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

2. RELATION TO PREVIOUS WORK

3. CHALLENGES IN THE DOMAIN OF UNDERSTANDING PATHOLOGICAL CONDITIONS

Theodora and Rahul

3.1. Defining pathological speech

3.2. Subjective Impressions

3.3. Speaker variability

3.4. Approaches by speech scientists

3.4.1. Obtaining data

Annotations, data demographics, data collection environment

3.4.2. Features

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 [1, 2, 3], while multi-scale spectro-temporal modulation indices attempt to represent the irregular spectral perturbations of pathological speech [4, 5, 6]. 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 [5, 7]. Other efforts have focused on developing distance measures between healthy and pathological speech [8]. These frame-level features can be incorporated into long-term measures through phone or utterance level functionals [9], contour parameterization [10], and other non-linear transformations [9, 11, 12].

ASR can yield confidence indices of normal speech through lattice posteriors and recognition accuracy metrics [9, 13, 14, 15, 16]. ASR output is further able to provide durational features at the syllable and word level that can be indicative of atypicality [11, 17]. 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 [18]. Previous studies have examined the role of fillers, pauses, and laughter in pathological speech [19, 18, 20].

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

3.4.3. Machine learning

4. CASE STUDIES

Jangwon, Naveen, Danny, Jimmy, Bo ....

4.1. 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 [22] 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 [23], 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 [24]. Modified LDA was tested to further reduce the feature dimension from initially reduced dimension by PCA in [25].

4.2. 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 [26] were tested for automatic severity rating. Rhythmic features adpoted from music information retrieval include beatspectrum [27] and spectral irregularity [28]. 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 [29, 30]. Using both acoustic and (predicted) articulatory information improved prediction accuracy [7].

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) [31], participatns tested various regressors, e.g. support vector regressor, random forest, deep neural networks and Gaussian staircase regression model [32]. The winning system [33] 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 [33] or the information of (predicted) prompt type [9] into clustering process improved severity rating accuracy as well as clustering accuracy significantly.

4.3. Depression work

3. Depression work

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

5. REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] . 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

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

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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.