PATHOLOGICAL SPEECH PROCESSING: STATE-OF-THE-ART, CURRENT CHALLENGES, AND FUTURE DIRECTIONS

Rahul Gupta, Theodora Chaspari, Jangwon Kim, Naveen Kumar, Daniel Bone, Shrikanth Narayanan

Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles, CA, USA

ABSTRACT

The study of speech pathology involves evaluation and treatment of speech production related disorders affecting phonation, fluency, intonation and aeromechanical components of respiration. Recently, speech pathology has garnered special interest amongst machine learning and signal processing (ML-SP) scientists. This growth in interest is led by advances in novel data collection technology, data science, speech processing and computational modeling. These in turn have enabled scientists in better understanding both the causes and effects of pathological speech conditions. In this paper, we review the application of machine learning and signal processing techniques to speech pathology and specifically focus on three different aspects. First, we list challenges such as controlling subjectivity in pathological speech assessments and patient variability in the application of ML-SP tools to the domain. Second, we discuss feature design methods and machine learning algorithms using a combination of domain knowledge and data driven methods. Finally, we present some case studies related to analysis of pathological speech and discuss their design.

***Index Terms*—** Pathological speech disorders, machine learning, signal processing

1. INTRODUCTION

The American Speech-Language Hearing Association (ASHA) [1] classifies pathological disorders into five categories, namely (i) speech disorders, (ii) language disorders, (iii) social communication disorders, (iv) cognitive communication disorders, and (v) swallowing disorders. With the evolution of scientific instruments and methodologies, pathologists have made large strides in the understanding of pathological disorders in all of these categories. Speech disorders are also of considerable interest to speech scientists, both for applying existing knowledge of human speech production and perception and for performing novel experiments. A specific cross-disciplinary approach that is increasingly popular utilizes machine learning and signal processing (ML-SP) tools to investigate patterns in pathological voices. In this work, we present a critical review of ML-SP applications in speech pathology, considering the methodological promises and pitfalls within this growing domain. We list recent advances made in speech pathology aided by ML-SP tools as well as problems that can be addressed in the near future. Specifically, we focus on three aspects: (i) challenges in the application of ML-SP tools within the domain; (ii) a review of proposed knowledge-based and data-driven approaches; and (iii) a few case studies of pathological speech processing. By addressing these topics, we aim to assess the current state of the art and inform future endeavors in the application of ML-SP tools to pathological speech.

1.1. Background

Amongst the pioneering studies in speech pathology, Van der Merwe [2] provides a sound theoretical foundation and emphasizes the need for developing a speech production framework for research and management of pathological disorders. To address this, she describes a four level framework characterizing pathological speech as a dysfunction at the levels of linguistic-symbolic planning and speech motor planning. Further in [2], Forrest et al. discuss the speech production mechanism in detail and Kent el al. describe assessment methods for motor speech disorders. Along similar lines of thought, the connection between speech production and pathological speech has also been discussed in [3, 4]. Particular emphasis has also been laid on specific speech and language disorders such as apraxia [5], dysarthria [6], and other voice disorders (e.g., hoarseness, spasmodic dysphonia) [7]. Apart from this, pathology researchers have also investigated speech disorders within specific population groups such as children with developmental disorders [8], people with schizophrenia [9] and Parkinson's disease [10].

Application of ML-SP techniques to the domain of speech pathology is not new. Some of the early works that used contemporary ML-SP tools towards the understanding of pathological disorders can be found in [11, 12]. Over the last decade, advances in the field of machine learning have led to several investigations laying special focus on detection of these speech disorders [13, 14]. Following this, various Interspeech challenges [15, 16] further attracted special interest in the application of ML-SP techniques to detection and analysis of pathological speech conditions. In this paper, we summarize some of these and various other inter-disciplinary approaches investigating pathological speech disorders using speech processing, acoustic signal processing and machine learning tools. We discuss the challenges, both knowledge and data driven approaches and case study designs. We draw specific examples in each of these topics and point out novel techniques as well as suggest future work. In the next section, we begin with stating the challenges in the application of ML-SP methods to study of pathological conditions. Sections 3 and 4 list a review of the state of the art methods and a few case studies in application of ML-SP tools to the study of pathological conditions. Finally, Section 5 presents our conclusions.

2. CHALLENGES AND OPPORTUNITIES

Advances in understanding the causes and characteristics of pathological conditions have allowed for a theoretically-grounded application of ML-SP techniques to the domain. However, the sensitive nature of medical research calls for thoroughly listing the objectives and limitations of the experiments being conducted. In this section, we discuss a few challenges that ML-SP researchers face in dealing with speech pathology data, and related application opportunities. We note that this list is by no means exhaustive, but certainly needs attention. For example, apart from the outlined challenges, ML-SP researchers also face questions regarding the choice of modeling techniques, variable recording conditions, and finding a balance between data-driven and knowledge-driven modeling (see Section 3).

2.1. Defining the scope of pathological speech study

Precise problem definition is essential in the formulation of ML-SP problems in order to achieve robustness and generalizability of the solutions. However, within the domain of speech pathology, the definition and the scope of the problems ML-SP researchers try to address are often inconsistent or too general. A part of this stems from the evolving nature of the pathological speech research. The Speech Pathology Association of Australia [17] cites several pathologists who describe the terminology in the field as being inconsistent, variable and inadequate. This poses a major challenge to ML-SP research as it could turn out to be inaccurate or irrelevant as the definitions change. Apart from this, ML-SP algorithms also need to be cautiously designed keeping in mind the spectrum of pathological speech conditions. For instance, just within aphasia, the severity could be categorized into anomic, Wernicke's, mixed non-fluent, Broca's, or global aphasia [18]. As there may or may not be a transfer of knowledge in understanding these conditions, the specificity and generality of symptoms being addressed should be laid out clearly for a larger impact and clearer understanding.

2.2. Subjective Impressions

Another challenge is due to the subjectivity of human perception of pathological speech. In the vast majority of cases, ML-SP algorithms are trained to model speech patterns using judgments from expert speech language pathologists. Although pathologists serve as the ground-truth, a single rater will be affected by their own personal experiences and training, as well as secondary features like their mood and attention. The issue of intra-rater variability is exacerbated by the previous argument relating to the evolving nature of definitions of pathological speech, wherein variability is increased due to the revised definitions. Therefore, an effective system will have to understand intra-rater variability. For example, recent research has sought to find the most reliable regions for individual raters [19]. This challenge also presents an opportunity. Ideally, the automatic system may predict a collective judgment made by many experts, which in effect could augment the perceptions of a single clinician.

2.3. Patient Variability

Another factor impacting the quality of ML-SP algorithms is the variability among patients within a population suffering from a specific pathological speech condition. As the goal of these algorithms often is to capture speech/vocal patterns for each pathological condition, patient specific variability serves as a source of noise. Moreover, this variability can occur both across speakers and within speakers. In terms of experimental setup, a common approach for dealing with inter-speaker variability is to control the speech content, particularly through reading tasks–the major drawback being that reading tasks lack spontaneous speech planning which may be relevant for certain disorders. There are also computational methods to discount the speaker-specific traits (e.g., speaker normalization or speaker-independent evaluation), but disassociating speaker-specific traits from the characteristics of a pathological condition remains challenging [20]. Intra-speaker variability, which can be due to performance fatigue or other passing factors, should also be modeled. In fact, finding a speaker's true baseline speaking attributes remains one of the most challenging paralinguistic tasks; yet, if that baseline can be established, one can see the great potential of speaker-specific models that can be used for tracking intervention outcome.

3. DOMAIN-KNOWLEDGE AND DATA-DRIVEN ANALYSIS OF PATHOLOGICAL CONDITIONS

Although accounting for all the discussed factors using ML-SP techniques can be fairly complex, researchers have made strides in mining intricate patterns from pathological speech. Given the complex nature of pathological conditions, researchers have explored various ML-SP modeling approaches. We discuss two aspects of previous research in the field: (i) feature design for pathological speech signals, and (ii) machine learning algorithms for capturing various feature patterns in pathological speech. These methods offer a combination of domain knowledge and data driven techniques and have shown promise in analysis of pathological speech.

3.1. Feature Design

Previous studies have attempted to capture the wide variety of pathological traits through various acoustic and phonological features, as well as non-verbal discourse markers. The link found in many psycholinguistic studies between abnormal prosody and various pathological cases has motivated the use of voice quality and prosodic features because of their high interpretability and computational efficiency [21, 22]. In a more computationally intense framework, multi-scale spectro-temporal modulation indices attempt to represent the irregular spectral perturbations and timing variations of pathological speech [23]. Motivated by irregularities in the motor function caused by vocal disorders [24, 25], vocal source excitation and articulatory features have been proposed in order to capture the malfunctioning of various parts of the speech production system. Other efforts have focused on developing distance measures between healthy and pathological speech [26]. These frame-level features can be incorporated into long-term measures through phone or utterance level functionals [27], contour parameterization [28], and other non-linear transformations [27, 29].

ASR can yield confidence indices of normal speech through lattice posteriors and recognition accuracy metrics [27, 30]. ASR output is further able to provide durational features at the syllable and word level that can be indicative of atypicality [29], such as stuttering or dysarthria. Despite the knowledge-driven nature of this approach, challenges of using ASR metrics include the potentially limited vocabulary size, the existence of sparse multilingual data, and the need for speaker-dependent acoustic models.

Non-verbal vocalizations are an essential part of spoken communication for regulating and coordinating discourse. Their atypical occurrence and expression has been related to various neurological and mental disorders [31]. Previous studies have examined the role of fillers, pauses, and laughters in pathological speech and have discussed how the absence or irregular occurrence of these non-verbal vocalizations can indicate pathological symptoms [32].

The inherently diverse information present in the speech signal, such as speaker traits, gender and age effects, environmental conditions, etc., makes it hard to disentangle actual pathology-dependent conditions from other factors. Although previous studies have indicated strong correlates of many of the aforementioned features to pathological constructs, careful methodological and experimental planning has to be conducted in order to make sure that the segmentation of the acoustic features space is performed in terms of the relevant pathological effects [33]. Towards this direction, ecological data capture procedures, reduced-size interpretable features, appropriate statistical analysis, and legitimate experimental validation are encouraged.

3.2. Machine Learning

From the point of view of machine learning methods in pathological speech, a major challenge is posed by the subjectivity of expert annotated labels. Several researchers have proposed novel methods to address this problem. For instance, Berisha et. al. [34] proposed a feature selection method using similarity labels with the annotators being asked to rate the similarity between utterances instead of individual intelligibility ratings. Wallen et al. [35] developed a screening test for speech pathology assessment by developing objective quality measures instead of using subjective judgments from humans. Saenz-Lechon et. al [36] further discuss issues in development of automatic systems for detection of pathological voice and focus on issues such as speaker variability, subjectivity, data sparsity. More recently, the authors in [37] proposed a reliability-aware intelligibility classification model that takes into account the subjective nature of annotations. Each annotation is decomposed into two components - a data dependent component representing the objective nature of the annotation and a data-independent component that models the annotator's own subjective bias. Another challenge in speech pathology results from the fact that depending on the source of pathological disorder the symptoms for reduced intelligibility might considerably differ. A mixture-of-experts approach was proposed in [38] to deal with this problem, by training multiple experts for these different conditions in a data-driven fashion.

Apart from these directed approaches towards addressing specific challenges in modeling pathological speech, several researchers have also resorted to transfer of existing methods to the domain of pathological speech processing. A few examples include using a combination of standard speech features and dimensionality reduction techniques in detecting pathological speech [39, 40]. Similarly, Chen et al. [41] and Oue et al. [42] used support vector machines and deep belief networks for identification of pathological voices and disfluency detection in dysarthric speech, respectively.

Generally for several speech pathology tasks (a few of which we review in section 4), annotation is an expensive process and the amount of data available is often not sufficient to learn reliable models given the large variability in the patterns of interest. This problem is further compounded when the system additionally lacks knowledge-driven reliable features for the task. Since several works consistently show that a few application specific hand-designed features outperform a set of generic audio features [16], we suggest a combination of directed feature design with a machine learning algorithm capable of handling annotator subjectivity, data sparsity and patient variability in the context of low-resource tasks in speech pathology.

4. CASE STUDIES OF COMPUTATIONAL PATHOLOGICAL SPEECH ANALYSIS

In this section, we focus on case studies of pathological speech analysis, listing a few studies which have earned considerable interest in the field. We analyze the characteristics of these case studies, discussing methodologies, limitations, and suggestions regarding design of future studies. In particular, we focus on (i) pathological speech sub-challenge, Interspeech 2012, (ii) Parkinson's condition sub-challenge, Interspeech 2015, and (iii) a study of pathological speech in developmental disorders. These case studies address various components of understanding pathological speech, such as assessing intelligibility of speech and analyzing speech quality, and offer piecemeal solutions towards the larger problem of complete quantitative understating of pathological speech conditions.

4.1. Pathological Speech Sub-Challenge, Interspeech 2012

There has been considerable interest to develop an automatic assessment system for speech intelligibility and quality that may offer more accurate, objective, and scalable engineering solutions in order to assist speech therapies in clinical practice. The Pathological Speech Sub-Challenge in Interspeech 2012 [15] was designed to draw more attention from speech signal processing and machine learning communities into this domain and promote technical advancement in this area. In particular, this challenge called for developing an automatic system to analyze and judge the speech intelligibility of patients suffering from pathological speech due to neck and head tumors. The challenge dataset contained speech audio of the patients recorded before and after a chemo-radiation treatment, in order to monitor the progress of their speech intelligibility during the period of the treatment and speech therapy. The speech audio consisted of Dutch read sentences; some of the speakers were second-language learners of Dutch. The intelligibility of individual speech utterances was judged on a scale of 1-7 by 13 expert raters.

This data collection design has several interesting implications. First of all, the longitudinal nature of the speech data allows for the evaluation of the progress of speech therapy toward improving speech intelligibility. This can be crucial in determining the course of treatment for a patient. Secondly, an attempt is made to reduce the subjectivity and variability of individual expert judgments by using many raters. For the purpose of the challenge, the final intelligibility score for each utterance was determined as the weighted sum of individuals' ratings. A combination of expert judgments has been shown to reduce annotator noise [43] and we encourage the application of recently developed multiple annotator schemes [19] used for determining the final intelligibility label. Another interesting aspect of the data is that the data contain speech audio of both native and non-native speakers. Although it fits a realistic scenario in clinical practice, this presents ML-SP methods the challenge of accounting for patient variability, as we discussed in section 2.3. The nativeness of a speaker affects the intelligibility of the speech sound as well [44], hence it is important to segregate the impact of the disease and the treatment from the nativeness of the speaker.

Within the scope of this challenge, one of the prominent machine learning approaches was the use of a large feature set coupled with dimensionality reduction techniques. For instance, Kim et al. [16] developed multiple expert subsystems on feature subsets, which were finally fused using Bayesian fusion models (Naive Bayes or Noise-Majority systems). Lu et al. [45] used sparse Gaussian processes and Huang et al. [46] used asymmetrical sparse partial least squares regression. On the other hand, some contributions designed acoustic features inspired by neurophysiology, biophysics, and psychoacoustics for intelligibility assessment [47]. Finally, Stark et al. [48] investigated the impact of speaker-trait and sentence-type on intelligibility assessments, probing questions related to generality versus specificity.

Overall, this challenge presented some unique data characteristics, and the researchers presented novel methods answering certain questions related to the application of ML-SP tools. In the next section, we describe another Interspeech challenge for automatic rating of Parkinson's disease severity from speech.

4.2. Parkinson's Condition Sub-Challenge, Interspeech 2015

Parkinson's disease is a prevalent neurological disorder that has a strong effect on speech, as noted in several works [10, 49]. The Parkinson's condition sub-challenge at Interspeech 2015 focused on the development of an automated rating system for Parkinson's severity using speech cues. This challenge focused on the cause of pathological speech, an aspect different from the previously discussed challenge, which instead focused on the symptom of pathological speech (intelligibility). Using the speech signal for monitoring the progression of Parkinson's disease is an attractive, useful approach for speech therapy, because it is a non-invasive, fast, easy-to-obtain, and cost-efficient.

The dataset of this challenge has several unique characteristics. First, training and test set partitions were recorded in dissimilar environmental noise conditions. This scenario is very realistic and calls for development of noise robust systems for a broader use. Second, each patient produced 42 speech utterances of five speaking tasks:

(i) reading single words, (ii) rapidly repeating syllables, (iii) reading sentences, (iv) reading a text, and (v) speaking spontaneously. This setup allowed the analysis of the impact of utterance type on different kinds of acoustic atypicality. Finally, the evaluation of Parkinson's severity was made using the standard Unified Parkinson's Disease Rating Scale (UPDRS) [50]. Availability of such carefully designed standards is crucial in training machine learning systems as they not only provide a ground truth for evaluation but also reduce ambiguity in problem definition while applying ML-SP techniques.

In this challenge, a lot of researchers focused on designing features to capture the spectral and prosodic characteristics of speech, while reducing the impact of differences in recording conditions. Foote el al. [51] and Jensen el al. [52] developed rhythmic features called beatspectrum and spectral irregularity, inspired from music information retrieval. Williamson et al. [53] focused on identifying Parkinson's severity based on channel-delay correlation and covariance matrices for the speech waveform, delta-MFCCs and formants, and the articulatory feature streams predicted using the Directions into Velocities of Articulators (DIVA) model. Hahm et al. [25] used an assembly of both acoustic and articulatory features to predict the UPDRS score. Lastly, Kim et al. [27] proposed an interesting approach to predict the UPDRS scores based on the five utterance types and performing a fusion to compute the final severity score for the speaker.

In summary, this challenge, along with the 2012 Interspeech Pathology Sub-Challenge, provided opportunities to investigate the cause and effect of pathological speech through objective signal processing. Also, the Challenges presented various tasks for ML-SP researchers, from accounting for speaker specificity to creating systems that are robust across diverse recoding conditions. In the next section, we present a summary of research on other mental health disorders that cause pathological speech.

4.3. Pathological Speech in Developmental Disorders

Certain types of pathological speech can indirectly result from another disorder. For instance, autism spectrum disorder (ASD) is a prevalent neuro-cognitive developmental disorder (1 in 68 in the United States) that affects social-communicative aspects of speech and language throughout the lifetime [54]. Speech pathologists play a critical role in both the assessment and treatment of ASD. In terms of acoustical analysis, speech scientists are focusing on speech prosody–the rhythm, stress, and intonation of speech–as a prime means for translational impact in assessment and intervention. This is because, although speech prosody is commonly referred to as *atypical* in autism, it is not currently utilized in the "gold-standard" diagnostic instruments. Interestingly, speech scientists may be able to develop computational models of *typical* prosody in order to infer *atypical* prosody, particularly those that are based on existing psychological and linguistic knowledge; this effort is in contrast to much of the present computational methods for pathological speech, which attempt to predict a gold-standard ground-truth.

For instance, Bone et al. [55] investigated prosodic cues of children with autism (and the interacting psychologist), finding interpretable acoustic measures that corroborate previous qualitative perceptions based on correlational analysis and linear-regression prediction of ASD-severity. Specifically, effects were found in the child's turn-end pitch slope that may relate to perceptions of *monotonous* voice, as well as effects in jitter and harmonics-to-noise ratio that could be perceived as *atypical*, *hoarse*, or *harsh* voice quality. These speech features were able to significantly predict the child's severity of ASD-related symptoms. Additionally, noting that not all individuals with autism have prosodic deficits, Bone et al [56] also incorporated global human judgments of prosodic "awkwardness" into their analysis of acoustic-prosodic measures. Findings suggested that the speech of individuals with ASD was perceived as more *awkward* and less *expressive*, and that the awkwardness could be quantified for this reading task through speech-rate and pausing features (both static and dynamic). Thus, we see in this case example that providing a computational measure which does not currently have an established ground truth could be of high value to the target domain, but also can be quite challenging.

5. CONCLUSIONS

In this paper, we present a glimpse of ongoing work in the application of machine leaning and signal processing tools to the domain of speech pathology. We specifically focus on three aspects of this cross-disciplinary study and point out the challenges in application of ML-SP to the domain; review a few knowledge-based and data-driven approaches and; summarize a few case studies directed towards pathological speech analysis. We present several examples addressing interesting aspects of pathological speech while also making contribution to the fields of machine learning and signal processing. We are also of the opinion that just within the purview of ML-SP, more steps have to be taken towards a more mature understanding of speech pathology. For instance, despite the quantity and quality of current work, they are limited in generalization due to various sources of variability. The need for careful design of studies conducted in speech pathology often leads to data sparsity, and a right balance has to be achieved between data driven and domain knowledge based methods during machine learning applications. We particularly encourage the development of domain knowledge inspired models, both for the sake of interpretability and parallel evolution with the domain.

It is encouraging to see the vast amount of attention that pathological speech processing has received from ML-SP researchers and the application has revealed many noteworthy conclusions. Our assessment is that the application of ML-SP to speech pathology is in an early stage and will only increase with time. As all of speech pathology, signal processing and machine learning are evolving domains, we anticipate that a continued collaboration between researchers in these fields will be mutually beneficial.

6. REFERENCES

[1] "Council for clinical certification in audiology and speech-language pathology," *American Speech-Language-Hearing Association and others, Retrieved September*, vol. 15, 2015.

[2] M. R. McNeil, *Clinical management of sensorimotor speech disorders*. Thieme, 2009.

[3] R. Netsell, *A neurobiologic view of speech production and the dysarthrias*. Singular, 1991.

[4] R. D. Kent, "Research on speech motor control and its disorders: A review and prospective," *Journal of Communication Disorders*, vol. 33, no. 5, 2000.

[5] J. L. Wambaugh, "A summary of treatments for apraxia of speech and review of replicated approaches." in *Seminars in Speech and Language*, vol. 23, no. 4, 2002.

[6] K. M. Yorkston, M. Hakel, D. R. Beukelman, and S. Fager, "Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic review," *Journal of Medical Speech-Language Pathology*, vol. 15, no. 2, 2007.

[7] A. E. Aronson and D. Bless, *Clinical voice disorders*. Thieme, 2011.

[8] D. C. Millar, J. C. Light, and R. W. Schlosser, "The impact of augmentative and alternative communication intervention on the speech production of individuals with developmental disabilities: A research review," *Journal of Speech, Language, and Hearing Research*, vol. 49, no. 2, 2006.

[9] L. E. DeLisi, "Speech disorder in schizophrenia: Review of the literature and exploration of its relation to the uniquely human capacity for language." *Schizophrenia Bulletin*, vol. 27, 2001.

[10] P. Lieberman, E. Kako, J. Friedman, G. Tajchman, L. S. Feldman, and E. B. Jiminez, "Speech production, syntax comprehension, and cognitive deficits in parkinson's disease," *Brain and Language*, vol. 43, 1992.

[11] N. S. Rees, "Auditory processing factors in language disorders: The view from procrustes' bed," *Journal of Speech and Hearing Disorders*, vol. 38, no. 3, 1973.

[12] S. B. Davis, "Acoustic characteristics of normal and pathological voices," *Speech and language: advances in basic research and practice*, vol. 1, 1979.

[13] E. S. Fonseca, R. C. Guido, P. R. Scalassara, C. D. Maciel, and J. C. Pereira, "Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders," *Computers in Biology and Medicine*, vol. 37, 2007.

[14] L. S. Chee, O. C. Ai, M. Hariharan, and S. Yaacob, "Mfcc based recognition of repetitions and prolongations in stuttered speech using k-nn and lda," in *Research and Development (SCOReD), 2009 IEEE Student Conference on*. IEEE, 2009.

[15] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. Van Son, F. Weninger, F. Eyben, T. Bocklet *et al.*, "The interspeech 2012 speaker trait challenge." in *INTERSPEECH*, 2012.

[16] J. Kim, N. Kumar, A. Tsiartas, M. Li, and S. S. Narayanan, "Automatic intelligibility classification of sentence-level pathological speech," in *Computer Speech and Language : Special Issue Next Generation Computational Paralinguistics*, 2014.

[17] "Criteria for the analysis of speech pathology terms: challenges and a methodology," *Melbourne: Speech Pathology Australia.*, vol. 16, 2009.

[18] M.-M. Mesulam and S. Weintraub, "Primary progressive aphasia: sharpening the focus on a clinical syndrome," in *Heterogeneity of Alzheimer's disease*. Springer, 1992.

[19] K. Audhkhasi and S. Narayanan, "A globally-variant locally-constant model for fusion of labels from multiple diverse experts without using reference labels," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 4, 2013.

[20] V. Berisha, S. Sandoval, R. Utianski, J. Liss, and A. Spanias, "Selecting disorder-specific features for speech pathology fingerprinting," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*. IEEE, 2013.

[21] J. Van Santen, E. Prud'hommeaux, L. Black, and M. Mitchell, "Computational prosodic markers for autism," *Autism*, 2010.

[22] A. Tsanas, M. Little, P. McSharry, J. Spielman, and L. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, 2012.

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

[24] T. H. Falk, W. 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, no. 5, 2012.

[25] S. Hahm and J. Wang, "Parkinson's condition estimation using speech acoustic and inversely mapped articulatory data," in *Proceedings of Interspeech*. Dresden, Germany: ISCA, September 2015.

[26] L. Gu, J. G. Harris, R. Shrivastav, and C. Sapienza, "Disordered speech assessment using automatic methods based on quantitative measures," *EURASIP Journal on Applied Signal Processing*, vol. 2005, 2005.

[27] J. Kim, M. Nasir, R. Gupta, M. V. Segbroeck, D. Bone, M. Black, Z. I. Skordilis, Z. Yang, P. Georgiou, and S. Narayanan, "Automatic estimation of parkinson's disease severity from diverse speech tasks," in *Proceedings of Interspeech*, Dresden, Germany, September 2015.

[28] J. Kim, N. Kumar, A. Tsiartas, M. Li, and S. Narayanan, "Automatic intelligibility classification of sentence-level pathological speech," *Computer speech & language*, vol. 29, no. 1, 2015.

[29] G. An, D. Brizan, M. Ma, M. Morales, A. Syed, and A. Rosenberg, "Automatic recognition of unified Parkinson's disease rating from speech with acoustic, i-vector and phonotactic features," in *Proc. Interspeech*, 2015.

[30] A. Zlotnik, J. M. Montero, R. San-Segundo, and A. Gallardo-Antolín, "Random forest-based prediction of Parkinson's disease progression using acoustic, ASR and intelligibility features," in *Proceedings of Interspeech*. Dresden, Germany: ISCA, September 2015.

[31] J. K. Lake, K. R. Humphreys, and S. Cardy, "Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders," *Psychonomic bulletin & review*, vol. 18, no. 1, 2011.

[32] P. Feyereisen and J. d. Lannoy, "Brain pathology, lateralization, and nonverbal behavior," *Fundamentals of nonverbal behavior*, 1991.

[33] D. Bone, T. Chaspari, K. Audhkhasi, J. Gibson, A. Tsiartas, M. Van Segbroeck, M. Li, S. Lee, and S. Narayanan, "Classifying language-related developmental disorders from speech cues: the promise and the potential confounds." in *Proc. Interspeech*, 2013.

[34] V. Berisha, J. Liss, S. Sandoval, R. Utianski, and A. Spanias, "Modeling pathological speech perception from data with similarity labels," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*. IEEE, 2014.

[35] E. J. Wallen and J. H. Hansen, "A screening test for speech pathology assessment using objective quality measures," in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, vol. 2. IEEE, 1996.

[36] N. Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiz, and P. Gomez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, no. 2, 2006.

[37] N. Kumar and S. Narayanan, "A discriminative reliability-aware classification model with applications to intelligibility classification in pathological speech," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[38] R. Gupta, K. Audhkhasi, and S. Narayanan, "A mixture of experts approach towards intelligibility classification of pathological speech," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*. IEEE, 2015.

[39] J. I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 10, 2006.

[40] K. Umapathy and S. Krishnan, "Feature analysis of pathological speech signals using local discriminant bases technique," *Medical and Biological Engineering and Computing*, vol. 43, no. 4, 2005.

[41] W. Chen, C. Peng, X. Zhu, B. Wan, and D. Wei, "Svm-based identification of pathological voices," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, 2007.

[42] S. Oue, R. Marxer, and F. Rudzicz, "Automatic dysfluency detection in dysarthric speech using deep belief networks."

[43] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.

[44] S. J. Van Wijngaarden, "Intelligibility of native and non-native dutch speech," *Speech communication*, vol. 35, no. 1, 2001.

[45] L. D. and F. Sha, "Predicting likability of speakers with gaussian processes," in *Proceedings of Interspeech*. Portland, OR: ISCA, September 2012.

[46] D.-Y. Huang, Y. Zhu, D. Wu, and R. Yu, "Detecting intelligibility by linear dimensionality reduction and normalized voice quality hierarchical features," in *Proceedings of Interspeech*. Portland, OR: ISCA, September 2012.

[47] X. Zhou, D. Garcia-Romero, N. Mesgarani, M. Stone, C. Espy-Wilson, and S. Shamma, "Automatic intelligibility assessment of pathologic speech in head and neck cancer based on auditory-inspired spectro-temporal modulations," in *Proceedings of Interspeech*. Portland, OR: ISCA, September 2012.

[48] A. P. Stark, A. Bayestehtashk, M. Asgari, and I. Shafran, "Interspeech pathology challenge: Investigations into speaker and sentence specific effects." in *INTERSPEECH*, 2012.

[49] D. Kempler and D. Van Lancker, "Effect of speech task on intelligibility in dysarthria: a case study of parkinson's disease," *Brain and language*, vol. 80, no. 3, 2002.

[50] G. T. Stebbins and C. G. Goetz, "Factor structure of the Unified Parkinson's Disease Rating Scale: motor examination section," *Movement Disorders*, vol. 13, no. 4, 1998.

[51] J. Foote, M. L. Cooper, and U. Nam, "Audio retrieval by rhythmic similarity," in *3rd International Conference on Music Information Retrieval*, 2002.

[52] K. Jensen, "Timbre models of musical sounds," Ph.D. dissertation, Department of Computer Science, University of Copenhagen, 1999.

[53] J. R. Williamson, T. F. Quatieri, B. S. Helfer, J. Perricone, S. S. Ghosh, G. Ciccarelli, and D. D. Mehta, "Segment-dependent dynamics in predicting Parkinson's disease," in *Proceedings of Interspeech*. Dresden, Germany: ISCA, September 2015.

[54] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders:: DSM-5*. American Psychiatric Association, 2013.

[55] D. Bone, C. Lee, M. Black, M. Williams, S. Lee, P. Levitt, and S. Narayanan, "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody," *Journal of Speech, Language, and Hearing Research*, vol. 57, 2014.

[56] D. Bone, M. P. Black, A. Ramakrishna, R. Grossman, and S. S. Narayanan, "Acoustic-prosodic correlates of 'awkward' prosody in story retellings from adolescents with autism," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.