AUTOMATIC AND PERCEPTUAL DISCRIMINATION BETWEEN DYSARTHRIA, APRAXIA OF SPEECH, AND NEUROTYPICAL SPEECH

Ina Kodrasi[*], *Michaela Pernon*[†,‡], *Marina Laganaro*[§], *Hervé Bourlard*[*]

[*]Speech and Audio Processing Group, Idiap Research Institute, Switzerland
[†]Department of Clinical Neurosciences, Geneva University Hospital, Switzerland
[‡]Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS-Université Sorbonne Nouvelle, France
[§]Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland

ABSTRACT

Automatic techniques in the context of motor speech disorders (MSDs) are typically two-class techniques aiming to discriminate between dysarthria and neurotypical speech or between dysarthria and apraxia of speech (AoS). Further, although such techniques are proposed to support the perceptual assessment of clinicians, the automatic and perceptual classification accuracy has never been compared. In this paper, we investigate a three-class automatic technique and a set of handcrafted features for the discrimination of dysarthria, AoS and neurotypical speech. Instead of following the commonly used One-versus-One or One-versus-Rest approaches for multi-class classification, a hierarchical approach is proposed. Further, a perceptual study is conducted where speech and language pathologists are asked to listen to recordings of dysarthria, AoS, and neurotypical speech and decide which class the recordings belong to. The proposed automatic technique is evaluated on the same recordings and the automatic and perceptual classification performance are compared. The presented results show that the hierarchical classification approach yields a higher classification accuracy than baseline One-versus-One and One-versus-Rest approaches. Further, the presented results show that the automatic approach yields a higher classification accuracy than the perceptual assessment of speech and language pathologists, demonstrating the potential advantages of integrating automatic tools in clinical practice.

Index Terms— dysarthria, apraxia of speech, support vector machine, hierarchical classification, perceptual classification

1. INTRODUCTION

Various conditions of brain damage may disrupt the speech production mechanism, resulting in motor speech disorders (MSDs) that encapsulate altered speech production in different dimensions. Two primary categories of MSDs are dysarthria and apraxia of speech (AoS). Although dysarthria and AoS arise from disruptions at different levels of the speech production mechanism, they manifest through overlapping clinical-perceptual characteristics such as articulation deficiencies, vowel distortions, reduced loudness variation, hypernasality, or syllabification [1, 2]. Diagnosing the presence of an MSD (i.e., discriminating between neurotypical and impaired speech) is crucial in clinical practice, since the presence of an MSD can be one of the earliest signs of several neurodegenerative disorders [3–5]. Further,

an accurate differential diagnosis of the MSD (i.e., discriminating between dysarthria and AoS) is also important, since it can provide clues about the underlying neuropathology [6, 7]. However, because of the difficulty of detecting clinical-perceptual characteristics by ear (particularly in the presence of mild impairments) and because the clinical-perceptual characteristics of dysarthria and AoS overlap, discriminating between dysarthria, AoS, and neurotypical speech is hard for non-experts and even expert inter-rater agreement can be low [8–10].

To complement the perceptual assessment of clinicians, automatic techniques based on pattern recognition models have been proposed. Typical automatic techniques operate on acoustic features which are handcrafted to reflect impaired speech dimensions. Many acoustic features have been successfully exploited to characterize impacted phonation and articulation, e.g., fundamental and formant frequencies, jitter, shimmer, Mel frequency cepstral coefficients, or temporal and spectral sparsity [11–17]. In an attempt to capture many impaired speech dimensions, also large-scale feature sets such as openSMILE have been used [18–20]. The extracted features are then used to train classifiers such as Support Vector Machines (SVMs) or Hidden Markov Models (HMMs).

The majority of state-of-the-art contributions deal with impaired speech arising due to dysarthrias or laryngeal disorders, with AoS being considered only in [17]. These contributions propose two-class techniques aiming to discriminate between dysarthria and neurotypical speech, laryngeal disorders and neurotypical speech, or dysarthria and AoS. Three-class techniques aiming to discriminate between dysarthria, AoS, and neurotypical speech have not been considered in the state-of-the-art literature. Multi-class techniques have seldom been proposed only in the context of laryngeal disorders [21–23]. In [21], three-class classification of nodular lesions, diffuse lesions, and neurotypical speech is achieved through multiple SVMs in One-versus-One (OvO) and One-versus-Rest (OvR) classification approaches. In [22], three-class classification of edema, nodules, and polyp is achieved through multiple SVMs in an OvO classification approach. In [23], five-class classification of laryngeal disorders is achieved through multiple HMMs in an OvR classification approach.

In this paper, we propose a three-class automatic technique for the discrimination of dysarthria, AoS, and neurotypical speech. Instead of following an OvO or OvR classification approach, we propose to follow a hierarchical classification approach with two SVMs [24, 25]. The first SVM discriminates between impaired and neurotypical speech whereas the second SVM discriminates between dysarthria and AoS. To characterize the different impaired speech dimensions, a 28–dimensional feature vector is constructed. Since the discrimina-

tive power of different features is expected to be different for different groups of speakers, two feature selection blocks are incorporated prior to the two SVMs.

To the best of our knowledge, although automatic techniques are proposed with the primary objective of complementing the perceptual assessment of clinicians, the automatic and perceptual classification accuracy have never been compared in the literature. In this paper, we also compare the classification accuracy of the proposed automatic technique to the classification accuracy achieved by speech and language pathologists (SLPs). A perceptual study is conducted where 20 SLPs are asked to listen to recordings of dysarthria, AoS, and neurotypical speech and decide which class the recordings belong to. The proposed automatic technique is evaluated on the same recordings and the automatic and perceptual classification performance are extensively compared.

The presented results on a French database of dysarthria, AoS, and neurotypical speech illustrate the advantages of the hierarchical classification approach in comparison to OvO and OvR approaches and to the perceptual assessment of SLPs.

2. AUTOMATIC CLASSIFICATION APPROACH

For the automatic classification of dysarthria, AoS, and neurotypical speech, we follow a hierarchical classification scheme with two classifiers as depicted in Fig. 1. The first classifier SVM_1 is trained to discriminate between neurotypical speakers and patients (dysarthria or AoS) whereas the second classifier SVM_2 is trained to discriminate between dysarthria and AoS. At test time, SVM_1 is first applied to decide whether the speaker is a neurotypical speaker or a patient. If the speaker is classified to be a patient, SVM_2 is applied to decide whether the patient suffers from dysarthria or AoS.

Depending on the available speech material, speakers under consideration, and the classification objective (i.e., classifying neurotypical speakers and patients or classifying dysarthria and AoS), the discriminative power of different acoustic features can be different. Hence, two feature selection blocks are incorporated. The first block selects a subset of features that are optimal for classifying neurotypical speakers and patients and the selected features are used to train SVM_1. The second block selects a subset of features that are optimal for classifying dysarthria and AoS and the selected features are used to train SVM_2.

In the following, additional details on the proposed approach are provided. Further, OvO and OvR classification approaches considered to be automatic baseline approaches are briefly described.

2.1. Acoustic features

The acoustic features proposed in this paper for discriminating between dysarthria, AoS, and neurotypical speech are motivated by the advantageous performance these features have shown in discriminating between dysarthria and neurotypical speech in [16] and in discriminating between dysarthria and AoS in [17]. In the following, a brief overview of these features is presented.[1]

Spectral sparsity. In [16], we have shown that spectral sparsity can successfully characterize imprecise articulation, abnormal pauses, and breathiness observed in dysarthria. Spectral sparsity describes the energy distribution of the speech spectral coefficients across time and is computed by i) transforming the signals to the short-time Fourier transform domain, ii) time-aligning all representations to a

---

[1]For additional details on the motivation behind these features and their computation, the interested reader is referred to [16, 17].
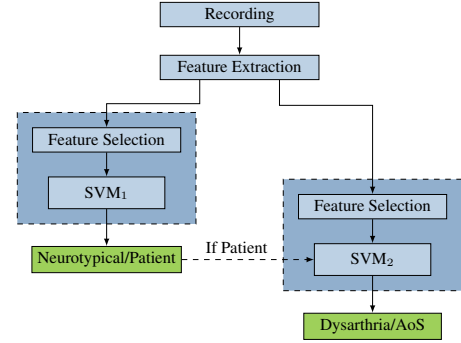
Fig. 1. Schematic representation of the proposed approach for the automatic discrimination of dysarthria, AoS, and neurotypical speech.

reference representation, and iii) computing the shape parameter of a Chi distribution best modeling the spectral magnitudes in each time frame [16]. To manage the (possibly) high dimensionality of such a feature vector when the number of time frames is large, in this paper we do not time-align representations. Instead, the shape parameter is computed for each time frame of the original representations and the used spectral sparsity feature vector $\mathbf{f}_1$ is a 4–dimensional vector constructed by taking the statistics (i.e., mean, standard deviation, kurtosis, and skewness) of the so-computed shape parameter across all time frames.

Formant frequencies and duration of continuously voiced regions. As in [17], vowel distortion and inappropriate vowel lengthening commonly observed in AoS are characterized by the 10–dimensional feature vector $\mathbf{f}_2$ constructed from the statistics (i.e., mean, standard deviation, kurtosis, and skewness) of the first and second formant frequencies across time and the statistics (i.e., mean and standard deviation) of the duration of continuously voiced regions.

Loudness peaks per second and long-term average speech spectrum. As in [17, 26, 27], abnormalities in loudness variation and hypernasality commonly observed in dysarthria are characterized by the 10–dimensional feature vector $\mathbf{f}_3$ constructed by computing the number of loudness peaks per second and the mean speech power across time in nine octave bands.

Temporal sparsity. In [15–17], we have proposed to use temporal sparsity to characterize syllabification. Temporal sparsity can be computed similarly as spectral sparsity, with the shape parameter modeling the speech spectral magnitudes in each frequency bin (rather than in each time frame). Hence, to characterize syllabification, we construct the 4–dimensional feature vector $\mathbf{f}_4$ by computing the statistics (i.e., mean, standard deviation, kurtosis, and skewness) of the shape parameter across all frequency bins.

Concatenating all previously described feature vectors into one vector, we obtain the 28–dimensional feature vector $\mathbf{f}$ extracted in the feature extraction block in Fig. 1, i.e.,

$$\mathbf{f} = [\mathbf{f}_1^T, \ \mathbf{f}_2^T, \ \mathbf{f}_3^T, \ \mathbf{f}_4^T]^T. \tag{1}$$

2.2. Feature selection

Although different feature selection approaches are applicable, statistical feature selection based on the ANOVA F-value is used in this paper [28]. Such an approach is independent of the used classifier and examines each feature individually. Features are grouped according to the class label in the training data, the mean value of each feature across the different groups is analyzed, and a (user-defined) number

of features showing the most significant differences between the different groups is selected. The first feature selection block in Fig. 1 selects optimal features for SVM_1 by analyzing whether there are significant differences in the mean values of the extracted features for neurotypical speakers and patients. The second feature selection block selects optimal features for SVM_2 by analyzing whether there are significant differences in the mean values of the extracted features for patients with dysarthria and AoS. The number of features selected by each block is determined based on the performance on the training set (cf. Section 3.3).

2.3. Baseline automatic classification approaches

As baseline classification approaches, we consider OvO and OvR classification. In the OvO approach, three SVMs are trained to discriminate between each class pair, i.e, neurotypical speech versus dysarthria, neurotypical speech versus AoS, and dysarthria versus AoS. At test time, the class which receives most votes is selected for the final prediction. In the OvR approach, three SVMs are trained to discriminate between each class and all other classes, i.e., neurotypical speech versus the rest, dysarthria versus the rest, and AoS versus the rest. At test time, the class predicted with the most confidence (i.e., largest distance from the separating hyperplane) is selected for the final prediction.

3. AUTOMATIC AND PERCEPTUAL CLASSIFICATION

In this section, the performance of the proposed automatic hierarchical approach is compared to using the baseline OvO and OvR approaches. To demonstrate the advantage of incorporating feature selection blocks, the performance when using all 28 acoustic features in $\mathbf{f}$ (i.e., without feature selection blocks) is also investigated. To demonstrate the advantage of using the proposed features $\mathbf{f}$, we also investigate the performance of all configurations (i.e., hierarchical with and without feature selection blocks, OvO, and OvR) using the openSMILE feature set from [29]. Finally, the automatic and perceptual classification accuracy are compared.

3.1. Database and preprocessing

We consider French recordings collected at Geneva University Hospitals and University of Geneva of 29 neurotypical speakers and 30 patients, with 20 patients diagnosed with dysarthria and the remaining 10 patients diagnosed with AoS. All patients with AoS have suffered a stroke, 14 of the patients with dysarthria suffer from Parkinson's disease, and the remaining 6 patients suffer from Amyotrophic Lateral Sclerosis. There are 19 female and 10 male speakers in the neurotypical group and 12 female (6 dysarthria, 6 AoS) and 18 male (14 dysarthria, 4 AoS) speakers in the patient group. The mean age of neurotypical speakers is 58.4 years old and the mean age of patients is 66.0 years old, with patients with dysarthria and AoS having a mean age of 72.8 and 52.5 years old, respectively.

The neurological diagnosis was established by neurologists, with the diagnosis of AoS based on the AoS rating scale [30]. The MSDs were assessed by an expert SLP using the perceptive score of BECD [31]. The BECD score reflects impairments in different dimensions such as voice quality, phonetic production, prosody, or intelligibility, and ranges from 0 (no impairment) to 20 (severe impairment). The mean BECD score of all patients is 7.0, where patients with dysarthria and AoS have a mean BECD score of 6.0 and 9.1, respectively.

For the results presented in the following, we consider recordings of two sentences at a sampling frequency of 44.1 kHz. To ensure that the phonetic content does not influence classification results, all speakers in the database utter the same sentences. After downsampling to 16 kHz and manually removing non-speech segments at the beginning and end of each sentence, the two sentences are concatenated and used to extract features (for automatic classification) or played back to judges (for perceptual classification). The mean length of the concatenated sentences for the neurotypical speakers, patients with dysarthria, and patients with AoS is 15.1 s, 16.4 s, and 33.8 s, respectively.

3.2. Performance evaluation

For automatic classification, the validation strategy is a stratified 5-fold cross-validation ensuring that each fold has a balanced number of neurotypical speakers and patients and preserving the imbalanced distribution of patients with dysarthria and AoS A single run of the 5-fold cross-validation procedure may result in a noisy estimate of the automatic model performance, with different data splits possibly yielding different results. For this reason, and in line with the number of perceptual evaluations available for each speaker (cf. Section 3.3), we repeat the 5-fold cross-validation procedure 10 times such that a different random split of the data is used each time. The performance is evaluated in terms of the mean and standard deviation of the balanced classification accuracy across all repetitions of the cross-validation procedure. To compute the balanced classification accuracy, the individual classification accuracy for each group of speakers is first computed, i.e.,

$$\text{Acc}_{\text{Group}} = \frac{\text{AP}_{\text{Group}}}{\text{T}_{\text{Group}}}, \qquad (2)$$

with Group $\in$ {Neurotypical, Dysarthria, AoS}, AP_{Group} denoting the number of accurately predicted speakers in the group, and T_{Group} denoting the total number of speakers in the group. The balanced classification accuracy is then defined as

$$\text{Acc}_{\text{Balanced}} = \frac{1}{3} \left(\text{Acc}_{\text{Neurotypical}} + \text{Acc}_{\text{Dysarthria}} + \text{Acc}_{\text{AoS}} \right). \qquad (3)$$

In addition to the individual and balanced classification accuracy, the classification accuracy for all patients $\text{Acc}_{\text{Patient}}$ is considered in Section 3.4. $\text{Acc}_{\text{Patient}}$ is defined as in (2), with $\text{AP}_{\text{Patient}}$ being the number of patients with dysarthria and AoS that are accurately predicted to be patients (independently of whether the exact label, i.e., dysarthria or AoS, is correct) and $\text{T}_{\text{Patient}} = \text{T}_{\text{Dysarthria}} + \text{T}_{\text{AoS}}$.

3.3. Automatic and perceptual classification settings

Automatic classification. For automatic classification, we use SVMs with a radial basis kernel function. To select the soft margin constant C and the kernel width γ for the SVMs, nested 5-fold cross-validation is performed on the training data in each fold, with $C \in \{10^{-2}, 10^4\}$ and $\gamma \in \{10^{-4}, 10^2\}$. To set the number of features n_f that the feature selection blocks should select, nested 5-fold cross-validation is performed on the training data in each fold, with $n_f \in \{5, 10, 15, 20\}$. The final hyper-parameters (i.e., C, γ, and n_f) used in each fold are selected as the ones resulting in the highest mean balanced accuracy on the training data.

Perceptual classification. For perceptual classification, 20 SLPs were recruited as judges. The judges were French native speakers and had on average 11 years of professional experience. The perceptual

Table 1. Mean and standard deviation of the balanced classification accuracy $\text{Acc}_{\text{Balanced}}$ [%] using several configurations: the proposed hierarchical approach with feature selection blocks, the hierarchical approach without feature selection blocks, and the baseline OvO and OvR approaches. The performance of all these configurations using the proposed handcrafted feature set **f** and the openSMILE feature set is presented.

Classification approach	Handcrafted **f**	openSMILE
Proposed (hierarchical with feature selection)	79.7 ± 4.0	74.4 ± 3.4
Hierarchical without feature selection	75.0 ± 3.4	58.8 ± 3.9
OvO	72.0 ± 3.8	57.3 ± 4.0
OvR	74.8 ± 2.4	60.8 ± 4.4

Table 2. Mean and standard deviation of the classification accuracy for different groups of speakers using automatic and perceptual classification. Automatic classification is done using the hierarchical approach proposed in Section 2. Perceptual classification is done by SLPs as described in Section 3.3.

Accuracy [%]	Automatic classification	Perceptual classification
$\text{Acc}_{\text{Balanced}}$	79.7 ± 4.0	68.9 ± 4.3
$\text{Acc}_{\text{Neurotypical}}$	82.1 ± 1.4	67.2 ± 9.2
$\text{Acc}_{\text{Patient}}$	91.7 ± 3.7	82.7 ± 7.0
$\text{Acc}_{\text{Dysarthria}}$	75.0 ± 7.7	64.5 ± 10.1
Acc_{AoS}	82.0 ± 7.5	75.0 ± 14.3

classification task was done following a similar methodology as for the automatic classification scheme in Fig. 1, i.e., judges listened to the available recordings and for each recording they were asked to decide: i) whether the recording belonged to a neurotypical speaker or a patient and ii) if the recording belonged to a patient, whether the patient suffered from dysarthria or AoS. To minimize the duration of the perceptual task for each judge, we split the available recordings into two groups, with one group containing the recordings of 15 neurotypical speakers, 10 patients with dysarthria, and 5 patients with AoS and the other group containing the remainder of the recordings (i.e., 14 neurotypical, 10 dysarthria, and 5 AoS). Consequently, 10 judges were asked to evaluate recordings belonging to one group and 10 judges were asked to evaluate recordings belonging to the other group. Hence, in line with the automatic classification results where each recording was evaluated by 10 different automatic models through repetitions of the cross-validation procedure, each recording was also perceptually evaluated by 10 different judges. The perceptual classification performance is then computed as described in Section 3.2, with the mean and standard deviation of the performance computed across judges.

3.4. Results

Table 1 presents the balanced classification accuracy obtained using all considered configurations with the handcrafted feature set **f** and the openSMILE feature set. It can be observed that the proposed classification approach using the handcrafted features **f** achieves the best performance, with the incorporation of feature selection blocks increasing the balanced classification accuracy from 75.0% to 79.7%. These results confirm that different subsets of features are optimal for different classifiers. The performance difference when incorporating feature selection blocks is even larger for the openSMILE feature set, since this is a high-dimensional feature vector (i.e., 6373) overfitting to the training data when feature selection is not used. Further, the presented results show the advantages of using the proposed hierarchical classification approach rather the OvO and OvR approaches.[2] In the following, the performance of the proposed hierarchical classification approach (using the handcrafted features **f** and feature selection blocks) is further analyzed and compared to the perceptual performance achieved by SLPs.

Table 2 presents the automatic and perceptual classification performance. It can be observed that automatic classification yields a higher performance than perceptual classification in terms of all considered accuracy measures. The difference in accuracy is particularly large for neurotypical speakers and patients with dysarthria. While the

classification accuracy for neurotypical speakers is 82.1% using automatic classification, the perceptual accuracy is only 67.2%. Further, while the classification accuracy for patients with dysarthria is 75.0% using automatic classification, the perceptual accuracy is only 64.5%. Since the impairment for patients with dysarthria can be milder than for patients with AoS (as shown by the lower mean BECD score presented in Section 3.1), judges often confuse neurotypical speakers for patients with dysarthria and conversely. Although the classification accuracy for neurotypical speakers is lower than for patients also for automatic classification (i.e., 82.1% versus 91.7%), confusion between neurotypical speakers and patients with dysarthria does not occur as often in the proposed automatic classification scheme.

In addition, the presented results show that the trend in performance for individual groups of speakers is similar for both automatic and perceptual classification approaches, i.e., $A_{\text{Patient}} > A_{\text{Neurotypical}}$ and $A_{\text{AoS}} > A_{\text{Dysarthria}}$. This similar trend is to be expected since the acoustic features used in the proposed automatic classification technique are motivated by the clinical-perceptual signs used to diagnose these MSDs and since the automatic and perceptual classification approaches follow a similar hierarchical methodology.

In summary, the presented results show that the proposed automatic classification scheme can be an advantageous tool to integrate in clinical practice. In addition, the presented results show that while automatic tools can achieve a high performance in discriminating patients from neurotypical speakers, the performance in discriminating subtypes of MSDs needs to be improved. Analyzing the generalisability of the presented results to other databases and analyzing the statistical significance of the performance differences between automatic and perceptual classification remain topics for future investigation.

4. CONCLUSION

In this paper, we have proposed a hierarchical three-class automatic technique operating on handcrafted acoustic features for the discrimination of dysarthria, AoS, and neurotypical speech. Two SVMs are used, with the first SVM discriminating between impaired and neurotypical speech and the second SVM discriminating between dysarthria and AoS. Since the discriminative power of different features is expected to be different for different groups of speakers, two feature selection blocks are incorporated prior to the two SVMs. The classification accuracy of this approach has been analyzed on a French database of dysarthria, AoS, and neurotypical speech. Additionally, a perceptual study has been conducted where SLPs are asked to discriminate between dysarthria, AoS, and neurotypical speech on the same database. The presented results have shown the advantages of the automatic classification technique, which yields a balanced classification accuracy of 79.7% in comparison to the balanced accuracy of 68.9% achieved in the perceptual assessment of SLPs.

---

[2]It should be noted that the proposed approach also outperforms using the OvO and OvR approaches with additional feature selection blocks. However, due to space constraints, these results are not presented here.

5. REFERENCES

[1] J. R. Duffy, *Motor speech disorders: substrates, differential diagnosis, and management*, Elsevier Mosby, Missouri, USA, 2003.

[2] W. Ziegler, I. Aichert, and A. Staiger, "Apraxia of speech: concepts and controversies," *Journal of Speech, Language, and Hearing Research*, vol. 55, no. 5, pp. 1485–1501, Oct. 2012.

[3] J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka, "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task," *Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2171–2181, Sept. 2013.

[4] P. Rong, Y. Yunusova, J. Wang, and J. Green, "Predicting early bulbar decline in Amyotrophic Lateral Sclerosis: A speech subsystem approach," *Behavioural Neurology*, vol. 2015, pp. 1–11, July 2015.

[5] J. M. Tracy, Y. Özkanca, D. C. Atkins, and R. Hosseini Ghomi, "Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease," *Journal of Biomedical Informatics*, vol. 104, April 2020.

[6] J. R. Duffy, *Parkinson's disease and movement disorders: diagnosis and treatment guidelines for the practicing physician*, chapter Motor speech disorders: Clues to neurologic diagnosis, pp. 35–53, Humana Press, New Jersey, USA, 2000.

[7] J. R. Duffy, R. K. Peach, and E. A. Strand, "Progressive Apraxia of Speech as a sign of motor neuron disease," *American Journal of Speech-Language Pathology*, vol. 16, no. 3, pp. 198–208, Aug. 2007.

[8] S. Fonville, H. B. van der Worp, P. Maat, M. Aldenhoven, A. Algra, and J. van Gijn, "Accuracy and inter-observer variation in the classification of dysarthria from speech recordings," *IEEE Transactions on Speech and Audio Processing*, vol. 255, no. 10, pp. 1545–1548, Oct. 2008.

[9] K. Bunton, R. Kent, J. R. Duffy, J. Rosenbek, and J. Kent, "Listener agreement for auditory-perceptual ratings of dysarthria," *Journal of Speech, Language, and Hearing Research*, vol. 50, no. 6, pp. 1481–1495, Jan. 2008.

[10] K. Haley, A. Jacks, J. Richardson, and J. Wambaugh, "Perceptually salient sound distortions and apraxia of speech: A performance continuum," *American Journal of Speech-Language Pathology*, vol. 26, no. 2S, pp. 631–640, June 2017.

[11] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, May 2012.

[12] J. R. Orozco-Arroyave, F. Hönig, J. Arias-Londoño, J. Bonilla, S. Skodda, J. Rusz, and E. Nöth, "Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinson's disease," in *Proc. Annual Conference of the International Speech Communication Association*, Dresden, Germany, Sept. 2015, pp. 95–99.

[13] D. Hemmerling, J. R. Orozco-Arroyave, A. Skalski, J. Gajda, and E. Nöth, "Automatic detection of Parkinson's disease based on modulated vowels," in *Proc. Annual Conference of the International Speech Communication Association*, San Francisco, USA, Sept. 2016, pp. 1190–1194.

[14] S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, "Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech," *Journal of Speech, Language, and Hearing Research*, vol. 53, no. 1, pp. 114–125, Feb. 2010.

[15] I. Kodrasi and H. Bourlard, "Super-Gaussianity of speech spectral coefficients as a potential biomarker for dysarthric speech detection," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Brighton, UK, May 2019, pp. 6400–6404.

[16] I. Kodrasi and H. Bourlard, "Spectro-temporal sparsity characterization for dysarthric speech detection," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 28, no. 1, pp. 1210–1222, Dec. 2020.

[17] I. Kodrasi, M. Pernon, M. Laganaro, and H. Bourlard, "Automatic discrimination of apraxia of speech and dysarthria using a minimalistic set of handcrafted features," in *Proc. Annual Conference of the International Speech Communication Association*, Shanghai, China, Oct. 2020.

[18] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in *Proc. ACM Multimedia*, Barcelona, Spain, Oct. 2018, pp. 835–838.

[19] T. Bocklet, S. Steidl, E. Nöth, and S. Skodda, "Automatic evaluation of Parkinson's speech - acoustic, prosodic and voice related cues," in *Proc. Annual Conference of the International Speech Communication Association*, Lyon, France, Sept. 2013, pp. 1149–1153.

[20] R. Norel, M. Pietrowicz, C. Agurto, S. Rishoni, and G. Cecchi, "Detection of Amyotrophic Lateral Sclerosis (ALS) via acoustic analysis," in *Proc. Annual Conference of the International Speech Communication Association*, Hyderabad, India, Sept. 2018, pp. 377–381.

[21] E. Vaiciukynas, A. Verikas, A. Gelzinis, M. Bacauskiene, and V. Uloza, "Exploring similarity-based classification of larynx disorders from human voice," *Speech Communication*, vol. 54, no. 5, pp. 601–610, June 2012.

[22] R. Behroozmand and F. Almasganj, "Comparison of Neural Networks and Support Vector Machines applied to optimized features extracted from patients' speech signal for classification of vocal fold inflammation," in *Proc. IEEE International Symposium on Signal Processing and Information Technology*, Athens, Greece, Jan. 2006, pp. 844–849.

[23] A. A. Dibazar, T. W. Berger, and S. S. Narayanan, "Pathological voice assessment," in *Proc. International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, USA, Sept. 2006, pp. 1669–1673.

[24] S. Kumar, J. Ghosh, and M. Crawford, "Hierarchical fusion of multiple classifiers for hyperspectral data analysis," *Pattern Analysis Applications*, vol. 5, pp. 210–220, June 2002.

[25] Y. Chen, M. M. Crawford, and J. Ghosh, "Integrating Support Vector Machines in a hierarchical output space decomposition framework," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, Anchorage, USA, Sept. 2004, pp. 949–952.

[26] R. Hummel, W.-Y. Chan, and T. Falk, "Spectral features for automatic blind intelligibility estimation of spastic dysarthric speech," in *Proc. Annual Conference of the International Speech Communication Association*, Florence, Italy, Aug. 2011, pp. 3017–3020.

[27] V. Berisha, R. Utianski, and J. Liss, "Towards a clinical tool for automatic intelligibility assessment," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 2013, pp. 2825–2828.

[28] M. Kuhn and K. Johnson, *Feature engineering and selection*, Taylor & Francis, Florida, USA, 2020.

[29] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism," in *Proc. Annual Conference of the International Speech Communication Association*, Lyon, France, Aug. 2013, pp. 148–152.

[30] E. A. Strand, J. R. Duffy, H. M. Clark, and K. Josephs, "The apraxia of speech rating scale : a tool for diagnosis and description of apraxia of speech," *Journal of Communication Disorders*, vol. 51, pp. 43–50, Sept. 2014.

[31] P. Auzou and V. Rolland-Monnoury, *Batterie d'évaluation clinique de la dysarthrie*, Ortho Édition, Isbergues, France, 2006.