A Review of Disorder Voice Processing Toward to Applications

Wenhua Xu¹, Runxuan Chen² and Xiaobo Chen³*
¹Sch. of Med. Inf. Eng., Guangdong Pharm. Univ., GuangZhou 510006, Chin
²Dept. of LNGS, Sch. Of LAM, FASS, Syd. Univ, Sydney 2006, Australia
³Dept. of Rehabil. The First Affiliated Hospital of Gannan Med. Univ., Ganzhou 341000, China

*Corresponding author email: gnmucxb@163.com

Abstract. Disorder voice processing has great potential to provide convenient, efficient and lowcost applications in medical diagnosis and treatment. For the purpose of a systematic summary of the research progress, this paper introduces it in three terms of research objects, acoustic parameters and features selection, and acoustic model and classification algorithms. It concludes that the appropriate feature selection, feature hybrid and feature offset, integrate with the deep learning frames are the future directions in disorder voice processing.

1. Introduction
The generation of speech and various acoustic events involve many organs of human body, including lung, trachea (vocal cord), mouth and nose. Meanwhile, the generation of speech is also related to human nervous system, mental state and personal characteristics[1-4]. Therefore, speech and acoustic events contain lots of biological, pathological or emotional information. When an organ has pathological lesions, it always produces pathological acoustic information. The studies on these pathological voices have potential benefit for depth analysis, disease screening or improvement of the quality of life(QoS) of patients[5, 6].

Compared with the existing clinical examination methods, such as CT scanning of the lung, MRI scanning of the brain, laryngoscopy and so on, pathological voice processing has unique advantages: non-invasive, simple data collection, low cost, no special requirements for the detection objects, and even can complete data collection under the condition that the patients can not cooperate[7]. Therefore, with the improvement of speech signal processing research, a number of researchers have carried out extensive and exploratory researches on pathological voice.

Based on these great benefits and unique advantages, this paper presents a bird-eye view of the disorder voice processing. The main contribution of this paper is the summary of the main features and algorithms that used at present and to clarify the direction of the processing in future. In the following part, the research object, acoustic parameters and features selection, and acoustic model and classification algorithms will be described.

2. Research Objects
In general, the research objects are disorder voices. But the mechanism of the disorder is very diversified, because there are many pathological factors leading to voice disorder, such as the excessive professional voice use (e.g., singers, lectures), lung cancer, Parkinson's, the autistic, the
depression, the hearing impairment, the pilots in weightless condition, the Alzheimer's syndrome, the aphasia and Amyotrophic Lateral Sclerosis (ALS), etc. For a special problem or clinic application, the research objects will be given, this means that the researchers will use different corpus according to the given purpose.

For the expression of disorder voice, the impairments of vocal organs and nervous system make symptoms very complex: articulatory disorders, laborious speech, paraphasia, word finding problem, verb stereotype, echolalia, perseveration, repetition, fluency, writing problem. All these symptoms are voice disorder, which is similar to typical voice events such as laugh, applause, whisper, but the processing of these disorder voice is more difficult due to the diversity and complex. Like the problems of word repeat, revise, and hesitation in the speech recognition in reality environments, such as multi-person meeting or phone talks, it is more difficult to deal with these events compared with the recognition of the isolated words or reading speeches. How to find the typical or special symptoms of a give disease is usually the key step to process the voices, this is because the symptoms always have special features and acoustic parameters. A wide range of parameters or features were employed to analyze, classify and recognize these symptoms and deal with given tasks.

To best of our knowledge, there is no system with the adaptive and general ability to deal with multiple symptoms, due to the complexity. The realistic approaches to processing these symptoms are to deal with it step by step, just like we are solving the speech recognition problem for non-specific large vocabulary continuous speech recognition in reality. However, all these symptoms are necessary to study, and at the same time, satisfy the clinic applications.

3. Acoustic Parameters and Features Selection

For most researches of the speech processing, there always is a hypothesis that the acoustic organs are healthy, but for disorder voices, it is necessary to consider the organ impairments (e.g., polyp, hare lip). Because the acoustic parameters traditionally calculated based on NORMAL organs, for example, The formant frequency calculation is based on the length of sound track, yet the impairment of the voice organs may affect the results. An earlier evidence of the effectiveness to the acoustic features due to the voice organs impairments was illuminated in [8]. It approximated the larynx disorders Mel-frequency cepstral coefficients (MFCCs) by sampling, averaging, and clustering by Gaussian mixture model, and then classified the samples to normal voice, nodular, and diffuse vocal fold lesion voice by using similarity-based classification of kernelized principal angle (KPA), support vector machines (SVMs) and Gaussian Mixture Models (GMM) model. The results showed that the MFCCs of nodular and lesion can be discriminated, which means that the MFCCs formula can’t be used directly to extract the MFCCs for disorder voice in practice. To reduce the effectiveness on the disorder voice features before further classification or recognition, one strategy is to offset the features [9, 10], the another one is to modify the speech models that simulated the pronunciation process. Based on the opinion of the nonlinearity of speech process, the irregular vibratory dynamics of vocal folds with a unilateral vocal polyp was discussed in paper [3, 11], they modeled the speech process by using a Volterra–Wiener–Korenberg (VWK) model and a biomechanical model of vocal cord with polyps and paralysis to assess the nonlinearity of a bio-mechanical voice production. Paper [5] introduced the common Pediatric dysphonia in children with vocal fold movement impairment, and suggested a cooperation of a pediatric otolaryngologist and a speech language pathologist with experience in the assessment and treatment of voice problems. The model difference of the speech not only were proved in disorder voice, but in gender. To investigate the gender difference of voice simulation, the relationship between biomechanical inputs (lung pressure and muscle activation levels) and acoustic outputs (fundamental frequency, F0, and sound pressure level, SPL) was modeled by using a global, Monte Carlo-based method [2]. Not only the organ impairments, but the mental/nervous factors affect the voice features, such as the famous Lombard effect. An earlier research [12] clarified that the differential TE0-Pitch feature changed in the condition of high-stress, such as aero-flight under G-force. In [13], the reflection of anxiety disorder in voices was investigated through acoustic parameters, it indicated that the subglottic pressure increases and the vocalization of the vowels decreases, and the MFCCs changes
also in the anxious state. Paper [14] compared the vocal differences between depressed and healthy people under various situations with irrelevant variables being regarded as covariate, the authors found that three acoustic features (loudness, MFCC5 and MFCC7) were consistently different between people with and without depression with large effect magnitude. It is difficult for a person with an articulation disorder resulting from athetoid cerebral palsy, because that the movements of such speakers are limited by their athetoid symptoms, and their consonants are often unstable or unclear. A phoneme categorized subdictionary and a dictionary selection method using Nonnegative Matrix Factorization (NMF) was applied to a voice from a speaker with an articulation disorder in [15]. The analysis of features that effected by mental or nervous extracted from the disorder voice is remain a challenging task, and highly related to the emotion analysis based on speech, which is a main direction in speech processing. Because the mental/nervous factors can be regarded as a kind of emotion, the emotion analysis methods reasonable can be use for reference in studies that will focus on these problems (such as the hysteric, the monotonous intonation of depression and AD).

With the development of the studies on effectiveness of the organ impairments and mental factors, more and more features were investigated for special diseases or tasks. An earlier paper employed the first rahmonic peak (R1) in the cepstrum to assess hoarse voice quality[16]. To characterize the mental disorders and personality traits, paper [4] used the fundamental frequency (F0), frame-to-frame Jitter Factor (LPJit), and the Glottal Flow Spectral Slope (Slope). In paper [6], the MFCCs, speech segmentation method voice activity detection (VAD) and linear thresholding were used to categorize the differences between inhale and exhale for the purpose of monitoring the sleep-breath movement of patients with breathing disorder. In paper[17], by averaged the features of Local Shimmer, Harmonic-to-noise ratio (HNR) and the Recurrence Period Density Entropy (RPDE), the university lectures’ voice risk was classified to Normal, medium and high risk by a Bayesian decision analysis approach. Paper [18] compared the existed popular speech features used in PD studies, Jitter, shimmer, F0, harmonicity parameters, recurrence period density entropy (RPDE), detrended fluctuation analysis (DFA) and pitch period entropy (PPE), meanwhile, mean and standard deviation of the original 13 MFCCs plus log-energy of the signal and their first–second derivatives were employed as features, and then extracted feature via a tunable Q-factor wavelet transform (TQWT) with high frequency resolution. The estimation of the F0 was computed and then used to estimate the dysphonia symptoms[19]. Paper [20]used a nonlinear energy difference ratio (NEDR) of voice to characterize voice disorders. In the study, an iteration method was used to calculate the energy distribution of a voice signal, the iteration was stopped when the deviation of the two nonlinear coefficients of energy is less then a given value. the paper discussed that the normal and abnormal voice have different NEDR. Paper [21] investigated the speech impairments of dysphonia in people with Parkinson’s disease (PD), dysphonia diagnosed in patients with different laryngeal pathologies (LP), and hypernasality in children with cleft lip and palate (CLP) were evaluated. Noise content measures, spectral-cepstral modeling, nonlinear features, and measurements were fed into SVM to make a decision of the speech. Different layer features were extracted in Paper [22], it investigated the central tendency and dispersion metrics of multiple types of the sustained vowels, words, and sentences for people with Parkinson’s disease.

In short, many features were used to analyze the voice disorder, but the following four features are typically used in medical examination of the voice lesions: harmonic noise ratio(HNR), frequency jitter, amplitude perturbation (shimmer), normalized noise energy (NNE). HNR is the ratio of harmonic component to noise component in speech. It is an objective index to detect abnormal voice and evaluate voice quality, which can effectively reflect the closure of glottis. It should be noted that the noise here is not the ambient noise, but the glottal noise caused by the incomplete closure of the glottis. Frequency perturbation is a physical quantity describing the basic frequency change of sound waves between adjacent periods. It mainly reflects the degree of rough sound, followed by the degree of hoarseness. The frequency perturbation in speech signal is consistent with the functional state of glottal region. During the normal voice cycle, there are many people with the same frequency and few people with different frequencies, so the frequency perturbation value is very small. When the vocal cord lesion occurs, the perturbation value
increases, making the sound rough. Amplitude perturbation describes the changes of acoustic amplitude between adjacent periods, mainly reflecting the degree of hoarseness. Jitter and shimmer together reflect the stability of vocal cord vibration. The more the value, the less the tiny change of acoustic signal in the process of sound production. It mainly calculate the energy of the glottal noise caused by the incomplete closure of the glottis when making a sound, and mainly reflects the degree of breath sound, followed by the degree of hoarseness. To some extent, it reflects the degree of glottis closing. It is very valuable for the analysis of pathological voice caused by organic or functional lesions of vocal cords.

Although the acoustic parameters and features used in the above researches achieved acceptable performance, all these parameters and features were used independently and without any selection or modification. Some novel methods to deal with the features were reported in recent years. Papers of [23, 24] proposed two similar composed and hybrid approaches to extract the temporal-frequency based (hybrid) features of acoustic signals, and achieved higher accuracy and low computational resource. Paper[9], similar to paper [25]examined the disorder voice impact of standard speech coders used in communication systems, to resolve the problem of disorder voice in speaker recognition, according to the distance between the disorder speech feature of MFCCs and the statistical normal MFCCs, the MFCCs of disorder speech were revised by using the distance as weighting coefficients. To deal with the problem of the degeneracy of the detection accuracy of multiple types of samples, paper [26]proposed a method ranked features by using chi-square statistical model, searched optimal subset of the ranked features and iteratively selected samples. Moreover, it built a smart healthcare framework using edge computing as an application framework. A Multiple Feature Evaluation Approach (MFEA) was proposed in [27] to select the features from disorder voice, and used to classify the health and PD patient by several classifiers. To recognize the PD patient and health people, paper [28] used the L1-norm SVM to select a new subset of features from the PD dataset based on a feature.

In applications, several prototype systems in multimodality were developed. The disorder voice group clinic applications of speech technology (CAST), has been focused on the clinic applications of speech technology and developed several systems named voice input voice output communication aid (VIVOCA) and speech recognition of people with severe dysarthria (STARDUST)[29]. By combine real cough audio data acquired from smartphone, the local Hu moments as a robust feature set with a variety of sounds including noise from both indoor and outdoor environments and noncough events (eg., sneeze, laugh, speech.) was analyzed [30]. By analyzed 14 questions form, a questionnaire data and twenty-three variables computed by the commercial “Dr.Speech” software from a digital voice recording of a sustained phonation of the vowel sound /a/ constitute a voice data vector, three classes of categorization, one is healthy and two disorders named diffuse and nodular, was classified in [31]. By an ambulatory voice monitor-surface electromyographic (sEMG) device, the vocal gestures signal from the human neck was collected and analyzed[32]. Wired accelerometer mounted on a silicone pad affixed to the anterior neck surface, the disorder voice of phonotraumatic vocal hyperfunction (PVH) was evaluated by two features: the first is non–zero-lag autocorrelation peak amplitude (normalized by the zero-lag amplitude) and the second is F0 in semitones (with reference to each subject’s week-long F0 mode) [33]. Another electroencephalogram (EEG) signals were analyzed to assess the voice disorder to voice pathology assessment [34]. Combined the hardware of ECG, thermistor, chest belt, accelerometer, contact, and audio microphones, the cough event was detected, and then the features were fed into the ANN[35]. By integrating head tracking, speech recognition, and tongue motion control, the researchers attempted to help disabilities communication with computer[36]. Collected data from mild traumatic brain injuries (mTBI) using mobile devices, the various of temporal and frequency metrics was analyzed statistically[37].

4. Acoustic Model and Classification Algorithm

In terms of acoustic model and classification algorithm, many machine learning algorithms were applied to the classification of these pathological and disorder voice. SVM, Decision Tree(DT), Bayesian Classification (BC), Logistic Model Tree (LMT) were employed to identify the voice
disorder by using the features of F0, Jitter, Shimmer, HNR and MFCCs in a hardware system [38]. The ordinal classification and Gaussian regression were employed in a system called Automatic Voice Condition Analysis (AVCA) [39], that aims at analyzing, classifying and quantifying the degree to which a patient is affected by a voice disorder. The Support Vector Machine, Random Forest, K-Nearest Neighbor, and Gradient Boosting were used to classify the disorder voice to four types (Normal, Neoplas, Phonotrauma and Vocal palsy), based on a database named Far Eastern Memorial Hospital (FEMH) [40]. A slow independent component analysis (ICA) algorithm was proposed to tackle the non-Gaussianity and disorder time series by concurrently considering the high-order statistic and slowness in paper [41]. In [42], the orthogonal wavelets method was used to sort the disorder voice to six various types: paralysis, nodules, polyps, edema, spasmodic dysphonia and keratosis were sorted. A five-band wavelet system and SVM were employed for feature extraction in Normal and pathological cases in [43]. The SVM was used for automatic pathology detection with continuous speech [44]. The discriminative paracognitive machine (DPM) investigated the speech pathologies by using the signal energy (SE), zero-crossing rates (ZCRs) and signal entropy (SH), from the sounds of sustained vowels, to provide a joint time-frequency-information map, and classify the voice signals [45]. Based on Pearson’s and Kendall’s Correlation Coefficient, Principal Component Analysis (PCA), Self-Organizing Map (SOM), and Artificial Neural Networks (ANNs), the acoustic features of Parkinson’s were extracted, and then input to classifiers of SVM, k-NN and ANN [46]. Four machine learning methods were used to classify data obtained from sustained phonation and speech to evaluate the use of eighteen feature extraction tasks [47].

As a state-of-art technology, the deep learning technology is being used more and more in these years. Based on the popular deep learning toolkit Kaldi, a type of features was derived from phone posterior probabilities in an automatic speech recognition (ASR) system. The research also classified the voice to three conditions: mild, moderate or severe [48]. By using an end-to-end deep neural network, the autism children’s speech and language abnormality was assessed. The utterance level high dimensional acoustic features were extracted in the Open SMILE toolkit, and modeled a typical prosody label from the constant Q transform spectrogram of Mandarin speech [49]. By using improved artificial neural network, pathological voice formants were repaired, and the Line spectral frequencies (LSF) of normal speech and pathological voice are extracted respectively, and then the pathological LSP was projected to the normal space so that to get the normal LSP [50].

Three machine learning algorithms, namely, deep neural network (DNN), SVM, and Gaussian mixture model were employed to classified and evaluated the voice disorder database of MEEI (Massachusetts Eye and Ear Infirmary), by extracting the MFCCs from 3-second samples of a sustained vowel [51]. A deep learning strategies for disorder detection was proposed in [52], to carry out a preliminary diagnosis of the samples from database of the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Labs to two classes—Normal and Pathological. A deep learning framework was proposed in [53], and the PD patient’s movements difficulties considering information from speech, handwriting, and gait was modeled by a methodology. The support vector regression, Gaussian process regression, and deep learning framework were used to investigate subjective and objective assessment of Parkinsonian speech quality, via the temporal, spectral, and/or cepstral parametrization feature vectors from the speech recordings, and then subsequently mapped to the predicted quality scores outputted the three classifiers [54].

5. Metrics for Evaluation of the Systems

The disorder voice classification or recognition problem is always an unbalance problem, and the main metrics to evaluate these recognition systems are: accuracy, recall, precision, F1 score, execution time, Receiver Operating Characteristic curves (ROC), Area Under Curve (AUC), Equal Error Rate (EER) and Detection Error Tradeoff (DET). The following four measurements are usually used to calculate the metrics:

- True Positive (TP): the voice sample is pathological, and the system recognizes this;
- True Negative (TN): the voice sample is healthy, and the system recognizes this;
- False Positive (FP): the voice sample is healthy, but the system recognizes it as pathological;
• False Negative (FN): the voice sample is pathological, but the system recognizes it as healthy.

The accuracy, that is the percentage of correctly classified instances, is defined as:

\[
\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}
\]

and the sensitivity and the specificity, that represent respectively the test's ability to detect positive results or the identification of negative results, are defined as:

\[
\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{and} \quad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}
\]

The ROC area is a goodness measure of a classification algorithm, which evaluated by plotting a curve representing the sensitivity versus the complementarity of specificity and measuring the area under this curve (AUC). The AUC can be interpreted as the average value of sensitivity for all the possible values of specificity. The maximum (AUC = 1) means that the algorithm is perfect in the classification between diseased and non-diseased voices. On the other hand, AUC = 0 means that the algorithm incorrectly classifies all subjects with diseases as negative and all healthy subjects as pathological. DET curve is a curve of bit error rate of binary classification system, which is the curve between false reject rate (FRR) and false accept rate (FAR) with the change of judgment threshold.

6. Conclusions

The disorder voice processing is a technology with unique advantages and has potential applications in clinic after these voices were affected by organ impairments and mental factors. The modifies, selections and combinations of features with the deep learning approaches which can extract different layers futures, will be the future directions in given clinic applications.

References

[1] Stamford, J.A., P.N. Schmidt and K.E. Friedl, What Engineering Technology Could Do for Quality of Life in Parkinson's Disease: A Review of Current Needs and Opportunities. IEEE Journal of Biomedical and Health Informatics, 2015. 19(6): p. 1862-1872.

[2] Smith, S.L., L. Maxfield and E.J. Hunter, Sensitivity analysis of muscle mechanics-based voice simulator to determine gender-specific speech characteristics. Biomechanics and Modeling in Mechanobiology, 2019. 18(2): p. 453-462.

[3] Zhang, Y., et al., Nonlinear detection of disordered voice productions from short time series based on a Volterra–Wiener–Korenberg model. Chaos, Solitons & Fractals, 2011. 44(9): p. 751-758.

[4] Guidi, A., et al., Analysis of speech features and personality traits. Biomedical Signal Processing and Control, 2019. 51(2019): p. 1-7.

[5] Hsu, A. and J. Ongkasuwan, Beyond Nodules—Diagnostic and Treatment Options in Pediatric Voice Disorders. Otolaryngologic Clinics of North America, 2019. 52(5): p. 949-956.

[6] Hector E. Romero, Ning Ma, and Guy J. Brown, et.al. Deep Learning Features For Robust Detection of Acoustic Events In Sleep-disordered Breathing. ICASSP2019, 2019: p. 810-814.

[7] Toledano, D.T., A.O. Giménez and A. Teixeira. Advances in Speechand Language Technologies for Iberian Languages. in IberSPEECH 2012 Conference. 2012. Madrid,Spain: Springer.

[8] Vaiciukynas, E., et al., Exploring similarity-based classification of larynx disorders from human voice. Speech Communication, 2012. 54(5): p. 601-610.

[9] J.He,Y.X.Li,and Q.H.He, An abnormal speech speaker recognition algorithm weighted by mutation features. Journal of South China University of Technology (NATURAL SCIENCE EDITION), 2012. 40(03): p.106-111.(in Chinese).

[10] J.WBi, J.Q.Han. A new method of modified mel frequency for recognition of stressed speech under G-force. Computer Scienc, 2002.20(12):p.150-154.(in Chinese).

[11] L.L.Gu, X.J.Zhang, and C.W.Huang. Study on the model of nonlinear dynamics phonation systme for the classificaiotn of ployps and paralylysis phonation. ACTA ACUSTICA, 2015. 40(06): p.878-885.(in Chinese).

[12] Y.L.Ma, J.Q.Han and L.Zhang, TEO-Pitch based classification of stressed speech under
G-force. ACTA ACUSTICA, 2002. 27(6): p.518-522. (in Chinese)

[13] Hilal I. KAHRAMANOZSEVEN, T., M. DU GENCI and A. DORUK, Voice Traces of Anxiety: Acoustic Parameters Affected by Anxiety Disorder. ARCHIVES OF ACOUSTICS, 2018. 43(4): p. 625-636.

[14] Afshar, AmberJinxi and Guo Soo, et al., Effectiveness of Voice Quality Features in Detecting Depression. Interspeech, 2018. 19(1): p. 1676-1680.

[15] Aihara, R., T. Takiguchi and Y. Ariki, Individuality-Preserving Voice Conversion for Articulation Disorders Using Phoneme-Categorized Exemplars. ACM Transactions on Accessible Computing, 2015. 6(4): p. 1-17.

[16] Alpan, A., et al., Assessment of disordered voice via the first rahmonic. Speech Communication, 2012. 54(5): p. 655-663.

[17] Rufo, M.J., et al., A Bayesian decision analysis approach to assess voice disorder risks by using acoustic features. Biometrical Journal, 2019. 61(3): p. 503-513.

[18] Sakar, C.O., et al., A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing, 2019. 74(2019): p. 255-263.

[19] Verde, L., G De Pietro and G Sannino, A methodology for voice classification based on the personalized fundamental frequency estimation. Biomedical Signal Processing and Control, 2018. 42(2018): p. 134-144.

[20] Liu, B., E. Polce and J. Jiang, An Objective Parameter to Classify Voice Signals Based on Variation in Energy Distribution. Journal of Voice, 2019. 33(5): p. 591-602.

[21] Orozco-Arroyave, J.R., et al., Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases. IEEE Journal of Biomedical and Health Informatics, 2015. 19(6): p. 1820-1828.

[22] Sakar, B.E., et al., Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings. IEEE Journal of Biomedical and Health Informatics, 2013. 17(4): p. 828-834.

[23] Ankışhan, H. and S.Ç. İnam, A hybrid measure for the discrimination of the acoustic signals: Feature matrix (FMx). Applied Acoustics, 2019. 152(2019): p. 88-100.

[24] Huang, Z., et al., A Hybrid Feature Selection Method Based on Binary State Transition Algorithm and ReliefF. IEEE Journal of Biomedical and Health Informatics, 2019. 23(5): p. 1888-1898.

[25] Parsa, V. and D.G. Jamieson, Interactions between speech coders and disordered speech. Speech Communication, 2003. 40(3): p. 365-385.

[26] Howard, D.M., E. Abberton and A. Fourcin, Disordered voice measurement and auditory analysis. Speech Communication, 2012. 54(5): p. 611-621.

[27] Mostafa, S.A., et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson’s disease. Cognitive Systems Research, 2019. 54(2019): p. 90-99.

[28] Haq, A.U., et al., Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson’s Disease Using Voice Recordings. IEEE Access, 2019. 7(2019): p. 37718-37734.

[29] Sheffield, T.U.O., https://www.sheffield.ac.uk/cast/projects. 2019.

[30] Monge-Alvarez, J., et al., Robust Detection of Audio-Cough Events Using Local Hu Moments. IEEE Journal of Biomedical and Health Informatics, 2019. 23(1): p. 184-196.

[31] Verikas, A., et al., Questionnaire- versus voice-based screening for laryngeal disorders. Expert Systems with Applications, 2012. 39(6): p. 6254-6262.

[32] Smith, N.R., et al., Detection of Simulated Vocal Dysfunctions Using Complex sEMG Patterns. IEEE Journal of Biomedical and Health Informatics, 2016. 20(3): p. 787-801.

[33] Ortiz, A.J., et al., Automatic speech and singing classification in ambulatory recordings for normal and disordered voices. The Journal of the Acoustical Society of America, 2019. 146(1): p. EL22-EL27.

[34] Deshpande, P.S. and M.S. Manikandan, Effective Glottal Instant Detection and
Electroglottographic Parameter Extraction for Automated Voice Pathology Assessment. IEEE Journal of Biomedical and Health Informatics, 2018. 22(2): p. 398-408.

[35] Drugman, T., et al., Objective Study of Sensor Relevance for Automatic Cough Detection. IEEE Journal of Biomedical and Health Informatics, 2013. 17(3): p. 699-707.

[36] Sahadat, M.N., A. Alreja and M. Ghovanloo, Simultaneous Multimodal PC Access for People With Disabilities by Integrating Head Tracking, Speech Recognition, and Tongue Motion. IEEE Transactions on Biomedical Circuits and Systems, 2018. 12(1): p. 192-201.

[37] Daudet, L., et al., Portable mTBI Assessment Using Temporal and Frequency Analysis of Speech. IEEE Journal of Biomedical and Health Informatics, 2017. 21(2): p. 496-506.

[38] Verde, L., G. De Pietro and G. Sannino, Voice Disorder Identification by Using Machine Learning Techniques. IEEE Access, 2018. 6(SPECIAL SECTION ON MOBILE MULTIMEDIA FOR HEALTHCARE): p. 16246-16255.

[39] Gómez-García, J.A., et al., Emulating the perceptual capabilities of a human evaluator to map the GRB scale for the assessment of voice disorders. Engineering Applications of Artificial Intelligence, 2019. 82(2019): p. 236-251.

[40] Pham, M., J. Lin and Y. Zhang, Diagnosing Voice Disorder with Machine Learning, in 2018 IEEE International Conference on Big Data. 2018, IEEE: Seattle, WA, USA. p. 5263-5266.

[41] Feng, L., C. Zhao and B. Huang, A slow independent component analysis algorithm for time series feature extraction with the concurrent consideration of high-order statistic and slowness. Journal of Process Control, 2019. 84(2019): p. 1-12.

[42] Erfanian Saeedi, N. and F. Almasganj, Wavelet adaptation for automatic voice disorders sorting. Computers in Biology and Medicine, 2013. 43(6): p. 699-704.

[43] Saidi, P. and F. Almasganj, Voice Disorder Signal Classification Using M-Band Wavelets and Support Vector Machine. Circuits, Systems, and Signal Processing, 2015. 34(8): p. 2727-2738.

[44] Ali, Z., et al., Voice pathology detection based on the modified voice contour and SVM. Biologically Inspired Cognitive Architectures, 2016. 15(2016): p. 10-18.

[45] Fonseca, E.S., et al., Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (DPM). Biomedical Signal Processing and Control, 2020. 55(2020): p. 101615.

[46] Berus, L., et al., Classifying Parkinson’s Disease Based on Acoustic Measures Using Artificial Neural Networks. Sensors, 2019. 19(1): p. 16.

[47] Almeida, J.S., et al., Detecting Parkinson’s disease with sustained phonation and speech signals using machine learning techniques. Pattern Recognition Letters, 2019. 125(2019): p. 55-62.

[48] Liu, Y., et al., Acoustical Assessment of Voice Disorder With Continuous Speech Using ASR Posterior Features. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019. 27(6): p. 1047-1059.

[49] Li, M., et al., An automated assessment framework for atypical prosody and stereotyped idiosyncratic phrases related to autism spectrum disorder. Computer Speech & Language, 2019. 56(2019): p. 80-94.

[50] L.J.Xue, B.Y.Sun, and X.J.Zhang, Pathological Voice Formants Repaired by Improved Artificial Neural Network. Chinese Journal of Electron Devices, 2019. 42(01): p.249-254.

[51] Fang, S., et al., Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. Journal of Voice, 2019. 33(5): p. 634-641.

[52] Guan, H. and A. Lerch, Learning Strategies for Voice Disorder Detection, in 2019 IEEE 13th International Conference on Semantic Computing. 2019, IEEE: Newport Beach, California, USA. p. 295-301.

[53] Vasquez-Correia, J.C., et al., Multimodal Assessment of Parkinson's Disease: A Deep Learning Approach. IEEE Journal of Biomedical and Health Informatics, 2019. 23(4): p. 1618-1630.

[54] Gaballah, A., et al., Objective and Subjective Speech Quality Assessment of Amplification Devices for Patients With Parkinson's Disease. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019. 27(6): p. 1226-1235.