Dysarthric speech recognition using dysarthria-severity-dependent and speaker-adaptive models

Conference Paper · August 2013

Myungjong Kim
University of Texas at Dallas

Dysarthric Speech Recognition Using Dysarthria-Severity-Dependent and Speaker-Adaptive Models

Myung Jong Kim, Joohong Yoo, Hoirin Kim
Department of Electrical Engineering
Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
myungjong@kaist.ac.kr, j9h2y1@kaist.ac.kr, hrkim@ee.kaist.ac.kr

Abstract

Dysarthria is a motor speech disorder that impairs the physical production of speech. Modern automatic speech recognition for normal speech is ineffective for dysarthric speech due to the large mismatch of acoustic characteristics. In this paper, a new speaker adaptation scheme is proposed to reduce the mismatch. First, a speaker with dysarthria is classified into one of the pre-defined severity-levels, and then an initial model to be adapted is selected depending on their severity-level. The candidates of an initial model are generated using dysarthric speech associated with their labeled severity-level in the training phase. Finally, speaker adaptation is applied to the selected initial model. Evaluation of the proposed method on a database of several hundred words for 31 speakers with moderate to mild dysarthria showed that the proposed approach provides substantial improvement over the conventional speaker-adaptive system when a small amount of adaptation data is available.

Index Terms: dysarthria, speech recognition, speaker adaptation, severity classification

1. Introduction

Dysarthria is a neuro-motor articulatory disorder that damages the physical production of speech, rendering it unintelligible. Dysarthria is often accompanied with a physical disability that limits the speaker's capability to communicate through computers and electronic devices, making keyboard typing about 300 times slower than for regular users [1]. However, dysarthric speech is at most about 15 times slower than regular speech [2]. Consequently, people with dysarthria tend to prefer spoken expression over other physical modes due to its relative naturalness and speed [1]. Although an automatic speech recognition (ASR) system is essential for dysarthria sufferers, current ASR systems for the general public are not well-suited to dysarthric speech because of their articulatory limitations.

There have been many studies for dysarthric speech recognition, and it can be categorized into two classes; feature-space and model-space approaches. Feature-space approaches focus on how to improve speech intelligibility using speech transformation techniques, such as formant and energy modification [3][4], or how to extract features appropriately capturing acoustic characteristics such as phonological features [5][6]. Model-space ones, on the other hand, deal with how to design an acoustic model. Ergodic topology [8], state-expanded [10], and state-interpolated [11] hidden Markov models (HMMs) were explored. Also, discriminative models [5][9], such as support vector machines (SVMs), neural networks, and latent dynamic conditional random fields, were applied to dysarthric speech recognition. To make the system more suitable for an individual, speaker-adaptive (SA) models, which are adjusted to a single user from speaker-independent (SI) initial models trained on a large population with normal speech, were investigated [14][15]. The studies recently report that SA models are more appropriate for mild to severe dysarthric speakers compared to speaker-dependent (SD) models, which are trained solely to the individual, and SI models, which are trained on normal speech.

Dysarthric speech deviates considerably from normal speech in various ways. Nonetheless, it can be characterized by highly consistent articulatory errors for each speaker [16]. Therefore, a speaker adaption technique would be a promising method. In adaptation methods such as maximum likelihood linear regression (MLLR) [17] or maximum *a posteriori* (MAP) [18], choosing an appropriate initial model to be adapted directly affects the overall performance [19]. However, existing speaker adaptation methods [14][15] used a normal SI model as the initial model. This can result in a less optimal solution due to considerable mismatch between the acoustic characteristics of normal and dysarthric speech.

In this paper, we propose a new speaker adaptation scheme to effectively reduce the mismatch. First, a speaker with dysarthria is automatically classified into one of the several pre-defined severity-levels using multiple-speech-dimension features [23]. Then, we select an initial model depending on the dysarthric speaker's severity-level. The candidates of an initial model are obtained using dysarthric speech associated with the identical severity-level graded by speech-language pathologists in the training phase. Finally, speaker adaptation is applied to the selected initial model. In this work, MLLR and MAP are sequentially used as the speaker adaptation method. While the current study is related to classical speaker adaptation methods [12]-[15], it tries to find an appropriate initial model depending on the dysarthric speaker's severity-levels, which has not been considered in the earlier studies. Also, compared with earlier studies, the proposed method was evaluated on a relatively large scale database of over several hundred utterances and of over 30 dysarthric speakers, in order to enhance the reliability of experimental results.

2. Data Description

A database of dysarthric speech was constructed for developing a speech recognizer individually customized for disabled persons with dysarthria under the Quality-of-Life Technology (QoLT) project in Korea [20]. In 2011, the database consisted of speech utterances from 130 subjects of which 100 (65 males and 35 females) are dysarthric and 30 (20 males and 10 females) are normal (non-dysarthric) speakers. 359 utterances were recorded from each dysarthric speaker and 595 utterances from each normal speaker. The database includes repetitions of 37 Assessment-of-Phonology-and-Articulation-for-Children (APAC) words [21] which are commonly used for assessing the articulation disorder for children in Korea, 100 command words, 36 Korean phonetic codes which are used for identifying the Korean alphabet letters in voice communication, and a subset from 452 Korean Phonetically Balanced Words (PBW). The APAC words are also phonetically balanced to assess partially the articulation ability on phone basis and are used for diagnosing the severity of dysarthric speech. This part is called the QoLT 2011 database. At present, the amount of the database is increased. 36 dysarthric speakers (22 speakers are selected from existing speakers and 14 speakers are newly added) were involved. 1,820 utterances were recorded from each dysarthric speaker. The database includes 10 repetitions of 37 APAC words, 100 command words, 36 Korean phonetic codes, and 9 words for measuring voice onset time. It is called the QoLT 2012 database. These are collected in multiple sessions.

All dysarthric participants have been diagnosed by a speech-language pathologist according to the percentage of consonant correct (PCC) [22], which is defined by the ratio of the number of correctly uttered consonants and the number of total consonants, using the APAC words. According to this assessment, among total 114 subjects, 77 subjects were graded as mildly dysarthric (PCC 85-100%), 23 subjects as mildly-to-moderately dysarthric (PCC 65-84.9%), 9 subjects as moderately-to-severely dysarthric (PCC 50-64.9%), and 5 subjects as severely dysarthric (PCC under 50%). In this work, we focus on mildly and moderately dysarthric speakers (except severely dysarthric speakers). In order to do that, mildly-to-moderately and moderately-to-severely dysarthric speakers were grouped into mildly-to-severely dysarthric speakers to simplify severity-levels as [7].

3. Proposed Method

3.1. Motivation and procedure

In general, a speaker adaptation technique modifies SI model parameters for a single speaker to make it more speaker-specific. An important issue that affects the overall performance is the choice of an initial model to be adapted [19]. However, earlier studies related to speaker adaptation for dysarthric speech recognition do not consider this issue [12]-[15]. They used a normal SI model as the initial model. This can result in a less optimal solution due to the considerable mismatch between the acoustic characteristics of normal and dysarthric speech. In this paper, a new speaker adaptation scheme is proposed to reduce the mismatch. In general, speech impairments may differ not only with dysarthria type, but also by the severity of the disorder [7]. The recent studies show that speech intelligibility predictors depending on severity-levels are more useful in assessing dysarthric speech intelligibility [26]. As such, the severity-dependent approach may be promising in speech recognition as well. The key idea of our approach is to select an appropriate initial model among candidate models trained by dysarthric speech associated with an identical severity-level, depending on the test speaker's severity. It is expected that initial acoustic model space is moved from normal acoustic space to dysarthric acoustic space, which results in making a general adaptation technique better fit for a dysarthric speaker.

The details of the proposed method shown in Figure 1 are as follows. First, a speaker with dysarthria is classified into one of the several pre-defined severity-levels using multiple-speech-dimension features [23], which will be discussed in Section 3.2. In this work, we used two severity-levels: mild and mild-to-severe levels. After that, an initial model to be

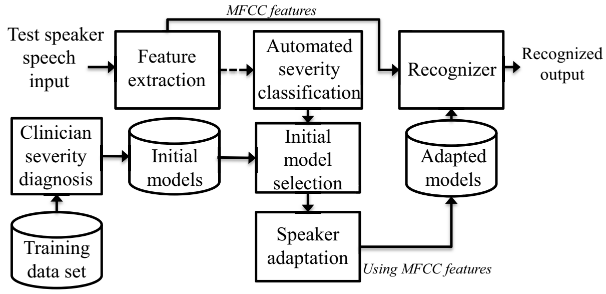

Figure 1: *Block diagram of the proposed method (actually the multiple speech dimension features as in [23] are used only for severity classification (- -) and MFCCs are used for speech recognition).*

adapted is chosen depending on the severity-level of a dysarthric speaker. The candidates of the initial model are obtained using dysarthric speech associated with the identical severity-level graded by speech clinicians in the training phase. In this work, the candidates include mild, mild-to-severe, and universal dysarthria-severity models. Here, the universal dysarthria-severity model is obtained using all dysarthric speech which is a combination of mild and mild-to-severe dysathria. The three dysarthria-severity models are all derived from the normal baseline SI model using adaptation data from either mild, mild-to-severe, or all dysarthric speakers separated from test speakers, respectively. Our preliminary study obtained the best performances on the mild dysarthria-severity model for mild speakers and on the universal dysarthria-severity model for mild-to-severe speakers, as reported in Section 4.2.1. Using the prior knowledge, a mild speaker was mapped to the mild dysarthria-severity model and a mild-to-severe speaker was mapped to the universal dysarthria-severity model as an initial model. Finally, speaker adaptation is applied to the selected initial model. In this work, MLLR and MAP are sequentially used and details will be explained in Section 3.3.

3.2. Automatic classification of dysarthria-severity-levels

The automatic classification of dysarthria-severity-levels was performed as a pre-step to make the proposed system fully automatic and flexible. In this work, we classify whether a test speaker is mild or mild-to-severe dysarthria. To this end, the best 10 speech features that are proper to predict the degrees of speech disorders from multiple speech dimensions composed of phonetic quality, prosodic quality, and voice quality are exploited [23]. The used features are as follows: *word recognition rates*, *percentage of voiced segments*, *mean of linear prediction residuals* (*LPRes*), *standard deviations of spectral roll-off and LPRes*, *zero-crossing rates* (*ZCR*) *of LPRes* (*LPResZCR*), *kurtosis of log-likelihoods*, *kurtosis of ratio of spectral flatness and centroid*, *kurtosis of LPResZCR*, and *skewness of ZCR*. For more details please refer to [23]. Then, a linear-kernel SVM [25] is adopted as a classifier to determine the severity-level of a test speaker.

3.3. Adaptation methods

We performed MLLR and MAP sequentially to adapt general acoustic models using adaptation data. In this work, the adaptation technique is applied to the construction of a dysarthria-severity model (e.g. mild and mild-to-severe models) using large population data associated with an identical dysarthria severity-level and to the construction of a speaker-specific model using data uttered by only single speaker (i.e. speaker adaptation).

MLLR [17] estimates linear transformations of model parameters to maximize the likelihood of the adaptation data. The component means in the initial model are modified by the transformations to reduce the mismatch between the model and the adaptation data. In this paper, we performed two-pass MLLR adaptation. Global adaptation is first applied and then it is used as an input transformation to compute more specific transforms using a regression class tree. Finally, MAP [18] adaptation was performed to maximize the posterior probability using the MLLR transformed models as the priors.

4. Experiments

4.1. Experimental setup

The acoustic features consist of 12 Mel-frequency cepstral coefficients (MFCCs), 1 energy term, and their dynamics corresponding delta and acceleration coefficients with frame size of 25 ms and shift size of 10 ms. The baseline SI model consists of 5392 tied-state left-to-right triphone HMMs, where each HMM has 3 states and each state is modeled with 16 Gaussian mixture components. Diagonal covariance matrices are used in all of the HMMs.

To construct the normal baseline SI model, we used the Korean Phonetically Optimized Words (KPOW) database [24] composed of 37,993 utterances of 3,848 Korean words, and a subset of normal speech data from the QoLT 2011 database described in Section 2, which is a total of 5,700 utterances including 100 control words, 36 Korean phonetic codes, and a subset of the PBW. Here, the number of males and females are balanced to avoid gender bias. We first trained the normal SI model using the KPOW data and then environmental adaptation was performed to the normal SI model using the normal speech data of the QoLT 2011 database to compensate the mismatch between the KPOW and the QoLT databases. For adaptation, MLLR and MAP described in Section 3.3 were sequentially performed. For MLLR, a regression class tree with 40 terminals was adopted.

To construct the dysarthria-severity model, we used the 19 mildly and 19 mildly-to-severely dysarthric speakers from the QoLT 2011 database including 6,000 utterances per each group. We trained the dysarthria-severity model by adapting the baseline SI model using MLLR and MAP adaptation methods for each group. For a mild severity model, only mildly dysarthric speech data were used, for a mild-to-severe severity model, only mildly-to-severely dysarthric speech data were used, and for a universal severity model, all dysarthric speech data were used.

To evaluate the proposed method, the QoLT 2012 database was used. The evaluation database consists of 18 mildly and 13 mildly-to-severely dysarthric speakers and consists of 5 repetitions of 100 command words and 36 Korean phonetic words (i.e. total 680 utterances) per each speaker. The repetitions are obtained in multiple sessions. For dysarthria-severity classification described Section 3.2, APAC words were used. For speaker adaptation, command words collected from another session were used. 10, 20, 50, and 100 adaptation words were evaluated. As adaptation methods, MLLR with a regression class tree of 4 terminals and MAP were sequentially performed. The speakers in the evaluation set are totally separated from the training set.

Table 1. *Word error rates (%) of the baseline and dysarthria-severity models according to the labeled severity-levels of test speakers (M-to-S is referred to as Mild-to-Severe).*

Severity-levels	Baseline model	Dysarthria-severity model		
		Universal	Mild	M-to-S
Mild	22.6	16.4	**15.8**	20.7
M-to-S	70.0	**60.6**	62.1	62.3
All	42.5	**34.9**	35.2	38.1

4.2. Experimental results

4.2.1. Effectiveness of dysarthria-severity models

Table 1 presents the performances of the baseline and several dysarthria-severity models according to the manual severity-levels of test speakers. The performances of all the three severity models were better than with the baseline model, relatively reducing the word error rate (WER) by 17.9%, 17.2%, and 10.4% for universal, mild, and mild-to-severe models on average, respectively. More specifically, the mild severity model was better fitted for mild speakers, achieving a WER of 15.8%, and the universal severity model was better fitted for mild-to-severe speakers, producing a WER of 60.6%. On the other hand, the mild-to-severe model was not well-suited to both mild and mild-to-severe speakers. This is because both the inter- and intra-speaker variations are too wide for most mildly-to-severely dysarthric speakers [7]. Therefore, the mild severity model was chosen for mild speakers as an initial SI model and the universal severity model was chosen for mild-to-severe speakers in the proposed method.

4.2.2. Effectiveness of adaptation to the baseline and dysarthria-severity models

Figure 2 compares the performances of speaker adaptation to the baseline and severity models by varying the number of adaptation data. In Figure 2, "SpkAdapt" means speaker adaptation; "universal" means universal severity model; "manual" and "automatic" indicate that manually labeled and automatically classified severity-levels are used in the test phase, respectively. First of all, the accuracy of automatic severity classification for test speakers, which is used as a pre-step of the proposed method, described in Section 3.2 was 80.6%. Nonetheless, the performance of the proposed method (fifth bar in Figure 2) is quite similar to the manually classified method (fourth bar in Figure 2) regardless of the amount of adaptation data. This indicates that the effect of misclassification is negligible. Compared with the baseline system (first bar in Figure 2), the proposed method largely reduces the WER by 32.9%, 45.2%, 51.3%, 56.9% when 10, 20, 50, and 100 adaptation data were used, respectively. Also, the proposed method outperformed the conventional speaker-adapted baseline system (second bar in Figure 2), achieving 11.2%, 8.3%, 5.5%, and 2.1% relative improvements in the WER reduction when using 10, 20, 50, and 100 adaptation data, respectively. These indicate that the dysarthria-severity model can be a good candidate of the initial model in providing better speaker-specific model through adaptation techniques. Since the performances of the speaker-adapted baseline system and proposed system are quite close when using over 50 adaptation data, McNemar's tests were performed to assess statistical significance of the WER differences [27]. It showed to be statistically significant, giving a p-value of 0.0017 and 0.0010 for 50 and 100 adaptation data, respectively. Also, it is statistically significant at the level of a 0.001 p-value for both 10 and 20 adaptation data. The proposed method also shows slight improvements compared with the speaker-adapted universal severity model (third bar in Figure 2) in all cases. Thus, selecting an appropriate initial model depending on the speaker's severity-level is more effective and promising for dysarthric speech recognition.

4.2.3. Effectiveness of the proposed method according to the dysarthric speaker's severity-levels

To see the effectiveness of the proposed method for the test speaker's severity-levels, further investigations are carried out, and the results are represented in Figure 3. In Figure 3, to compare the methods in the same condition, the automatic severity classification method was used to decide the speaker's severity-level. As shown in Figure 3, as the amount of adaptation data is decreased, the performance gap between the proposed method and speaker-adapted baseline tends to be growing. For mild speakers, 5.9%, 8.0%, 11.6%, and 13.9% relative WER reductions were obtained when using 100, 50, 20, and 10 adaptation data, respectively. For mild-to-severe speakers, 1.5%, 3.4%, 6.8%, and 9.9% relative improvements were obtained on the same condition. We also obtained p-values of 0.0805 and 0.0070 for mild speakers and mild-to-severe speakers using 100 adaptation data, respectively. Although the performances between the speaker-adapted baseline system and the proposed system are quite similar, it showed to be statistically significant. From these results, we can conclude that the proposed method is especially successful on highly small adaptation data (less than 20) for both mild and mild-to-severe speakers. Since the initial model is influential when using small adaptation data, it proves that the mismatch between the initial model and adaptation data is greatly reduced.

4.2.4. Consistency of recognition accuracy

One of the factors which give rise to performance degradation is the consistency of speech [7]. To observe the consistency of speech, our evaluation set was divided into five sets in which each set consists of 136 words collected in the same session. That is, the five sets have same vocabulary but separate session. Recognition accuracy was then computed using the proposed method on 100 adaptation data for each set, and finally the standard deviation of the recognition rates for the five sets is calculated for each speaker. Surprisingly, the standard deviation is 2.0 on average for mild speakers, on the other hand, 4.6 for mild-to-severe speakers, which is more than twice that of mild speakers. This implies that the intra-speaker variation is too wide to be adapted by the available adaptation data and it limits the performance for mild-to-severe speakers. We found out that the causes which lead to the wide intra-speaker variation are articulatory errors as well as involuntary breathing, stuttering, and accidental pauses between syllables. Therefore, further works include the investigation to deal with the problems.

5. Conclusion

We proposed a new speaker adaptation scheme to improve the performance of dysarthric speech recognition. First, the severity of a speaker with dysarthria is automatically classified into one of two severity-levels (i.e. mild and mild-to-severe levels) and then an initial model to be adapted is selected depending on their severity-level. The candidates of an initial model are mild, mild-to-severe, and universal severity models. For mild speakers and mild-to-severe dysarthric speakers, mild and universal severity models are chosen as an initial model, respectively. Finally, MLLR and MAP speaker adaptation methods are sequentially applied to the selected initial model. The proposed method is evaluated on a relatively large scale database of over several hundred utterances and of over 30 dysarthric speakers. Experimental results showed that the proposed speaker adaptation approach provides significant improvement over the conventional approach when a small amount of adaptation data is available. Dysarthric speakers are easily tired by prolonged speech activity, so minimizing data required for a valid training set is an important consideration.

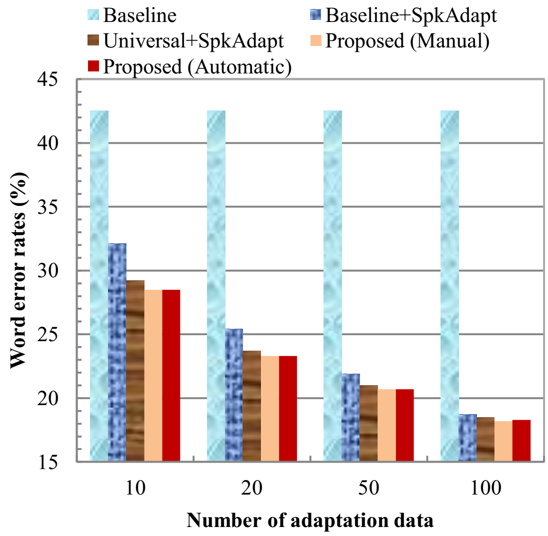

Figure 2: *Performance comparison of speaker adaptation to the baseline and severity models by varying the number of adaptation data.*

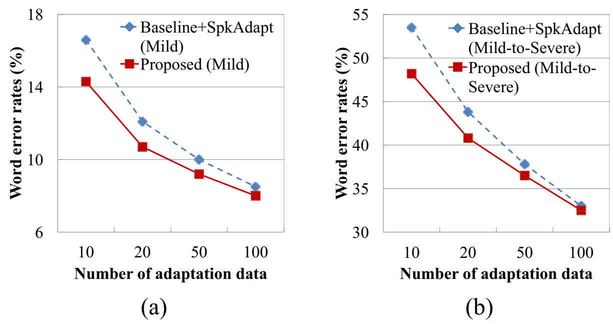

Figure 3: *Performance comparison of the speaker-adapted baseline and proposed method with the automatic severity classification according to the number of adaptation data for (a) mild and (b) mild-to-severe dysarthria speakers.*

6. Acknowledgements

This work was supported by the Technology Innovation Program of the Ministry of Trade, Industry & Energy. [10036461, Development of an embedded key-word spotting speech recognition system individually customized for disabled persons with dysarthria]

7. References

[1] J. –P. Hosom, A. B. Kain, T. Mishra, J. P. H. van Santen, M. Fried-Oken, and J. Staehely, "Intelligibility of modifications to dysarthric speech," in *Proc. IEEE Int. Conf. Acoust. Speech, and Signal Process.*, Apr. 2003, pp. 924–927.

[2] F. Rudzicz, "Learning mixed acoustic/articulatory models for disabled speech," in *Proc. Workshop on Machine Learning for Assistive Technologies on Nueral Information Processing Systems*, Dec. 2010, pp. 70–78.

[3] Al. Kain, X. Niu, J. –P. Hosom, Qi Miao, and J. van Santen, "Formant re-synthesis of dysarthric speech," in *Proc. IEEE Workshop on Speech Synthesis*, 2004, pp. 25–30.

[4] H. Tolba and A. S. EL Torgoman, "Towards the improvement of automatic recognition of dysarthric speech," in *Proc. IEEE Int. Conf. Comput. Science and Inform. Technology*, Aug. 2009, pp. 277–281.

[5] F. Rudzicz, "Articulatory knowledge in the recognition of dysarthric speech," *IEEE Trans. Audio, Spcch, Lang. Process.*, vol. 19, no. 4, May 2011, pp. 947–960.

[6] F. Rudzicz, "Phonological features in discriminative classification of dysarthric speech," in *Proc. IEEE Int. Conf. Acoust. Speech, and Signal Process.*, Taipei, Apr. 2009, pp. 4605–4608.

[7] 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: The Official Journal of RESNA*, vol. 22, no. 2, pp. 99–112, 2010.

[8] P. D. Polur and G. E. Miller, "Experiments with fast fourier transform, linear predictive and cepstral coefficients in dysarthric speech recognition algorithms using hidden Markov model," *IEEE Trans. Neural Sys. Rehab. Eng.*, vol. 13, no. 4, Dec. 2005, pp. 558–561.

[9] M. Hasegawa-Johnson, J. Gunderson, A. Perlman, and T. Huang, "HMM-based and svm-based recognition of the speech of talkers with spastic dysarthria," in *Proc. IEEE Int. Conf. Acoust. Speech, and Signal Process.*, 2006, pp. 1060–1063.

[10] P. Green, J. Carmichael, A. Hatzis, P. Enderby, M. Hawley, and M. Parker, "Automatic speech recognition with sparse training data for dysarthric speakers," in *Proc. Interspeech 2003*, Geneva, Switzerland, Sep. 2003.

[11] H. V. Sharma and M. Hasegawa-Johnson, "State-transition interpolation and MAP adaptation for HMM-based dysarthric speech recognition," in *Proc. NAACL HLT 2010 Workshop on Speech and Lang. Process. Assistive Tech.*, Jun. 2010, pp. 72–79.

[12] P. Raghavendra, E. Rosengren, and S. Hunnicutt, "An investigation of different degrees of dysarthric speech as input to speaker-adaptive and speaker-dependent recognition systems," *AAC: Augmentative and Alternative Communication*, vol. 17, no. 4, pp. 265–275, 2001.

[13] F. Rudzicz, "Comparing speaker-dependent and speaker-adaptive acoustic models for recognizing dysarthric speech," in *Proc. ASSETS*, Tempe, Arizona, USA, Oct. 2007, pp. 255–256.

[14] K. T. Mengistu and F. Rudzicz, "Adapting acoustic and lexical models to dysarthric speech," in *Proc. IEEE Int. Conf. Acoust. Speech, and Signal Process.*, 2011, pp. 4924–4927.

[15] H. Christensen, S. Cunningham, C. Fax, P. Green, and T. Hain, "A comparative study of adaptive, automatic recognition of disordered speech," in *Proc. Interspeech 2012*, Portland, Oregon, Sep. 2012.

[16] K. M. Yorkston, D. R. Beukelman, and K. R. Bell, *Clinical management of dysarthric speakers*, San Diego, CA: College-Hill Press, 1988.

[17] C. J. Legetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models," *Computer Speech and Language*, vol. 9, no. 2, pp. 171–185, 1995.

[18] J. –L. Gauvain and C. –H. Lee, "Maximum a posteriori estimation of multivariate Gaussian mixture observations of markov chains," *IEEE Trans. Speech and Audio Process.*, vol. 2, no. 2, pp. 291–298, Apr. 1994.

[19] X. Huang, A. Acero, and H. –W. Hon, *Spoken language processing*, Prentice-Hall, 2001.

[20] D. –L. Choi, B. –W. Kim, Y. –J. Lee, Y. Um, and M. Chung, "Design and creation of dysarthric speech database for development of QoLT software technology," in *Proc. Speech Database and Assessment* (*Oriental COCOSDA*), Hsinchu City, Taiwan, Oct. 2011, pp. 47–50.

[21] M. J. Kim, S. Pae, and C. Park, *Assessment of phonology and articulation for children*, Human Brain Research & Consulting, 2007.

[22] L. D. Shriberg and J. Kwiatrkowski, "Phonological disorders III: A procedure for assessing severity of involvement," *Journal of Speech and Hearing Disorders*, vol. 47, no. 3, pp. 256–270, 1982.

[23] M. J. Kim and H. Kim, "Combination of multiple speech dimensions for automatic assessment of dysarthric speech intelligibility," in *Proc. Interspeech 2012*, Portland, Oregon, Sep. 2012.

[24] Y. Lim and Y. Lee, "Implementation of the POW (phonetically optimized words) algorithm for speech database," in *Proc. IEEE Int. Conf. Acoust. Speech, and Signal Process.*, Detroit, Mich, USA, May 1995, pp. 89–92.

[25] C. M. Bishop, *Pattern recognition and machine learning*, Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006.

[26] M. S. Paja and T. H. Falk, "Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech," in *Proc. Interspeech 2012*, Portland, Oregon, Sep. 2012.

[27] L. Gillick and S. J. Cox, "Some statistical issues in the comparison of speech recognition algorithms," in *Proc. IEEE Int. Conf. Acoust. Speech, and Signal Process.*, Glasgow, Scotland, May 1989, pp. 532–535.