, we alleviate some of the previously mentioned issues with the Neymours dataset by using the TORGO database [11]. The TORGO database has a more diverse set of speakers and a larger set of speech samples including spontaneous speech. Second, we utilize the Quality of Life (QoLT) database which is composed of Korean speakers with dysarthria along with healthy counterparts [12]. These two data sets allow us to examine the generalizability of rhythm metrics to English and Korean dysarthric speech. Lastly, we include a wide range of prosodic measures related to speech rate, pitch, voice quality and rhythm.

The paper is organized as follows: Section 2 introduces the full prosodic feature set including rhythm. Section 3 describes the TORGO and QoLT corpora in more detail. Section 4 provides the training procedures for both detection and assessment. The results for all experiments are presented in section 5. Lastly, Section 6 concludes the paper with a discussion of the results and possible directions for future work.

Table 1: Prosodic features used for classification.

Speech rate	Pitch	Voice Quality	Rhythm
# of pauses	F0 mean	Shimmer	%V
Speaking rate	F0 std	Jitter	ΔV, ΔC
Articulation rate	F0 median/ min/max	% of voice breaks	Varcos
Duration	F0 25% quantile	# of voice breaks	rPVIs
	F0 50% quantile	HNR	nPVIs

2. Prosodic Features

As mentioned in the introduction, speakers with dysarthria often have irregular prosodic measurements when compared to healthy speakers. In this study, we extract a range of prosodic features both related to rhythm and unrelated to rhythm which we separate into 4 categories; speech rate, pitch, voice quality, and rhythm. The full feature list can be seen in Table1.

2.1. Speech Rate

In general speakers with dysarthria show a reduced tempo, and slower speech rate than healthy speakers [13,14]. Therefore, we included a number of measurements which reflect speech rate. First, we include the number of pauses which is higher in dysarthric speech. Speaking rate is the number of syllables per seconds including pauses, while articulation rate is the number of syllables not including pauses. In both cases, speakers with dysarthria tend to have a lower rate since they often have prolonged pronunciations. Lastly, we include speaking duration as speakers with dysarthria tend to have higher durations given their tendency for prolonged pronunciation.

2.2. Pitch

Pitch is a commonly studied cue of dysarthria, showing differences not only with healthy speakers but also between speakers of different severity levels. Mild dysarthric speakers tend to be more monotonic while severe speakers often have significantly higher F0 values than both mild and healthy speakers [15]. We include, standard pitch measurements such as mean, median, minimum and maximum F0 along with standard deviation, 25% and 50% quantiles.

2.3. Voice Quality

Voice quality refers to the properties of speech related to the vocal folds. Individuals with dysarthria tend to have less control over their vocal folds leading to irregular measurements [16]. Voice quality features have been shown to be useful in sentence-level classification of impaired speech [17]. Jitter represents the variations of F0 within a time period. Shimmer is like jitter except that perturbation falls in the amplitude domain. Harmonics to noise ratio (HNR) refers to the periodicity of a speech signal over noise. Speakers with dysarthria tend to have trouble maintaining the phonation of voiced segments. Therefore, we extracted two voice break related measures. The first being the number of voice breaks. Second, we include degree of voice breaks, which is total duration of the breaks over the signal, divided by the total duration, excluding silence at the beginning and end of the sentence.

2.4. Rhythm Metrics

Unlike pitch or voice quality measures, rhythm does not have a specific acoustic cue. Instead, linguists have proposed several durational measures of vocalic and intervocalic segments. These measures have been shown to be correlates of rhythm [18-20]. All rhythm metrics are extracted using the software Correlatore 2.3.4 [21].

2.4.1. Deltas

One of the first rhythm metrics was proposed in [18], based on divisions of speech into vocalic and consonantal parts. The proportion of the vocalic intervals of a sentence (%V), the standard deviation of vocalic intervals (ΔV) and the standard deviation of consonantal intervals (ΔC) are the proposed metrics. It was found that the proportion of time of vocalic intervals in the sentence (%V) and the standard deviation of intervocalic intervals (ΔC) was the best correlate for distinguishing different rhythm classes.

2.4.2. Varcos

Researchers have tried to normalize delta values in order to reduce the interaction between speech rate and deltas [19]. This study proposed a method where the values of deltas are divided by the mean duration of (vocalic or consonantal) intervals, then multiplied by 100. We use both VarcoV and VarcoC in our feature set.

2.4.3. Pairwise Variability Index (PVI)

Another approach to rhythm was taken in [20], where the temporal succession of the vocalic and consonantal intervals is taken into consideration instead of joining all the values and calculating the standard deviation. The influence of speech rate variation can be controlled by calculating the normalized PVI, which calculates the mean absolute normalized difference between durations of neighboring interval pairs. We include the following PVI measurements: CrPVI, VrPVI, CnPVI and VnPVI, where C refers to consonants and V for vowels.

3. Corpus

3.1. TORGO

The TORGO dataset contains 15 speakers, 8 with dysarthria (5 males, 3 females) and 7 healthy controls (4 males, 3 females). Four speakers were considered to have mild dysarthria using the Frenchay assessment [22]. Two fell into the severe group, one into the moderate group and one was considered moderate/severe. We group this speaker into the moderate category while conducting our severity-based assessment. All 15 speakers had utterances in training and test sets but with varying sentences such that no sentence in the test set was seen in the training set. Sentences come from a mixture of read passages and spontaneous speech elicited from an image description task. For this study we looked at 160 unique sentences split between training and test sets. In total this led to 577 total utterances: 156 healthy utterances and 421 dysarthric utterances.

Table 2: Mean values of rhythm metrics for all speaker groups.

Speaker Group	%V	ΔV	ΔC	varco-V	varco-C	Vrpvi	Crpvi	Vnpvi	Cnpvi
English Healthy	41.72	60.70	73.28	53.18	50.89	66.20	81.85	55.85	56.89
English Dysarthric	43.54	93.30	107.56	50.66	55.07	102.86	116.34	54.03	58.58
Korean-Healthy	54.37	65.69	51.79	57.59	55.23	67.52	65.78	61.51	70.36
Korean-Dysarthria	57.83	139.57	96.65	58.43	60.05	148.56	110.56	60.90	69.18

3.2. QoLT

The continuous section of the QoLT dataset has 90 dysarthric speakers and 10 healthy controls (5 males, 5 females). To have more balance group sets we only examined 28 dysarthric speakers (15 males, 13 females): 8 with severe dysarthria, 10 moderate and 10 mild. All speakers recorded 5 sentences twice leading to 280 utterances for dysarthric speakers and 100 utterances for healthy speakers. While we could not have separate unique sentences for training and test sets as we did for the TORGO dataset, we instead chose to build speaker-independent models where no speaker in the test set appeared in the training set.

3.3. Corpus Comparison using Rhythm Metrics

To examine any group differences, we take mean scores of rhythm metrics for all speaker groups which can be seen in Table 2. As seen from the table 2, healthy English speakers have a higher ΔC but lower %V compared to Korean speakers. English speakers also have lower varco and nPVI means compared to Korean speakers. Speakers with dysarthria from both language groups have overall higher means for deltas and rPVI metrics. This is likely due to difficult in articulating, leading to highly variable durations of consonantal and vocalic intervals. Both Varcos and nPVI measures show minimal difference between healthy and dysarthric speakers.

4. Training Procedure

4.1. Classifiers

Dysarthria detection is implemented by training features on random forest (RF), support vector machine (SVM) and feed-forward neural network (MLP) classifiers. For severity assessment we use SVM and MLP classifiers. SVM's have often been used for impaired speech classification tasks as they consistently perform well even with small datasets [23-26]. Our baseline models use all non-rhythm prosodic features, while the proposed model used all prosodic features including rhythm metrics.

4.2. Dysarthria Detection

For detection we build binary classifiers. Each model had their hyperparameters optimized by applying a grid search. The number of trees and maximum depth are optimized for the random forest classifier, while margin parameter C and gaussian kernel γ is optimized for the SVM using values between 10^{-4} to 10^{3}. Lastly, we optimize our MLP by finetuning the learning rate, hidden layer size, number of neurons, epochs, activation function and optimization algorithm.

4.2.1. TORGO

We balance the data so that both healthy and dysarthric groups have around 200 utterances for training. A k-fold cross validation where k=10 is implemented during training for hyperparameter fine tuning. A separate test set is used for the final evaluation where about 140 utterances containing difference sentences is used.

4.2.2. QoLT

A similar training procedure as above is used with the exception of the data split. Instead of splitting training and test sets by sentences we split them by speakers. For training we use 6 healthy speakers and 17 dysarthric speakers. The test set contains 4 healthy speakers and 11 dysarthric speakers. Furthermore, for each set we balance the number of utterances to have equal amount of data in both groups.

4.3. Severity Assessment

Severity assessment is treated as a multiclass classification problem. This is inherently handled by our MLP when applying a softmax function instead of a sigmoid as the output function. For our SVM we apply a one-verses-one approach where if n is the number of classes, then n * (n - 1) / 2 classifiers are constructed and each one trains data from two classes. We constructed four severity levels, healthy, mild, moderate and severe. For the TORGO database each group has an average of 85 utterances for training and 60 utterances each for testing. The QoLT data has around 60 utterance in each group for training and 35 for testing.

5. Results

5.1. Dysarthria Detection

As seen in Table 3, including rhythm improves the performance of TORGO detection results for all three classifiers. MLP classifiers saw an accuracy increase from 82.3% to 85.7% and a precision increase from 83.1% to 86.4%. Recall and F1-scores saw similar improvements. From Table 6 we see relative increases of 3.3%, 1%, and 4.1% for RF, SVM and MLP classifiers respectively.

In Table 4 an accuracy increases from 94.4% to 97.8% is seen for the QoLT data when including rhythm metrics. Furthermore, a precision of 100% and F1-score of 98% can be reached when using an RF classifier. Lastly, Table 6 shows relative increases of 3.6%, 4.9% and 7.5% for RF, SVM, and MLP classifiers respectively.

Table 3: *Detection results in % for TORGO data. First row only uses non-rhythm features. Second row using all features including rhythm.*

Classifier	Accuracy	Precision	Recall	F1-score
RF	78.9	76.9	83.3	80
SVM	81.5	**86.5**	75	80.4
MLP	**82.3**	83.1	**85**	**82.4**
RF	81.5	78.8	**86.7**	82.5
SVM	82.3	81	85	82.9
MLP	**85.7**	**86.4**	86	**85.7**

Table 4: *Detection results in % for QoLT data. First row only uses non-rhythm features. Second row using all features including rhythm.*

Classifier	Accuracy	Precision	Recall	F1-score
RF	**94.4**	**97.9**	**92**	**94.9**
SVM	88.9	97.6	82	89.1
MLP	87.8	91.5	86	88.7
RF	**97.8**	**100**	**96**	**98**
SVM	93.3	97.8	90	93.8
MLP	94.4	97.9	92	94.8

5.2. Severity Assessment

Table 5 displays the accuracy scores for both TORGO and QoLT datasets when using SVM and MLP classifiers. An accuracy improvement of 66.8% from 64.73% is achieved with the TORGO data when using a rhythm based SVM classifier. From Table 6 we see that the rhythm based SVM provides a 3.2% relative increase in accuracy. However, including rhythm in an MLP classifier does not improve accuracy. When implementing MLP classifiers we see an accuracy of 64.35% when using non-rhythm features and an accuracy of 63.9% when using rhythm.

Similar to the TORGO dataset, the QoLT dataset also shows higher accuracy scores with the SVM classifier compare to MLP classifiers. When using rhythm with SVM classifiers we see an accuracy of 63.6% while an accuracy of 59.1% was seen for the MLP classifier. However, a large relative increase is seen for both SVM (15% relative increase) and MLP classifiers (13.4% relative increase) when including the rhythm metrics. See Table 6 for a full comparison of relative increase for both tasks from both datasets.

Table 5: Assessment *Accuracy in % for QoLT and TORGO data.*

Corpus	Classifier	Accuracy w/o rhythm	Accuracy w/ rhythm
TORGO	SVM	64.7	**66.8**
	MLP	64.3	63.9
QoLT	SVM	55.3	**63.6**
	MLP	52.1	59.1

Table 6: Relative % accuracy improvement when using rhythm.

Corpus	Experiment	Model	Relative increase (%)
TORGO	Detection	RF	3.3
		SVM	1
		MLP	**4.1**
	Assessment	SVM	**3.2**
		MLP	-0.6
QoLT	Detection	RF	3.6
		SVM	4.9
		MLP	**7.5**
	Assessment	SVM	**15**
		MLP	13.4

6. Discussion and Conclusion

In this paper we examine the effects of including rhythm-based metrics for both detecting dysarthria and assessing severity level. We extracted the following rhythm measures: %V, ΔV/C, varco-V/C, V/Crpvi, and V/Cnpvi and used them as input along with standard prosodic measures to RF, SVM and MLP classifiers. The results suggest that rhythm metrics can be used along with prosody features to improve dysarthria detection and severity assessment for both English and Korean speakers. For detection, larger benefits were seen for the Korean QoLT dataset with a 7.5% relative increase in accuracy for detection compared to the English TORGO dataset with a 4.1% increase.For assessment, a 3.2% and 15% relative increase was seen for English and Korean respectively.

The differences seen between English and Korean results suggest that rhythm metrics may be more useful in detecting dysarthria with Korean speech. More research will need to be conducted on other languages or with larger datasets to make any claims for specific language benefits. Future research should include a more thorough investigation on the specific rhythm metrics and their individual contribution towards detection or assessment.

7. Acknowledgements

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2020.

8. References

[1] Bunton, K., Kent, R. D., Kent, J. F., & Rosenbek, J. C. (2000). Perceptuo-acoustic assessment of prosodic impairment in dysarthria. *Clinical Linguistics & Phonetics*, *14*(1), 13-24.

[2] Lowit-Leuschel, A., & Docherty, G. J. (2001). Prosodic variation across sampling tasks in normal and dysarthric speakers. *Logopedics Phoniatrics Vocology*, *26*(4), 151-164.

[3] Shin, H. B., & Ko, D. H. (2017). An aerodynamic and acoustic characteristics of Clear Speech in patients with Parkinson's disease. *Phonetics and Speech Sciences*, *9*(3), 67-74.

[4] Kang, Y., Yoon, K., Seong, C., & Park, H. (2012). A Preliminary Study of the Automated Assessment of Prosody in Patients with Parkinson's Disease. *Communication Sciences & Disorders*, *17*(2), 234-248.

[5] Nam, H. W., & Kwon, D. H. (2005). Prosodic characteristics in the persons with spastic and athetoid cerebral palsy. Journal of Speech and Hearing Disorders, 14(2), 111-127

[6] Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of speech, language, and hearing research*.

[7] Selouani, S. A., Dahmani, H., Amami, R., & Hamam, H. (2012). Using speech rhythm knowledge to improve dysarthric speech recognition. *International Journal of Speech Technology*, *15*(1), 57-64.

[8] Dahmani, H., Selouani, S. A., O'shaughnessy, D., Chetouani, M., & Doghmane, N. (2013). Assessment of dysarthric speech through rhythm metrics. *Journal of King Saud University-Computer and Information Sciences*, *25*(1), 43-49.

[9] Kadi, K. L., Selouani, S. A., Boudraa, B., & Boudraa, M. (2013). Discriminative prosodic features to assess the dysarthria severity levels. In *Proceedings of the World Congress on Engineering* (Vol. 3).

[10] Menendez-Pidal, X., Polikoff, J. B., Peters, S. M., Leonzio, J. E., & Bunnell, H. T. (1996, October). The Nemours database of dysarthric speech. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* (Vol. 3, pp. 1962-1965). IEEE.

[11] Rudzicz, F., Namasivayam, A. K., & Wolff, T. (2012). The TORGO database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation*, *46*(4), 523-541.

[12] Choi, D. L., Kim, B. W., Kim, Y. W., Lee, Y. J., Um, Y., & Chung, M. (2012, May). Dysarthric Speech Database for Development of QoLT Software Technology. In *LREC* (pp. 3378-3381).

[13] Le Dorze, G., Ouellet, L., & Ryalls, J. (1994). Intonation and speech rate in dysarthric speech. *Journal of communication disorders*, *27*(1), 1-18.

[14] Ackermann, H., & Hertrich, I. (1994). Speech rate and rhythm in cerebellar dysarthria: An acoustic analysis of syllabic timing. *Folia Phoniatrica et Logopaedica*, *46*(2), 70-78.

[15] Schlenck, K. J., Bettrich, R., & Willmes, K. (1993). Aspects of disturbed prosody in dysarthria. *Clinical linguistics & phonetics*, *7*(2), 119-128.

[16] Dogan, M., Midi, I., Yazıcı, M. A., Kocak, I., Günal, D., & Sehitoglu, M. A. (2007). Objective and subjective evaluation of voice quality in multiple sclerosis. *Journal of Voice*, *21*(6), 735-740.

[17] Kim, J., Kumar, N., Tsiartas, A., Li, M., & Narayanan, S. S. (2015). Automatic intelligibility classification of sentence-level pathological speech. *Computer speech & language*, *29*(1), 132-144.

[18] Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292.

[19] Dellwo, V., & Wagner, P. (2003). Relationships between speech rate and rhythm. In *Proceedings of the ICPhS*.

[20] Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, *7*(515-546).

[21] Mairano, P., & Romano, A. (2010). Un confronto tra diverse metriche ritmiche usando Correlatore 1.0. *La dimensione temporale del parlato*, *427*, 44.

[22] Enderby, P. (1980). Frenchay dysarthria assessment. *British Journal of Disorders of Communication*, *15*(3), 165-173.

[23] Orozco-Arroyave, J. R., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Daqrouq, K., Skodda, S., ... & Nöth, E. (2016). Automatic detection of Parkinson's disease in running speech spoken in three different languages. *The Journal of the Acoustical Society of America*, *139*(1), 481-500.

[24] López, J.V.E., Orozco-Arroyave, J.R., Gosztolya, G. (2019) Assessing Parkinson's Disease from Speech Using Fisher Vectors. Proc. Interspeech 2019, 3063-3067

[25] Kodrasi, I., & Bourlard, H. (2019, May). Super-gaussianity of Speech Spectral Coefficients as a Potential Biomarker for Dysarthric Speech Detection. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6400-6404). IEEE.

[26] A. Tripathi, S. Bhosale and S. K. Kopparapu, "Improved Speaker Independent Dysarthria Intelligibility Classification Using Deepspeech Posteriors," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 6114-6118, doi: 10.1109/ICASSP40776.2020.9054492.