

Joint Application of Audio Spectral Envelope and Tonality Index in an E-Asthma Monitoring System

Marcin Wiśniewski, Tomasz P. Zieliński

Abstract— This paper presents in detail a recently introduced highly efficient method for automatic detection of asthmatic wheezing in breathing sounds. The fluctuation in the Audio Spectral Envelope (ASE) from the MPEG-7 standard and the value of the Tonality Index (TI) from the MPEG-2 Audio specification are jointly used as discriminative features for wheezy sounds, while the Support Vector Machine (SVM) with a polynomial kernel serves as a classifier. The advantages of the proposed approach are described in the paper (e.g. detecting weak wheezes, very good ROC characteristics, independence from noise color). Since the method is not computationally complex it is suitable for remote asthma monitoring using mobile devices (personal medical assistants). The main contribution of this paper consists of presenting all the implementation details concerning the proposed approach for the first time, i.e. the pseudo-code of the method, and adjusting the values of the ASE and TI parameters after which only one (not two) FFT is required for analysis of a next overlapping signal fragment. The efficiency of the method has also been additionally confirmed by the AdaBoost classifier with a built-in mechanism to feature ranking, as well as a previously performed minimal-redundancy-maximal-relevance (mRMR) test.

Index Terms—asthma, monitoring, wheezes, recognition, Discrete Fourier Transform (DFT), audio spectral envelope, tonality index.

I. INTRODUCTION

THE capability for patients to monitor conditions at home, exploiting e-health technology and personal medical assistants, is a very important and promising trend in contemporary medicine. Heart disease, diabetes and asthma are well known examples of chronic illnesses requiring permanent home monitoring over some periods of time (e.g. after intensification of disease symptoms or dangerous attacks).

Manuscript received November 7, 2013.

Manuscript revised August 1, 2014.

This work has been (partially) supported by the Polish Ministry of Science and Higher Education under the European Regional Development Fund, Grant No. POIG.01.01.02-00-045/09-00 *Future Internet Engineering*.

Marcin Wiśniewski is with the Department of Telecommunications, AGH University of Science and Technology, Kraków, Poland (e-mail: wisniewski@kt.agh.edu.pl).

Tomasz P. Zieliński is with the Department of Telecommunications, AGH University of Science and Technology, Kraków, Poland (e-mail: tzielin@agh.edu.pl).

In order to increase patient comfort and device functionality, the mobile medical devices used in the patient's native environment should be very small and characterized by low power consumption ensuring long battery life. For this reason, computational algorithms designed for mobile e-health applications should be efficient but not very computationally intensive.

This paper addresses the issue of remote monitoring of patients with asthma. Automatic recognition of asthmatic wheezes has a very long history, starting from the early 1980s [1, 2]. Many different methods relevant to this purpose have been developed and reported to the present day. Since excellent review papers and books have already been written on this subject [3, 4, 5] and are still current, only a brief introduction to this issue will be given below.

A medical introduction to lung sound classification can be found in [6, 7, 8]. The asthmatic wheeze recognition task can be summarized as detecting whether bursts of summations of pure tones are present in the breathing sound (as colored noise) or not [9]. In the past many methods for automatic wheeze recognition have been proposed. Most of these make use of a variety of frequency analysis methods for sound feature extraction (e.g. short-time Fourier transform [9], mel-frequency spectral coefficients (MFCC) [10], wavelet transform [11], or auto-regressive analysis [12]) that are combined with different tools for data *modeling* (vector quantization, Gaussian mixture models (GMM)) and *classification* (k-nearest neighbors, support vector machine (SVM) and artificial neural networks (ANN)). After testing several approaches it was indicated in [10] that superior results are ensured by combining MFCC with GMM. MFCC coefficients are used also with success in the classification of breathing movements (inspiratory and expiratory) in real-time [13].

In [14], an interesting algorithmic perspective on asthma telemonitoring by means of low-power mobile systems and wireless networks is described. In this application scenario, the computational expense of the algorithm is of great importance. This goal is addressed, for example, with the following simple wheeze detection methods using: kurtosis (K), Renyi entropy and frequency ratio f_{50}/f_{90} from [15], and spectral peak entropy (SPE) from [16].

In this paper, a new low-cost computer algorithm for automatic recognition of asthmatic wheezing in breathing sounds is proposed. The method makes use of a set of two acoustic sound features, i.e. the Audio Spectral Envelope (ASE) descriptor, adopted from the MPEG-7 multimedia standard [17], and the Tonality Index (TI) taken from the MPEG-2 Audio specification [18], and described in detail in [19, 20]. As a classifier, the Support Vector Machine (SVM) [21] with a polynomial 3rd order kernel function is used. The paper describes how, using real lung sound recordings, the new method represents a very flexible solution for mobile real-time asthma monitoring: it is very sensitive while not computationally demanding, has very good ROC characteristics, and is independent of noise color.

This paper presents the final results of the research activity already reported in the series of papers [22-28]. First, we presented state-of-the-art solutions used in e-asthma mobile systems [22]. Then we introduced an application of the Tonality Index (TI) from the MPEG-2 Audio standard for wheeze detection [23] and compared its efficiency [23, 24] with some other sound features (e.g. kurtosis in the time and frequency domains, spectral flatness, spectral peak entropy, and frequency ratio f_{50}/f_{90}). Next, we built a computer model of normal and wheezy breath that we used in our simulations [25] and tested the application of some other sound features (linear prediction error and two new tonality measures based on autocorrelation function and eigenvalue decomposition) [26]. Finally, in [27, 28] we introduced the application of an audio spectral envelope (ASE) from the MPEG-7 standard and exploited the minimal-redundancy-maximal-relevance (mRMR) method [29] to confirm the superiority of the dual feature set {ASE+TI}. Additionally in [28] we demonstrated that this set of features also guarantees wheeze recognition independently of noise color. In the present paper we provide for the first time a pseudo-code for the developed solution (in detail) in which ASE calculation is significantly simplified. Additionally, both ASE and TI features now have parameters jointly optimized (unified) for fast real-time sliding-window implementation (now only one 256-point FFT per new overlapping signal fragment is required). We also show some new results including a joint 2D scatter plot for the ASE+TI algorithm and ROC characteristics of several tested methods. Finally, we present new classification (ranking) of all methods using the AdaBoost algorithm.

The structure of the paper is as follows. In the second section the detection problem is explained in more detail. In the third section the calculation of the ASE and TI tonality measures is described. In the fourth section this method is experimentally compared with some alternative solutions. The paper ends traditionally with conclusions and references.

II. WHEEZE DETECTION PROBLEM

From the signal processing perspective, asthmatic wheeze recognition relies on detecting **bursts** of oscillatory tones (usually from 1 to 6, typically 2-4, with frequencies in the

range from 100 to 1200 Hz and lasting at least 80 milliseconds) that are embedded in additive colored noise – the breathing sound [6, 7]. The typical breath signal taken from the chest wall, as in our case, is a **colored noise** whose spectrum decreases exponentially in frequency (but can be detected near 1 kHz by sensitive microphones as well). Figures 1 and 2 present time waveforms, Fourier spectra energy $|X(f)|^2$ in decibels (computed via FFT) and spectrograms (joint energy distribution in the time-frequency domain calculated by means of a short-time Fourier transform) for a normal breath (colored, mainly low-frequency noise) and a wheezy breath (multi-tones embedded in colored noise) respectively. Our aim is to distinguish clear inspiration from wheezy. However, the situation is not always as clear as presented: normal breath can have a wider spectrum (with higher frequencies) and low-frequency tones, not only high-frequency ones, can also be present in a wheeze.

Since the background noise is colored not white, the described above recognition task cannot be solved by means of known optimal detectors [30, 31] developed for additive white Gaussian noise. Such optimal detectors are used in sonar systems [32] for finding the parameters of linear mixtures of cosines, and in nuclear magnetic spectroscopy [33] for the measurement of mixtures of exponentially damped cosines. Unfortunately, they are useless in our detection problem due to the noise color.

Origin of the proposed ASE+TI wheeze detection method can be explained as follow. Asthmatic wheezes (multi-tones) occur from time to time in breathing sounds and their frequency spectra have higher peaks at this moments. Length of the spectral envelope is sensitive to existence of such peaks, especially when the background noise floor is exponentially decaying in frequency (colored). Therefore the ASE tonality measure has larger values for a wheeze than for a normal breath. The same time spectra of tonal components (in contrary to noise) are not changing so rapidly for consecutive/overlapping signal fragments so they are predictable. Since the TI tonality descriptor measures a spectral unpredictability, after combing it with ASE we are tracking both, how much the spectrum is peaky and how much it is not varying.

III. THE PROPOSED METHOD

The application of an audio spectral envelope (ASE) [17] for wheeze detection was proposed for the first time in [27, 28]. ASE calculation relies on: 1) extracting an m -th overlapping signal fragment $x_m(n)=x(m \cdot M+n)$, $n=0,1,\dots,N-1$ (where N , M denote observation window length and shift respectively), 2) multiplying it with a window function $w(n)$: $x_m(n) \leftarrow x_m(n) \cdot w(n)$ (in our case with a Kaiser window with shape parameter $\beta = 7$, chosen experimentally for maximization of recognition accuracy), 3) calculation of the discrete Fourier transform (DFT) $X_m(k)$, $k=0,1,2,\dots,N-1$, and the power spectrum $P_m(k)=|X_m(k)|^2$ of the windowed signal, 4) performing a weighted summation of the $P_m(k)$ over k in predefined frequency bands b :

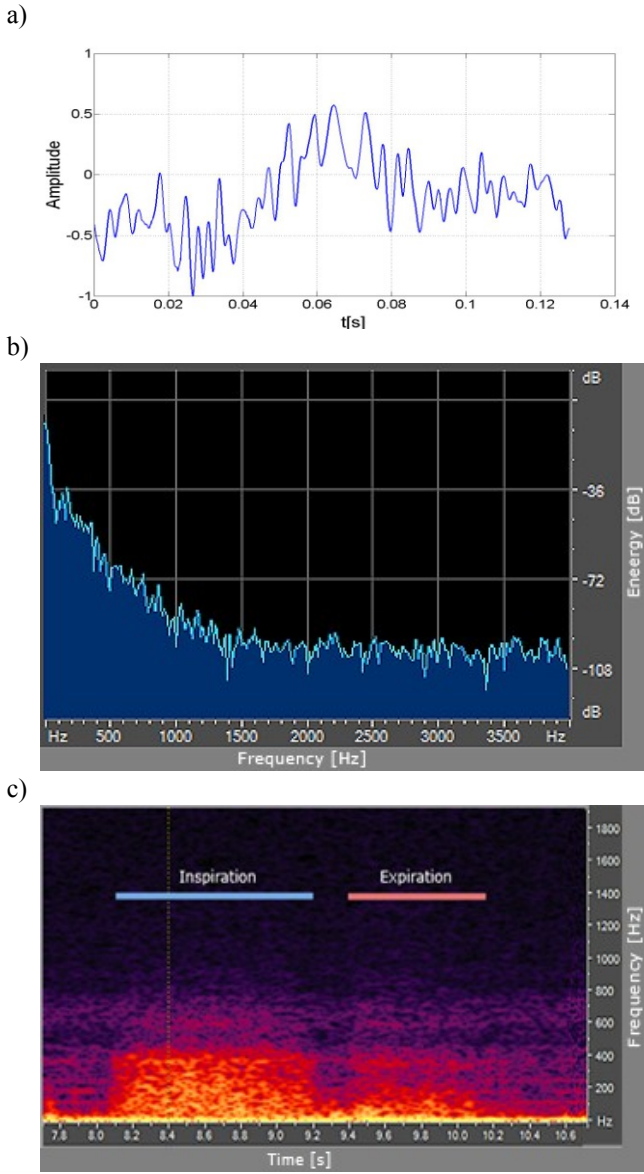


Fig. 1. Sound of a normal breath: a) time waveform, b) energy spectrum, c) spectrogram (i.e. change of energy spectrum in time – brighter colors represent higher energy values).

$$ASE_m(b) = \sum_{k \text{ for given } b} P_m(k), \quad b=0,1,2,\dots,B \quad (1)$$

The bands are logarithmically distributed between a predefined low f_{low} and high f_{high} frequency edge with the chosen resolution r being equal to the number of bands per octave (e.g. 8, 16, 32):

$$f_b = f_{low} 2^{b\Delta}, \quad \Delta = 1/r, \quad b=0,1,2,\dots,r \cdot \log_2(f_{high}/f_{low}) \quad (2)$$

In our previous papers [27, 28], lung sounds were sampled with frequency $f_{sample}=8000$ Hz, DFT had length $N=256$ samples, and window shift equaled $M=64$ samples. In consequence, the DFT frequency resolution was equal to $df=f_{sample}/N=31.25$ Hz, and the DFT spectral bins from $k=4$ (k_{min}) to $k=38$ (k_{max}) covered the frequency range of interest [125-1187.5Hz] during the asthmatic wheeze detection.

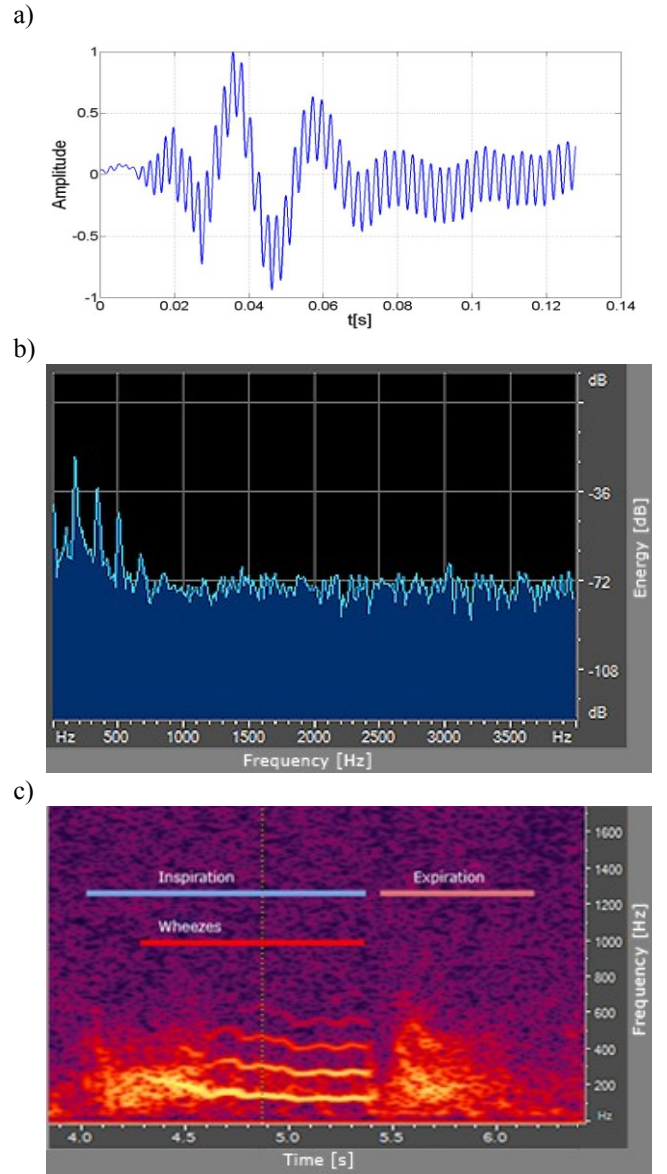


Fig. 2. Sound of a wheezy breath: a) time waveform, b) energy spectrum, c) spectrogram (i.e. change of energy spectrum in time – brighter colors represent higher energy values).

In this study the following values for the ASE parameters were used (according to ISO/IEC defaults [17]):

$$f_{low} = 62.5 \text{ Hz}, \quad f_{high} = 4000 \text{ Hz}, \quad r = 32 \text{ bands per octave.}$$

In consequence, equation (1) takes the very simple form $ASE_m(k) = P_m(k)$ (with no summation) because inside the range [100-1200] Hz of wheeze presence only one frequency DFT bin lay in each frequency band (2) in the range $f_b^{min} = 106.5$ Hz ($b=23$) and $f_b^{max} = 1189.2$ Hz ($b=137$). For this reason, the ASE calculation turned out to be easy and therefore appropriate for low-power implementation on a mobile device.

The ASE-based algorithm for detection of signal tonality is summarized in table I. From power spectra $P_m(k)$, calculated by means of DFT/FFT, only bins from $k_{min}=4$ to $k_{max}=38$ are taken, next these spectra fragments are locally averaged in time ($L_{ASE}=13$) and normalized into the range [0, 1]. Finally,

the spectrum fluctuation $fluctASE_m$ is calculated, which serves next as a local tonality measure (descriptor). The $fluctASE_m$ takes significantly smaller values for normal breaths than for wheezes due to peaks existing in frequency spectra of wheeze sounds.

During numerous experiments performed on real as well as on modeled (synthesized) data it turned out that the proposed wheeze detector based on ASE is an absolute winner (among front-end simple approaches) in the situation where multi-tones (>1) to be found are embedded in colored (not white!) noise with a strongly exponentially damped spectrum (see fig. 1b) – this is the case for asthmatic wheezes. When the background noise spectrum is close to white, the superiority of the ASE detector vanishes. It was also observed that the ASE works less effectively for single-tone wheezes.

With the goal of minimizing the aforementioned drawbacks of the separate ASE usage and increasing the field of application of the ASE wheeze detector, we proposed in [27, 28] combining the ASE with the tonality index (TI) descriptor originating from the MPEG audio standard [18, 19, 20]. The optimality of the choice of {ASE+TI} was confirmed in [27] by the minimal-redundancy-maximal-relevance- (mRMR) test [29]. Since the use of the TI in asthmatic wheeze detection has been widely reported in our previous papers [23, 24, 26], only the final method's equations are given in table I. This descriptor also makes use of the signal DFT/FFT spectrum, but now the spectrum unpredictability is of interest (which is high for a normal noisy breath and small for a tone-like wheeze). One can observe that now, contrary to the standard [18]: 1) no psychoacoustically motivated spreading function is used, 2) all DFT-based parameters are only calculated for spectral lags changing from k_{min} to k_{max} , 3) energy and weighted spectral unpredictability are only found in this range and, next, 4) averaged for last $L_{TI}=7$ signal frames.

It is important to stress that in table I a new version of the method is presented in which ASE and TI both exploit the results of the same $N=256$ -point FFT (only one, not two), which significantly reduces the computational cost (expense). Previously in the TI calculation we used the following parameter values [23-28]: $N=512$, $M=64$, $k_{min}=6$, $k_{max}=83$, $L_{TI}=7$ and Kaiser window with $\beta=12$. Comparison of computational costs between the old and new methods for real-time sliding-window implementation is performed in table II, where the number of required real-value arithmetic operations is given. We can conclude that the present version of the algorithm requires about 3 times fewer real-value multiplications and additions than before the optimization.

It is important to stress that in this study we analyze 1024-point signal fragments but we: 1) cut them into consecutive, overlapping portions/frames (having $N=256$ or $N=512$ samples and shifted by $M=64$ samples in respect to each other), 2) do the same calculations for each signal frame separately and 3) average partial results (summations over l in table II) and calculate final tonality measures. Since the developed method addresses the real-time continuous asthma monitoring, in table

II only these calculations are taken into account which are required when a new signal frame is available. It is assumed that some temporary results (e.g. $P_l(k)$, e_l , $c_l^{(w)}$) are stored in memory, and this way the expense of averaging is significantly reduced.

TABLE I. ASE+TI WHEEZES DETECTION ALGORITHM

► **INITIALIZATION:**

$f_{sampling} \leftarrow 8000\text{Hz}$ ▷ sampling frequency
 $N \leftarrow 256$ ▷ DFT: window length
 $M \leftarrow 64$ ▷ STEP: window shift
 $k_{min} \leftarrow 4, k_{max} \leftarrow 38$ ▷ DFT: lags numbers of interest
 $L_{ASE} \leftarrow 13, L_{TI} \leftarrow 7$ ▷ orders of time averaging

► **MAIN LOOP - repeat for each new (m -th) data frame:**

REPEAT for $m = 0, 1, 2, 3, \dots$

► **Data pre-processing** -----

repeat for $n = 0, 1, 2, \dots, N-1$:

$x_m(n) \leftarrow x(m \cdot M + n)$ ▷ take new frame
 $x_m(n) \leftarrow x_m(n) - \text{mean}(x_m(n))$ ▷ subtract mean
 $x_m(n) \leftarrow \frac{x_m(n)}{\max_n(|x_m(n)|)}$ ▷ normalize

end repeat

► **DFT/FFT with windowing** -----

$$X_m(k) \leftarrow \sum_{n=0}^{N-1} x_m(n) w(n) e^{-j \frac{2\pi}{N} kn}$$

► **ASE calculation** -----

repeat for $k = k_{min}, \dots, k_{max}$:

$P_m(k) \leftarrow |X_m(k)|^2$ ▷ power
 $ASE_m(k) \leftarrow \frac{1}{L_{ASE}} \sum_{l=m-L_{ASE}+1}^m P_l(k)$ ▷ mean
 $ASE_m^{norm}(k) \leftarrow \frac{ASE_m(k) - \min_k(ASE_m(k))}{\max_k(ASE_m(k)) - \min_k(ASE_m(k))}$ ▷ norm

end repeat

$$fluctASE_m \leftarrow \sum_{k=k_{min}+1}^{k_{max}} |ASE_m^{norm}(k) - ASE_m^{norm}(k-1)| \quad \triangleright \text{final}$$

► **TI calculation** -----

repeat for $k = k_{min}, \dots, k_{max}$:

$r_m(k) \leftarrow |X_m(k)|, \varphi_m(k) \leftarrow \text{angle}(X_m(k))$
 $\overline{r}_m(k) \leftarrow 2r_{m-1}(k) - r_{m-2}(k)$ ▷ predict magnitude
 $\overline{\varphi}_m(k) \leftarrow 2\varphi_{m-1}(k) - \varphi_{m-2}(k)$ ▷ predict phase
 $A_m(k) \leftarrow r_m(k) \cdot \cos(\varphi_m(k)) - \overline{r}_m(k) \cdot \cos(\overline{\varphi}_m(k))$
 $B_m(k) \leftarrow r_m(k) \cdot \sin(\varphi_m(k)) - \overline{r}_m(k) \cdot \sin(\overline{\varphi}_m(k))$

$$c_m(k) \leftarrow \frac{\sqrt{A_m^2(k) + B_m^2(k)}}{r_m(k) + |r_m(k)|} \quad \triangleright \text{spectral unpredictability}$$

end repeat

$$e_m \leftarrow \sum_{k=k_{\min}}^{k_{\max}} r_m^2(k) \quad \triangleright \text{energy}$$

$$c_m^{(w)} \leftarrow \sum_{k=k_{\min}}^{k_{\max}} c_m(k) r_m^2(k) \quad \triangleright \text{weighted spectral unpredictability}$$

$$\bar{e}_m \leftarrow \frac{1}{L_{TI}} \sum_{l=m-L_{TI}+1}^m e_l, \quad \bar{c}_m^{(w)} \leftarrow \frac{1}{L_{TI}} \sum_{l=m-L_{TI}+1}^m c_l^{(w)} \quad \triangleright \text{mean values}$$

$$TI_m \leftarrow \log_{10} \left(\bar{c}_m^{(w)} / \bar{e}_m \right) \quad \triangleright \text{final}$$

END REPEAT

TABLE II. NUMBER OF REQUIRED REAL-VALUE ARITHMETIC OPERATIONS

Operation	*	+, -	/	$\sqrt{}$	sin, atan
Old method	15086	22336	999	156	234
New method	4842	7145	400	70	105
Speed-up	$3.12 \times$	$3.13 \times$	$2.5 \times$	$2.2 \times$	$2.2 \times$

IV. RESULTS

In this section we present some results demonstrating the efficiency of the proposed wheeze detection method: in the beginning results obtained on real data, which in our opinion are more reliable, and then results for modeled data (when we are checking the tonality descriptors' behavior in more controlled conditions/setup).

A. Real Data

During the tests, human lung sound databases available online were mainly used [34, 35]; however, several of our own recordings of normal and wheezy breaths were proportionally appended to these (the recorder and its parameters are described in section V). From longer recordings chosen by medical doctors we carefully selected 130 short records representing normal and 130 records of wheezy breath, each with 1024 samples. About 30% of these came from our own recordings. The database created for testing had all the different recognition cases (very easy, typical, and very difficult to interpret) previously noted by the doctors.

In figure 3, histograms of the *fluctASE* and *TI* tonality measure values are shown. They demonstrate that both features discriminate asthmatic wheezes (multi-tones) from normal breathing well. We can observe that both descriptors take significantly smaller values for normal breaths than for wheezes.

In figure 4, a two-dimensional plot of *fluctASE* vs. *TI* is presented, showing that it is not difficult to separate a 2D plane into two regions of values typical of normal breaths ("+" , left down corner) and wheezes ("•" , upper part).

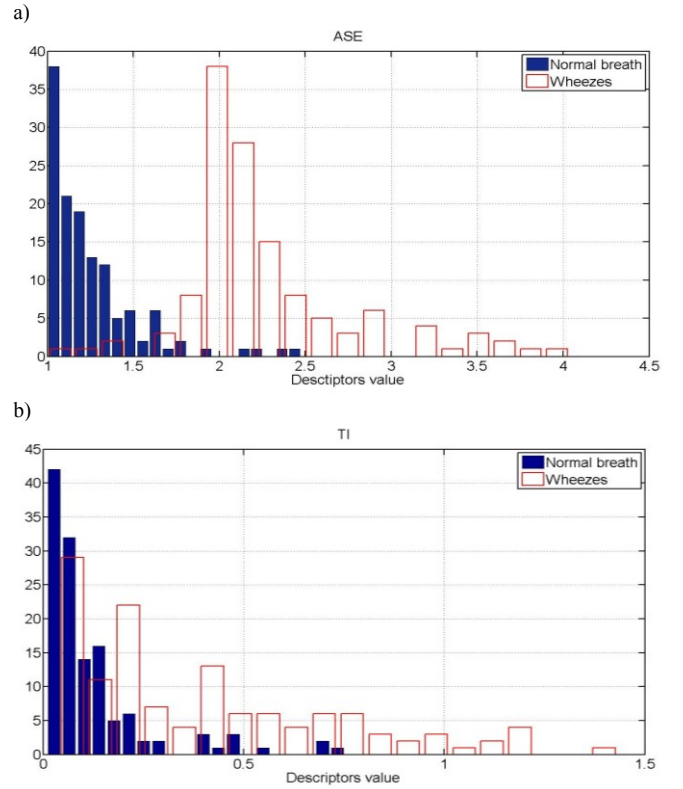


Fig. 3. Histogram of *fluctASE* (a) and *TI* (b) values calculated for normal breath (red/dark) and wheezes (white).

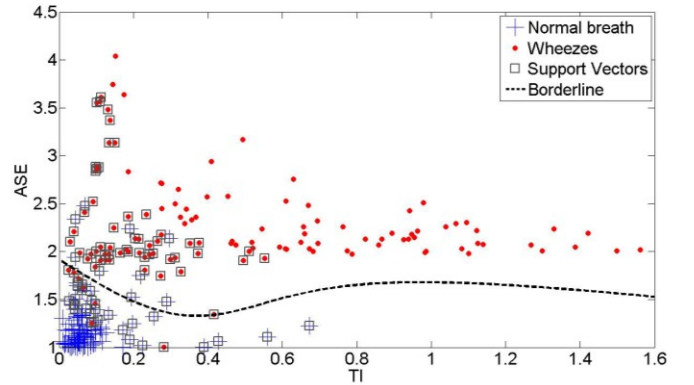


Fig. 4. Joint 2D scatter of *fluctASE* and *TI* values calculated for normal breath ("+") and wheezes ("•").

The dashed border line plotted in this figure, separating the two classes, was calculated using an SVM classifier with a polynomial 3rd order kernel.

Due to the relatively small data set, all recognition experiments were conducted using *k*-fold, in our case 10-fold, methodology: recordings of both types were randomly divided into 10 blocks having 13 sound fragments each, then nine blocks from each class were used for training the classifier (SVM with polynomial kernel with order 3, chosen experimentally) and the remaining one block was exploited for testing; the experiments were repeated 10 times and recognition results were averaged. The following measures of wheeze recognition efficiency were used:

- positive predictive value: $PPV = \frac{TP}{TP + FP} \cdot 100\%$

- negative predictive value: $NPV = \frac{TN}{TN + FN} \cdot 100\%$

- sensitivity: $Sens = \frac{TP}{TP + FN} \cdot 100\%$

- specificity: $Spec = \frac{TN}{TN + FP} \cdot 100\%$

- accuracy: $AC = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100\%$

where: TP – true positive (percentage of wheezes correctly recognized), TN – true negative (percentage of wheezes not recognized), FP – false positive (percentage of normal breaths correctly recognized) and FN – false negative (percentage of normal breaths not recognized).

Figure 5 presents the recognition accuracy for different tonality measures:

- joint application of *fluctASE*+*TI*,
- separate usage of *fluctASE* and *TI*;
- some other reference methods that were used as background in our previous research [23–28]. Below, the reference methods are listed in order of decreasing significance: autocorrelation tonality measure (CTM) [24, 26], spectral energy ratio (ER) [26], spectral peak-to-mean ratio (P2M) [26], optimum constant false alarm rate (CFAR) detector originated from sonar applications [32], linear-prediction (LP) measure [12, 26], comparison (tracking) of vector of spectral coefficients (VCMP) [27], spectral peaks entropy (SPE) [16, 23, 24], diagonal of Fourier bispectrum (DBS) [36], and spectral flatness (SF) [23].

Figure 5 takes the form of a Matlab boxplot [37]: the central mark denotes the median; the edges of the box represent the 25th and 75th percentiles; the whiskers extend to the most extreme data points, while ‘×’ and ‘*’, if they are marked, denote the median value of *PPV* and *Sens* respectively. We can observe the superior efficiency of the *fluctASE* tonality measure, the high performance of the autocorrelation-based tonality descriptor (CTM) and the better than average behavior of the tonality index (TI). Joint usage of *fluctASE* and *TI* does not improve the recognition accuracy significantly; however, this combination does ensure the robustness of the detector against changes in noise color, as will be shown below. Combining *fluctASE* and *TI* was suggested by the mRMR method [27]. This pair of tonality descriptors is also better than the $\{fluctASE + CTM\}$ pair from the point of view of lower computational cost.

The experimental receiver operating characteristics (ROC, sensitivity as a function of 1-specificity) obtained for the tested methods are presented in figure 6. The ROC curves were calculated for the whole dataset of signals without cross

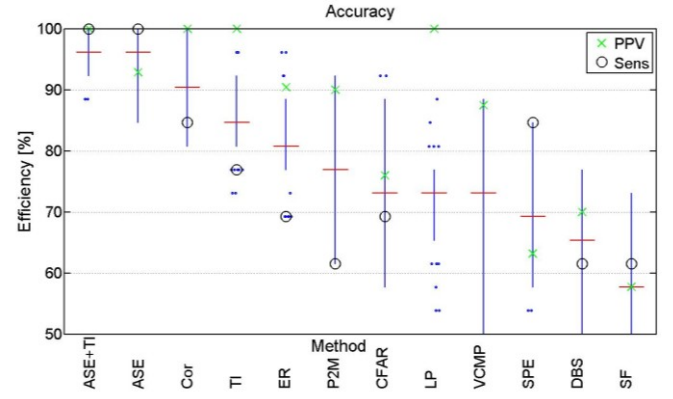


Fig. 5. Experimental wheeze recognition accuracy for different methods.

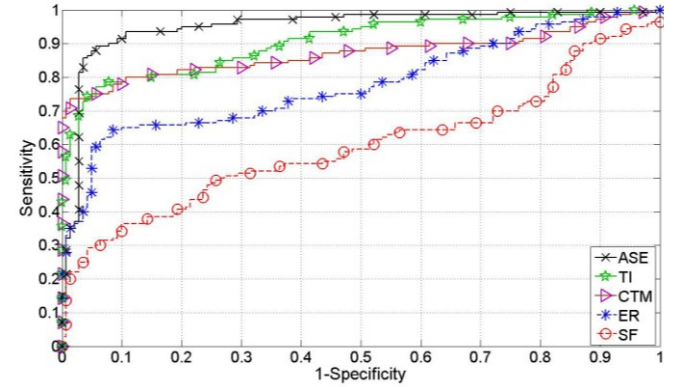


Fig. 6. ROC of different tonality (wheeze) recognizers.

TABLE III. AREA UNDER CURVE FOR SELECTED FEATURES FOR REAL DATA

Method	ASE	TI	CTM	ER	SF
AUC	0.951	0.905	0.871	0.779	0.595

TABLE IV. DESCRIPTORS WEIGHTS CALCULATED BY ADABOOST ALGORITHM

ASE	TI	CTM	P2M	ER	Others
2.87	1.44	0.74	0.61	0.37	0

validation. For better clarity, only the three best descriptors (solid lines: ASE, CTM, TI) are shown, together with the best and the worst method from the rest of the testing set (dashed lines: ER, SF). In table III, the area under each curve (AUC) is given. Again *fluctASE* outperforms the other methods with its very high sensitivity and AUC value.

As mentioned above, we have already confirmed the optimality of the ASE+TI choice in [27] using the mRMR method. In the mRMR test, the following feature ranking was obtained: ASE, TI, CTM, VCMP, SPE, ER, K, CFAR, SF, EVD, P2M, LP. Now we would like to present the new verification results obtained using the AdaBoost feature evaluation methodology [38]. This makes use of a simple classifier (thresholding in our case) and allows evaluation of the significance of particular features by means of their weights calculated during the training process. The weights obtained are shown in table IV; they are largest for ASE and TI. The overall recognition ACC was equal 93%.

B. Simulated/Modeled Data

The results presented so far were obtained using the database of real recordings. Since it was impossible to control the tonal components of the wheezes (e.g. the number of components, their frequencies and power) and to observe the sensitivity of the method we developed for different signal features, we analyzed the available real data and on this basis built two signal models – for a normal breath and a wheezy breath [25]. We calculated the mean spectrum envelope (EN) of the normal breath, characterized by approximately exponential damping in frequency, and a spectral background envelope (EW) for the wheezy breath. A wheezy breath typically had 2-4 tones with frequencies in the range [100-1200] Hz as expected. We discovered that the distribution of typical breath samples is approximately Gaussian. With the model, we generated many synthetic signals representing normal-like and wheeze-like breaths: Gaussian noise was smoothed by an appropriate filter (with the specified EN or EW frequency response) and, next, from 1 to 6 random tones were added to it (with pre-defined probability). During training and testing, the signal-to-noise (wheeze-to-breath) ratio was chosen as a random number from the range [-20, 20] dB.

First (scenario #1), normal-like breath signals were obtained by filtering the Gaussian noise with the EN filter. In turn, for generation of the wheezy-like breath signals, the EW filter was used, and then the multi-tones were added. The joint 2D scatter of *fluctASE* and *TI* values calculated for these modeled signals is presented in figure 7 (compare it with fig. 4). The recognition efficiency obtained in this case is presented in figure 8, which is very similar to figure 5. This confirms, in our opinion, the validity of our signal model.

Next (scenario #2), during the generation of the pseudo-wheezes, the EN filter (rather than the EW filter) was used. This signal modeling is typically recommended in the literature [9]. The recognition results are shown in figure 9. As we can see, the efficiency of the methods is different; however, ASE+TI, ASE, CTM and TI are again among the winners, which shows that they are robust against changes in the background spectrum of the wheezes. In turn, the ER and SF tonality descriptors are very sensitive to this background. Therefore in the following part of the paper we will present some additional results obtained for the second recommended modeling scenario only. We will investigate the sensitivity of the proposed ASE+TI method as well as its robustness against breathing noise color.

Sensitivity. Using the model, we trained the SVM classifier with synthetic normal and wheezy breaths. We start by training using mixed SNR data and testing on a single SNR. During this training, the signal-to-noise (wheeze-to-breath) ratio was chosen as a random number from the range [-20, 20] dB. During testing we tried to recognize different modeled wheezes with the same SNR. These experiments were repeated many times for an SNR changing during testing from -40 dB to 20 dB. Figure 10 presents recognition accuracy as a function of wheeze-to-breath energy ratio for the different methods

tested (ASE+TI, ASE, CTM, TI, VCMP). Once again, in order to provide better clarity, only the best three single descriptors are shown (solid lines: ASE, CTM, TI) together with the worst method from the remainder of those tested (dashed lines: VCMP). In addition, the joint application of ASE+TI is also presented (wide dotted line). Among the single-feature detectors the ASE method is the most efficient for low SNR – only CTM can compete with it. However, ASE does not work as well for higher SNR values. The joint application of ASE+TI is only slightly better than the usage of ASE alone for low SNR, although it is significantly better for higher SNR values.

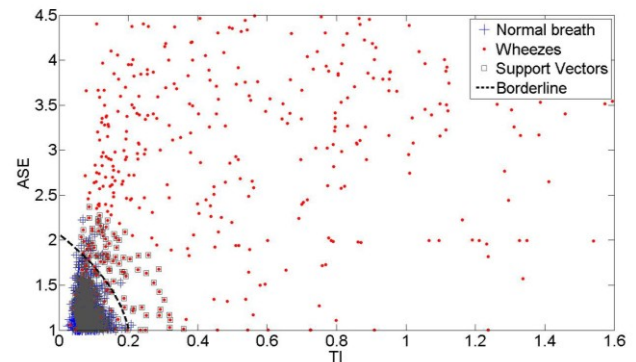


Fig. 7. Joint 2D scatter of *fluctASE* and *TI* values calculated for **modeled signals** (scenario #1) normal breath (“+”) and wheezes (“•”).

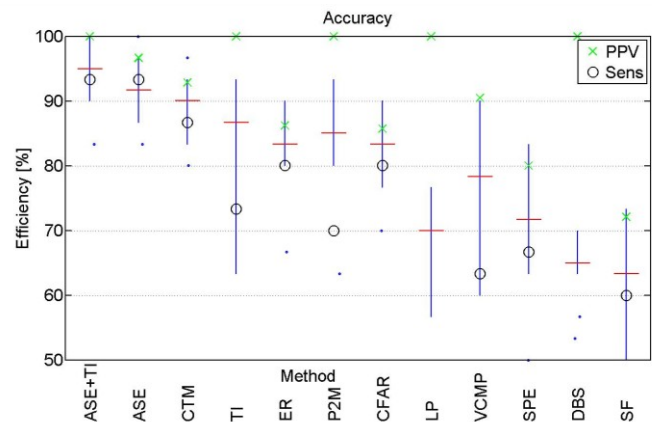


Fig. 8. Recognition accuracy for **modeled signals** (scenario #1).

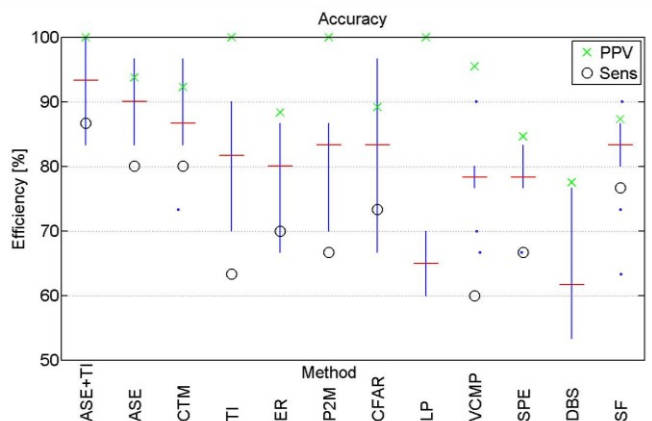


Fig. 9. Recognition accuracy for **modeled signals** (scenario #2).

Robustness to noise color. Additional simulations showed that ASE does not work very well for single tone wheezes and in the situation where the noise background is white not colored. The addition of TI was found to be a remedy to both problems. As an example, figure 11 presents results from synthetic wheeze recognition by means of ASE and ASE+TI when the noise changed smoothly from white (0 dB, no increase in near-DC frequencies) to strongly colored (60 dB increase in near-DC frequencies, then exponential decay and 0 dB for 1200 Hz – typical case of an asthmatic wheeze), see details in [28]. As we can see, the ASE method depends strongly on noise color while the ASE+TI does not. So the use of ASE+TI is noise-type robust and only slightly more complex computationally (only slightly since both methods make use of the same single FFT).

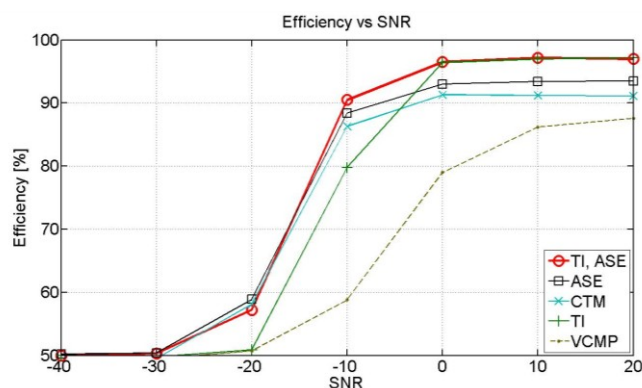


Fig. 10. Recognition accuracy as a function of SNR (wheeze-to-breath) ratio (scenario #2).

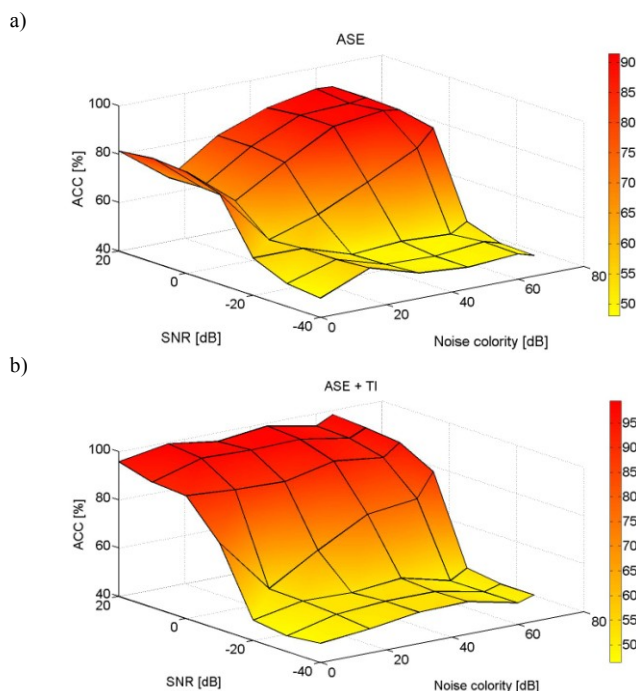


Fig. 11. Recognition accuracy as a function of SNR (wheeze-to-breath) ratio and level of noise color (increase of near-DC frequency components artificially added to initially white background noise and expressed in decibels) for scenario #2: a) ASE, b) ASE+TI.

V.E-ASTHMA MONITORING SYSTEM

The wheeze detection method presented in this paper was developed for the e-asthma monitoring system built by the POIG Polish project entitled “Future Internet Engineering” [39]. The system consists of [40]: a 4-channel wheeze recorder (shown in figure 12, with Panasonic WM-61 microphones, ARM LPC2468 ARM7TDMI-S processor, and Bluetooth transmitter), a smartphone (HTC HD2 mini) with dedicated Windows Mobile application and AMONIT (Asthma MONIToring) medical server. A simplified system diagram is presented in figure 13. Connections which have already been implemented are marked with solid lines.

Microphones for breathing sound recordings are placed on the patient's chest wall. Data are sampled with 16-bit resolution and a frequency of 8 kHz. The patients were not moving during the building our own database of breathing sounds. Patients and recordings were selected by medical doctors but the data were separated by our research team. Our own data took up about 30% of the database used for training and testing.

The smartphone is the heart (master) of the mobile part of the system. A built-in Bluetooth interface is used to connect it, in a star topology, with the following external devices: pulsometer sensor, peak flow meter and wheeze recorder. The results of measurements are sent to the e-health server continuously or on demand. Using a built-in GPS receiver, air quality measurements can be obtained by a patient more precisely from an external internet pollution database. All devices used in the mobile subsystem guarantee high power

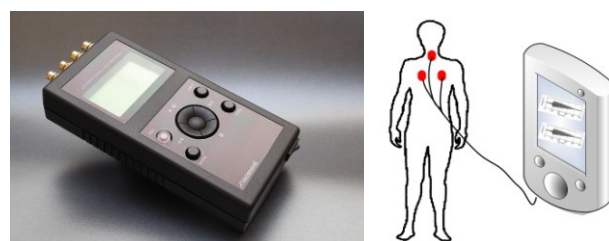


Fig. 12. Prototype of 4-channel breath sound recorder.

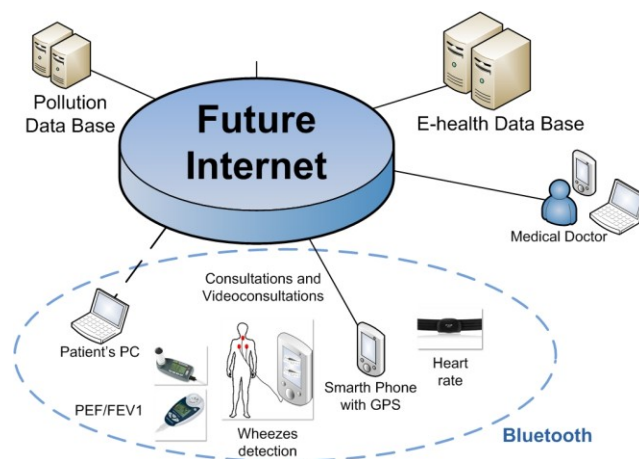


Fig. 13. Developed E-asthma monitoring system.

efficiency, and therefore battery charging is not required every day.

To increase the application range, a stationary system extension is offered. This exploits a PC (netbook) for periodic patient self tests (e.g. peakflowmetry, spirometry and wheeze detection). Mostly these tests are performed at home, but the Bluetooth interface present in the peak flow meter and wheeze recorder enables tests to be conducted elsewhere too. In this case, the data are transferred to external servers using the mobile phone's Wi-Fi interface or a GSM modem. The patient may be supervised by a medical doctor during the self tests.

VI. CONCLUSION

In this paper we have summarized the final results of our long study of the issue of asthmatic wheeze detection. More than 20 different tonality measures have been tested (typical and proposed, based on linear prediction, bi-spectrum, Fourier spectrum energy, signal statistics, etc.), real and synthetic data were used, and thousands of experiments were performed. Finally, the "winner" was the fluctuation of Audio Spectral Envelope descriptor adopted from the MPEG-7 standard. This ensures two very important features: high recognition accuracy and low computational complexity (after optimization of parameter values). Therefore, this is a very good candidate for implementation in mobile asthma monitoring devices. Combining ASE with the second tonality measure, i.e. the Tonality Index from the MPEG audio standard, makes the developed method even more robust (to noise color and number of tones), but at the cost of slightly increased computational complexity. Since the ASE+TI detector is immune to noise color, it can be used in different applications, and not only for recognition of asthmatic wheezes.

The authors are aware that the main drawback of this study is our limited access to large lung sound repositories. However, we would like to stress in this context that the database of real recordings we used, though not very large, was carefully annotated and contained signals that were very differentiated: easy, midrange and difficult to interpret. It is also important to note that all the tested methods were faced with exactly the same level of difficulty, and it is interesting to observe that in our tests the ASE method overcame some very strong and well known competitors such as the linear prediction method (preferred in [12, 14]) and the optimal CFAR tone detector (proposed in [32]).

REFERENCES

- [1] G. Charbonneau, J. Racineux, M. Sudraud, E. Tuchais, "Digital processing techniques of breath sounds for objective assistance of asthma diagnosis", *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '82*, pp. 736 - 738, 1982
- [2] T. R. Fenton, H. Pasterkamp, A. Tal, and V. Chernick, "Automatic spectral characterization of wheezing in asthmatic children," *IEEE Transaction on Biomedical Engineering*, vol. BME-32, no. 1, pp. 50-55, 1985.
- [3] S. Reichert, R. Gass, Ch. Brandt, and E. Andr s, "Analysis of respiratory sounds: State of the art," *Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine*, no. 2, pp. 45-58, 2008.
- [4] L. J. Hadjileontiadis, *Lung Sounds: An Advanced Signal Processing Perspective*, Morgan & Claypool Publisher, 2009.
- [5] R. Palaniappan, K. Sundaraj, N.U. Ahamed, A. Arjunan, S. Sundaraj, "Computer-based respiratory sound analysis: A systematic review," *IETE Technical Review*, vol. 30, iss. 3, pp. 248-256, 2013.
- [6] H. Pasterkamp, S.S. Kraman, G. R. Wodicka, "Respiratory sounds. Advances beyond the stethoscope," *American Journal of Respiratory and Critical Care Medicine*, vol. 156, no. 3, pp. 974-987, 1997.
- [7] A.R.A. Sovijarvi, L.P. Malmberg, G. Charbonneau J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, and J.R. Earis, "Characteristics of breath sounds and adventitious respiratory sounds," *European Respiratory Review*, vol. 77, no. 10, pp. 591-596, 2000.
- [8] Special issue on respiratory sound analysis, *European Respiratory Review*, vol. 77, no. 10, 2000.
- [9] S. A. Taplidou and L. J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds," *Computers in Biology and Medicine*, vol. 37, no. 8, pp. 1073 - 1083, 2007.
- [10] M. Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Computers in Biology and Medicine*, vol. 39, pp. 824 - 843, 2009.
- [11] S. A. Taplidou, L. J. Hadjileontiadis: „Analysis of wheezes using wavelet higher order spectral features", *IEEE Transaction on Biomedical Engineering*, vol. 57, no. 7, pp. 1596-1610, 2010.
- [12] B. Sankur, Y. P. Kahya, E. Cagatay Guler, and T. Engin, "Comparison of ar-based algorithms for respiratory sounds classification," *Computers in Biology and Medicine*, vol. 24, no. 1, pp. 67 - 76, 1994.
- [13] A. Abushakra, and M. Faezipour, "Acoustic Signal Classification of Breathing Movements to Virtually Aid Breath Regulation," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, No. 2, pp. 493-500, 2013.
- [14] D. Oletic, "Wireless sensor networks in monitoring of asthma," http://www.fer.unizg.hr/_download/repository/Dinko_Oletic_KDI.pdf.
- [15] S. Aydo , I. Sen, Y.P. Kahya, and M.K. Mihcak, "Classification of respiratory signals by linear analysis," *IEEE International Conference on Engineering in Medicine and Biology*, pp. 2617-2620, 2009.
- [16] J. Zhang; W. Ser; J. Yu, and T.T. Zhang, "A novel wheeze detection method for wearable monitoring systems," *IEEE International Symposium on Intelligent Ubiquitous Computing and Education*, pp. 331 - 334, 2009.
- [17] ISO/IEC MPEG, "Information Technology — Multimedia Content Description Interface", Part 4: Audio, ISO/IEC JTC 1/SC 29, 2001.
- [18] ISO/IEC MPEG, "IS11172-3 Coding of moving pictures and associated audio for digital storage media up to 1.5 Mbit/s", Part 3: Audio, Annex D. ISO/IEC JTCL, 1992.
- [19] A.J.S. Ferreira, "Tonality detection in perceptual coding of audio," *Proceedings Audio Engineering Society Conference*, Paris, paper no 3947, 1995.
- [20] T. Painter, A. Spanias: "Perceptual coding of audio signals", *Proceedings IEEE*, vol. 88, pp. 451-513, 2002.
- [21] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer 2006.
- [22] M. Wisniewski, and T.P. Zielinski, "Digital analysis methods of wheezes in asthma," *Proceedings International Conference on Signals and Electronic Systems ICSES-2010, Krakow (Poland) 2010* (IEEE Xplore).
- [23] M. Wisniewski, and T.P. Zielinski, "Application of tonal index to pulmonary wheezes detection in asthma monitoring," *Proc. 19th European Signal Processing Conference EUSIPCO-2011, Barcelona (Spain) 2011*.
- [24] M. Wisniewski, and T.P. Zielinski, "Fast and robust method for wheeze recognition in remote asthma monitoring," in *Information Technologies in Biomedicine*, vol. 2, Springer Serie: Advances in Intelligent and Soft Computing, 2012.
- [25] M. Wisniewski, and T.P. Zielinski, "Analysis and modeling of lung sound," *PAK Publishing House, PAK*, no. 4, pp. 383-386, 2012 (in Polish).
- [26] M. Wisniewski, and T.P. Zielinski, "Tonality detection methods for wheezes recognition system," *Proceedings International Workshop on Signal and Image Processing IWSSIP-2012, Vienna (Austria) 2012* (IEEE Xplore).
- [27] M. Wisniewski, and T.P. Zielinski, "MRMR-based feature selection for automatic asthma wheezes recognition," *Proceedings International*

- Conference on Signals and Electronic Systems ICSES-2012*, Wroclaw (Poland) 2012 (IEEE Xplore).
- [28] M. Wisniewski, and T.P. Zielinski, "Detection of multi-tones in white and colored noise with application to asthmatic wheezes detection," *Proceedings International Conference on Signals and Electronic Systems ICSES-2012*, Wroclaw (Poland) 2012 (IEEE Xplore).
 - [29] H. Peng, F. Long, C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp.1226-1238, 2005.
 - [30] S. M. Kay, *Modern Spectrum Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
 - [31] L. Marple, *Digital Spectrum Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
 - [32] Q. Wang, C.-R. Wan: "A novel CFAR tonal detector using phase compensation", *IEEE Journal of Oceanic Engineering*, vol. 30, no. 4, pp. 900-911, 2005.
 - [33] J.-B. Pouillet, D. M. Sima, and S.e Van Huffel, "MRS signal quantitation: A review of time- and frequency-domain methods," *Journal of Magnetic Resonance*, 195, pp. 134–144, 2008.
 - [34] Lung sounds repository, http://solutions.3m.com/wps/portal/3M/en_US/Littmann/stethoscope/education/lung_sounds/.
 - [35] Lung sounds repository, http://faculty.etsu.edu/arnall/www/public_html/heartlung/breathsounds/contents.html.
 - [36] Chang G-C, Cheng Y-P, "Investigation of noise effect on lung sound recognition", *International Conference on Machine Learning and Cybernetics*, Kunming, pp. 1298 – 1301, 2008.
 - [37] <http://www.mathworks.com/help/toolbox/stats/boxplot.html>
 - [38] Y. Freund, and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.
 - [39] Project financed by the European Union: "Future Internet Engineering": <https://www.iip.net.pl/>.
 - [40] M. Wisniewski, et al., "Remote monitoring of patient with asthma disease," *Przegląd Telekomunikacyjny, Wiadomości Telekomunikacyjne*, vol. 84, no. 8-9, pp. 750-752, 2011 (in Polish).