PAPER FOR PATHOLOGY REVIEW

SAILORS, Shrikanth Narayanan

Signal Analysis and Interpretation Lab (SAIL),
University of Southern California, Los Angeles, California, USA

ABSTRACT

Index Terms— Pathological speech, intelligibility

1. INTRODUCTION

American Speech-Language Hearing Association (ASHA) [1] categorizes pathological disorders into five categories, namely, (i) speech disorders (ii) language disorders (iii) social communication disorders, (iv) cognitive communication disorders and, (v) swallowing disorders.

2. PATHOLOGICAL SPEECH DISORDERS

Within the characterization of speech pathological disorders, Van der Merwe [2, 3] provides a sound theoretical foundation. Citing several speech language pathologists [4, 5, 6], she emphasizes on the need for 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 [3], Kent, Ballard et al. and Forrest et al. describe assessment/examination methods for motor speech disorders and speech production mechanism. Later, Kent [7] reviewed research on speech motor control and its disorders. Particular emphasis has also been laid on specific speech language and disorders such as apraxia [8], dysarthria [9], sluttering [10], and voice disorders (e.g., hoarseness, spasmodic dysphonia) [11]. Apart from this, research has also investigated speech disorders in specific population groups such as children with developmental disorders [12, 13], people with schizophrenia [14] and Parkinson's desease [15].

In this work, we specifically address the challenges in the interdisciplinary approach of investigating pathological speech disorders using speech processing, acoustic signal processing and machine learning tools. Although, a considerable amount of research has presented novel Signal Processing (SP)/Machine Learning (ML) schemes with relation to pathological speech, we investigate three factors: (i) challenges in application of (ML-SP) to the domain of pathological speech understanding (ii) domain knowledge driven analysis (ex: feature design, machine learning algorithms) of pathological speech characteristics and, (iii) design of specific case studies to investigate pathological speech. Over the course of this paper, we draw specific examples in each of these topics and point out novel techniques as well as suggest future work. We discuss each of these topics in detail below.

3. CHALLENGES IN THE DOMAIN OF UNDERSTANDING PATHOLOGICAL CONDITIONS

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 a thoroughly listing of the objectives and limitations for the experiments being conducted. In this section, we list out a few challenges that ML-SP researchers face in dealing with pathology data. We would like to point out that this list by no means is exhaustive but certainly needs attention.

3.1. Defining pathological speech

Several ML-SP algorithms require a crisp definition of what is being modeled/investigated. However due to the evolving nature of the study of pathological speech, definitions are being formulated or revised. The Speech Pathology Association of Australia [**?**] cites several pathologists who describe the terminology in the field as being inconsistent, variable and inadequate. This poses a major challenge in terms of 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 []. As there may/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 of pathological conditions.

3.2. Subjective Impressions

This challenge in training ML-SP algorithms follows from the previous argument relating to the evolving nature of definitions of pathological speech. Often, training ML-SP algorithms use judgements from trained pathologists [] to model speech patterns. Although pathologists can provide the best assessment of pathological conditions, there could be variations within their evaluations. This variability stems from changing definitions as well as subjectivity involved in making the decisions. We would also like to point out that standards for clinical competence in speech-language pathology [] as laid by ASHA are also scrutinized and revised periodically adding another source of variability to the professional judgements. Whereas speech pathologists are the most reliable source of diagnosis used by ML-SP algorithms, one should be careful about the subjectivity which can be reduced by using assesments from multiple pathologists using joint annotator modeling schemes [].

3.3. Patient variability

Another factor impacting the quality of ML-SP algorithms is the patient specific variability in modeling population suffering from a specific pathological condition. As often the goal of these algorithms is to capture speech/vocal patterns for each pathological condition, patient specific variability serves as a source of noise. Although there

are methods to discount the speaker specific traits (e.g., speaker normalization, speaker independent evaluation), disassociating speaker specific traits from the characteristics of a pathological condition is challenging both in terms of modeling and analysis. On the other hand, one can also argue in the favor of patient specific models opening up the questions of model specificy vs generality.

Apart from these challenges, ML-SP research also faces the questions regarding the choice of modeling techniques, recording conditionsand finding the right balance between data-driven and knowledge-driven modeling. Although accounting for all these factors in understanding pathological speech using ML-SP techniques could be fairly complex, researchers have made succesful strides in mining intricate patterns in the domain of pathological speech. In the next section, we list a few approaches using combinations of domain knowledge as well as the data driven learning in modeling patterns in pathological speech.

4. DOMAIN KNOWLDEGE + DATA DRIVEN ANALYSIS OF PATHOLOGICAL CONDITIONS

Given the intricate nature of pathological conditions, several researchers have explored different ML-SP modeling aspects. In this section, we focus on two approaches addressing: (i) feature design from pathology speech signals and, (ii) machine learning algorithms capturing various feature patterns in pathological speech.

4.1. Feature design

Previous studies have attempted to capture the wide variability of pathological speech through various acoustic and phonological features, as well as non-verbal discourse markers.

Voice quality and prosodic features have been extensively used because of their high interpretability and computational efficiency [16, 17, 18], while multi-scale spectro-temporal modulation indices attempt to represent the irregular spectral perturbations of pathological speech [19, 20, 21]. Vocal source excitation and articulatory features have been proposed in order to capture the malfunctioning of various parts of the speech production system caused by vocal disorders [20, 22]. Other efforts have focused on developing distance measures between healthy and pathological speech [23]. These frame-level features can be incorporated into long-term measures through phone or utterance level functionals [24], contour parameterization [25], and other non-linear transformations [24, 26, 27].

ASR can yield confidence indices of normal speech through lattice posteriors and recognition accuracy metrics [24, 28, 29, 30, 31]. ASR output is further able to provide durational features at the syllable and word level that can be indicative of atypicality [26, 32]. 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 [33]. Previous studies have examined the role of fillers, pauses, and laughter in pathological speech [34, 33, 35].

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 [36]. Towards this direction, ecological data capture procedures, reduced-size interpretable features, appropriate statistical analysis, and legitimate experimental validation are encouraged.

4.2. Machine learning

Naveen

5. DESIGN OF CASE STUDIES FOR PATHOLOGICAL SPEECH ANALYSIS

5.1. Pathological speech sub-challenge

Recently, an automatic assessment system for speech intelligiblity and quality has been obtaining lots of attention for assisting speech therapy. Since manual evaluation by human experts is costly, time-consuming, and subjective, there has been research effort to develop an automatic system to analyze and judge the intelligibility and quality of patients' speech. The pathology challenge in Interspeech 2012 is the special session where various acoustic features and algorithms were proposed.

The winner of the challege [37] developed multiple expert subsystems which were fused for the final label by using Bayesian fusion models (Naive Bayes or Noise-Majority system). Individual subsystem focuses on particular aspects of speech, e.g., acoustic similarity to normal speech, prosody, intonation, voice quality and pronunciation quality. Acoustic features include pitch stylization parameters (quadratic polynomials) for pitch trajectory, speaking rate, harmony-noise ratio, jitter, shimmer, Mel-Frequency Cepstral Coefficients (MFCCs), formants, and phoneme probability feature driven from Automatic Speech Recognition (ASR) lattice. They employed joint classification scheme: an ad-hoc way of utilizing the high similarity of intelligibility score for the speech audio closely located in the acoustic space.

Ways to handle high feature dimensionality were examined in this challenge. In order to capture a variety of atypicality in pathological speech, a large number of acoustic features is initially extracted. Hence, some feature reduction techniques were applied for achieving high accuracy on the test set. Sparse Gaussian process was introduced in [38], where the original features were transformed to a lower dimensional feature space using kernel PCA. Asymmetric sparse partial least squares regression were also tested in [39]. Modified LDA was tested to further reduce the feature dimension from initially reduced dimension by PCA in [40].

5.2. Parkinson's condition sub-challenge

Parkinson's disease is one of the most common neurological disorders. In clinical practice, it is important to track the severity of its symptom. Using speech signal for monitoring the pregression of Parkinson's disease is an attractive approach, because it is non-invasive, fast, easy-to-obtain and cost-efficient as well as useful for speech treatment. The task of the Parkinson's condition challenge in Interspeech 2015 was to develop an automatic system to predict the severity of Parkinson disease using a set of speech signals of the patients.

Various features on top of the standard baseline features extracted using openSMILE [41] were tested for automatic severity rating. Rhythmic features adpoted from music information retrieval

include beatspectrum [42] and spectral irregularity [43]. The structure of correlations among frame-level speech features were also captured using channel-delay correlation and covariance matrices on the speech waveform, delta-MFCCs and formants, and the articulatory feature streams predicted using the Directions into Velocities of Articulators (DIVA) model [44, 45]. Using both acoustic and (predicted) articulatory information improved prediction accuracy [22].

Since the goal was to achieve the best Spearman correlation between predicted and true labels, that is the Unified Parkinson's Disease Rating Scale (UPDRS) [46], participatns tested various regressors, e.g. support vector regressor, random forest, deep neural networks and Gaussian staircase regression model [47]. The winning system [48] in this challenge used a deep neural network regressor on the average of multiple predictors' scores - the predictors were diversified with different hyperparameter values.

In the challenge dataset, the UPDRS score was assigned to each speaker, not each utterance. Hence, judging the final severity score of individual utterances based on all utterances of each speaker (or each UPDRS-score cluster) improved prediction accuracy significantly. Specifically, incorporating feature selection [48] or the information of (predicted) prompt type [24] into clustering process improved severity rating accuracy as well as clustering accuracy significantly.

5.3. Depression work

3. Depression work

4. Other work in ASD, addiction etc. Danny, Jimmy, Bo

6. REFERENCES

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

[2] Anita Van der Merwe, "Characterization of pathological speech," *Clinical management of sensorimotor speech disorders*, p. 1, 1997.

[3] Malcolm R McNeil, *Clinical management of sensorimotor speech disorders*, Thieme, 2009.

[4] Malcolm R McNeil and Raymond D Kent, "Motoric characteristics of adult aphasic and apraxic speakers," *Advances in Psychology*, vol. 70, pp. 349–386, 1990.

[5] RD Kent and MR McNeil, "Relative timing of sentence repetition in apraxia of speech and conduction aphasia," *Phonetic approaches to speech production in aphasia and related disorders*, pp. 181–220, 1987.

[6] Thomas P Marquardt and Harvey Sussman, "The elusive lesion-apraxia of speech link in broca's aphasia," *Apraxia of speech: Physiology, acoustics, linguistics, management*, pp. 91–112, 1984.

[7] Ray D Kent, "Research on speech motor control and its disorders: A review and prospective," *Journal of Communication disorders*, vol. 33, no. 5, pp. 391–428, 2000.

[8] Julie L Wambaugh, "A summary of treatments for apraxia of speech and review of replicated approaches.," in *Seminars in Speech and Language*, 2002, vol. 23, pp. 293–308.

[9] Kathryn M Yorkston, Mark Hakel, David R Beukelman, and Susan 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, pp. XI–XXXVI, 2007.

[10] Anne K Bothe, Jason H Davidow, Robin E Bramlett, and Roger J Ingham, "Stuttering treatment research 1970–2005: I. systematic review incorporating trial quality assessment of behavioral, cognitive, and related approaches," *American Journal of Speech-Language Pathology*, vol. 15, no. 4, pp. 321–341, 2006.

[11] Arnold E Aronson and Diane Bless, *Clinical voice disorders*, Thieme, 2011.

[12] Diane C Millar, Janice C Light, and Ralf 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, pp. 248–264, 2006.

[13] Ralf W Schlosser and Oliver Wendt, "Effects of augmentative and alternative communication intervention on speech production in children with autism: A systematic review," *American Journal of Speech-Language Pathology*, vol. 17, no. 3, pp. 212–230, 2008.

[14] Lynn 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, no. 3, pp. 481, 2001.

[15] EM Critchley, "Speech disorders of parkinsonism: a review.," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 44, no. 9, pp. 751–758, 1981.

[16] J.P.H. Van Santen, E.T. Prud'hommeaux, L.M. Black, and M. Mitchell, "Computational prosodic markers for autism," *Autism*, 2010.

[17] A. Tsanas, M. 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, 2012.

[18] D. Bone, C.C. Lee, M.P. Black, M.E. 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, no. 4, pp. 1162–1177, 2014.

[19] 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.

[20] 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, no. 5, pp. 622–631, 2012.

[21] J.R. Williamson, T.F. Quatieri, B.S. Helfer, J. Perricone, S.S. Ghosh, G. Ciccarelli, and D.D. Mehta, "Automatic recognition of unified Parkinson's disease rating from speech with acoustic, i-vector and phonotactic features," in *Proc. Interspeech*, 2015.

[22] Seongjun Hahm and Jun Wang, "Parkinson's condition estimation using speech acoustic and inversely mapped articulatory data," in *Proceedings of Interspeech*, Dresden, Germany, September 2015, pp. 513 – 517, ISCA.

[23] Lingyun Gu, John G Harris, Rahul Shrivastav, and Christine Sapienza, "Disordered speech assessment using automatic methods based on quantitative measures," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 1400–1409, 2005.

[24] Automatic estimation of Parkinson's disease severity from diverse speech tasks, "Jangwon kim and md nasir and rahul gupta and maarten van segbroeck and daniel bone and matthew black and zisis iason skordilis and zhaojun yang and panayiotis georgiou and shrikanth narayanan," in *Proceedings of Interspeech*, Dresden, Germany, September 2015, pp. 914 – 918, ISCA.

[25] . Kim, N. Kumar, A. Tsiartas, M. Li, and S.S. Narayanan, "Automatic intelligibility classification of sentence-level pathological speech," *Computer speech & language*, vol. 29, no. 1, pp. 132–144, 2015.

[26] G. An, D.G. Brizan, M. Ma, M. Morales, A.R. 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.

[27] C. Middag, T. Bocklet, J.P. Martens, and E. Nöth, "Combining phonological and acoustic asr-free features for pathological speech intelligibility assessment.," in *Proc. Interspeech*, 2011, pp. 3005–3008.

[28] Alexander Zlotnik, Juan M. Montero, Rubén San-Segundo, and Ascensión Gallardo-Antolín, "Random forest-based prediction of Parkinson's disease progression using acoustic, ASR and intelligibility features," in *Proceedings of Interspeech*, Dresden, Germany, September 2015, pp. 503 – 507, ISCA.

[29] A. Maier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, and E. Nöth, "PEAKS - A system for the automatic evaluation of voice and speech disorders," *Speech Communication*, vol. 51, no. 5, pp. 425–437, 2009.

[30] H.V. Sharma, M. Hasegawa-Johnson, J. Gunderson, and A. Perlman, "Universal access: Preliminary experiments in dysarthric speech recognition," in *Proc. Interspeech*, 2009, p. 4.

[31] C. Middag, J.P. Martens, G. Van Nuffelen, and M. De Bodt, "Automated intelligibility assessment of pathological speech using phonological features," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 3, 2009.

[32] D. Duez, "Consonant and vowel duration in parkinsonian french speech," in *Proceedings of Speech Prosody*, 2006, pp. 101–105.

[33] Johanna K Lake, Karin R Humphreys, and Shannon Cardy, "Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders," *Psychonomic bulletin & review*, vol. 18, no. 1, pp. 135–140, 2011.

[34] P.A. Heeman, R. Lunsford, E. Selfridge, L. Black, and J. Van Santen, "Autism and interactional aspects of dialogue," in *Proc. Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2010, pp. 249–252.

[35] R. Gupta, P.G. Georgiou, D. Atkins, and S.S. Narayanan, "Predicting client?s inclination towards target behavior change in motivational interviewing and investigating the role of laughter," in *Proc. InterSpeech*, 2014.

[36] D. Bone, T. Chaspari, K. Audhkhasi, J. Gibson, A. Tsiartas, M. Van Segbroeck, M. Li, S. Lee, and ShS.rikanth Narayanan, "Classifying language-related developmental disorders from speech cues: the promise and the potential confounds.," in *Proc. Interspeech*, 2013, pp. 182–186.

[37] Jangwon Kim, Naveen Kumar, Andreas Tsiartas, Ming Li, and S. Shrikanth Narayanan, "Intelligibility classification of pathological speech using fusion of multiple subsystems," in *Proceedings of Interspeech*, Portland, OR, September 2012, ISCA.

[38] Lu D. and F. Sha, "Predicting likability of speakers with gaussian processes," in *Proceedings of Interspeech*, Portland, OR, September 2012, pp. 286 – 289, ISCA.

[39] Dong-Yan Huang, Yongwei Zhu, Dajun Wu, and Rongshan Yu, "Detecting intelligibility by linear dimensionality reduction and normalized voice quality hierarchical features," in *Proceedings of Interspeech*, Portland, OR, September 2012, pp. 546 – 549, ISCA.

[40] 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, September 2012, pp. 542 – 545, ISCA.

[41] Florian Eyben, Martin Wöllmer, and Björn Schuller, "OpenSMILE: The Munich versatile and fast open-source audio feature extractor," in *Proceedings of the international conference on Multimedia*, Firenze, Italy, October 2010, pp. 1459 – 1462, ACM.

[42] Jonathan Foote, Matthew L Cooper, and Unjung Nam, "Audio retrieval by rhythmic similarity," in *3rd International Conference on Music Information Retrieval*, 2002.

[43] Kristoffer Jensen, *Timbre models of musical sounds*, Ph.D. thesis, Department of Computer Science, University of Copenhagen, 1999.

[44] Frank H Guenther, Satrajit S Ghosh, and Jason A Tourville, "Neural modeling and imaging of the cortical interactions underlying syllable production," *Brain and language*, vol. 96, no. 3, pp. 280–301, 2006.

[45] James R. Williamson, Thomas F. Quatieri, Brian S. Helfer, Joseph Perricone, Satrajit S. Ghosh, Gregory Ciccarelli, and Daryush D. Mehta, "Segment-dependent dynamics in predicting Parkinson's disease," in *Proceedings of Interspeech*, Dresden, Germany, September 2015, pp. 518 – 522, ISCA.

[46] Glenn T Stebbins and Christopher G Goetz, "Factor structure of the Unified Parkinson's Disease Rating Scale: motor examination section," *Movement Disorders*, vol. 13, no. 4, pp. 633–636, 1998.

[47] James R. Williamson, Thomas F. Quatieri, Brian S. Helfer, Rachelle Horwitz, Bea Yu, and Daryush D. Mehta, "Vocal biomarkers of depression based on motor incoordination," in *Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, New York, NY, USA, 2013, AVEC, pp. 41–48, ACM.

[48] Tamás Gróz, Róbert Busa-Fekete, Gábor Gosztolya, and László Tóth, "Assessing the degree of nativeness and Parkinson's condition using Gaussian processes and deep rectifier neural networks," in *Proceedings of Interspeech*, Dresden, Germany, September 2015, pp. 919 – 923, ISCA.