Nonavinakere Prabhakera, Narendra; Alku, Paavo

Dysarthric speech classification using glottal features computed from non-words, words and sentences

Published in:
Proceedings of Interspeech

DOI:
10.21437/Interspeech.2018-1059

Published: 02/09/2018

Document Version
Publisher's PDF, also known as Version of record

Interspeech 2018
2-6 September 2018, Hyderabad

Dysarthric speech classification using glottal features computed from non-words, words and sentences

N P Narendra, Paavo Alku

Aalto University, Department of Signal Processing and Acoustics, Espoo, Finland
{narendra.prabhakera, paavo.alku}@aalto.fi

Abstract

Dysarthria is a neuro-motor disorder resulting from the disruption of normal activity in speech production leading to slow, slurred and imprecise (low intelligible) speech. Automatic classification of dysarthria from speech can be used as a potential clinical tool in medical treatment. This paper examines the effectiveness of glottal source parameters in dysarthric speech classification from three categories of speech signals, namely non-words, words and sentences. In addition to the glottal parameters, two sets of acoustic parameters extracted by the openSMILE toolkit are used as baseline features. A dysarthric speech classification system is proposed by training support vector machines (SVMs) using features extracted from speech utterances and their labels indicating dysarthria/healthy. Classification accuracy results indicate that the glottal parameters contain discriminating information required for the identification of dysarthria. Additionally, the complementary nature of the glottal parameters is demonstrated when these parameters, in combination with the openSMILE-based acoustic features, result in improved classification accuracy. Analysis of classification accuracies of the glottal and openSMILE features for non-words, words and sentences is carried out. Results indicate that in terms of classification accuracy the word level is best suited in identifying the presence of dysarthria.
Index Terms: Dysarthric speech, glottal source, glottal parameters, openSMILE, support vector machines

1. Introduction

Dysarthria is a neuro-motor disorder resulting from neurological damage of motor component of speech production [1]. Dysarthria is generally a result of either a neurological injury (i.e., cerebral palsy, brain tumor, brain injury, stroke) or a symptom of a neurodegenerative disease (i.e., Parkinsons's disease, Amyotrophic lateral sclerosis, Huntington's disease). Speech disorders due to dysarthria are associated with reduced vocal tract volume and tongue flexibility, atypical speech prosody, imprecise articulation and variable speech rate, factors that all lead to poor speech intelligibility [2]. As dysarthric speech is distinguishable from healthy speech, the assessment of speech can be carried out for the identification of dysarthria. The speech assessment can be conducted by speech-language pathologists through intelligibility tests to judge the presence of dysarthria [3]. Subjective intelligibility tests are costly, laborious, and frequently prone to intrinsic biases of pathologists due to their familiarity with patients and their speech disorders [4][5]. This motivates to design an objective assessment method that can distinguish dysarthric voices from healthy speech.

For objective assessment, a dysarthric speech classification system is used which is basically a data-driven model trained on a dysarthric speech database. The data-driven model establishes a mapping between speech features and labels (dysarthric/healthy) determined by speech-language pathologists. In order to develop an efficient dysarthric speech classification system, the existing works mainly focus on extraction of acoustic features capable of capturing wide variabilities of sources and patterns in pathological speech [6][7][8]. Previous works have explored a range of features including spectral features (e.g., Mel-frequency cepstral coefficients, formants), prosody features (e.g., fundamental frequency, pitch contour, phone duration, RMS energy), voice quality features (e.g., jitter, shimmer, harmonic to noise ratio), perceptual features and phonological features [7][8] [9][10][11]. In addition to these widely explored acoustic features, few studies have utilized glottal parameters for detecting the presence of dysarthria [12]. In [12], glottal parameters are utilized in combination with acoustic parameters for developing cross-database models (training on one database and testing on another database) for identification of dysarthria. Existing works use glottal flow waveforms which are estimated using known glottal inverse filtering (GIF) methods such as Iterative Adaptive Inverse Filtering (IAIF) [13] and Rank-Based Glottal Quality Assessment (RBGQA) [14]. However, recent efficient GIF methods, such as Quasi-Closed Phase Analysis (QCP) [15], have not been explored in objective assessment of dysarthric speech. In addition, most of the previous studies on dysarthric speech classification have addressed a single type of speech signal such as either modulated vowels [16], words [7] or sentences [8]. To our knowledge, there are no previous comprehensive investigations to study the effectiveness of a set of proposed features and classifiers on different categories of speech signals, for example, vowels, words and sentences.

The main goal of this research work is to explore glottal parameters, extracted from the voice source signal estimated using the recently proposed QCP method [15], for dysarthric speech classification in three different speech signal categories, namely, non-words, words and sentences. Two sets of acoustic features extracted with the widely explored openSMILE toolkit [6] are used as baseline features. Support vector machine (SVM) classifier is trained using features extracted from each of the speech utterance and its corresponding label indicating dysarthria/healthy. This work explores the effectiveness of the glottal features, when used individually and combined with the baseline openSMILE features, in classification of dysarthric speech. Another important contribution of the current work which has not been explored before is the analysis of classification performance on different categories of speech signals (non-words, words and sentences) under a common framework. The paper is organized as follows. Description about the proposed dysarthric speech classification system is given in Section 2. The details about the speech database, experimental setup and results are provided in Section 3. The summary of the present work and discussion are given in Section 4.

10.21437/Interspeech.2018-1059

2. Proposed method

In order to classify dysarthric voices from healthy speech, a new dysarthric speech classification system is proposed. The training phase of the proposed system is shown in Figure 1. First, a multi-speaker dysarthric speech database is considered for training the classifier (described in Section 3.1). From every speech utterance of the database, time-domain and frequency-domain glottal parameters are extracted. In order to extract parameters, glottal flow waveform is estimated from the speech utterance by using the QCP method [15] which was found to be the best performing GIF method in comparison to existing algorithms [13][17][18] in [15] (details are provided in Section 2.1). Two sets of acoustic features are also extracted from every speech utterance using openSMILE [6] (described in Section 2.2) which is a widely used toolkit in paralinguistic speech processing tasks. Generally, sizes of acoustic and glottal feature sets are large. To avoid the risk of over-fitting, the feature set size is reduced by using the sequential forward feature selection (SFFS) algorithm [19]. The SFFS algorithm selects a subset of features from the feature set that results in the best classification accuracy. Starting from an empty feature set, SFFS creates candidate feature subsets by sequentially adding each of the features and each candidate feature subset is evaluated by computing the classification accuracy with 10-fold cross-validation. Using the features extracted from every speech utterance as input and corresponding dysarthric/healthy labels as output, a classifier is trained. Separate classifiers are trained using reduced and non-reduced feature sets for openSMILE, glottal features and their combination. In this work, SVMs are used as classifiers. SVMs are widely used in pathological speech classification and they are validated with consistent performance even for small amount of speech data in contrast to other techniques such as deep neural nets which require large amount data for proper training [8][20].

After completing the training, the SVM classifiers can be used to identify the presence of dysarthria in the input speech utterance. The same set of speech features which were used during training are extracted from the speech utterance, and the extracted features are fed to the SVM classifier, which outputs the dysarthric/healthy labels.

2.1. Glottal parameters extraction

In dysarthria, as the motor component of speech production is affected, vibration of the vocal folds will change compared to healthy speech. The difference in vocal fold vibration between dysarthric and healthy speech production cannot be characterized completely by the *rate* of vibration (i.e., pitch information). Instead, the *mode* of vibration of the vocal folds will be affected also. Therefore, the waveform of the acoustic speech excitation generated by the vocal folds, the glottal flow, may have useful discriminating information for dysarthric speech classification. In order to parameterize the glottal source, the flow waveform must be estimated first with GIF from the speech signal. In this work, we use QCP [15] as the GIF method in the estimation of the glottal flow.

2.1.1. QCP

QCP [15] is one of the recently proposed GIF methods to estimate the glottal source from speech. The QCP method is based on the principles of the closed phase analysis (CP) [17] which estimates the vocal tract response using the covariance method of linear prediction from few speech samples located in closed

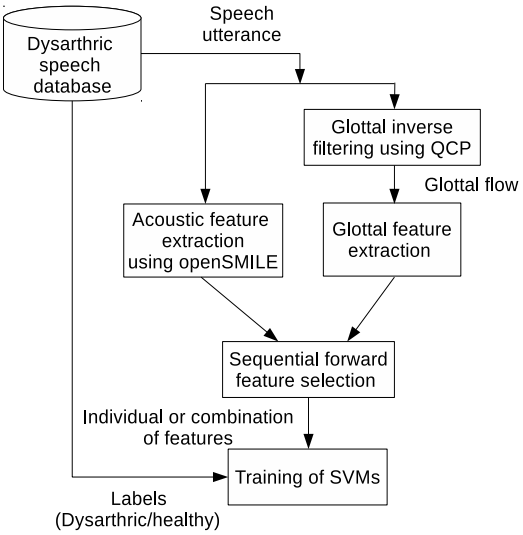

Figure 1: *Training phase of the proposed dysarthric speech classification system.*

phase of the glottal cycle. In contrast to the CP method, QCP creates a specific temporal weighting function, called the Attenuated Main Excitation (AME) function, using glottal closure instants (GCIs) estimated from speech. The AME function is used to attenuate the contribution of the (quasi-) open phase in the computation of the Weighted Linear Prediction (WLP) coefficients, which results in good estimates of the vocal tract transfer function. Evaluation results in [15] show that the accuracy of QCP is better than that of CP [17], IAIF [13] and complex cepstral decomposition (CCD) [18]. Hence, in this work, the glottal flow waveform is estimated using the QCP method. From the estimated glottal flow waveforms, time- and frequency-domain glottal parameters are extracted.

2.1.2. Time- and frequency-domain glottal parameters

The glottal flow computed by QCP is parameterized with a glottal parameter set consisting of 12 time- and frequency-domain parameters which characterize various aspects of the glottal flow waveform [21][22]. These parameters are extracted using APARAT Toolbox [23]. The time- and frequency-domain glottal parameters are listed in Table 1. H1H2 and HRF are obtained in the dB scale, and other parameters are obtained in a linear scale. The glottal parameters are computed in 30-ms frames. H12 and HRF are computed pitch-asynchronously once per frame whereas the rest of the parameters are computed pitch-synchronously once per glottal cycle and then averaged over the frame. The glottal parameters computed from all voiced frames of the input speech signal form finally the glottal parameter vector of the utterance. The following 8 statistical measures are computed from the glottal parameter vector as well as from its delta vector: mean, median, min, max, standard deviation, range, skewness, and kurtosis. This results in $(12 + 12) \times 8 = 192$ parameters representing the glottal feature set.

2.2. Acoustic parameters extraction using openSMILE

Acoustic parameters are extracted from speech using openSMILE, a freely available feature extraction toolkit [6]. The openSMILE features have been used as baselines for differ-

Table 1: *Time- and frequency-domain glottal parameters. For more details, see [23]*

Time-domain parameters	
OQ1	Open quotient, computed from primary glottal opening
OQ2	Open quotient, computed from secondary glottal opening
NAQ	Normalized amplitude quotient
AQ	Amplitude quotient
ClQ	Closing quotient
OQa	Open quotient, derived from the LF model
QOQ	Quasi-open quotient
SQ1	Speed quotient, computed from primary glottal opening
SQ2	Speed quotient, computed from secondary glottal opening
Frequency-domain parameters	
H12	Difference between first two glottal harmonics
PSP	Parabolic spectrum parameter
HRF	Harmonic richness factor

ent paralinguistic challenges from INTERSPEECH 2009 [24]. Some examples of paralinguistic challenges are recognition of emotion, speaker traits and states, and speech pathology. The acoustic features extracted by openSMILE mainly represent spectrum, prosody and voice quality. In this work, two sets of acoustic features defined in the openSMILE toolkit are used for dysarthric speech classification. The first set (referred in this work as openSMILE-1) is INTERSPEECH 2009 Emotion Challenge [24] feature set consisting of 384 features. This feature set consists of 16 acoustic features extracted from every frame (described in Table 2). The set of 16 acoustic features along with their derivatives obtained from all frames of a speech utterance forms the acoustic feature vector. 12 statistical functionals (shown in Table 2) are computed from the acoustic feature vector of the utterance to obtain (16 + 16) $\times$ 12 = 384 features representing the openSMILE-1 feature set.

The second set (referred in this work as openSMILE-2) is the large openSMILE emotion feature set consisting of 6552 features. This is the largest feature set in terms of the number of features in the openSMILE toolkit. The largest feature set is chosen to involve as much acoustic information as possible which may be helpful in dysarthric speech classification. A set of 56 acoustic features (given in Table 2) are extracted from every frame. 56 acoustic features along with their first and second order derivatives form the frame-level acoustic features. As in openSMILE-1, statistical functionals are applied on the acoustic feature vectors which are extracted from all frames of the speech utterance. Instead of 12, 39 statistical functionals (shown in Table 2) are applied to obtain (56 + 56 + 56) $\times$ 39 = 6552 features representing the openSMILE-2 feature set.

3. Experiments

The experiments conducted in this study evaluate the effectiveness of the glottal parameters in dysarthric speech classification. The classification accuracies of the combination of the glottal and openSMILE features are analyzed separately on non-words, words and sentences.

3.1. TORGO database

To develop the dysarthric speech classification system, the TORGO database [25] was utilized. This database contains speech recordings from seven patients (three females and four males), diagnosed with cerebral palsy or amyotrophic lateral sclerosis and speech recordings from seven healthy control speakers (three females and four males). The age range of patients is from 16 years to 50 years. The database includes speech signals in three categories, namely non-words, words and sentences. Non-words consist of 5-10 repetitions of */iy-p-ah/*, */ah-p-iy/*, and */p-ah-t-ah-k-ah/* and high- and low-pitched vowels maintained over 5 s (e.g., "Say 'eee' in a high pitch for 5 s"). Text prompts used to record short words include 50 words from the word intelligibility section of the Frenchay Dysarthria Assessment [26] and 360 words from the word intelligibility section of the Yorkston-Beukelman Assessment of Intelligibility of Dysarthric Speech [27]. Sentences comprise three pre-selected phoneme-rich sentences sets: Grandfather passage from the Nemours database [28], 162 sentences from sentence intelligibility section of the Yorkston-Beukelman Assessment of Intelligibility of Dysarthric Speech [27], 460 sentences from the MOCHA database [29] and spontaneously elicited descriptive texts.

In this study, utterances in all three speech signal categories of TORGO, recorded by an array microphone with 16-kHz sampling, are used. In each of the three categories, speech samples of seven patients (three females and four males) with dysarthria and seven healthy speakers (three females and four males) are considered. For speech signals at the level of words and sentences, 80 utterances from each speaker are used (except for two dysarthric speakers at sentence-level, only 23 and 28 utterances are used due to lack of availability of recordings) and for non-words, 8-9 utterances from each speaker (which are available in the database) are used in dysarthric speech classification.

3.2. Experimental setup

The speech data is processed in 30-ms frames at 15-ms intervals in the dysarthric speech classification system. Using openSMILE, two sets of acoustic features (openSMILE-1 and openSMILE-2) are extracted from every speech utterance of the TORGO database. Every frame of speech utterance is inverse filtered using QCP to obtain the glottal flow estimate. From glottal flow waveforms of every utterance, time- and frequency-domain glottal parameters are extracted using the APARAT toolbox. Both acoustic and glottal features are individually normalized by subtracting the global mean and dividing by the global standard deviation. The sizes of the feature sets are reduced by the SFFS algorithm. The process of feature extraction and reduction is carried out separately for the speech utterances of non-words, words and sentences. Separate set of SVM classifiers are developed for each category using acoustic and glottal feature sets both individually and combined. Also, SVM classifiers are developed for both the non-reduced and reduced feature sets. The SVM classifiers are trained using Gaussian, radial basis function kernel. The optimal values of kernel parameter γ and penalty parameter C are chosen based on grid search with C and γ varying from 10^{-3} to 10^{3} in multiples of 10. The pair (C,γ) is selected which resulted in the highest classification accuracy on test data. A leave-one-subject-out (LOSO) cross validation strategy is used to determine the classification accuracy. In this strategy, one speaker is used at every fold for validation and all other speakers are used for training. The cross-validation process is then repeated with each of the speaker used exactly once as the validation data. The classification accuracies obtained at all folds are averaged to obtain the final accuracy. The classification accuracy (also called the unweighted average recall) is computed as the ratio of number of correctly classified speech utterances to the total number of speech utterances.

Table 2: *Two openSMILE feature sets. For more details, see [6]*

Feature sets	Acoustic features	Statistical functionals
openSMILE-1	RMS-energy, MFCCs (12), zero-crossing rate, pitch, voicing probability	min (or max) value and its relative position, median, range, standard deviation, skewness, kurtosis, 2 linear regression coeff. and quadratic error
openSMILE-2	log-energy, MFCCs (13), Mel-spectrum (26), zero-crossing rate, pitch, jitter, shimmer, voicing probability, spectral flux, roll-off points, spectral centroid, position of spectral maximum and minimum	min (or max) value and its relative position, median, range, standard deviation, skewness, kurtosis, 2 linear regression coeff., linear and quadratic errors, 3 quartiles, 2 percentiles (95% & 98%), 3 inter-quartile errors, number of peaks, mean of peaks, mean distance between peaks, arithmetic, geometric and quadratic means

Table 3: *Classification accuracies obtained for each of three speech signal categories using both reduced and non-reduced feature sets.*

	Classification Accuracy	
Feature set (Non-words)	Without Feature selection (%)	With Feature selection (%)
OpenSMILE-1	60.11	84.46
OpenSMILE-2	70.07	89.23
Glottal	69.34	78.16
OpenSMILE-1 + glottal	69.53	88.41
OpenSMILE-2 + glottal	67.75	93.52
Feature set (Words)	Without Feature selection (%)	With Feature selection (%)
OpenSMILE-1	78.84	88.39
OpenSMILE-2	80.36	93.39
Glottal	68.30	72.77
OpenSMILE-1 + glottal	77.14	92.77
OpenSMILE-2 + glottal	82.32	94.29
Feature set (Sentences)	Without Feature selection (%)	With Feature selection (%)
OpenSMILE-1	69.39	87.08
OpenSMILE-2	76.77	90.87
Glottal	61.08	71.86
OpenSMILE-1 + glottal	64.31	87.56
OpenSMILE-2 + glottal	74.63	91.38

3.3. Results

Table 3 shows the average classification accuracies of leave-one-subject-out cross validation for non-words, words and sentences using both the reduced and non-reduced feature sets. In comparing the classification accuracies for all types of feature sets in each of the three categories, it can be observed that the usage of reduced feature sets results in better accuracy compared to the non-reduced feature sets. From the table, it can be observed that with more than 80 % classification accuracies, two sets of openSMILE based features have better classification accuracies than glottal parameters after feature selection. The classification accuracies of the glottal parameters is more than 70 % after feature selection. This indicates that the glottal parameters contain discriminative information important for the classification of dysarthric speech. Most importantly, by combining the glottal parameters with the openSMILE features, the classification accuracies improve in all three speech signal categories after feature selection. This shows that the glottal parameters contain complementary information which results in the improvement of accuracies when combined with the widely used openSMILE features.

On comparing the classification accuracies in the three speech signal categories, word-level signals show the highest score for both of the openSMILE feature sets and for both of the combined feature sets. Word-level utterances are short speech segments which provide useful segmental information related to speech pathology and feature extraction can be performed on word-level utterances with less complexity. Hence, it can be concluded that word-level speech signals are best suited for performing dysarthric speech classification. However, classification accuracies obtained using the glottal parameters alone were highest for non-words. This suggests that even though non-words do not carry any linguistically important information, speech production is mostly voiced in this category and therefore glottal features contain relevant information related to speech disorder classification. Sentence-level signals have slightly better accuracies compared to non-words for two sets of the openSMILE features. The amount of improvement in accuracy by adding the glottal parameters to the openSMILE features is very small (about 0.5 %) for sentences compared to words and non-words. The feature characteristics and robustness that are observed for shorter segments (e.g., words) might not be consistent with sentence-level speech data due to high variability and complexity of sentence-level speech production.

4. Conclusions

This paper proposes a dysarthric speech classification system using glottal features and evaluates its performance separately on non-words, words and sentences. Two openSMILE-based feature sets are used as baselines. SVM classifiers are trained to predict dysarthria/healthy labels using features extracted from speech. Experiments show that the glottal parameters resulted in fairly good classification accuracies (around 70%) in all three speech signal categories. Results also show that combining the glottal parameters with the openSMILE features results in improved classification accuracies. Among the three speech signal categories, word-level signals leads to the highest classification accuracies for the two openSMILE feature sets and for the combination of the openSMILE and glottal features.

To the best of our knowledge, the current study is the first investigation in which effectiveness of glottal parameters in dysarthric speech classification is compared between non-words, words and sentences. Possible future works are as follows. Dysarthric speech classification using glottal parameters can be explored using other databases such as the Universal access speech database [30]. In addition to the standard time- and frequency-domain glottal parameters, other methods of parameterization can be explored. The effectiveness of glottal parameters in classification of dysarthria can be investigated under different realistic scenarios such as presence of environment noise, band-pass filtering and speech coding.

5. Acknowledgements

This research has been funded by the Academy of Finland (project no. 312490).

6. References

[1] P. C. Doyle, H. A. Leeper, A. L. Kotler, N. Thomas-Stonell, C. O'Neill, M. C. Dylke, and K. Rolls, "Dysarthric speech: a comparison of computerized speech recognition and listener intelligibility," *Journal of Rehabilitation Research and Development*, vol. 34, no. 3, pp. 309–316, 1997.

[2] J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, 3rd ed. Elsevier Health Sciences, 2012.

[3] R. D. Kent, *Intelligibility in Speech Disorders: Theory, Measurement, and Management*. John Benjamins Publishing, 1992, vol. 1.

[4] M. S. De Bodt, M. E. Hernández-Díaz Huici, and P. H. Van De Heyning, "Intelligibility as a linear combination of dimensions in dysarthric speech," *Journal of Communication Disorders*, vol. 35, no. 3, pp. 283–292, 2002.

[5] G. Van Nuffelen, C. Middag, M. De Bodt, and J.-P. Martens, "Speech technology-based assessment of phoneme intelligibility in dysarthria," *International Journal of Language and Communication Disorders*, vol. 44, no. 5, pp. 716–730, 2009.

[6] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in *Proc. ACM International Conference on Multimedia*, 2013, pp. 835–838.

[7] T. H. Falk, W.-Y. Chan, and F. Shein, "Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility," *Speech Communication*, vol. 54, pp. 622–631, 2012.

[8] J. Kim, N. Kumar, A. Tsiartas, M. Li, and S. S. Narayanan, "Automatic intelligibility classification of sentence-level pathological speech," *Computer Speech and Language*, vol. 29, pp. 132–144, 2015.

[9] A. A. Dibazar, S. Narayanan, and T. W. Berger, "Feature analysis for automatic detection of pathological speech," in *Proc. Joint EMBS/BMES Conference*, 2002, pp. 182–182.

[10] F. Rudzicz, "Phonological features in discriminative classification of dysarthric speech," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 4605–4608.

[11] J. M. Liss, S. LeGendre, and A. J. Lotto, "Discriminating dysarthria type from envelope modulation spectra," *Journal of Speech, Language, and Hearing Research*, vol. 53, no. 5, pp. 1246–1255, 2010.

[12] S. Gillespie, Y.-Y. Logan, E. Moore, J. Laures-Gore, S. Russell, and R. Patel, "Cross-database models for the classification of dysarthria presence," in *Proc. Interspeech*, 2017, pp. 3127–3131.

[13] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Communication*, vol. 11, no. 2-3, pp. 109–118, 1992.

[14] E. Moore and J. Torres, "A performance assessment of objective measures for evaluating the quality of glottal waveform estimates," *Speech Communication*, vol. 50, no. 1, pp. 56–66, 2008.

[15] M. Airaksinen, T. Raitio, B. Story, and P. Alku, "Quasi closed phase glottal inverse filtering analysis with weighted linear prediction," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 3, pp. 596–607, 2014.

[16] K. L. Lansford and J. M. Liss, "Vowel acoustics in dysarthria: Speech disorder diagnosis and classification," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 1, pp. 57–67, 2014.

[17] D. Wong, J. Markel, and A. Gray Jr, "Least squares glottal inverse filtering from the acoustic speech waveform," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 4, pp. 350–355, 1979.

[18] T. Drugman, B. Bozkurt, and T. Dutoit, "Complex cepstrum-based decomposition of speech for glottal source estimation," in *Proc. interspeech*, 2009, pp. 116–119.

[19] J. Reunanen, "Overfitting in making comparisons between variable selection methods," *Journal of Machine Learning Research*, vol. 3, pp. 1371–1382, 2003.

[20] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londono, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, , and E. Nöth, "Automatic detection of Parkinsons disease from words uttered in three different languages," in *Proc. Interspeech*, 2014, pp. 1573–1577.

[21] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parameterization of the glottal flow," *Journal of the Acoustical Society of America*, vol. 112, no. 2, pp. 701–710, 2002.

[22] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *Journal of the Acoustical Society of America*, vol. 90, no. 5, pp. 2394–2410, 1991.

[23] M. Airas, H. Pulakka, T. Bäckström, and P. Alku, "A toolkit for voice inverse filtering and parametrisation," in *Proc. Interspeech*, 2005, pp. 2145–2148.

[24] B. Schuller, S. Steidl, and A. Batliner, "The INTERSPEECH 2009 emotion challenge," in *Proc. Interspeech*, 2009, pp. 312–315.

[25] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," *Language Resources and Evaluation*, vol. 46, no. 4, pp. 523–541, 2012.

[26] P. M. Enderby, *Frenchay dysarthria assessment*. San Diego: College Hill Press, 1983.

[27] K. M. Yorkston and D. R. Beukelman, *Assessment of intelligibility of dysarthric speech*. Tigard, OR: C.C. Publications, 1981.

[28] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzjo, and H. Bunnell, "The Nemours database of dysarthric speech," in *Proc. International Conference on Spoken Language Processing (ICSLP)*, 1996, pp. 1962–1965.

[29] A. Wrench. (1999) The MOCHA-TIMIT articulatory database. [Online]. Available: http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html

[30] H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. Huang, K. Watkin, and S. Frame, "Dysarthric speech database for universal access research," in *Proc. Interspeech*, 2008, pp. 1741–1744.