CNN based Parkinson's Disease Assessment using Empirical Mode Decomposition

Ayush Tripathi[a], Sunil Kumar Kopparapu[a]

[a] TCS Research & Innovation - Mumbai, Tata Consultancy Services Limited, Maharashtra, India.

Abstract

Parkinson's Disease (PD) is a neuro-degenerative disorder which is caused by a decrease in dopamine producing neurons in the human body and affects the body's motor system. In addition to affecting several motor and non-motor activities of a person's day to day life, PD patients have difficulty in speech production due to reduced coordination of the muscles that control breathing, phonation, articulation and prosody. Analyzing speech allows clinicians to objectively measure the severity of PD in a non-invasive way. In this work, we propose an effective method to discriminate between PD and healthy control (HC) subjects by utilizing a technique to decompose a speech signal into simpler Intrinsic Mode Functions called the Empirical Mode Decomposition. We train a Convolutional Neural Network (CNN) to learn significant properties from raw IMFs for the purpose of PD-HC classification. We evaluate our technique on sustained phonations speech from the Italian Parkinson's Voice and Speech database. Experimental results show that significant characteristics of Parkinsonian dysarthria can be learnt by using the raw IMFs and the need for explicitly extracting handcrafted features could be mitigated.

Keywords

Parkinson's speech, Empirical Mode Decomposition, Intrinsic Mode Function, sustained phonation

1. Introduction

Parkinson's Disease (PD) is a neuro-degenerative disorder which is caused by a decrease in dopamine producing neurons in the human body and affects the body's motor system [1]. PD affects 1-2 per 1000 of the population at any time. The prevalence of PD increases with age and it affects roughly 1% of the population above 60 years [2]. Normal respiratory and well controlled articulatory movements are fundamental for producing well-coordinated normal speech. The common signs and symptoms of PD such as tremor, bradykinesia, rigid muscles and akinesia hamper the ability of an individual to precisely control the speech producing organs which leads to disordered speech. This manifests in PD patients in the form of soft voice, monotone, breathiness, hoarse voice quality, imprecise articulation and a decrease in naturalness while speaking [3].

In the absence of any specific laboratory test or instruments to measure or monitor the evolution and treatment response of PD, it is extremely crucial to track the motor functions such as gait freezing and speech analysis to examine the disease. Importantly, the signs of PD are often confused with those of natural aging hence making the diagnosis even more challenging. Clinicians widely use the Unified Parkinson's Disease Rating Scale (UPDRS) [4] for evaluation of PD. The evaluation is carried out through face to face interviews and clinical observations using a set of questions to evaluate: (a) non-motor experiences of daily living, (b) motor experiences of daily living, (c) motor examination, and (d) motor complications.

Naturally spoken speech can be analyzed in a non-invasive manner and hence the study of changes in acoustic properties of speech are a center-point of research for the measurement of symptomatic changes in PD [5]. Articulation, voice intensity, frequency spectrum, and speech intelligibility are the main acoustic parameters observed for tracking changes in speech. It was observed [6] that PD patients suffer from reduction in the range of articulatory movement which in turn leads to impaired vowel articulation. The production of vowels is a complicated process that involves precise control over the movements of the tongue, lips and jaw, creating oropharyngeal resonating cavities, which amplify certain frequency bands of the voice spectrum called formants. The possibility of using sustained phonation /a/ for discriminating PD from healthy subjects was first proposed in [7].

A set of 13 features describing different aspects of Parkinsonian speech for the task was suggested in [8]. Phonation and rhythm features [9] and other vowel features [10] to capture characteristics of PD dysarthria have been proposed in literature. An extensive feature analysis followed by a 2 stage feature selection to rep-

Proceedings of the CIKM 2020 Workshops, October 19-20, Galway, Ireland. Editors of the Proceedings: Stefan Conrad, Ilaria Tiddi
EMAIL: t.ayush@tcs.com (A. Tripathi);
sunilkumar.kopparapu@tcs.com (S.K. Kopparapu)
URL: https://www.tcs.com (S.K. Kopparapu)
ORCID: 0000-0002-7944-2260 (A. Tripathi); 0000-0002-0502-527X (S.K. Kopparapu)

resent physiological aspects of PD obtained from sustained vowel /a/ and DDK task was proposed in [11]. A set of frame-level features was used to construct a Fischer Vector representation of the speech sample along with a Support Vector Machine classifier in [12]. An i-vector based approach along with a large set of acoustic features was used in [13] in order to identify the most relevant features for characterizing the disorders in speech of PD patients. Voxtester [14], is a system for assessing PD related impairment by using a wide set of parameters including: voice spectrum, formants, DDK rate, voice intensity and vocal sound pressure level.

With the advent of machine learning in all spheres of processing the trend has been to extract more and more features using signal processing in order to discriminate PD and HC subjects. In this paper, we propose a method to classify PD and HC by decomposing the speech utterance by using the Empirical Mode Decomposition (EMD) technique. EMD is the process of decomposing non-stationary time series into simpler Intrinsic Mode Functions (IMF) in the time domain. This technique has had various applications in the speech domain such as enhancement, denoising [15], formant tracking [16], pathological voice analysis [17], emotion recognition [18], glottal activity detection [19] etc. In these studies, the emphasis has been on extracting temporal and spectral features using the IMFs which are then used for classification tasks. However, to the best of our knowledge, employing raw IMFs for classification of pathological speech has not been studied. The main contribution of this paper lies in using a Convolutional Neural Network architecture to learn these features from raw IMFs without the need of explicitly extracting handcrafted features for the purpose of PD-HC classification. The approach is validated on the Italian Parkinson's Voice and Speech database. The rest of the paper is organized as follows: Section 2 describes the database used for the experiments; we provide the description of the proposed approach in Section 3 while Section 4 details achieved results. We discuss the salient aspects of the proposed approach while also providing an analogy to the traditional feature extraction based methods in Section 5 and conclude in Section 6

2. Dataset

The Italian Parkinson's Voice and Speech database [20] consists of recordings from 28 (19 Male, 9 Female) speakers with Parkinson's Disease aged between 40 and 80 years and 22 (10 Male, 12 Female) healthy controls (HC) aged between 60 and 77 years. The utterances have been recorded in a warm, echo free and quiet room at a sampling frequency of 16 kHz by keeping the microphone at a distance of 15 to 25 centimeters from the subject. The speech intelligibility of the patients was perceptually assessed on a 5-point scale based on the UPDRS protocol. The following reading tasks were performed by the subjects:

Table 1
Number of 1 second utterances for PD and HC categories in the dataset.

Phonation	PD	HC
/a/	390	269
/e/	385	290
/i/	403	297
/o/	400	284
/u/	379	305
Total	1957	1445

- 2 phonations each of the vowel /a/, /e/, /i/, /o/, /u/
- execution of syllable /pa/ and /ka/ (5 sec)
- 2 readings of a phonetically balanced text
- reading of phonetically balanced words and phrases

In our study, we use a subset of this dataset, namely the sustained phonations (/a/, /e/, /i/, /o/, /u/). Depending on the severity of the condition and the speaker, the amount of time a subject can sustain a phonation is different and subsequently the length (in seconds) of the audio recordings are unequal. As will be discussed in Section 3, we segment the unequal length speech samples into non-overlapping segments (utterance) of each 1 second duration. In all there were 1957 utterances from PD and 1445 utterances from HC (see Table 1); this forms the data in all our experiments on the phonation data for PD-HC classification. For complete information on the recording protocol, the subjects and the tasks, please refer to [21].

3. Proposed Approach

The proposed PD diagnosis system consists of two major parts. First, the raw speech utterance of 1 second duration is decomposed into its Intrinsic Mode Functions (IMFs) by using the Empirical Mode Decomposition (EMD) technique. A 1D-CNN model is then trained using the raw IMFs as input for classifying the speech utterance into one of the two categories, namely, HC or PD. We now describe the signal decomposition

process and the architecture of the 1D-CNN model used in our experiments.

3.1. Empirical Mode Decomposition

Empirical Mode Decomposition is an adaptive, data driven technique used to decompose non-stationary and non-linear signals into Intrinsic Mode Functions of a signal, in the time-domain itself without the requirement of any a priori basis [22]. Any function that satisfies the following two conditions is categorized as an Intrinsic Mode Function:

1. The number of extrema and the number of zero crossings in the signal must be either equal or differ at most by one, and
2. The mean value of the envelope defined by joining the points of local minima and local maxima must be zero.

In order to decompose a signal $s[n]$ into its corresponding IMFs, the signal is subjected to a sifting process, namely,

1. For the signal $s[n]$, find the locations of all local maxima and minima. Define initial residue as, $r_0[n] = s[n]$
2. Connect all the local maxima (minima) by applying a cubic spline interpolation to obtain upper (lower) envelope E_{upper} (E_{lower}).
3. Compute the mean $E_{mean} = \frac{(E_{upper}+E_{lower})}{2}$
4. Update initial residue $r_0[n] \longleftarrow r_0[n] - E_{mean}$
5. Repeat Steps 1 - 4 until $r_0[n] = s[n]$ gets reduced to a function $h_1[n]$ which satisfies the properties of an IMF.
6. Obtain the first residue $r_1[n] = r_0[n] - h_1[n]$
7. Repeat Steps 1-6 with the residue $r_1[n]$ as the initial residue to find all the IMFs $h_i[n]$ $i = 1, 2, \cdots, K$.
8. Stop the process when the residue $r_K[n]$ becomes either monotonic, or a function with single maxima and minima or is a constant.

By performing the decomposition process, the signal $s[n]$ can be represented as a sum of IMFs and the final residue, namely,

$$s[n] = r_K[n] + \sum_{i=1}^{K} h_i[n] \tag{1}$$

Figure 1 depicts the IMFs obtained as a result of decomposing a natural speech utterance of one second duration, where the decomposition is curtailed at $K = 9$. Note that the process of decomposing a signal into

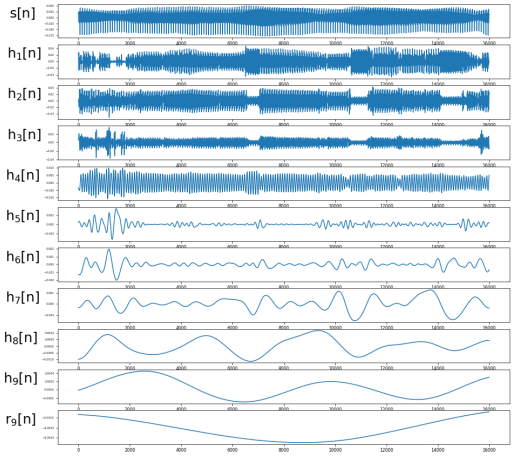

Figure 1: Empirical Mode Decomposition of a 1 second sample.

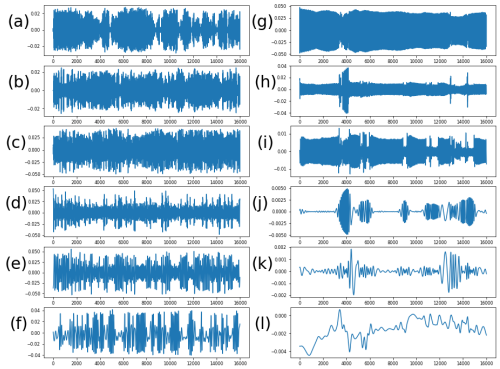

Figure 2: IMFs for PD and HC, (a)-(f) ((g)-(l)) represent first 5 IMFs and residue for PD (HC) speech of phonation /a/.

IMFs and then representing the IMFs using the instantaneous amplitude and frequency is termed as Hilbert Huang Transform (HHT). Features extracted from the IMFs can be used as complimentary features to the standard signal processing practices. In this regard, HHT can be understood as a generalized Fourier Transform that represents the signal in terms of a finite number of components [23].

In general, healthy speech is more coherent than the speech of a PD patient and as a result HC speech is decomposed faster (smaller K) than PD speech. This observation forms the hypothesis of our work. Previ-

ous studies have focused on using handcrafted spectral and temporal features extracted from these IMFs in order to discriminate between healthy and pathological speech (see [11, 24]). In this paper, we propose a machine learning approach to use the raw IMFs in order to diagnose the presence of Parkinson's disease. The first set of results are on the sustained phonations from both PD and HC. We consider the first five IMFs, namely, $h_1[n]$ to $h_5[n]$ and the residue, $r_5[n]$ as the input to our classifier.

Figure 2 depicts the first 5 IMFs and the final residue corresponding to the sustained phonation /a/ spoken by a HC ((a)-(f)) and a PD ((g)-(l)) subject. Clearly, one can visually notice the difference between the IMFs and the residue for HC and PD speech sample. These IMFs capture the characteristics of the parent signal and hence can be employed to extract information useful for pathological speech classification. This is the difference we wish to exploit to discriminate speech uttered by PD and speech uttered by HC.

3.2. Experimental Setup

The architecture of the 1D-CNN model used for the classification task is shown in Figure 3. The input to the 1D-CNN model is the raw IMF signal. The 1D-CNN was trained using Keras [25] deep learning library with Tensorflow [26] backend. We use speech signal (as mentioned in Table 1) of 1 second which corresponds to 16000 samples. Each of the 1 second speech utterance is subject to the EMD process and the first 5 IMFs ($h_1[n], h_2[n], \cdots, h_5[n]$) were extracted along with the final residue ($r_5[n]$). These are then fed as input to a multiple-input 1D-CNN network. Thus, the input to the network is a set of 6, 16000 dimensional vector (time series). We set the kernel size for the CNN to be 320 with a stride of 160 and the number of filters is chosen by performing a grid search to optimize the classification accuracy. The output of the CNN is then concatenated after a Global MaxPooling operation and is fed to a fully connected layer with ReLU activation function, while the number of neurons is optimized by using a grid search. For the output layer, softmax activation function is used with the output dimensions being the two classes, namely, HC and PD. The target to the model was one-hot encoding of the health state of the individual. We trained the network using binary cross-entropy loss with Adam optimizer. We set the learning rate to the default value of 0.001. In order to obtain speaker independent results which can be scaled to populations outside the training set, we perform a leave-one-speaker-out validation of the model wherein utterances corresponding

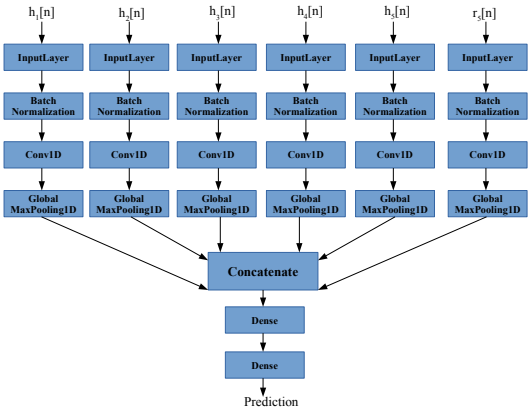

Figure 3: Proposed 1D-CNN Architecture.

to 49 speakers are used for training the model and the model is tested on the left out speaker. For all experiments, 20% of the training data is randomly chosen for the purpose of validating the model. For the test speaker, the posterior probabilities obtained from the model output for each 1 second utterance was averaged for classification. Note that Italian PD dataset is not very large (as is common with any pathological speech databases) to define separate train, test and validation sets, using leave one out mechanism allows predictions for all the speakers without relying on any sort of speaker specific information.

4. Results

The experimental results using 1D-CNN obtained for leave-one-speaker-out for different phonations are tabulated in Table 2. In order to account for variations in outcomes due to random weight initialization of the 1D-CNN, we repeat the experiment 5 times and report the average accuracy obtained in Table 2. We also report the specificity and sensitivity which is defined as the percentage of correctly classified HC and PD utterances respectively. The confusion matrix for 5 individual runs for the phonation /a/ is also shown in Table 3, as can be observed the number of correctly recognized subjects are not significantly different; the variation between different runs is ±2. As can be observed in Figure 2, the final residue ($r_5[n]$) is most reflective of the difference between PD and HC speech samples followed by IMFs $h_4[n]$ and $h_5[n]$. To evaluate if $r_5[n]$ by itself independently captures the discriminating properties between HC and PD, we trained a single input 1D-CNN model using $r_5[n]$ as the input,

Table 2
Accuracies for Phonation tasks (proposed approach).

Phonation	Accuracy	Specificity	Sensitivity
/a/	76.00	80.00	72.86
/e/	76.40	78.57	73.64
/i/	72.00	68.57	76.36
/o/	72.40	68.57	77.27
/u/	72.00	70.00	74.55
Average	73.76	73.14	74.94

Table 3
Confusion matrix for 5 runs for the phonation /a/ (proposed approach).

	PD	HC
PD	**21**, 20, 20, **21**, 20 (72.86%)	7, 8, 8, 7, 8 (27.14%)
HC	5, 3, 4, 5, 5 (20.0%)	17, **19**, 18, 17, 17 (80.0%)

Table 4
Accuracies for Phonation tasks (using $h_4[n]$, $h_5[n]$ and $r_5[n]$).

Phonation	Accuracy	Specificity	Sensitivity
/a/	69.6	61.82	75.00
/e/	72.8	71.82	73.57
/i/	59.6	51.82	65.71
/o/	66.4	60.00	71.43
/u/	62.4	52.73	70.00
Average	66.16	59.64	71.14

namely, all inputs were 0 except the last residue input $r_5[n]$ in Figure 3. We perform a similar analysis by training another model with inputs as signals $h_4[n]$, $h_5[n]$ and $r_5[n]$. The results obtained by using these approaches are reported in Tables 4 and 5. Clearly, the performance detoriates (it can be observed that for the phonation /a/ there is drop in accuracy from 76% to 69.6% and 64.4%) compared to when all the IMFs and residue are used together. Further, we combine the results obtained by using each of the individual phonations by taking a majority vote on the predictions obtained by each of the 5 different models. The class confusion matrix using this approach is presented in Table 6. We achieve an average accuracy of 85%, while the specificity and sensitivity values are 81.82% and 87.5% respectively.

The use of IMFs signals as raw features in a 1D-CNN classifier shows promise to be able to discriminate PD and HC as can be seen in Table 2. To the best of our knowledge, a study on classification of PD

Table 5
Accuracies for Phonation tasks (using only residue).

Phonation	Accuracy	Specificity	Sensitivity
/a/	64.4	52.72	73.57
/e/	67.2	74.54	61.43
/i/	56.4	51.82	60.00
/o/	62.4	55.45	67.86
/u/	61.2	57.27	64.29
Average	62.32	58.36	65.43

Table 6
Class confusion matrix for the classification system by using majority voting across all 5 sustained phonations.

	PD	HC
PD	87.5	12.5
HC	18.18	81.82

and HC using the Italian Parkinson's Voice and Speech has not been attempted earlier. However, our results are comparable to the state-of-the art measures which have been validated on other datasets, for example [11, 12, 13, 27]. Note that we did not have access to these datasets to make a direct comparison. On closer observation, we observed that most of the misclassified PD patients by our proposed approach belong to the class of 11 (of the 28) PD patients in the database who were rated 0 (namely, having no speech problems) on the UPDRS test scale by the clinicians. This is consistent with the fact that assigning a precise rating (PD or HC) for these *boundary* cases is challenging even for the trained experts which translates to misclassification of these samples.

5. Discussion

EMD is a popular decomposition technique used to analyze non-stationary and non-linear signals. The IMFs can be used to extract features like instantaneous amplitude and frequency, marginal spectrum etc which are relevant for pathological speech classification However, in this paper we propose a deep architecture in the form of 1D-CNN which allows us to use raw IMF signal instead of having to select and extract explicit features useful for pathological speech classification. It is commonly assumed that neural networks are black boxes that are unable to interpretable results. We attempt to explain the performance of the proposed architecture.

For the 1D-CNN, we used a kernel size of 320 with a stride of 160. In the hindsight this is equivalent to

extracting features from 20 ms of speech with a shift of 10 ms which is common practice in speech processing owing to the non-stationary nature of the speech signal. Further,

- The 1D-CNN network can be assumed to be a feature extraction mechanism which, given a raw IMF (or residue), extracts a set of discriminative features. The number of filters may be interpreted as the number of features extracted from a particular input signal.
- The extracted features from input signals $h_1[n]$ - $h_5[n]$ and $r_5[n]$ are then concatenated to form a feature vector.
- The Dense layers then act as a simple binary classifier with the input as the concatenated feature vector.

As one can observe, the use of raw IMFs mitigates the need to explicitly extract handcrafted features from the IMFs, the 1D-CNN architecture learns discriminating features from the raw signal to distinguish between PD and HC speech samples. For the purpose of decomposing the signal, the speech sample is segmented into fixed durations of 1 second each. This duration is long enough to capture the non-stationary aspect of speech as well as the dynamics involved in the phonation of vowel sounds.

6. Conclusion

Parkinson's Disease is a chronic neuro-degenerative disease which is difficult to diagnose. The symptoms of PD can be mistaken with natural aging, thereby making the diagnosis very very challenging. Tracking changes in speech has proven to be a useful tool for establishing non-invasive approach to early detection of PD. In this work, we propose an efficient technique to discriminate PD and HC patients by analyzing their speech samples of sustained phonation. Traditional approaches have focused on experimenting with handcrafted spectral and temporal features. In this paper, however, we focus on *machine learning* the discriminating features of speech associated with PD patients and healthy control from the raw IMF signals. We train a 1D-CNN model using these raw IMFs to learn the discriminating properties in the signals to classify PD and HC subjects.

References

[1] M. Hoehn, M. Yahr, Parkinsonism: onset, progression and mortality, Neurology 17 (1967) 427–442.

[2] O. Tysnes, A. Storstein, Epidemiology of parkinson's disease, Journal of Neural Transmission 124 (2017) 901–905. doi:10.1007/s00702-017-1686-y.

[3] A. K. Ho, R. Iansek, M. C., B. J.L., G. S., Speech impairment in a large sample of patients with parkinson's disease, Behavioral Neurology 11 (1998/1999) 131–137.

[4] S. Fahn, R. L. Elton, Unified parkinsons disease rating scale, Recent Developments in Parkinsons Disease,Macmillan Health Care Information 2 (1987) 153–163.

[5] H. Cohen, Disorders of speech and language in parkinson's disease, Mental and Behavioral Dysfunction in Movement Disorders, M. A. Bédard,Y. Agid, A. D. Korczyn, P. Lesperance, and S. Chouinard, Eds. New York,NY, USA: Humana Press, (2003) 125–134.

[6] A. K. Ho, R. Iansek, M. C., B. J.L., G. S., Motor instability in parkinsonian speech intensity, Neuropsychiatry, Neuropsychology and Behavioral Neurology 14 (2001) 109–116.

[7] M. A. Little *, P. E. McSharry, E. J. Hunter, J. Spielman, L. O. Ramig, Suitability of dysphonia measurements for telemonitoring of parkinson's disease, IEEE Transactions on Biomedical Engineering 56 (2009) 1015–1022.

[8] M. Novotný, J. Rusz, R. Čmejla, E. Růžička, Automatic evaluation of articulatory disorders in parkinson's disease, IEEE/ACM Transactions on Audio, Speech, and Language Processing 22 (2014) 1366–1378.

[9] J. Rusz, R. Cmejla, Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease, Journal of Acoustical Society of America 129 (2011) 350.

[10] J. Rusz, R. Cmejla, Imprecise vowel articulation as a potential early marker of parkinson's disease: Effect of speaking task, Journal of Acoustical Society of America 134 (2013) 2171.

[11] A. Rueda, J. Vásquez-Correa, C. D. Rios-Urrego, J. R. Orozco-Arroyave, S. Krishnan, E. Noeth, Feature Representation of Pathophysiology of Parkinsonian Dysarthria, in: Proc. Interspeech 2019, 2019, pp. 3048–3052. URL: http://dx.doi.org/10.21437/Interspeech.2019-2490. doi:10.21437/Interspeech.2019-2490.

[12] J. V. E. López, J. R. Orozco-Arroyave, G. Gosztolya, Assessing Parkinson's Disease from Speech Using Fisher Vectors, in: Proc. Interspeech 2019, 2019, pp. 3063–3067. URL: http://dx.doi.org/10.21437/Interspeech.2019-2217. doi:10.21437/Interspeech.2019-2217.

[13] Y. Hauptman, R. Aloni-Lavi, I. Lapidot, T. Gurevich, Y. Manor, S. Naor, N. Diamant, I. Opher, Identifying Distinctive Acoustic and Spectral Features in Parkinson's Disease, in: Proc. Interspeech 2019, 2019, pp. 2498–2502. URL: http://dx.doi.org/10.21437/Interspeech.2019-2465. doi:10.21437/Interspeech.2019-2465.

[14] G. Dimauro, D. Caivano, V. Bevilacqua, F. Girardi, V. Napoletano, Voxtester, software for digital evaluation of speech changes in parkinson disease, in: 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2016, pp. 1–6.

[15] G. Rilling, P. Flandrin, P. Goncalves, Empirical mode decomposition, fractional gaussian noise and hurst exponent estimation, in: Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., volume 4, 2005, pp. iv/489–iv/492 Vol. 4.

[16] A. Bouzid, N. Ellouze, Voiced speech analysis by empiricalmode decompositio, Advances in Nonlinear Speech Pro-cessing, Springer (2007).

[17] B. Mijović, M. Silva, V. den B. R. H. Bergh, K. Allegaert, J. M. Aerts, D. Berckmans, V. S. Huffel, Assessment of pain expression in infant cry signals using empirical mode decomposition., Methods Inf Med 49(05) (2010).

[18] L. Xiang, X. L., Speech emotion recognition using novel hht-teo based features, Journal of Computers 6 (2011).

[19] R. Sharma, S. R. Mahadeva Prasanna, Characterizing glottal activity from speech using empirical mode decomposition, in: 2015 Twenty First National Conference on Communications (NCC), 2015, pp. 1–6.

[20] G. D. F. Girardi, Italian parkinson's voice and speech, 2019. URL: http://dx.doi.org/10.21227/aw6b-tg17. doi:10.21227/aw6b-tg17.

[21] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, F. Girardi, Assessment of speech intelligibility in parkinson's disease using a speech-to-text system, IEEE Access 5 (2017) 22199–22208.

[22] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, H. H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454 (1998) 903–995. doi:10.1098/rspa.1998.0193.

[23] R. Sharma, L. Vignolo, G. Schlotthauer, M. Colominas, H. L. Rufiner, S. Prasanna, Empirical mode decomposition for adaptive am-fm analysis of speech: A review, Speech Communication 88 (2017) 39 – 64. URL: http://www.sciencedirect.com/science/article/pii/S0167639316302370. doi:https://doi.org/10.1016/j.specom.2016.12.004.

[24] M. Kaleem, B. Ghoraani, A. Guergachi, S. Krishnan, Pathological speech signal analysis and classification using empirical mode decomposition, Med Biol Eng Comput 51 (2013).

[25] F. Chollet, et al., Keras, https://keras.io, 2015.

[26] M. Abadi, et al., TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/, software available from tensorflow.org.

[27] N. Garcia, J. C. Vásquez Correa, J. R. Orozco-Arroyave, E. Nöth, Multimodal i-vectors to detect and evaluate parkinson's disease, in: Proc. Interspeech 2018, 2018, pp. 2349–2353. URL: http://dx.doi.org/10.21437/Interspeech.2018-2295. doi:10.21437/Interspeech.2018-2295.