Exploiting voice signal decomposition in expert system for Parkinson's disease detection

Aivaras Šimulis
Department of Information Systems
Kaunas University of Technology
Kaunas, Lithuania
e-mail: aivaras.simulis@ktu.edu

Evaldas Vaičiukynas
Department of Information Systems
Kaunas University of Technology
Kaunas, Lithuania
e-mail: evaldas.vaiciukynas@ktu.lt

***Abstract*—The goal of this research is the robust detection of Parkinson's disease by acoustic analysis of sustained voice recordings. Application of signal decomposition into intrinsic mode functions (IMFs) is investigated as a novel type of audio features and a custom solution for decision-level fusion, employing statistical functionals to compress decisions from all IMFs. Proposed audio features are perceptual linear predictive cepstral coefficients (PLPCCs) estimated on the extracted components from one or several equally-spaced windows of audio signal. Decompositions used are empirical mode decomposition (EMD) and variational mode decomposition (VMD). Random forest (RF) is used as a base detector as well as meta learner for decision-level fusion. Cost of log-likelihood ratio and equal error rate (EER) were used to measure goodness-of-detection. Baseline solution using PLPCCs from all frames was compared to several types of decision-level fusion (EMD, VMD, and EMD+VMD) using 1–3 windows of various sizes (10–100 ms). Decomposition-based PLPCCs and EMD+VMD fusion from three 30 ms sized windows resulted in detection performance with an average EER of 6.5% and clearly outperformed the baseline solution of decomposition-less PLPCCs, having EER of 32.9%. Variable importance from meta RF found both decompositions as useful and variance-related statistics of base RF decisions as the most important.**

Keywords*— *Parkinson's disease; voice analysis; empirical mode decomposition; variational mode decomposition; PLPCC; random forest; medical decision support

I. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease after Alzheimer's [1] and it is expected that the prevalence of PD is going to increase due to population ageing. If it is detected early, the progression of PD may be more researched using neuroprotective strategies, which could result in increased life span and improved living conditions.

Amongst many various symptoms, PD induces speech disorders, which can be observed as early as 5 years before the diagnosis [2]. Investigations show that Parkinsonian vocal dysfunction can be characterized by: reduced vocal tract volume and reduced tongue flexibility, significantly narrower pitch range, longer pauses and smaller variations in pitch range, voice intensity level, and articulation rate. Therefore, acoustic analysis is considered by many researchers as an important non-invasive approach for PD detection. Detailed review of the related work can be found in [3].

Most of the studies use audio features obtained from the entire record (global attributes) directly or calculated from short-term frames (local attributes). Frame-based features usually are compressed with the statistical functionals or the Gaussian mixture model [3]. Some studies use large feature sets aiming to obtain comprehensive characterization of the voice signal, while others rely on "clinically useful" set of measures or perform feature selection to collect a compact set of audio descriptors. In this work, we investigate if applying signal decomposition of a few short-term frames of a sustained voice recording and calculating perceptual linear predictive cepstral coefficients (PLPCCs) for extracted components can outperform a simple decomposition-less approach of using PLPCCs from all frames.

Small size of previously used databases and data samples (less than 60 PD subjects) is a major deficiency resulting in unreliable estimates of reported performance. Incorrect assessment of accuracy is another common deficiency – studies often lack conformity to leave-one-subject-out [3] or leave-one-individual-out [4] validation scheme. The need for such scheme arises when subject has several recordings, where all recordings of a subject should be included either in a training or in a testing fold, but not in both.

This work explores application of two techniques for voice signal decomposition into intrinsic mode functions, namely, empirical mode decomposition (EMD) and variational mode decomposition (VMD), introduces a novel decision-level fusion approach using EMD and VMD base detectors and addresses the aforementioned deficiencies by using relatively large database and leave-one-subject-out validation for a task of PD detection.

The organization of this paper is as follows: voice recordings database is described in section 2, feature extraction in section 3, detection methodology is presented by section 4, results of experiments are in section 5, while conclusions are drawn in section 6.

II. Feature Extraction

Audio features of choice were PLPCC, which were extracted either from all frames of recording and compressed using statistical functionals (baseline solution) or from a few frames after applying decomposition and extracting IMFs (researched solution). Two approaches for decomposition were considered – empirical mode decomposition (EMD) and variational mode decomposition (VMD). Therefore, the researched solution proposes EMD-PLPCC and VMD-PLPCC as audio descriptors.

A. Signal decomposition

Empirical mode decomposition is a recent method for non-stationary signal analysis which has found extensive application in many areas of science and engineering [5, 6]. EMD decomposes signals into functions of time, from which spectral information may also be obtained. Therefore, EMD lends itself well to extraction of temporal as well as spectral descriptors from the original signal [7]. Application of EMD to analysis of speech signals is scarce. It is an adaptive technique that allows decomposition of non-linear and non-stationary data into intrinsic mode functions. An intrinsic mode function (IMF) satisfies the following two conditions [8]:

1. The number of maxima, which are strictly positive, and the number of minima, which are strictly negative, for each IMF, are either equal, or differ at most by one.
2. The mean value of the envelope, as defined by the maxima and the minima, for each IMF, is zero.

The technique for decomposition of the data into IMFs is known as sifting, a brief description of which follows:

1. For a given discrete time signal $x[n]$, all the local minima and maxima of $x[n]$ are identified.
2. The upper envelope E_U is calculated by using a cubic spline to connect all the local maxima. Similarly, the lower envelope E_L is calculated from the local minima. The upper and lower envelopes should cover all the data in $x[n]$ between them.
3. The mean $E_{mean} = E_U + E_L/2$ of the upper and lower envelopes is calculated, and $x[n]$ is updated by subtracting the mean from it $x[n] \leftarrow x[n] - E_{mean}$.
4. The previous three steps are executed till $x[n]$ is reduced to an IMF $c_1[n]$, which conforms to the properties of IMFs described previously. The first IMF contains the highest oscillation frequencies found in the original data $x[n]$.
5. The first IMF $c_1[n]$ is subtracted from $x[n]$ to get the residue $r_1[n]$.
6. The residue $r_1[n]$ is now taken as the starting point instead of $x[n]$, and the previously mentioned steps are repeated to find all the IMFs $c_1[n]$ so that the final residue r_K either becomes a constant, a monotonic function, or a function with a single maximum and minimum from which no further IMF can be extracted.

Therefore, at the end of the decomposition, we can represent $x[n]$ as the sum of K IMFs and a residue r_K:

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

The EMD algorithm variant used is ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [7], which automatically selects the number of intrinsic mode functions (IMFs). According to the selected number of IMFs for EMD, the same number of IMFs was used in the VMD algorithm.

Variational mode decomposition [9] decomposes the signal into various modes or intrinsic mode functions using calculus of variation. Each mode of the signal is assumed to have compact frequency support around a central frequency. VMD tries to find out these central frequencies and intrinsic mode functions centered on those frequencies concurrently using an optimization methodology called alternate direction method of multipliers (ADMM13). The original formulation of the optimization problem is continuous in time domain. The constrained formulation is given in [9].

B. PLPCC

The idea of a perceptual front end for determining linear prediction cepstral coefficients (PLPCCs) has been applied in different ways to improve speech detection and coding, as well as noise reduction, reverberation suppression, and echo cancellation. Linear prediction of a signal is done via autoregressive moving average (ARMA) modeling of the time series. In an ARMA model, the current sample is expressed as:

$$y[n] = \sum_{p=0}^{P} a_p x[n-p] + \sum_{q=1}^{Q} b_q y[n-q] \tag{2}$$

where $x[n]$ is the current input signal, and $y[n]$ is the current output. The perceptual linear prediction coefficients are created from the linear prediction coefficients by performing perceptual processing before performing the autoregressive modeling [10]. The main sequence of steps for PLPCC calculation is provided by a diagram in Fig. 1.

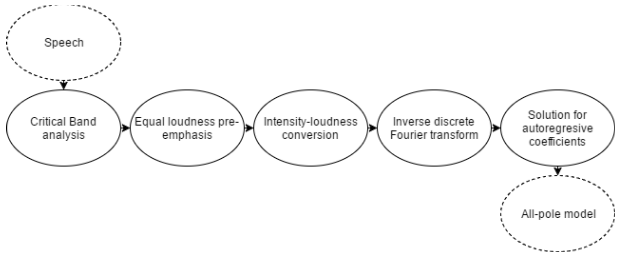

Fig. 1. Calculating perceptual linear prediction cepstral coefficients.

After this processing, cepstral conversion is performed. This is because linear prediction coefficients are very sensitive to frame synchronization and numerical error. In other words, the linear prediction cepstral coefficients are much more stable than the linear prediction coefficients themselves [10].

For a baseline solution, PLPCCs extracted from all frames were compressed using the following statistical functionals: min, max, mean, median, trimean, standard deviation, inter-quartile range (IQR), lower quartile (Q_{lo}), upper quartile (Q_{up}), lower range (IQ_{lo}), upper range (IQ_{up}), skewness, and kurtosis.

For a researched solution, only one or a few frames were used, and PLPCCs were calculated both for unprocessed frame and for each of IMFs after signal decomposition. Depending on the number of IMFs the same number of feature vectors, containing concatenation of original PLPCCs with IMF-based PLPCCs, are constructed.

III. Voice Database

Voice database had 383 speakers (141 men and 242 women), where each speaker was represented by 3 recordings (419 recordings for men and 715 recordings for women) of sustained voicing of vowel /a/, each at least 2 seconds in length. Speaker age range was from 16 to 82 for HC and from 39 to 85 for PD. Detailed summary of voice recording database is in Table I. Recordings were collected in a sound-proof chamber using acoustic cardioid microphone (AKG Perception 220) placed at a distance of 10 cm from the mouth. Audio format was ".wav" (16-bit mono PCM with 44.1 kHz sampling frequency). HC voice subgroup encompassed healthy volunteer individuals who considered their voice as normal, had no complaints concerning their voice and no history of chronic laryngeal diseases or other long-lasting voice disorders. Voices of these individuals were also confirmed as healthy by clinical specialists. Furthermore, no pathological alterations in the larynx of the subjects from HC voice subgroup were found during video laryngostroboscopy.

TABLE I. SUMMARY OF VOICE RECORDINGS DATABASE

Recordings	**Parkinson (PD)**	**Healthy (HC)**	**Total**
Men	36 (107)	105 (312)	141 (419)
Women	39 (116)	203 (599)	242 (715)
Total:	75 (223)	308 (911)	383 (1134)

IV. Detection Methodology

A. Random Forest

Random forest (RF) [11] is a well-known pattern recognition algorithm, suitable for a detection task. RF is a committee of decision trees, where the final decision is obtained by majority voting. The core idea of RF is to combine many (B in total) decision trees, built using different bootstrap samples of the original data set, and a random subset (of predetermined size q) of features $x^1, \dots, x^p$. RF is known to be robust against over-fitting and as the number of trees increases, the generalization error converges to a limit [11]. For our experiments B was set to 5000, several specific values of q ($\sqrt{p}$, $2 \cdot \sqrt{p}$, $\frac{1}{2} \cdot \sqrt{p}$) were tested and the best performing q setting retained.

The generalization performance of RF was evaluated using internal out-of-bag (OOB) validation, where each observation is classified only by the trees which did not have this observation in bootstrap sample during construction. It is well known that OOB validation provides an unbiased estimate of a test set error, similar to leave-one-out scheme. Because of the "repeated measures" aspect in voice data, where each subject is represented by several recordings of sustained vowel, sampling part of the RF algorithm [12] had to be modified to ensure that all recordings of each subject are either included in a bootstrap sample or left aside as OOB. Such modification corresponds to leave-one-subject-out scheme, which helps to avoid speaker detection intermingling with pathology detection. Additionally, RF setting of stratified sampling was configured to preserve class ratio and gender balance of the full dataset in each drawn bootstrap sample.

B. Decision-level Fusion

Individual RFs were built independently for each decomposition variant (EMD or VMD) and each frame location (center frame only or 3 frames in equally-spaced locations) and decisions of these individual experts were combined in a meta-learner fashion. RF was used both as a base learner and as a meta learner. This implies that outputs from RF models from the first stage after compression of decisions for all IMFs of a single recording using statistical functionals are treated as inputs (meta-features) for another RF in the second stage.

For the detection task, a decision from base RF is the difference between class posteriori. Given a trained RF, this difference or variant of soft decision is estimated as:

$$d(t_1, \dots, t_L) = \frac{\sum_{i=1}^{L} f(t_i, x, q = 2)}{L} - \frac{\sum_{i=1}^{L} f(t_i, x, q = 1)}{L}, \qquad (3)$$

where x is the object being classified, L is the number of trees $t_2, \dots, t_L$ in the RF for which observation x is OOB, q is a class label (label number 1 corresponds to HC and 2 to PD), and $f(t_i, x, q)$ stands for the qth class frequency in the leaf node, into which x falls in the ith tree t_i of the forest:

$$f(t_i, x, q) = \frac{n(t_i, x, q)}{\sum_{j=1}^{Q} n(t_i, x, q)}, \qquad (4)$$

where Q is the number of classes and $n(t_i\ x, q)$ is the number of training data from class q and falling into the same leaf node of t_i as x.

For a baseline solution, there was no need of decision-level fusion and a single base RF was enough. For a researched solution, decision-level fusion used decisions from the decomposition-based base RFs. Base RF was constructed using all components (IMFs) resulting from EMD or VMD in a specific frame. The number of decisions from a base RF corresponded to the number of extracted components. This varied number of base decisions was compressed into meta-features using the following statistical functionals: min, max, mean, median, trimean, standard deviation, inter-quartile range (IQR), lower quartile (Q_{lo}), upper quartile (Q_{up}), lower range (IQ_{lo}), upper range (IQ_{up}), skewness, and kurtosis.

Meta-features in fusion RF were also investigated by performing permutation-based variable importance analysis using mean decrease in accuracy as the variable importance measure. Values of each meta-feature are permuted several times and the mean difference in fusion RF performance on OOB data is estimated.

C. Assesing Detection

To evaluate the goodness of detection, detector's scores for OOB data were used. Votes of RF were converted to a proper score vector by normalizing votes for a specific class through division by the total number of times the case was OOB, as in formula (3). A quick way to compare detectors is the equal error rate (EER). The cost of log-likelihood-ratio (C_{llr}) is a comprehensive detection metric, used here as the main criterion for model selection. The log-likelihood-ratio is the logarithm of the ratio between the likelihood that the target (PD) produced the signal and the likelihood that a non-target (HC) produced the signal. EER and C_{llr} measures were estimated using the ROC convex hull method, available in the BOSARIS toolkit [13]. A well-calibrated and useful detector should have $C_{llr} < 1$ and EER < 50 %.

V. Experiments

A. Experimental Setup

For our frame-based features several window sizes were tested: 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 milliseconds. For a researched solution, different quantity of windows were extracted and detection performance compared: 1 window from the center of recording versus 2 or 3 windows placed at equally-spaced locations to evenly cover all recording. For each setting of window size and quantity, three decision-level fusion variants were tested: EMD, VMD and EMD+VMD. The number PLPCCs used was 12 in both baseline and researched solutions.

From the results of initial experiments, the single best window size was selected. Sensitivity analysis was then performed by repeating decomposition for each recording 5 times. Each of the 5 different collections of IMFs, converted into EMD-PLPCCs or VMD-PLPCCs, were used in base RF and construction of base RF repeated 5 times. Finally, each of 25 results from decomposition and base RF runs was fused using meta RF and the fusion was repeated 5 times. All runs resulted in 125 detection performance measures (EER and C_{llr}) and statistical testing for equality of central tendencies (mean and median) was performed to compare the setting of using 1 central window only versus using 3 windows at equally-spaced locations. Parametric independent samples t-test was used to compare means. Non-parametric Wilcoxon rank-sum test (Mann-Whitney U-test) was used to compare medians.

B. Results

From the results of initial experiment with various window sizes in Table II we decide that 30 ms window size, 1 or 3 windows and fusion of EMD+VMD could be a recommended setting for a researched solution. Both window sizes had same EER of 2.22%, but C_{llr} was slightly better for 3 windows (0.079) than 1 window (0.085). VMD consistently outperformed EMD, but the best overall performance was found when fusing EMD+VMD. The best window size was difficult to identify, because the results change from one window size to another sometimes rather erratically. This could be due to stochastic nature of EMD and VMD, where optimization results in sub-optimal solutions and not identical IMFs when run repeatedly.

TABLE II. Detection performance using various window sizes, window quantities and type of decomposition for fusion by meta RF.

Size (s)	**Quantity**	**Type of decomposition used in decision-level fusion**					
		EMD		***VMD***		***EMD+VMD***	
		Cllr	***EER***	***Cllr***	***EER***	***Cllr***	***EER***
0.01	1	0.718	26.23	0.237	6.65	0.215	6.10
	2	0.617	20.83	0.754	26.20	0.565	19.12
	3	0.115	3.03	0.245	6.71	0.079	2.24
0.02	1	0.448	13.96	0.682	23.03	0.375	12.42
	2	0.385	13.08	0.610	20.32	0.336	10.45
	3	0.321	9.70	0.499	16.66	0.255	7.72
0.03	1	0.456	13.92	***0.100***	***2.65***	0.085	***2.22***
	2	0.306	10.07	0.503	14.95	0.264	8.27
	3	0.342	11.32	0.115	3.33	0.079	***2.22***
0.04	1	0.419	11.77	0.586	19.08	0.359	11.51
	2	0.379	11.69	0.541	17.01	0.313	10.05
	3	***0.010***	***2.90***	0.451	13.70	0.100	2.87
0.05	1	0.355	10.71	0.583	19.02	0.303	8.53
	2	0.347	10.92	0.539	16.67	0.300	9.73
	3	0.216	5.80	0.449	13.36	0.180	5.37
0.06	1	0.316	10.08	0.555	17.16	0.266	7.88
	2	0.328	10.07	0.441	14.23	0.265	8.41
	3	0.244	7.39	0.419	13.24	0.220	6.54
0.07	1	0.405	11.84	0.561	17.15	0.349	10.61
	2	0.450	14.68	0.474	15.48	0.380	12.73
	3	0.236	6.07	0.356	12.15	0.190	5.40
0.08	1	0.379	12.09	0.529	16.85	0.309	9.93
	2	0.323	11.27	0.505	15.91	0.288	9.78
	3	0.297	9.87	0.448	14.11	0.275	9.13
0.09	1	0.443	13.34	0.591	19.59	0.368	11.24
	2	0.349	10.56	0.504	16.10	0.307	9.84
	3	0.317	10.11	0.458	13.90	0.278	9.28
0.10	1	0.363	10.78	0.495	16.21	0.252	7.66
	2	0.390	12.17	0.509	15.64	0.359	12.07
	3	0.254	7.21	0.109	3.86	***0.074***	2.73

The summary of results from Table II for 1 and 3 windows is illustrated in Fig 2–3. Both settings of window quantity were able to provide the lowest EER of 2.2% with 30 ms window size, but performance when using 3 windows appear to be more stable and doesn't fluctuate that much with respect to window sizes.

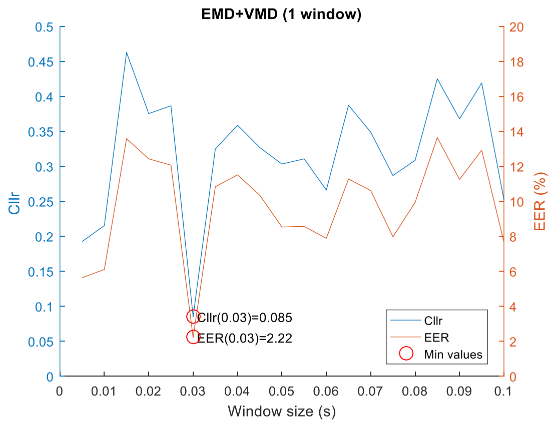

Fig. 2. EMD+VMD detection performance by EER and C_{llr} using 1 window.

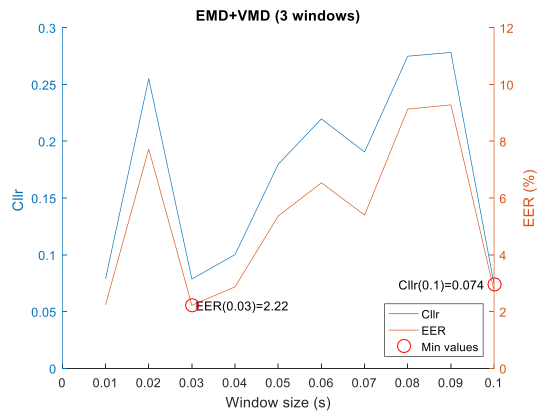

Fig. 3. EMD+VMD detection performance by EER and C_{llr} using 3 windows.

Results of the baseline solution (see Fig. 4), when extracting PLPCCs from all frames (without any decomposition) and compressing with statistical functionals, were rather stable irrespective of the window size. The best detection performance was 32.9% by EER and 0.895 by C_{llr}.

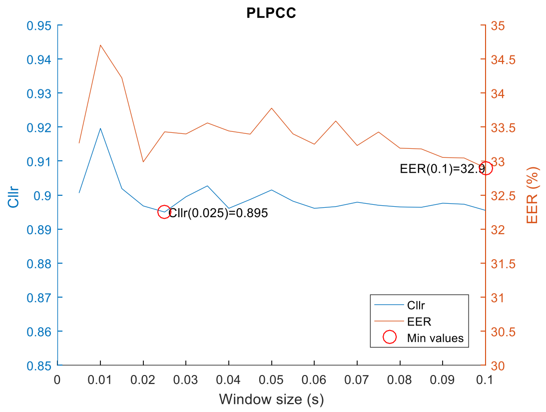

Fig. 4. Detection performance by EER and C_{llr} for the baseline solution.

Variable importance from the meta RF when using 3 windows and EMD+VMD type of decision-level fusion is shown in Fig 5. Both EMD and VMD types of decomposition are useful, but from statistical functionals variance-related measures (standard deviation and inter-quartile range with its lower part) of base decisions appear to be the most important for the detection.

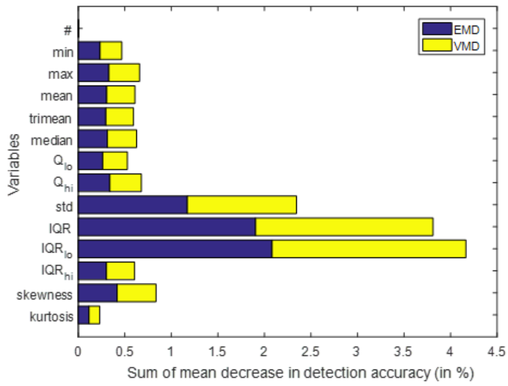

Fig. 5. Variable importance from fusion RF (EMD+VMD) using 3 windows.

Sensitivity analysis was performed by repeating detection task using 30 ms window size and EMD+VMD fusion. Distribution of the resulting C_{llr} and EER measures is illustrated by boxplots in Fig 6 – 7.

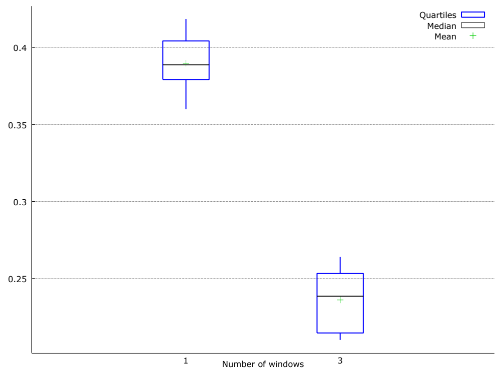

Fig. 6. Distribution of C_{llr} after 125 repetitions of the detection task, visualized by boxplot: 1 (***left***) vs 3 windows (***right***).

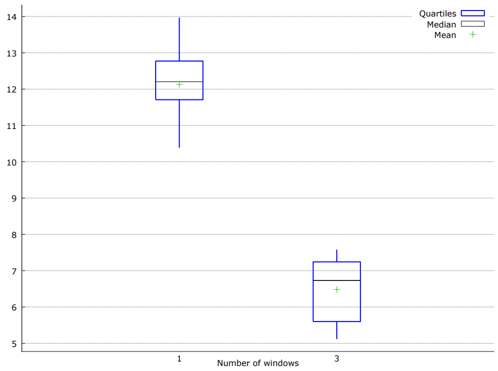

Fig. 7. Distribution of EER after 125 repetitions of the detection task, visualized by boxplot: 1 (***left***) vs 3 windows (***right***).

TABLE III. TEST OF CLLR FOR NORMALITY OF 1 AND 3 WINDOWS

Test name	**Test statistic**		**p-value**	
	1 window	**3 windows**	**1 window**	**3 windows**
Doornik-Hansen	12.3997	34.4653	0.00202975	3.28061e-008
Shapiro-Wilk W	0.926049	0.863831	3.66809e-006	2.39331e-009
Lilliefors	0.119188	0.197622	~= 0	~= 0
Jarque-Bera	7.1929	13.3053	0.0274209	0.00129062

TABLE IV. TEST OF EER FOR NORMALITY OF 1 AND 3 WINDOWS

Test name	**Test statistic**		**p-value**	
	1 window	**3 windows**	**1 window**	**3 windows**
Doornik-Hansen	11.2948	52.6303	0.00352675	3.72794e-012
Shapiro-Wilk W	0.952697	0.840665	0.00025031	2.69682e-010
Lilliefors	0.0835316	0.204496	~= 0.03	~= 0
Jarque-Bera	6.0258	15.6308	0.049149	0.00040348

Due to the lack of normality in C_{llr} and EER measures (see Tables III – IV), non-parametric statistical testing was performed. Medians of C_{llr} and EER were compared between 1 window and 3 windows setting using Wilcoxon rank-sum test and results are provided in Table V. Statistically significant difference with 99% confidence (p-value < 0.01) was indicated in favor of using 3 windows.

TABLE V. HYPOTHESIS OF EQUAL CENTRAL TENDENCIES TESTING FOR C_{LLR} AND EER USING WILCOXON RANK-SUM TEST

	Rank-sum test results for C_{llr}
Null hypothesis	The two medians are equal
Cllr statistics	n1 = 125, n2 = 125 w (sum of ranks, sample 1) = 23500 z = (23500 - 15687.5) / 571.684 = 13.6658 P(Z > 13.6658) = 0, Two-tailed **p-value = 0**
EER statistics	n1 = 125, n2 = 125 w (sum of ranks, sample 1) = 23500 z = (23500 - 15687.5) / 571.684 = 13.6658 P(Z > 13.6658) = 0, Two-tailed **p-value = 0**

VI. CONCLUSIONS

Decomposition-based PLPCC features, namely, EMD-PLPCC and VMD-PLPCC, were found useful for building expert system for Parkinson's detection from sustained voice. The baseline solution without decomposition and decision-level fusion, where PLPCCs were obtained from all frames and compressed using statistical functionals, resulted in the lowest EER of 32.90% for 30 ms windows size. The researched solution using decision-level fusion of PLPCCs from EMD and VMD, obtained from 3 windows at equally-spaced locations, resulted in EER as low as 2.22% for 30 ms window size.

Sensitivity analysis was performed by fixing window size at 30 ms and repeating detection task 125 times to choose between fusion of components from 1 or 3 windows. Fusion of EMD+VMD from 1 window provided an average EER of 12.13%, whereas fusion of EMD+VMD from 3 windows provided an average EER of 6.48%. Detection performance when using decision-level fusion of decomposition results from 3 evenly located windows was better than using 1 central window and the difference in detection performance was statistically significant, as indicated by statistical tests.

Main challenge remains lack of numerical stability in decomposition output, where extracted IMFs were not identical between runs. This limitation leaves the choice of the window size questionable, but future work could exploit a multitude of sub-optimal IMFs and use them for boosting data amounts when building detectors. Not only the first stage detectors could be built on more data, but also statistical functionals when compressing decisions of base detectors for the second stage would certainly benefit from increased number of decisions – statistics for decision-level fusion would be obtained not from 6–12 components, as in a current solution, but from several times more. We speculate that meta-leaner in the second stage could achieve robustness by using all IMFs of several repeated decomposition runs.

ACKNOWLEDGMENT

Voice database was collected at Lithuanian University of Health Sciences (LUHS) under a grant (No. MIP-075/2015) from the Research Council of Lithuania. Authors would like to thank LUHS specialists – otorhinolaryngologist Evaldas Padervinskis and neurologist Jolita Čičelienė.

REFERENCES

[1] M. C. de Rijk, L. J. Launer, K. Berger, M. M. B. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. M. Martınez-Lage, C. Trenkwalder, A. Hofman, "Prevalence of Parkinson's disease in Europe: a collaborative study of population-based cohorts," Neurology, vol. 54(11 Supply 5), S21-S23, 2000.

[2] B. Harela, M. Cannizzaro, P. J. Snyder, "Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study," Brain and Cognition, vol. 56 , pp. 24-29, 2004.

[3] J. R. Orozco-Arroyave, F. Honig, J. D. Arias-Londono, J. F. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, E. Noth, "Automatic detection of Parkinson's disease in running speech spoken in three different languages," J. Acoust. Soc. Am., vol. 139(1), pp. 481-500, 2016.

[4] C. O. Sakar, O. Kursun, "Telediagnosis of Parkinson's disease using measurements of dysphonia," J. Med. Syst., vol. 34(4), pp. 591-599, 2010.

[5] M. F. Kaleem, L. Sugavaneswaran, A. Guergachi, S. Krishnan, "Application of empirical mode decomposition and Teager energy operator to EEG signals for mental task classification," Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology (EMBC), pp. 4590-4593, 2010.

[6] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, S. Van Huffel, "Source separation from single-channel recordings by combining empirical mode decomposition and independent component analysis," IEEE Trans. Biomed. Eng., vol. 57(9), pp. 2188-2196, 2010.

[7] M. E. Torres, M. A. Colominas, G. Schlotthauer, "A complete Ensemble Empirical Mode decomposition with adaptive noise," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-11), pp. 4144-4147, 2011.

[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. Yen, C. C. Tun, H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. Lond., A 454(1971), pp. 903-995, 1998.

[9] K. Dragomiretskiy, D. Zosso, "Variational Mode Decomposition," IEEE Trans. Signal Process, vol. 62(3), pp. 531-544, 2014.

[10] H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of Speech," J. Acoust. Soc. Am., vol. 87(4), pp. 1738-1752, 1990.

[11] L. Breiman, "Random forests," Mach. Learn., vol. 45, pp. 5-32, 2001.

[12] A. Jaiantilal, "Random forest (regression, classification and clustering) implementation for Matlab (and standalone)," 2012. http://code.google.com/archive/p/randomforest-matlab/.

[13] N. Brummer, E. de Villiers, "The BOSARIS toolkit: Theory, algorithms and code for surviving the new DCF". arXiv 1304(2865v1), pp. 1–23, Presented at the NIST SRE 2011 Analysis Workshop, Atlanta, December 2011. http://sites.google.com/site/bosaristoolkit.