Features extraction for the automatic detection of ALS disease from acoustic speech signals

Maxim Vashkevich*, Elias Azarov*, Alexander Petrovsky*,
Yuliya Rushkevich†
*Belarusian State University of
Informatics and Radioelectronics
vashkevich@bsuir.by
†Republican Research and Clinical Center of
Neurology and Neurosurgery

***Abstract*—The paper presents a features for detection of pathological changes in acoustic speech signal for the diagnosis of the bulbar form of Amyotrophic Lateral Sclerosis (ALS). We collected records of the running speech test from 48 people, 26 with ALS. The proposed features are based on joint analysis of different vowels. Harmonic structure of the vowels are also taken into consideration. We also presenting the rationale of vowels selection for calculation of the proposed features. Applying this features to classification task using linear discriminant analysis (LDA) lead to overall correct classification performance of 88.0%.**

***Index Terms*—speech analysis, formants, ALS.**

I. Introduction

Perceptible changes in speech are inherent to many neurological diseases. Bulbar motor changes (i.e. difficulty with speech or swallowing) are the first symptoms in approximately 30% of persons with amyotrophic lateral sclerosis (ALS) [1]. In most cases detection of speech motor involvement in ALS currently are based on subjective assessment of clinicians' auditory perceptions. However auditory-perceptual judgment as a tool for classifying speech disorders are susceptible to a variety of sources of error and bias [2]. Some symptoms of speech motor changes in ALS cannot be easily detected without instrumentation [3] especially at the beginning of the disease. In turn, late detection of voice pathology can lead to late detection of ALS. Advanced assessment strategies of speech motor changes are needed for early disease detection and optimizing the efficacy of therapeutic ALS drug trials [4].

One of the problem of ALS detection is there is no standardized speech diagnostic procedure. For detecting neurological diseases many vocal test have been proposed. Some of them include *sustained phonations* [5, 6], where the patient is instructed to produce a single vowel and hold the pitch of it as constant as possible, for as long as possible. In *running speech* tests patients are instructed to speak a standard sentence that is constructed to contain a representative sample of linguistic units [7]. Another approach is to use rapid repetitions of syllables, which is referred to as a *diadochokinetic task* (DDK). During this speaking test patients are asked to produce the maximum number of syllable (e.g., "tah" or "pah") as rapidly and accurately as possible in a single breath [1].

Currently for detecting neurological diseases using vocal tests different time, frequency or time-frequency features are used. They are extracted form the speech signal using either linear or non-linear processing techniques. Features based on linear processing include F0 (the fundamental frequency of vocal oscillation), absolute sound pressure level, *harmonics-to-noise ratio* (HNR) [8], *jitter* (the degree of variation of F0 from cycle to cycle), *shimmer* (the degree of variation in speech amplitude from cycle to cycle) [7], Mel-frequency cepstral coefficients (MFCC) [9]. The main drawback of the mentioned features is that many of them does not specifically designed for detection of voice disorders and therefore their performance is limited. Also they can be well applied to sustained phonation test but not to running speech test.

More recently, several new measurement methods have been proposed to assess dysphonic symptoms in speech [6]. Those methods are based on nonlinear time series analysis. The most popular among them *detrended fluctuation analysis* (DFA) and *recurrence period density entropy* (RPDE) [7]. The drawback of the nonlinear measurement methods is their much higher sensitivity to noise and other environmental factors.

As a rule all extracted feature vectors with corresponding labels are used to obtain classifier based on supervised learning. Linear discriminant analysis (LDA) along with support vector machine (SVM) are the most frequently used classification tools in tasks of neurological diseases diagnosis [6, 9, 10].

The aim of this work is to present new features that based on linear processing techniques and designed specially for detecting dysphonic/dysathic speech pathology of patients with ALS. The proposed features are robust to uncontrolled variation in acoustic environment, have clear rationale and possibly can be used in telemedicine systems. Using the proposed features classifier for the diagnosis of ALS based on LDA have been obtained.

II. Foundation of vowels selection

The ALS results in changes of the stimulation of the muscles in general and the muscles of the tongue in particular. The position of the tongue's surface is manipulated by large and powerful muscles in its root, which move it within the mouth [11]. From the speech production point of view important horizontal position of the tongue surface (front ⇔ back) and vertical position (high ⇔ low). Fig. 1 shows a schematic characterization of some Russian vowels in terms of relative tongue positions.

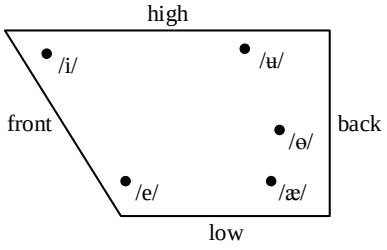

Fig. 1: Relative tongue positions of Russian vowels (in International Phonetic Alphabet (IPA) representation) [11]

For detecting symptoms of bulbar form of ALS from the acoustic speech signal, it is advisable to select the vowels /æ/ and /i/, since for their pronouncing requires a considerable activity of tongue muscles.

In our experiments we use running speech as more realistic test of impairment in actual everyday life. Records with

counting form 1 to 10 (in Russian) were used as a materials for experiments. For the analysis we have selected close in time fragments of speech signal containing the vowels /æ/ and /i/, (as a rule, sounds were selected from the words “od**in**”, “dv**æ**”, “t**ri**”). An example of the formant structure of vowels /æ/ and /i/, produced by a healthy person is shown in Fig. 2.

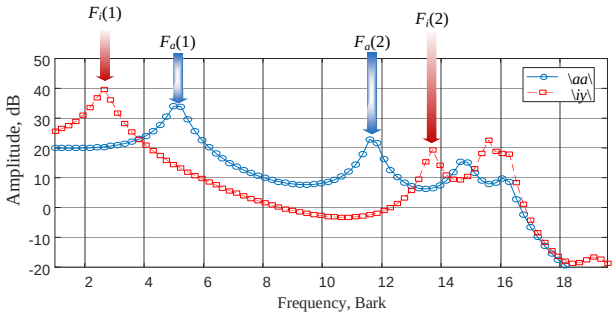

Fig. 2: Formant structure of the vowels /æ/ and /i/ (healthy person)

Visual analysis of envelopes in Fig. 2 shows that formants significantly spaced in frequency domain and are arranged in the following order: $F_i(1) < F_a(1) < F_a(2) < F_i(2)$. As a rule pathological changes in speech are perceived aurally therefore it is meaningful to use psychoacoustically motivated Bark scale for improving correlation between perceptual and acoustic data [2]. As it will be shown in the following convergence of the formants can mean existence of pathological abnormalities. In this regard, Bark scale allows to unify distance between firsts and seconds formants of the vowels /æ/ and /i/.

III. Features for automatic detection

A. Distance between envelopes

Joint analysis of envelopes of vowels /æ/ and /i/ of persons with ALS have revealed an increased similarity between the shapes of these envelopes. A typical example of envelopes with a high degree of similarity is given in Fig. 3.

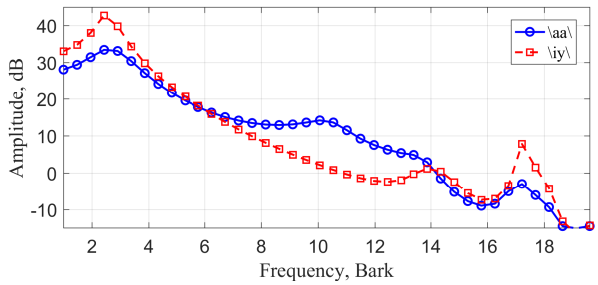

Fig. 3: Similarity of the envelope sounds /æ/ and /i/ (for patient with ALS)

To quantify the differences between the envelopes of vowels /æ/ and /i/ it is suggested to use the l_1-norm distance measure

$$d_1(E_i, E_a) = \sum_{k=1}^{P} |E_i(k) - E_a(k)|, \qquad (1)$$

where $E_i(k)$ is envelope of the vowel /i/, $E_a(k)$ – envelope of the vowel /æ/, P – the number of points in the Bark frequency domain in which envelope is defined.

B. Mutual location of the formant frequencies

As mentioned in section II formants of vowels /æ/ and /i/ have fixed order in normal case. However, in patients with ALS mutual location of the formant frequencies can be violated. Fig. 4 shows an example of the envelopes of vowels /æ/ and /i/, pronounced by a patient with ALS (voice disorder was perceivable).

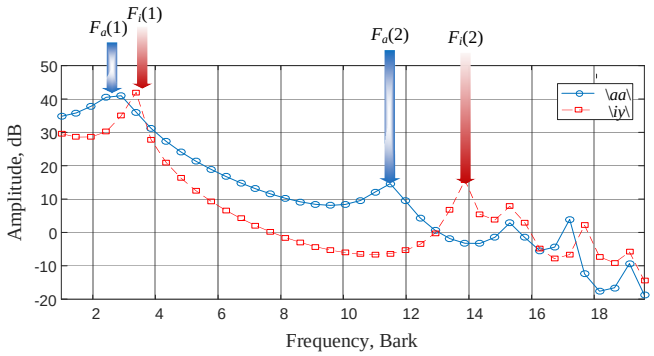

Fig. 4: Abnormal mutual location of the formant frequencies in patient with ALS

In the case when the normal order is not violated, there is often a significant convergence of the formant frequencies as shown in Fig. 5.

To quantify the degree of violation of the mutual formant structure of vowels /æ/ and /i/ feature $fmt_{err}(F_i, F_a)$ is proposed (see eq. (2)). This expression returns value in the range $[0,\ 2]$. The fmt_{err} is equal to 2 when the normal mutual formant structure is violated (i.e. either $F_i(1) > F_a(1)$, or $F_a(2) > F_i(2)$). For normal mutual formant location fmt_{err} returns 0. It has been noticed that the distance between the formants of vowels /æ/ and /i/ for healthy person is more than 2 Bark, therefore the degree of convergence of the formants is estimated by the function $fmt_{err}(F_i, F_a)$ in cases when distance between the first formants and/or between the second formants of vowels /æ/ and /i/ are less than 2 Barks.

C. Difference in the amplitudes of the harmonics.

Analysis of harmonic structure of vowel /æ/ in persons with ALS have revealed that disphonic disorders have affect on fist three harmonic components. Fig. 6 and Fig. 7 shows some representative examples.

To quantify the degree of deviation in amplitude structure of harmonics of vowels /æ/ the following measure is proposed

$$harm_{diff}(A_1, A_2, A_3) = \max(A_1, A_2) - A_3, \qquad (3)$$

where A_i – amplitude of i-th harmonic in dB.

$$fmt_{err}(F_i, F_a) = \begin{cases} 2 & \text{if } F_i(1) > F_a(1) \text{ or } F_a(2) > F_i(2) \\ 2 - \frac{F_a(1)-F_i(1)}{2} - \frac{F_i(2)-F_a(2)}{2}, & \text{if } F_a(1) - F_i(1) < 2 \text{ and } F_i(2) - F_a(2) < 2 \\ 1 - \frac{F_a(1)-F_i(1)}{2}, & \text{if } F_a(1) - F_i(1) < 2 \\ 1 - \frac{F_i(2)-F_a(2)}{2}, & \text{if } F_i(2) - F_a(2) < 2 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

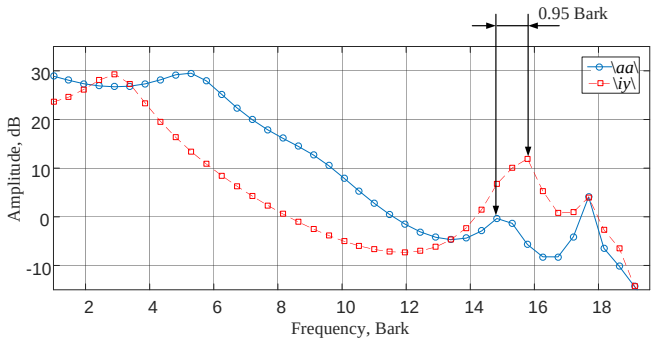

Fig. 5: Convergence of formant frequencies of the vowels /æ/ and /i/ (patient with ALS)

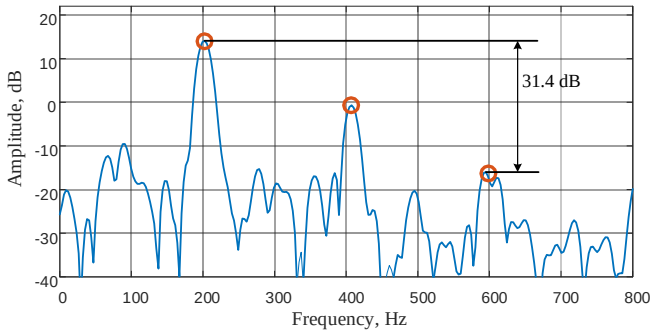

Fig. 6: First three harmonics of vowel /æ/ (ALS patient). Difference between amplitudes of 1st and 3rd harmonic is 31 dB

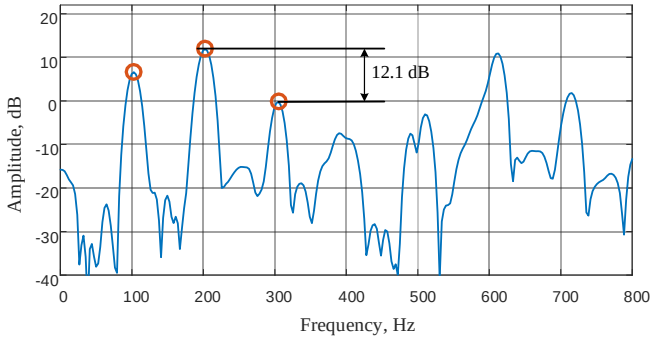

Fig. 7: First three harmonics of vowel /æ/ (ALS patient). Difference between amplitudes of 2nd and 3rd harmonics is 12 dB

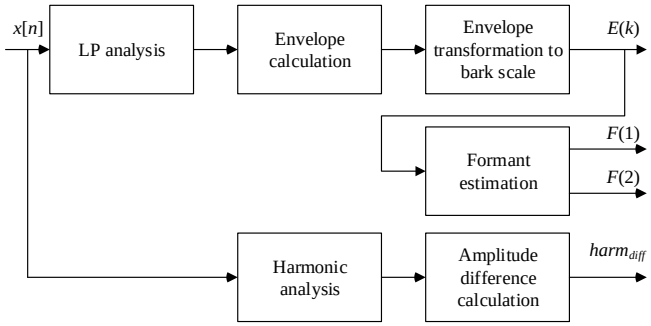

Fig. 8: Scheme of speech signal analysis

IV. Classification

A. Scheme for features extraction

General scheme for features extraction for automatic detection of bulbar ALS is given in in Fig. 8. For the analysis, segments of speech signal with a duration of 150-200 msec containing vowels /æ/ and /i/ were selected. LP-analysis is done using traditional algorithms, while harmonic analysis is performed using technique described in [12]. For each pair of vowels from the dataset features (1), (2) and (3) are extracted and concatenated into vector $\mathbf{x} = [d_1(E_i, E_a) \;\; fmt_{err}(F_i, F_a) \;\; harm_{diff}(A_1, A_2, A_3)]^T$.

B. Linear discriminant analysis

In order to discriminating between the two classes of normal and pathological cases, linear discriminant analysis (LDA) with Fisher criterion was used [13]. The idea of linear discriminant analysis (LDA) lies in the search for such a hyperplane $\mathbf{w}$ in the feature space, so that the projection of all training vectors onto it minimizes the within-class variation and maximizes the between-class variation:

$$\mathbf{w} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}\mathbf{S}_B\mathbf{w}^T}{\mathbf{w}\mathbf{S}_W\mathbf{w}^T}, \quad (4)$$

where $\mathbf{S}_B$ – between class scatter matrix and $\mathbf{S}_W$ – within class scatter matrix. In turn these matrices are calculated as follows

$$\mathbf{S}_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T, \quad (5)$$

$$\mathbf{S}_W = \sum_{j=1}^{2} \sum_{\mathbf{x}} (\mathbf{x} - \mu_j)(\mathbf{x} - \mu_j)^T \quad (6)$$

where μ_1 – mean value of feature vector for healthy people and μ_2 – mean value of feature vector for people with ALS. The solution of (4) can be found via the generalized eigenvalue problem

$$\mathbf{S}_B\mathbf{w} = \lambda\mathbf{S}_W\mathbf{w}, \quad (7)$$

where the maximum eigenvalue λ and its associated eigenvector gives the quantity of interest and the projection basis. More detailed description of LDA is given in [13].

V. Experimental results

A. Data collection

To validate proposed new features, real world clinical samples were used. Speech recording of Russian speaking patients with ALS was carried out in Republican Research and Clinical Center of Neurology and Neurosurgery (Minsk, Belarus). A total of 48 speakers were recorder, with 22 healthy speakers (15 males, 7 females) and 26 speakers (14 males, 12 females) having been diagnosed with ALS. The average age in the healthy group was 36.3 years (SD 9.5, Min 22, Max 81) and the average age in the ALS group was 56.5 years (SD 10.5, Min 36, Max 82). The samples recorded at 44.1 kHz using smartphone with a standard headset and stored as 16 bit uncompressed PCM files.

For classification purpose 106 pairs of vowels /æ/ and /i/ were manually pre-segmented prior to feature extraction (61 – healthy, 45 – pathology).

B. Statistical analysis of features

In order to gain a preliminary understanding of the statistical properties of the features we compute their distributions estimated using Gaussian kernel density.

Figure 9 shows the density function for distance between envelopes $d_1()$. This feature shows a considerable distinction between healthy controls and people with ALS.

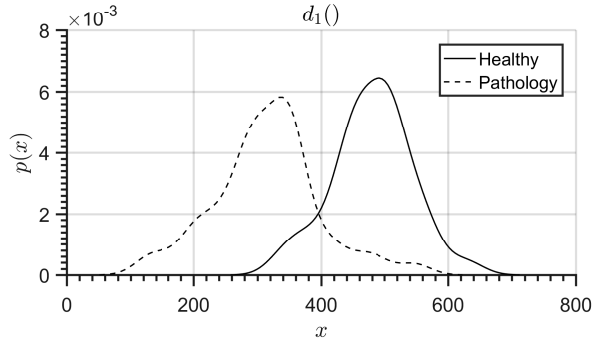

Fig. 9: Probability density of distance between envelopes $d_1(E_i, E_a)$

The probability density of fmt_{err} feature is shown in Fig. 10. This results show that there are violations of mutual

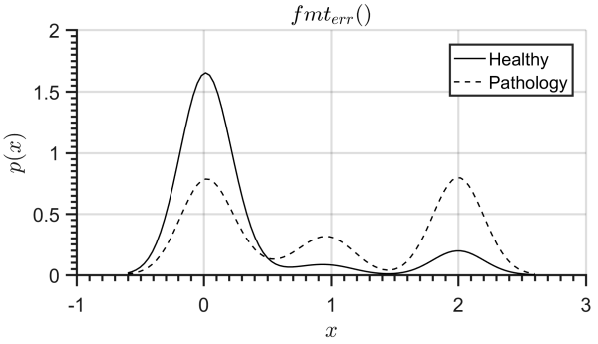

Fig. 10: Probability density for $fmt_{err}(E_i, E_a)$

location of the formant frequencies for some samples from healthy control group. However, this could appear due to inaccuracy of algorithm of formant frequencies detection.

The feature $harm_{diff}$ is also shows a good separation between healthy and pathological groups (Fig.11).

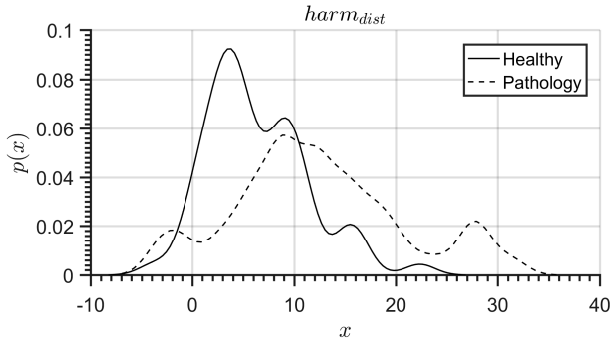

Fig. 11: Probability density for difference between harmonic amplitudes $harm_{diff}$

For comparison reason we have computed a density function for widely used HNR feature (Fig. 12). Although HNR is

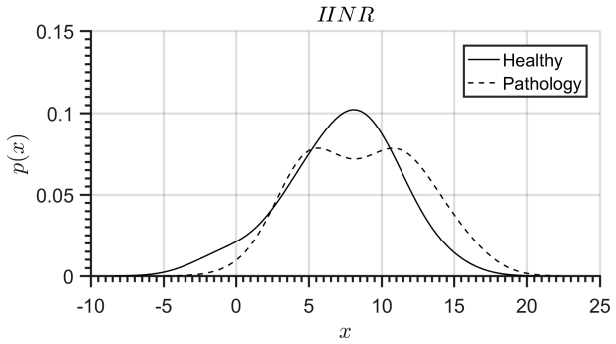

Fig. 12: Probability density for HNR computed for vowel /æ/ taken from running speech test

quite effective for sustained phonation test [7] it is not so good when used for analysis of short vowels (<200 ms).

C. Classification results

Using collected base of 106 train samples LDA was performed based on the following steps:

- Between class scatter matrix $\mathbf{S}_B$ calculation using (5);
- Within class scatter matrix $\mathbf{S}_B$ calculation using (6);
- Solving the eq. (7) by calculating matrix $\mathbf{S_W^{-1}S_B}$ and performing its eigenvalue decomposition. To maximize Fisher's criterion (4) projection hyperplane $\mathbf{w}$ is determined by eigenvector associated with maximum eigenvalue λ.

Classification is performed using following equation

$$p = \text{sign}(\mathbf{w}^T\mathbf{x} - b) \tag{8}$$

where b is boundary, if $p = -1$ then vector $\mathbf{x}$ classified as healthy, if $p = 1$ then vector $\mathbf{x}$ classified as pathology.

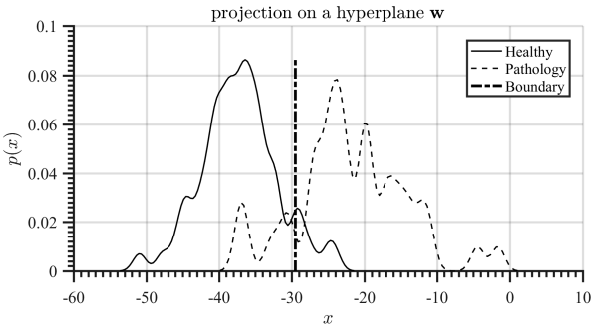

Fig. 13: Probability density function for projection on hyperplane $\mathbf{w}$ of all train vectors

In Fig. 13 kernel density function for projection on hyperplane of all train vectors is shown. Overall classification accuracy is equal to 88.0%, true positive 90.5% and true negative 84.6%.

VI. Conclusion

The paper presents a several new features that can be calculated form running speech test for ALS diagnosis. New features are based on 1) analysis of envelopes of vowels /æ/ and /i/ and 2) analysis of mutual formant structure of vowels /æ/ and /i/. The vowels /æ/ and /i/ were selected as the most suitable because their pronouncing requires a considerable work of tongue muscle (the bulbar symptoms of ALS include tongue atrophy). Another one feature is based on analysis of harmonic structure of vowel /æ/, the statistical analysis have shown that for pathological cases difference between first two and third amplitudes of harmonics lager then in healthy control group. Further work is necessary to improve classification result. Usage of presented feature with LDA-based classier allows to achieve overall classification accuracy of 88%.

Acknowledgment

This work was supported by the Belarusian Fundamental Research Fund (F17U003).

References

[1] T. Spangler, N. V. Vinodchandran, A. Samal, and J. R. Green, "Fractal features for automatic detection of dysarthria," in *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, Feb 2017, pp. 437–440.

[2] R. D. Kent, "Hearing and believingsome limits to the auditory-perceptual assessment of speech and voice disorders," *American Journal of Speech-Language Pathology*, vol. 5, no. 3, pp. 7–23, 1996.

[3] Y. Yunusova, J. S. Rosenthal, J. R. Green, S. Shellikeri, P. Rong, J. Wang, and L. Zinman, "Detection of bulbar ALS using a comprehensive speech assessment battery," in *Models and analysis of vocal emissions for biomedical applications: 8th international workshop*, Dec 2013, pp. 217–220.

[4] J. R. Green, Y. Yunusova, M. S. Kuruvilla, J. Wang, G. L. Pattee, L. Synhorst, L. Zinman, and J. D. Berry, "Bulbar and speech motor assessment in als: Challenges and future directions," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 14, no. 7-8, pp. 494–500, 2013.

[5] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice, 2nd ed.* San Diego: Singular Thomson Learning, 2000.

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

[7] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinsons disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, April 2009.

[8] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Institute of Phonetic Sciences, University of Amsterdam, Proceedings 17*, 1993, pp. 97–110.

[9] A. Benba, A. Jilbab, and A. Hammouch, "Discriminating between patients with parkinson's and neurological diseases using cepstral analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 10, pp. 1100–1108, Oct 2016.

[10] M. Little, P. McSharry, I. Moroz, and S. Roberts, "Nonlinear, biophysically-informed speech pathology detection," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 2, May 2006, pp. II–II.

[11] X. Huang, A. Acero, and G.-W. Hon, *Spoken language processing: a guide to theory, algorithm, and system development*. Upper Saddle River, New Jersey, USA: Prentice Hall PTR, 2001.

[12] A. Petrovsky and E. Azarov, "Instantaneous harmonic analysis: Techniques and applications to speech signal processing," in *Speech and Computer*, A. Ronzhin, R. Potapova, and V. Delic, Eds. Springer International Publishing, 2014, pp. 24–33.

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.