Robust Detection of Audio-Cough Events using local Hu moments

Jesús Monge-Álvarez[1], Carlos Hoyos-Barceló[1], Paul Lesso[2], Pablo Casaseca-de-la-Higuera[1,3*]

***Abstract*— Telehealth has shown potential to improve access to health-care cost-effectively in respiratory illness. However, it has failed to live up to expectation, in part because of poor objective measures of symptoms such as cough events, which could lead to early diagnosis or prevention. Considering the burden that these conditions constitute for national health systems, an effort is needed to foster telehealth potential by developing low cost technology for efficient monitoring and analysis of cough events. This paper proposes the use of local Hu moments as a robust feature set for automatic cough detection in smartphone-acquired audio signals. The final system feeds a *k*-Nearest Neighbors classifier with the extracted features. To properly evaluate the system in a diversity of noisy backgrounds, we contaminated real cough audio data with a variety of sounds including noise from both indoor and outdoor environments, and non-cough events (sneeze, laugh, speech, etc.). The created database allows flexible settings of Signal to Noise Ratio (SNR) levels between background sounds and events (cough and non-cough). This evaluation was complemented using real patient data from an outpatient clinic. The system is able to detect cough events with high sensitivity (up to 88.51%) and specificity (up to 99.77%) in a variety of noisy environments, overcoming other state-of-the-art audio features. Our proposal paves the way for ubiquitous cough monitoring with minimal disruption in daily activities.**

***Index Terms*— Cough Detection, Respiratory Illness, mHealth, Hu moments, *k*-NN, SVM.**

I. Introduction

Cough is one of the commonest symptoms causing patients seek medical advice. Cough can be understood as a natural reflex physiologically aiming at clearing the lower airways of debris, especially mucus. It is thus a defense mechanism for ejecting foreign material out of the respiratory system [1], [2]. From the signal processing perspective, an audio cough event is a non-stationary signal without a clear formant structure and composed of three phases: the explosive phase, the intermediate phase and the voice phase. The average duration is approximately 300 ms. Its spectrum exhibits a high-energy peak around 400 Hz and a secondary peak between 1000 and 1500 Hz [3].

Over one hundred pathological conditions are associated with cough [4]. Many of them are respiratory illnesses such as pneumonia, asthma, laryngitis or chronic obstructive pulmonary disease, while others are more generic (cold, *flu*, allergies, etc.). In addition, cough can be associated to life style (smokers, sedentary people, etc.). Cough treatments constitute a significant burden for national health systems – an estimation of £100 million/year cost for NHS Scotland [5] and $40 billion per annum in the USA from direct and indirect costs of the common cold [6] – and economies, with an average yearly productivity loss cost of £2176 per patient [7].

Despite the fact that cough sounds convey vital information of the state of the respiratory system, there are no gold standard methods to objectively assess cough [8]. This explains why, until recent years, the study of cough has been restricted to subjective measurement tools: the practitioner usually asks the patient to provide his/her own appreciation of the frequency and severity of their coughs and how they affect their quality of life. Cough scores, diaries and symptom questionnaires are typically used in this process [9], [10]. However, this approach presents some drawbacks that can lead to misinterpretation of cough symptoms [11]. First, the actual limitations of the human hearing system and other tools employed (e.g., stethoscopes), which behave as low-pass filters [12]. Secondly, there exists inter-expert variability [13]. Finally, secondary aspects of the underlying diseases like their physical and psychological comorbidity: urinary incontinence, chest pain, sleep disturbance, relationship difficulties, social embarrassment or depression [8], [14].

To overcome these limitations, governmental institutions have highlighted the potential of telemedicine in the management of respiratory conditions [15]. Even though the first cough monitors arose in the 1950s, it was not until the development of the new digital devices and processing techniques when the measurement of cough was rigorously undertaken [11], [16]. Current systems rely on pattern recognition engines primarily based on features extracted from cough sounds. Most of the so far proposed cough detectors suffer from some limitations which make them unsuitable for

This paragraph of the first footnote will contain the date on which you submitted your paper for review. This work was supported by the Digital Health & Care Institute Scotland as part of the Factory Research Project SmartCough/MacMasters. Thanks are given to University of the West of Scotland for partially funding J. Monge-Álvarez and C. Hoyos-Barceló studentships. UWS acknowledges the financial support of NHS Research Scotland (NRS) through Edinburgh Clinical Research Facility. Acknowledgement is extended to Cancer Research UK for grant C59355/A22878.

**Asterisk indicates corresponding author*

J. Monge-Álvarez, C. Hoyos-Barceló, and P. Casaseca-de-la-Higuera* are with the Center for Artificial Intelligence, Visual Communications and Networking (AVCN) School of Engineering and Computing, University of the West of Scotland, Paisley Campus, Paisley, PA1 2BE, United Kingdom. P. Casaseca-de-la-Higuera is also with the Laboratory of Image Processing (LPI), ETSI Telecomunicación, Universidad de Valladolid. (email: pablo.casaseca@uws.ac.uk, casaseca@lpi.tel.uva.es).

P. Lesso is with Cirrus Logic, Edinburgh, EH112QB, United Kingdom.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2800741, IEEE Journal of Biomedical and Health Informatics

real-time monitoring in real-life situations. Some rely on complex recording systems (low-noise microphones, pre-amplifiers, etc.) and have only been tested in quiet and controlled environments [17]. Others focus on a very specific population [18], [19] (infants, patients with a particular pathology, etc.) and thus present lack of generalization. On the other hand, some methods were conceived to solve a wider problem than cough detection [20], [21] and fail to achieve optimal performance. Finally, some approaches have not been designed with efficiency in mind (large feature sets or many classifiers [22] iterative algorithms [23] , etc.) and may not be advisable in real-time situations. Apart from the limitations mentioned above, these bespoke systems can be considered as expensive and uncomfortable (i.e., non-wearable during daily activity) solutions at a time when telehealth has moved towards generic readily available sensors.

The recent advances in smartphone and watch technology additionally allow employing these daily use devices as intelligent cough monitoring systems since they feature a number of embedded sensors able to measure cough sounds and related movement. Moreover, the computational capability of these devices is increasingly growing while, at the same time, they feature real-time connectivity to offload complex operations to higher performance computing systems.

The proposal from Larson et al. [18] processed the audio signal acquired from lapel microphones and a consumer-grade recorder worn in a fanny pack. This solution forced the user to carry multiple devices and upload the data to a server for analysis. If the device was carried in the pocket with just the specific application running on it at full functionality or with slight seamless modifications to its configuration, the impact on their activity would be minor, and the patient would be less conscious of the medicalization of their life. This raises important research challenges in using a smartphone as a medical device, namely the necessity to deal with noisy inputs in mobile environments as well as battery consumption issues related to continuous sensor monitoring and computing [24]. Efficient and robust signal processing methods to deal with continuous monitoring of noisy inputs from the mobile microphone or misaligned acceleration signals due to carrying the device need to be investigated.

Our preliminary work in [25] showed the promising applicability of local Hu moments for automatic segmentation of cough events. This feature set was recently imported from the image processing field to speech emotion recognition [26]. Assessing emotions in speech requires characterizing subtle differences within the signal, which to some extent, is equivalent to distinguishing two different signals with comparable acoustic properties.

On this basis, this paper proposes the use of local Hu moments as robust feature set for an automatic cough detection system based on smartphones. The proposed cough detector is evaluated using two signal databases. In the first one, we combined different types of real cough sounds (male/female, adult/children/babies, smokers/non-smokers, etc.) with noisy signals from a variety of indoor and outdoor environments and non-cough events (sneezing, laughing, speech, snore, etc.). The created database allows flexible setting of the Signal to Noise Ratio (SNR), defined as the ratio between the average power of the background sounds and the average power of the foreground sounds/events (cough and non-cough). Further evaluation was performed over twenty-six hours of ambulatory patient audio recordings. The acquisition protocol leading to this second database simulated different environments and daily life activities.

The finally proposed system relies on a *k*-Nearest Neighbor (*k*-NN) classifier using Hu moments of the audio signal as inputs. The system is able to detect cough events with high sensitivity and specificity in a variety of noisy environments. To demonstrate the robustness of the study, we have performed a comparison of the proposed feature set with a number of different audio features. These have been employed in fields such as speech processing, automatic music classification, asthma wheeze recognition, speech emotion recognition, among others. Similarly, two extended classifiers have been used: a Support Vector Machine (SVM) and a *k*-NN classifier. Derived from the problem context, our study also analyzes the trade-off between performance and efficiency (measured as CPU execution time) to inform the decision on the final implementation on a smartphone.

Compared to our preliminary study, the work in [25] employed a more limited data source for evaluation, both in terms of quantity (number of signals) and quality (diversity of noisy sounds and foreground events). The only noisy source therein was the friction between the embedded microphone and the fabric when the smartphone was carried in the pocket. In addition, the main objective in [25] was to analyze the suitability of importance sampling techniques to cope with the class-unbalance in the cough detection problem. Finally, no comparison of the proposed system with other feature sets and classifiers was provided in [25] as opposed to this paper.

The rest of the paper is organized as follows. Section II summarizes the state-of the art in cough detection and further motivates our proposal. Section III describes the methodology of the proposed cough detector. Section IV is devoted to the experimental results, discussed in Section V. Finally, Section VI outlines some future research lines and the conclusions of the study.

II. State-of-the-art

The commercial Lifeshirt monitor (no longer available since the company was liquidated in 2009) was based on a wireless health monitoring system integrating electrocardiogram, respiratory inductance plethysmography, 3-axis accelerometer, and a contact microphone placed on the throat. It achieved a sensitivity of 78.2% in laboratory conditions [27]. The Hull Automated Count Counter relies on audio recordings fed to an adaptive neural network. It offered a sensitivity of 80% measured in a group of 33 patients [28]. The Leicester Cough Monitor performs a preliminary detection of the events by means of Hidden Markov Models (HMM) followed by a semiautomatic classification stage. In the analysis of the recording of 26 subjects, it reached a sensitivity of 85.7% [29]. The VitaloJAK employs a contact microphone placed on the chest wall to detect cough sounds. A preliminary study over a small patient group achieved 98% sensitivity [8]. The PulmoTrack-CC launched by Karmelsonix in 2010 used a combination of sounds recorded from the neck and a movement sensor placed in the chest wall, and achieved a sensitivity around 96% in counting voluntary cough events [30].

From the strict point of view of signal processing, there are also recent studies which have focused on audio cough signals. They exploit several features and classifiers with the aim of cough counting or cough assessment. Amrulloh et al. [17] performed cough segmentation within pediatric wards using Shannon entropy, Mel Frequency Cepstral Coefficients (MFCC) and a non-gaussianity measure as features. After classification with an Artificial Neural Network (ANN), they achieved a sensitivity of 93%. Larson et al. [18] assessed the recovery of pulmonary tuberculosis by analyzing cough sound recordings. They respectively employed MFCC and Sequential Minimal Optimization as features and classifier, achieving a sensitivity of 75.5%. Yatani and Truong [20] developed a wearable acoustic sensor which records the sounds produced in the user’s throat area for activity recognition (including coughing). Using features such as spectral roll-off, spectral flux, spectral centroid or MFCC, which fed a SVM, their sensitivity in cough detection was between 62% and 74%. They also analyzed the performance of two other classifiers: a Naïve Bayes classifier and a *k*-NN classifier. No specific results of cough classification performance were reported for these two cases. Drugman et al. used a set of 50 features (after dimensionality reduction by feature selection) and two ANN to create a system for automatic, objective and reliable detection of cough events. The set of features included MFCC, a measurement of loudness in the Bark scale and several parameters describing the audio spectral shape. They achieved average sensitivity was 94.7% [22]. Matos et al. [19] evaluated the intensity and frequency of occurrence of cough events for the assessment of patients with chronic disease. They followed a keyword-spotting approach using a HMM classifier, which resulted in an average detection rate of 82%. Finally, this problem has also been tackled in the field of audio event-detection. Drugman presented a new technique consisting of an iterative process to synchronize features and cough labels. This was applied to cough detection and results showed improvements both in feature selection and detection capabilities compared to more classical approaches [23]. Ezgi and Sert [21] proposed an optimized MFFC-SVM approach to recognize events such as cough, throat clearings, speech, knockings, etc. within an office live environment. Sensitivity values of 63.6% were reported for cough events. You et al. [31] provided and ad-hoc feature extraction method for cough detection based on non-negative matrix factorization. They achieved sensitivity and specificity values around 85% on a database encompassing signals from 18 patients (80 min. of recording each).

Other approaches based on Convolutional Neural Networks (CNN) and deep neural networks have also been explored [32], [33]. Amoh and Odame [33] employed CNN and a Recurrent Neural Network (RNN) to perform cough segmentation. Both networks offered sensitivity around 83%, whereas the specificity of the CNN was better (93%) than the RNN one (75%). Approaches which are based on Wavelet transform [34] or time domain analysis [3] have been also explored. Finally, some patents four cough analysis have been recently presented, e.g. [35].

Following the European Respiratory Society (ERS) guidelines on the assessment of cough [6], cough monitors should be capable of digitally capturing and processing 24-hour recordings. Likewise, Smith and Woodcock [8] established other critical and desirable characteristics:

a) Differentiation of cough from background noise

b) Differentiation of cough from other sounds produced by the patient such as laugh, speech, throat clearing, etc.

c) Dealing with the variability of cough acoustics: both within and between individuals, as well as the additional complexity of different respiratory diseases.

Most of the reported systems and methods for cough analysis have been tested in idyllic conditions where noise was present at low level or even absent. Moreover, the use of smartphones implies environmental changes from time to time depending on the daily life activities of the user/patient. Accordingly, the captured signals may be a mixture of background sounds – e.g. babble noise, music, environmental noise, footsteps, or even noise generated from the smartphone moving inside the pocket of the user/patient. – together with cough and non-cough events. In addition, some of the non-cough events – e.g., throat clearing and sneeze events – have very similar acoustical characteristics to cough. Thus, features that simulate the cochlea response such as MFCC may struggle to detect cough in noisy environments. This makes exploring more robust alternatives advisable.

III. Methodology

A. Overview of the system

Fig. 1 depicts the pipeline of our cough detection system. It is composed of four blocks namely, pre-processing, feature extraction, classification and post-processing.

The pre-processing module separates the signal into frames by means of a Kaiser window with $\beta = 3.5$. This window showed the best tradeoff between spectral resolution and leakage among other evaluated windows (Kaiser with b=1.5, Taylor, and Hamming). As we showed in [36], the frequency band between 0 and 2 kHz is sufficient to detect cough events. Thus, there is no need to keep the original sampling frequency 44.1 kHz (see section IV.A). So, we downsampled the acquired signals lowering the sampling frequency to 8820 Hz. The window length is 50 ms (*N*=441 samples) and the window shift 25 ms (221 samples). As a starting point for most of the computed feature sets, the power spectral density of each window (*PSD*[*k*]) was estimated as the Fourier transform of the autocorrelation function, according to the Wiener-Khinchin-Einstein theorem [37]. Later, each PSD was normalized using the following factor derived from the Kaiser:

$$U = (1/N)\cdot\sum_{n=1}^{N}\left|w[n]\right|^2 \qquad (1)$$

where $w[n]$ is temporal shape of the Kaiser window. Finally, the one-sided PSD was selected:

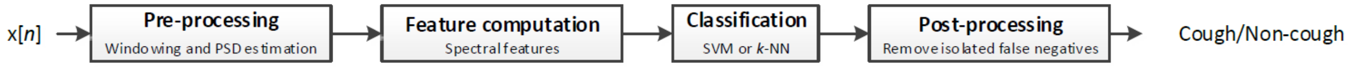

Fig. 1. Pipeline of our system for cough detection.

$$PSD[k] = \begin{cases} PSD[k], & k = 1 \\ 2 \cdot PSD[k], & k = 2, \ldots, Nend - 1 \\ PSD[k], & k = Nend \end{cases} \qquad (2)$$

where *Nfft* is the number of FFT points and *Nend* = $(Nfft + 1)/2$ for odd *Nfft* and $(Nfft/2)+1$ otherwise.

Different feature sets and classifiers have been developed and compared to find the most suitable combination for the final implementation of the system. The following subsections describe them.

B. Evaluated Feature sets

1) Multidimensional spectral features

A number of features aiming at the recognition of specific types of audio signals do exist in the literature. Two of the most employed are MFCC and Linear Prediction Cepstral Coefficients (LPCC). They were initially designed for automatic speech recognition but, over time, they were used for other purposes. MFCC account for the non-linear response of the human ear across the audio spectrum and are obtained using a frequency transform of the log spectrum [38] whereas LPCC are an extension of linear prediction via autoregressive modeling in the cepstral domain [39]. Derived from their success, MFCC became a *de facto* standard, so other features based on the same philosophy were proposed. Among this group, the following can be highlighted: GammaTone Cepstral Coefficients (GTCC), Normalized Audio Spectral Envelope (NASE), Octave Spectral Contrast (OSC) or Spectral Subband Centroid Histograms (SSCH).

GTCC – together with MFCC – have been the most widely used in cough detection [40], even though they have other uses like non-speech audio classification [41]. Others have been employed in music genre classification (OSC, [42]) whereas NASE was defined in the MPEG-7 standard for sound classification [42], [43]. Finally, SSCH can be considered as a more noise-robust improved version of MFCC [44].

The underlying rationale of these features is the characterization of the signal spectrum in different frequency bands. The main differences among them lie in the scale employed in the frequency representation – e.g. cepstral scale [38]-[41], octave scale [42], [43] and Bark scale [44] – or in the type of filters defining those frequency bands – e.g. triangular filters [38], biologically inspired gammatone filters [41] and highly overlapped rectangular filters [44] – as well as the metrics to apply in each frequency band – e.g. energy as in [38], [40], [43], mean power and frequency centroids as in [44] or the peaks and valleys of the spectrum as in [42]. The dimensionality of these features directly depends on the value of their inner parameters. We have implemented and tested all of these features in our study. The configuration of the inner parameters for each feature set is summarized in Table I.

2) Unidimensional spectral features

We analyzed a set of features which have shown to be meaningful in the biomedical signal processing field and had never been used in this problem (to our knowledge). To this end, we grouped several unidimensional features into a feature set with a comparable dimension to the above described (13 for all of them except 12 for OSC).

In particular, we computed the following thirteen features (henceforth referred as SpecBlock13):

- Spectral Centroid (SpecCen): center of gravity of the magnitude spectrum [45].
- Spectral Bandwidth (SpecBand): a measure of the spectral dispersion [45].
- Spectral Crest Factor (SpecCresFac), a measure of tonality [45].
- Spectral Turbulence (SpecTurb), which quantifies variations over time in the spectral content [46].
- Spectral Flux (SpecFlux): this measure also enables detecting variations over time in the spectral content [47].
- Ratio f50 vs f90 (Ratiof50f90): Ratio between f50 and f90, frequencies for which the concentrated energy below them is 50% and 90%, respectively [48].
- Spectral Roll-off (SpecRolloff): it accounts for the frequency below which, 85% of the energy is concentrated [47].
- Spectral Standard Deviation (SpecSD), Spectral Skewness (SpecSkew) and Spectral Kurtosis (SpecKurto) aim to distinguish spectra on the basis of their shape. For example, the kurtosis describes how the spectrum in concentrated around the mean whereas

TABLE I
ALGORITHMS AND CONFIGURATION OF INNER PARAMETERS FOR MFCC, LPCC, OSC, SSCH AND GTCC FEATURES

Feature	Algorithm	Parameters
MFCC	[38]	• Filterbank edges: [0 2000] Hz • Number of filters: 26 • Number of lifter coefficients: 22 • Number of DCT coefficients: 13 • *Nfft* = 1024
LPCC	[39]	• Number of coefficients: 13 • *Nfft* = 1024
NASE	[43]	• Frequency limits: [0 2000] Hz • *Nfft* = 8192
OSC	[42]	• Frequency limits: [0 3200] Hz • $\alpha = 0.2$ • *Nfft* = 1024 • 6 contrast + 6 valleys were chosen
SSCH	[44]	• Filterbank edges: [0 2000] Hz • Number of filters: 26 • Number of DCT coefficients: 13 • Width of each filter: 3 Barks • Number of bins in the histogram: 38 • *Nfft* = 2048
GTCC	[40]	• Filterbank edges: [0 2000] Hz • Number of DCT coefficients: 13 • *Nfft* = 2048

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2800741, IEEE Journal of Biomedical and Health Informatics

skewness is a measure of asymmetry. We computed these features using logarithmic units [48].

- Spectral Peak Entropy (SpecPeakEn): it is a measure based on the local maxima of the spectrum [48].
- Renyi Entropy (RenyiEn): It can be considered as an estimation of the irregularity of the spectrum [49].
- Tsallis Entropy (TsallisEn): It is a non-logarithmic entropy to explore the properties of a spectral probability distribution in a different scale [49].

The computational details for the aforementioned features are presented in Appendix A.

3) *Local Hu moments*

Finally, we calculated local Hu moments as a robust candidate feature set for cough detection in noisy environments. To do so, the following steps were carried out [26]:

First, the PSD for each window was obtained using 4096 points in the FFT algorithm.

Second, we computed the logarithm of the spectral energies for every window in a series of bands defined by a filterbank in the Mel scale:

$$E_k(m) = \log\left(\sum_{f=f_{\min}}^{f_{mas}} PSD_k[f]\cdot H_m[f]\right) \quad 0 \le m < M \tag{3}$$

where k refers to the *k-th* window and m denotes each filter within the filterbank. $f_{\min}$ and $f_{\max}$ are 0 and 2 kHz, respectively. The filterbank in the Mel scale is defined as:

$$H_m(f) = \begin{cases} 0, & f < C(m-1) \\ \dfrac{2\cdot(f - C(m-1))}{(C(m+1)-C(m-1))(C(m)-C(m-1))}, & C(m-1) \le f < C(m) \\ \dfrac{2\cdot(C(m+1)-f)}{(C(m+1)-C(m-1))(C(m+1)-C(m))}, & C(m) \le f < C(m+1) \\ 0, & f \ge C(m+1) \end{cases} \tag{4}$$

C(*m*) $0 < m < M$ are the centers of each filter in the filterbank [Hz], uniformly spaced between $f_{\min}$ and $f_{\max}$ in the Mel scale. The equations to convert natural frequencies to the Mel scale and viceversa are shown below:

$$f[Mel] = 2595\cdot\log_{10}(1+f[Hz]/700) \tag{5}$$

$$f[Hz] = 700\cdot\left(10^{f[Mel]/2595} - 1\right) \tag{6}$$

The total number of filters was $M = 75$. Consequently, after performing this step for all the signal windows, a $(K \times (M-1))$ matrix was obtained, with *K* the number of signal windows.

Next, we computed the local Hu moments of the energy matrix *E*. To do so, we divided *E* into $(K \times ((M/w)-1))$ blocks B_{ij}, with *w* the block size. In our calculation, we used $w = 5$ as in [26]:

$$B_{ij} = \begin{pmatrix} E_i(w\cdot j) & \cdots & E_i(w\cdot j + w - 1) \\ \vdots & \ddots & \vdots \\ E_{i+w-1}(w\cdot j) & \cdots & E_{i+w-1}(w\cdot j + w - 1) \end{pmatrix} \tag{7}$$

$$i = 1\ldots K \quad j = 1\ldots((M/w)-1)$$

The latest $(w-1)$ blocks, corresponding to $i = K - w + 2, \ldots, K$, are padded with zeros up to the size $(w \times w)$.

We got the first invariant moment θ of each B_{ij} as:

$$\theta = \eta(p=2, q=0) + \eta(p=0, q=2) \tag{8}$$

$$\eta(p,q) = \frac{\mu(p,q)}{(\mu(0,0))^{\rho}}, \quad \rho = (p+q+2)/2 \tag{9}$$

$$\mu(p,q) = \sum_{u=1}^{w}\sum_{v=1}^{w}(u-\bar{u})^p\cdot(v-\bar{v})^q\cdot g(u,v) \tag{10}$$

$$g(u,v) = B_{ij}(u,v) \quad p,q = 0,1,2,\ldots$$

In (10), $\bar{u}$ and $\bar{v}$ are $\bar{u} = \varphi(p=1, q=0)/\varphi(p=0, q=0)$ and $\bar{v} = \varphi(p=0, q=1)/\varphi(p=0, q=0)$, with:

$$\varphi(p,q) = \sum_{u=1}^{w}\sum_{v=1}^{w} u^p \cdot v^p \cdot g(u,v) \tag{11}$$

To finish this step, all θ are used to construct a real $(K \times ((M/w)-1))$ matrix, *Q*.

To conclude, the discrete cosine transform (DCT) is computed for each row in *Q* and coefficients 2-14 are finally kept. The result is a $(K \times 13)$ matrix *TQ*, being the rows of this matrix the local Hu moments for each window in the signal. Fig. 2 depicts a diagram of the latest local Hu moments computation steps, for the sake of clarity.

C. *Classifiers*

To finally achieve cough event detection, the 50 ms windows feed a classifier after feature computation. We compared two classifiers, namely SVM [50] and *k*-NN [51].

SVM and *k*-NN were selected as the most prominent classifiers in a wide range of *machine hearing* problems. Simpler classifiers such as decision trees, discriminant analysis or logistic regression have shown poor performance in such problems [52] whereas other solutions such as ensemble classifiers or random forests could lead to complex final implementations in mobile devices.

The classifiers were trained using 60% of the observations, and tested using 30% of them. The remaining 10% were used to validate the inner configuration of each classifier. SVMs with 2nd-5th order polynomial, linear, Gaussian, and radial basis function kernels were evaluated to finally select a SVM with 4th order polynomial as best performing on the validation set. As for *k*-NN classifiers, we tested *k*={1,3,5} and different distance metrics: standardized Euclidean, Chebychev, cityblock, cosine, and Mikowski. The best performing over the validation set used standardized Euclidean distance with the inverse of the distance as weighting function, exhaustive computation of all the distances, and *k*=1. Prior to classification, all the feature sets were normalized to have zero-mean and unitary standard deviation.

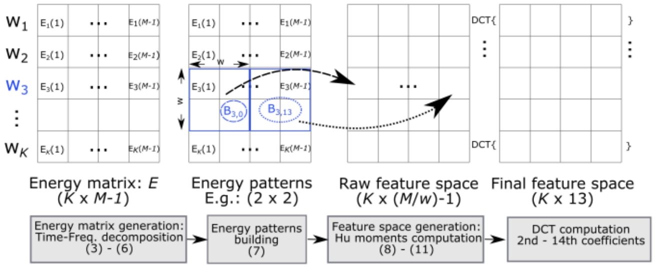
W1
W2
W3
⋮
WK
E1(1) ... E1(M-1)
E2(1) ... E2(M-1)
E3(1) ... E3(M-1)
EK(1) ... EK(M-1)
w
w
B3,0
B3,13
...
DCT{ }
DCT{ }
Energy matrix: E (K x M-1)
Energy patterns E.g.: (2 x 2)
Raw feature space (K x (M/w)-1)
Final feature space (K x 13)
Energy matrix generation: Time-Freq. decomposition (3) - (6)
Energy patterns building (7)
Feature space generation: Hu moments computation (8) - (11)
DCT computation 2nd - 14th coefficients

Fig. 2. Example of local Hu moments computation.

D. Post-processing

To improve sensitivity, we carried out a simple post-processing task avoiding isolated false negatives by setting every non-cough window surrounded by cough ones to actual coughs.

IV. Experiments

A. Materials

In order to assess the performance of the system in a variety of noise conditions, we designed two experiments leading to the corresponding signal databases. The following subsections describe them in detail.

1) Real cough sounds in changing noisy environments

The first database included a wide range of real cough sounds which were artificially contaminated using noise from different environments and non-cough events. The noise signals were added at different levels, thus enabling full control of the SNR as a parameter. This way, the performance of different classifiers and feature sets could be assessed as a function of the SNR this enabling an informed decision on the most suitable method. The following paragraphs describe the database creation procedure:

1. We collected or recorded the raw signals one by one – cough events, non-cough events and background sounds, all of them acquired at 44.1 kHz, with 16 bits per sample, and a lossless format. We used publicly-available audio signals databases [53], [54].
2. Due to the diversity of origins of the raw sounds and the uncontrolled recording conditions, prior to the synthesis, we equalized all the raw signals to have the same average power.
3. After that, we synthesized the signals for different SNR values. For the particular experiments in this paper, we used eight SNR values: -6, -3, 0, 3, 6, 9, 12 and 15 dB. To do so, we firstly selected the foreground events and the background sounds that would compose each final signal. The foreground events were collated one after the others in a larger signal. Between each foreground event, zero samples with random duration between 0.25 and 1 s were inserted. The reason why we included these gaps is due to the fact that two foreground events of different nature are very unlikely to occur one immediately after the other. Next, we calculated a gain value, *G*, to be applied to the background sounds signal to get the desired SNR (12). Finally, both the event and background signals are added. Fig. 3 shows eight SNR versions of one of the synthesized signals.

$$SNR_{dB} = 10 \cdot \log_{10}\left(1/G\right) \Rightarrow G = 10^{-\frac{SNR_{dB}}{10}} \qquad (12)$$

The first database is composed of 26 signals with durations between 15 and 155 s. The total duration of each SNR version of the database is 1245 s. Thus, the overall length of the signal database for the 8 SNR values evaluated in this paper is 8x1245=9960 s. As far as possible, we tried to define each signal with the greatest realism. For instance, one of the samples replicates a situation in which a person is jogging in a park. The background sounds include steps of the jogger, wind, etc. As for the foreground events, they come up in the following order: normal breathing, sounds of breathless breathing, a cough episode and finally throat clearing. Neither any foreground event nor background sound was used more than once in the synthesis. Background sounds cover both indoor (air conditioning, an office, the subway, a supermarket, toilets, a crowded restaurant, the indoor of an airport, a classroom during a lecture, a hall of a train station, a buffet restaurant, a casino, a court house, a post office, a museum or the corridor of a hospital, etc.) and outdoor (breeze, strong wind, rain under an umbrella, a crowded street, a park with children playing, a quiet residential area, a street with traffic, an open-air market, etc.) environments. Among the non-cough foreground events, the database includes throat clearing, sniffing, sneezing, burping, breathing, breathless breathing, laughs (male and female), speech (male and female), blowing nose, snoring or swallowing.

2) Ambulatory patient recordings

The second database includes ambulatory recordings emulating the functional conditions of a smartphone-based cough detector. We recruited thirteen adult patients from the Outpatient Chest Clinic, Royal Infirmary of Edinburgh (UK), all presenting cough as a symptom from a variety of conditions (see Table II).

One hour of audio was acquired from each patient, divided in three parts:

- The first part simulates a low-noise environment. In this situation, the patient is sitting and is requested to speak or read aloud. From time to time, we asked the patient to produce other foreground events such as throat clearing, swallowing (by drinking a glass of water), blowing nose, sneezing, breathless breathing or laugh (by reading a joke or a humor comic).
- The second part emulated a noisy environment with a external source of contamination, i.e., the noisy background sounds are not produced by the patient. To do so, we repeated the same experiment as in part one with either a television set or radio player on. Besides, the door of the room was left open so that noisy sounds from the corridor of the hospital were recorded as well.

TABLE II
Basic clinic information of the patient population

Patient	Age	Gender	Pathology
1	70	Female	Bronchiectasis
2	45	Male	Asthma
3	69	Female	COPD*
4	48	Male	COPD
5	48	Female	Bronchiectasis
6	72	Female	Asthma
7	66	Female	COPD
8	66	Female	Bronchiectasis
9	61	Female	COPD
10	68	Female	Bronchiectasis
11	65	Female	COPD
12	72	Female	Asthma
13	67	Male	COPD

*COPD: Chronic Obstructive Pulmonary Disease

These included trolleys, phones ringing, babble noise, typing noise, etc.

- Finally, the third part of the protocol was designed to represent noisy environments where the own patients become also a source of contamination because of their movements and other activities. In this case, the patient could move freely around the room while we asked her to perform some activities like turning on/off the radio, opening/closing the window, opening/closing a drawer, moving a chair, washing hands, lying on the bed and standing up immediately, typing, putting on the coat and taking it off immediately, picking up something from the floor, etc. As in part two, the door was left open. Equally, while the patient was performing these activities, we requested her to produce other foreground events as in the first and second part.

Each hour was doubly recorded by two smartphones. The first smartphone (Samsung Galaxy S6 Edge running Android 5.1.1) was placed on a table in the center of the room. The second one (Sony Xperia Z2 with Android 5.1.1) was placed into the pocket or the handbag of the patient. When placed in the handbag, the patient carried it during the third part of the protocol. The acquisition parameters were the same as in the first database. Overall the patient database contained 78 signals lasting 1560 minutes. The percentage of cough samples ranges between 5% and 18% depending on the specific patient.

B. Performance metrics

The performance in our two-class classification problem can be summarized using the confusion matrix in Table III. Given the imbalance between classes, our segmentation process will be mainly assessed by means of the following metrics:

Sensitivity (SEN), as a metric quantifying the capacity of the system to detect a true positive (cough events): $SEN = TP/(TP+FN)$.

Specificity (SPE), as a metric quantifying the performance in detecting a true negative (non-cough events and isolated background sounds): $SPE = TN/(TN+FP)$.

Matthew Correlation Coefficient (MCC), as a metric of the whole performance of a classification process, i.e. equivalent to the accuracy – $ACC = (TP+TN)/(TP+TN+FP+FN)$ – when the classes are unbalanced:

$$MCC = \frac{((TP\cdot TN)-(FP\cdot FN))}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}.$$

Additionally, we provide the positive and negative predictive values (PPV and NPV), since they are the probabilities that the system correctly predicts a randomly-chosen positive or negative sample, respectively. They both depend on the prevalence of the classes: $PPV = TP/(TP+FP)$ and $NPV = TN/(TN+FN)$. MCC, PPV, and NPV provide reliable measures of the performance of the system even in cases where there is clear unbalance between the positive and negative classes.

C. Results

1) Analysis over different SNR values

After training the SVM and the *k*-NN classifier using the training data in the first database, we assessed their performance for detection of cough events using the test group. Classification results are presented in Fig. 4. To test the statistical significance in the comparison between *k*-NN and SVM we ran Mann Whitney's U tests [55] on the test group using 10 different random partitions of the datasets for each SNR and feature set. The obtained *p*-values for sensitivity and specificity are presented in Table IV.

Considering SVM we can see that performance improves for all features as SNR increases. In particular, Hu moments offered the best PPV for all SNRs and their MCC results are also the best ones between -6 dB and 6 dB, although they are outperformed by MFCC for the most favorable SNRs. On the other hand, the sensitivity of Hu moments is the worst for SNRs equal to 0, 6, 9, 12 and 15 dB. In any case, the values of sensitivity are quite low even in the case of the highest SNR: the best sensitivity, around 75%, is reached by OSC for the best SNR.

The rest of features exhibit similar tendencies. For instance, LPCC reported medium values of sensitivity but its specificity is the best just behind the Hu moments, with the exception of SNR equal to 12 dB, in which MFCC are slightly better. The most remarkable aspect in OSC results is their superiority in terms of sensitivity. However, OSC is among the features with lower specificity, which reduces the global classification performance (ACC and MCC). Finally, it is worth highlighting that the feature sets with higher improvements in performance depending on the SNR are SSCH and SpecBlock13.

As for the *k*-NN classifier results, the superiority of Hu moments in all metrics is the major evidence. Their sensitivity is above 80% for all SNRs. The same behavior is observed for specificity, which is above 96% for all SNRs as well. MCC and ACC are also high as could be expected. If we compare Hu moments with the remaining features, the smallest sensitivity difference is 18.59% (LPCC and SNR equal to 15 dB). This pattern is maintained for all the other of metrics, being LPCC, MFCC and GTCC slightly better than the rest of features.

From this analysis, using Hu moments together with a *k*-NN classifier is the best choice. The performance of this combination is the highest for all the metrics and SNR values, with statistical significance on the superiority of using *k*-NN vs. SVM according to Table IV (*p*-value <1.8·10-4 for all SNR values).

TABLE III
CONFUSION MATRIX OF OUR TWO CLASSES CLASSIFICATION PROBLEM

		Predicted Class	
		Non-cough	*Cough*
Real Class	*Non-cough*	True Negative (TN)	False Positive (FP)
	Cough	False Negative (FN)	True Positive (TP)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2800741, IEEE Journal of Biomedical and Health Informatics

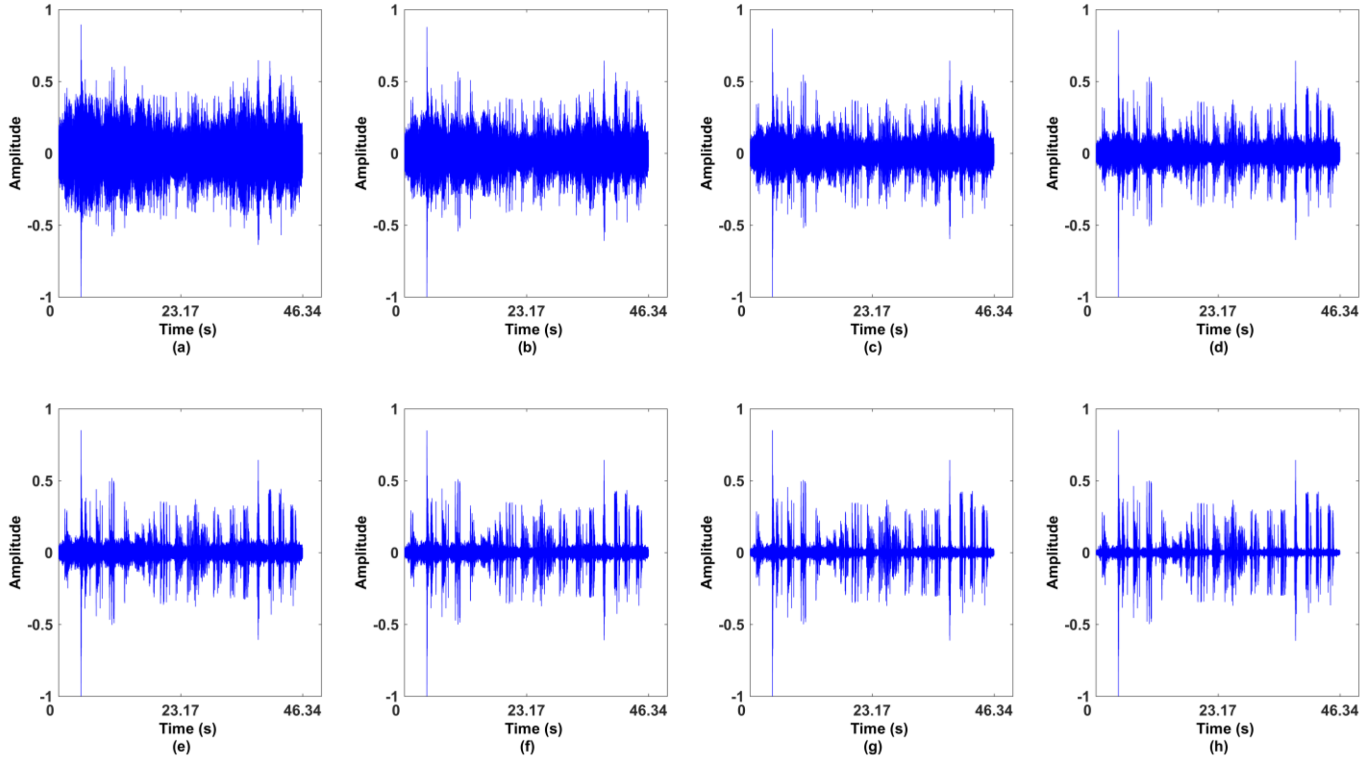

Fig. 3. Representation of the eight SNR versions of one of the synthesized signals: (a) -6 dB; (b) -3 dB; (c) 0 dB; (d) 3 dB; (e) 6 dB; (f) 9 dB; (g) 12 dB; (h) 15 dB.

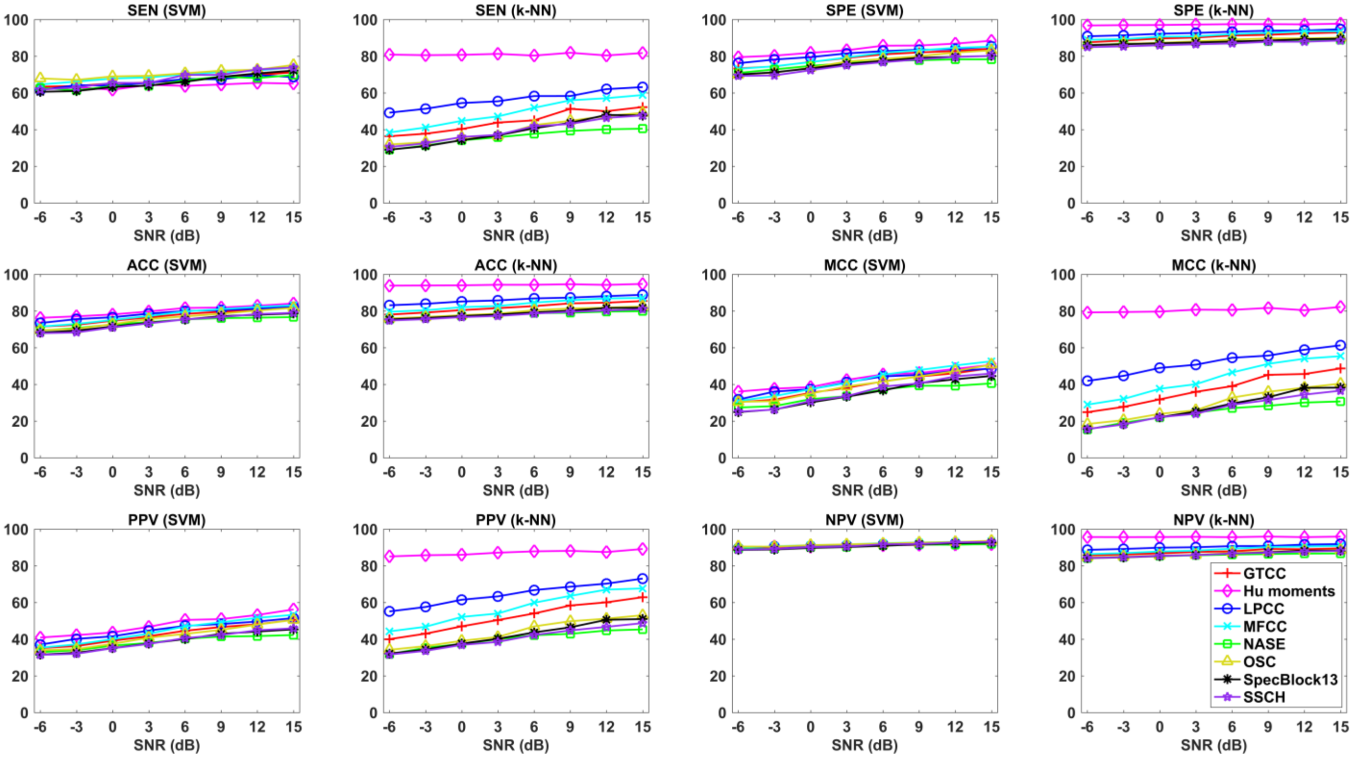

Fig. 4. Pairwise comparison of *k*-NN and SVM in terms of all the performance metrics described in Section IV.B.

2) *Computational load*

Regarding computational efficiency, Table V shows that the feature set demanding less computational resources is LPCC whereas, alternatively, Hu moments require the largest computing time. Concerning the two classifiers, both display the same behavior among all the features. In general terms, the classification task using SVM needs approximately 5 s, whereas with *k*-NN classifier around 7 s. These results are based on a PC with a processor Intel(R) Core(TM) i7-3930K CPU @ 3.20 GHz, 64 GB of RAM and running Windows 7 Enterprise SP1.

TABLE IV
P-VALUES OBTAINED FROM MANN-WHITNEY'S U TEST ON SENSITIVITY AND SPECIFICITY VALUES FOR ALL FEATURE SETS AND SNR VALUES. TESTS FAILING TO REJECT THE NULL HYPOTHESIS AT A=0.05 CONFIDENCE LEVEL ARE SHOWN IN LIGHTER FONT.

SENSITIVITY

Feature set	**SNR (dB)**							
	-6	*-3*	*0*	*3*	*6*	*9*	*12*	*15*
GTCC	$1.82\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.80\cdot 10^{-4}$	$1.70\cdot 10^{-3}$
Hu Moments	$1.80\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.80\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.79\cdot 10^{-4}$
LPCC	0.821	0.473	0.472	$2.70\cdot 10^{-3}$	$5.77\cdot 10^{-4}$	$3.60\cdot 10^{-3}$	0.053	$2.44\cdot 10^{-4}$
MFCC	$1.80\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$2.44\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$3.28\cdot 10^{-4}$	0.028	0.015	0.212
NASE	$1.82\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	0.473	$1.70\cdot 10^{-4}$	0.037	0.0623	0.017	0.025
OSC	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.78\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$
SpecBloc13	$1.82\cdot 10^{-4}$	$2.80\cdot 10^{-3}$	$4.60\cdot 10^{-4}$	0.052	$1.82\cdot 10^{-4}$	$9.10\cdot 10^{-3}$	0.241	$3.60 \cdot 10^{-3}$
SSCH	$1.81\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$

SPECIFICITY

Feature set	**SNR (dB)**							
	-6	*-3*	*0*	*3*	*6*	*9*	*12*	*15*
GTCC	$1.83\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$
Hu Moments	$1.79\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.78\cdot 10^{-4}$	$1.80\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$
LPCC	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.80\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$
MFCC	$1.79\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.79\cdot 10^{-4}$
NASE	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$
OSC	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$
SpecBloc13	$1.77\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	0.473	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$
SSCH	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.81\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$	$1.82\cdot 10^{-4}$

TABLE V
COMPUTING TIME THE EXTRACTION OF EACH FEATURE AND THE CLASSIFICATION TASK BASED ON SVM AND *K*-NN CLASSIFIERS

Feature	Feature computation (s)	SVM classification (s)	*k*-NN classification (s)
GTCC	50.44	5.19	7.18
Hu moments	323.03	5.25	7.19
LPCC	4.10	5.10	7.12
MFCC	23.17	5.55	7.95
NASE	71.90	5.14	7.22
OSC	33.44	5.84	7.28
SpecBlock13	187.2	5.71	7.61
SSCH	130.7	4.95	7.03

3) Superiority of Hu moments in a real scenario with varying noise-levels

In order to verify that our results were indeed true, we carried out a validation process. It aimed to assess the superiority of Hu moments in noisy-background environments. To do so, we used a mixed group of 26 signals from the first database with different SNR values in the range [-6, 15] dB. After computing the Hu moments of this new group we created a partition with the previous percentages of observations for training (60%), test (30%), and validation (10%) and kept the ratio between classes in each group. The obtained results after classification are shown in Table VI. Again, the *k*-NN classifier outperforms the SVM. Moreover, the obtained results are aligned with the ones computed in the previous section, where each SNR version of the database was classified separately. Thus, we confirm the preliminary conclusion that Hu moments are by far the most robust against noisy-background environment among the studied features.

4) Additional improvements from post-processing: final implementation

Table VII presents the classification results in a real scenario with post-processing, showing an improvement on both sensitivity and overall performance.

5) Analysis over ambulatory patient recordings

To show the performance of the final system over real patient data, we separately evaluated the best performing approach above (Hu moments and *k*-NN classifier) using the patient database for each of the three parts of the protocol. The inner parameters of the Hu moments algorithm and the configuration of the *k*-NN classifier were the same as in previous experiments.

For each part of the protocol, the dataset was divided in two groups: 60% of observations for training and 40% for testing. The experiment was based on a repeated random hold-out validation of five experiments, as in other sound event classification problems [56], being the feature space partition of each run different. Likewise, we also applied the post-processing technique described in Section III.D. The final

TABLE VI
CLASSIFICATION RESULTS IN A REAL SCENARIO

Class.	SEN (%)	SPE (%)	ACC (%)	MCC (%)	PPV (%)	NPV (%)
k-NN	82.49	97.25	94.51	81.49	87.23	96.06
SVM	62.83	85.41	81.22	44.21	49.55	90.97

TABLE VII
CLASSIFICATION RESULTS IN A REAL SCENARIO WITH POST-PROCESSING

Class.	SEN (%)	SPE (%)	ACC (%)	MCC (%)	PPV (%)	NPV (%)
k-NN	88.42	96.85	95.28	84.54	86.47	97.35
SVM	66.55	84.04	80.79	45.21	48.73	91.68

classification average results are shown in Table VIII (smartphone on the table) and Table IX (smartphone in the pocket or handbag). In both tables, the standard deviation is always below 2%. Based on the aforementioned tables, the best classification results are obtained in the first part of the protocol (quieter environment) whereas the lowest performance is obtained in the third part (daily activities), both when the smartphone is on the table and when it is in the pocket or handbag. As could be expected, when the smartphone is on the table, the classification results are better (approximately 5% of improvement in terms of SEN) than when it is in the pocket or handbag. Finally, from a general perspective, these classification results are in line with the ones achieved for the synthetic-signal database.

V. Discussion

We studied a variety of feature sets to characterize audio cough signals. Next, a SVM and a *k*-NN classifier were used to separate coughs from non-cough events and background noise.

The most remarkable result is the symbiosis between Hu moments and the *k*-NN classifier. This combination reported results that confirm our hypothesis and make Hu moments the most robust feature against noisy-background environments in comparison with the others. It is worth noting that the achieved classification metrics are significantly high even for low SNR values and also remain almost constant between -6 dB and 15 dB. This means that Hu moments are able to extract practically the same information from cough events despite the degree of contamination due to background sounds. An explanation to this fact may lie in some of the properties of Hu moments as a feature set. Hu moments have successfully been applied for object recognition in image processing. This requires features to be invariant with respect to translation, scaling, and rotation. Hu moments do have these properties [26]. Thus if we understand the variations introduced by the background sounds as the equivalent for signals to translations, scaling and rotation in images, our results confirm that these properties translate to noise robustness in our problem. Likewise, this is also positive to deal with inter- and intra-variability of cough events depending on the user/patient.

In the same line, another explanation is derived from the particular extension of Hu moments to 1D signal processing that Sun et al. carried out [26]. The first steps of the algorithm are shared with a widely used feature set in audio signal processing: MFCC, which actually emulates the response of the cochlea. The difference is that, after getting the energy matrix in each frequency band, MFCC does not consider the relationship between the frequency bands within a window and between windows, whereas Hu moments does it using block processing (see (7)).

Unfortunately, the combination of Hu moments and *k*-NN classifier is the worst in terms of efficiency. However, in our opinion, their large performance outweighs their inefficiency, becoming them also the most cost-effective feature. We hold this opinion since the lack of efficiency can be treated from other perspectives. For instance, today graphics processing units (GPU) are available in many smartphones. This allows to parallel compute a lot of simple operations, exactly what must be done with B_{ij} blocks in the algorithm of Hu moments. Therefore, a GPU-implementation of the Hu moments would increase the opportunities for technological transfer of our research. Alike, other more efficient implementations of the *k*-NN classifier should be tested. We have used an exhaustive *k*-NN classifier, i.e. the distance between each new observation and previous observations is computed at high computational cost. Nevertheless, an efficient implementation of the *k*-NN classifier based on vantage trees can work well without a severe degradation of the classification metrics. Finally, as we proved in [36], some signal processing techniques such as downsampling beyond the Nyquist limit or compressive sensing could additionally be explored to enlarge Hu moments efficiency.

When comparing our results to previous studies, we observe that some of them achieved higher sensitivity values [17], [22], [19], [57] although they cannot be directly compared to our work. For example, Martinek et al. [57] created a monitoring system to distinguish between voluntary cough sounds and speech in healthy volunteers. They achieved sensitivity close to 98% but their spectral analysis was based on windows of 512 samples (45 ms) with a shift of 5 samples. This shift may introduce such amount of correlation between windows that the principle of independency between observations which is supposed in pattern recognition can be questioned. Matos et al. [19] achieved average sensitivity values between 50% and 99%. Our signal database, however, has been specifically designed to emulate multiple noisy background environments with minimal disturbance for the users – e.g. without a microphone attached to the patients' chest, as in [19], so our experimental scenario is more challenging in terms of noise content..

Even though the *k*-NN classifier showed to be the best option for Hu moments, for the rest of features it was the SVM with statistical significance for most of them (see Table IV). Focusing on the SVM results, MFCC presents the best performance in general terms or, in order words, they have acceptable values of both sensitivity and specificity. This can also be appreciated in the MCC results. These results are in accordance with other studies such as Ezgi and Sert [21], where SVM is usually the best performing classifier. Lastly, we would like to discuss the behavior of SSCH. This feature was designed

TABLE VIII
Average classification results of patient-signal database with post-processing (smartphone on the table)

Part	SEN (%)	SPE (%)	ACC (%)	MCC (%)	PPV (%)	NPV (%)
1st	88.51	99.72	99.72	86.85	87.51	99.78
2nd	87.37	99.77	99.59	85.96	84.99	99.81
3rd	86.41	99.70	99.49	84.44	83.02	99.77

TABLE IX
Average classification results of patient-signal database with post-processing (smartphone in the pocket or handbag)

Part	SEN (%)	SPE (%)	ACC (%)	MCC (%)	PPV (%)	NPV (%)
1st	84.27	99.73	99.44	84.63	85.56	99.70
2nd	83.98	99.77	99.53	84.15	84.80	99.86
3rd	79.04	99.69	99.38	79.14	79.87	99.68

to improve the robustness against noise of MFCC. Surprisingly, MFCC resulted to be more robust than SSCH for all SNR and according to all metrics. We believe that this effect is due to different conceptions of noise. Gajic and Paliwal [44] considered only three types of noise: white Gaussian noise, factory noise and babble noise, whereas our database encompasses more noise types. The fact that speech is the target of detection in [44] but here is actually something to discard may also influence. On the other hand, SSCH experience the greatest improvement when increasing SNR.

Regarding performance on real patient data, our results confirm three main points: (1) the suitability of using the first database to identify the best performing method for audio signals; (2) the superiority of Hu moments plus *k*-NN respect to so far-employed audio features and classifiers; (3) the internal coherence of the acquisition protocol for ambulatory recordings.

The first and the second points are confirmed since the results over both databases are strongly aligned, in particular when the smartphone is placed into the pocket or handbag. This is equivalent to lower SNR values in the synthetic database, as opposed to higher in quieter environments. The third point is supported from better classification results in the first part of the protocol (quite environment) compared to the second part (only external source of auditory contamination). These latter are also better than the ones from the third-part, where external noise and daily activity were the sources of contamination). This behavior is observed for both smartphones (the one on the table and the one in the pocket or handbag). In any case, from a holistic perspective, all the performance figures are high for the range of analyzed situations and SNR values, which confirm the suitability of our final proposal for continuous monitoring of audio cough events using a smartphone.

VI. Conclusions

In this work, we present a suitability analysis on the use of several spectral features and two classifiers for audio-cough detection in noisy environments. The analysis led to the proposal of a system using local Hu moments as feature set and a *k*-NN classifier as the best, featuring sensitivity and specificity values up to 88.51% and 99.77% respectively. The evaluation has been carried out using a novel synthesized database including a variety of environment noise types which can be tested with flexible SNR settings and sixteen hours of ambulatory recordings from real respiratory patients.

From the medical point of view, cough is not generally a serious symptom, so patients can self-manage their own respiratory diseases [58]. The availability of a reliable monitoring device can be very helpful to track the evolution of these people, avoiding unreported or fabricated symptoms. The outcome of our research paves the way to create a device which will be convenient and minimally disruptive for patients and in which practitioners can rely on. Besides, thanks to these devices the number of hospitalizations and consultant referrals from respiratory disease may be reduced. This would significantly decrease costs for national health systems.

Appendix. Computation of Spectral Features

The following summarizes the computation of unidimensional spectral features presented in Section III.B.2).

A. Spectral Centroid

$$SpecCen = \sum_{k=1}^{Nend} f[k]\cdot PSD[k] \Big/ \sum_{k=1}^{Nend} PSD[k] \qquad (13)$$

with *f*[*k*]: vector of discrete frequencies.

B. Spectral Bandwidth

$$C_b[k] = f[k] - SpecCen\left(PSD[k]\right) \qquad (14)$$

$$SpecBand = \sum_{k=1}^{Nend} \left(C_b[k]\right)^2 \cdot PSD[k] \Big/ \sum_{k=1}^{Nend} PSD[k] \qquad (15)$$

C. Spectral Crest Factor

$$C_c = \max\left\{ f[k]\right\} - \min\left\{ f[k]\right\} + 1 \qquad (16)$$

$$SpecCresFac = \max\left\{PSD[k]\right\} \Big/ \left(1/C_c\right)\cdot \sum_{k=1}^{Nend} PSD[k] \qquad (17)$$

D. Spectral Turbulence

$$SpecTurb = corr\left\{PSD^{i}[k], PSD^{i-1}[k]\right\} \qquad (18)$$

where $PSD^{i}[k]$ is the PSD of the *i*th signal window and *corr* the correlation coefficient.

E. Spectral Flux

$$SpecFlux = \sum_{k=1}^{Nend} \left(PSD^{i}[k] - PSD^{i-1}[k]\right)^2 \qquad (19)$$

F. Ratio f50 vs f90

$$\sum_{k=1}^{k_{50}} PSD[k] = 0.5\cdot \sum_{k=1}^{Nend} PSD[k] \qquad (20)$$

$$\sum_{k=1}^{k_{90}} PSD[k] = 0.9\cdot \sum_{k=1}^{Nend} PSD[k] \qquad (21)$$

$$f_{50} = f(k_{50}) \text{ and } f_{90} = f(k_{90}) \qquad (22)$$

$$f50\,f90\,Ratio = f_{50}/f_{90} \qquad (23)$$

G. Spectral Roll-off

$$\sum_{k=1}^{k_{85}} PSD[k] = 0.85\cdot \sum_{k=1}^{Nend} PSD[k] \qquad (24)$$

$$SpecRolloff = f(k_{85}) \qquad (25)$$

H. Spectral Standard Deviation, Spectral Skewness and Spectral Kurtosis

$$C_{H1}[k] = 10\cdot\log_{10}\left(PSD[k]\right) \qquad (26)$$

$$C_{H2} = \left(1/Nend\right)\cdot \sum_{k=1}^{Nend} C_{H1}[k] \qquad (27)$$

$$SpecSD = SD\left\{C_{H1}[k]\right\} = C_{H3} \qquad (28)$$

where *SD* refers standard deviation.

$$SpecSkew = \left(1/Nend\right)\cdot\sum_{k=1}^{Nend}\left(C_{H1}[k]-C_{H2}\right)^{3} \Big/ C_{H3}^{\;3} \qquad (29)$$

$$SpecKurto = \left(1/Nend\right)\cdot\sum_{k=1}^{Nend}\left(C_{H1}[k]-C_{H2}\right)^{4} \Big/ C_{H3}^{\;4} \qquad (30)$$

I. Spectral Peak Entropy

The local maxima (lm) of the PSD are first sought to subsequently compute:

$$P = PSD[k_{lm}] \Big/ \sum PSD[k_{lm}] \qquad (2)$$

k_{lm} refers to the discrete frequency at which the lm are found.

$$SpecPeakEn = (-1)\cdot\sum P\cdot\log_{10}\left(P\right) \qquad (31)$$

J. Renyi Entropy

$$RenyiEn = \left(1/(1-q)\right)\cdot\log\left\{\sum_{k=1}^{Nend}\left(PDS[k]\right)^{q}\right\},\ q=2 \qquad (32)$$

K. Tsallis Entropy

$$C_{K1}[k] = PSD[k]-\left(PSD[k]\right)^{q},\ q=2 \qquad (33)$$

$$C_{K2} = 1/q-1 \qquad (34)$$

$$TsallisEn = C_{K2}\cdot\log\left\{\sum_{k=1}^{Nend} C_{K1}[k]\right\} \qquad (35)$$

1024 points were used in the FFT algorithm for the computation of these unidimensional features.

Acknowledgements

The authors would like to thank the Smartcough clinical team at the University of Edinburgh (Prof. Brian McKinstry, Dr. Hilary Pinnock, Dr. Roberto Rabinovich, and Dr. Lucy McCloughan) for their help and support in clinical matters. Thanks are also given to Chest Heart and Stroke Scotland (Lorna Stevenson, Dave Bertin, and Jill Adams) for her support in setting up the patient panel for the Smartcough project.

References

[1] K. K. Lee and S. S. Birring, "Cough," *Medicine*, vol. 40, no. 4, pp. 173-176, Apr. 2012.

[2] G. A. Fontana, "Before we get started: what is a cough?," *Lung*, vol. 186, Suppl. 1, pp. S3-S6, Oct. 2007.

[3] S. A. Walke and V. R. Thool, "Differentiating nature of cough sounds in time domain analysis," *International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 1022-1026.

[4] G. A. Fontana and J. Widdicombe, "What is cough and what should be measures?," *Pulm Pharmacol Ther*, vol. 20, no. 4, pp. 307-312, Dec. 2007.

[5] Clinical standards for chronic obstructive pulmonary disease services. National Health Service. Edinburgh. NHS quality improvement Scotland, 2010.

[6] A. H. Morice, *et al.*, "ERS guidelines on the assessment of cough," *Eur Respir J*, vol. 29, no. 6, pp. 1256-1276, Jun. 2007.

[7] M. J. Fletcher, *et al.*, "COPD uncovered: an international impact of the chronic obstructive pulmonary disease [copd] on a working age population," *BMC Public Health*, vol. 1, no. 11, pp. 612, Aug. 2011.

[8] J. Smith and A. Woodcock, "New developments in the objective assessment of cough," *Lung*, vol. 186, Suppl. 1, pp. S48-S54, Dec. 2007.

[9] C. T. French, *et al.*, "Evaluation of a cough-specific quality-of-life questionnaire," *Chest*, vol. 121, no. 4, pp. 1123-1131, Apr. 2002.

[10] S. S. Birring, *et al.*, "Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ)," *Thorax*, vol. 58, no. 4, pp. 339-343, Apr. 2003.

[11] K. F. Chung, "Measurement of cough," *Respir Physiol Neurobiol*, vol. 152, no. 3, pp. 329-339, Jul. 2006.

[12] K. Kosasih, *et al.*, "High frequency analysis of cough sounds in pediatric patients with respiratory diseases," in *Proc IEEE Annu. Int. Conf. Eng. Med. Boil. Soc.*, 2012, pp. 5654-5657.

[13] J. A. Smith, *et al.*, "The description of cough sounds by healthcare professionals," *Cough,* vol. 2, no. 1, Jan. 2006.

[14] L. PA. McGarvey, *et al.*, "Prevalence of psychomorbidity among patients with chronic cough," *Cough,* vol. 2, no. 4, Jun. 2006.

[15] Communication from the Commission of the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on telemedicine for the benefit of patients, healthcare systems and society, 2008.

[16] J. Smith, "Monitoring chronic cough: current and future techniques," *Expert Rev Resp Med*, vol. 4, no. 5, pp. 673-683, Oct. 2010.

[17] Y. A. Amrulloh, *et al.*, "Automatic cough segmentation from non-contact sound recordings in pediatric wards," *Biomed Signal Process Control*, vol. 21, pp. 126-136, Aug. 2015.

[18] S. Larson, *et al.*, "Validation of an automated cough detection algorithm for tracking recovery of pulmonary tuberculosis patients," *PLoS One*, vol. 7, no. 10, pp. e46229, Oct. 2012.

[19] S. Matos, *et al.*, "Detection of cough signals in continuous audio recordings using Hidden Markov Models," *IEEE Trans Biomed Eng.*, vol. 53, no. 6, pp. 1078-1083, Jun. 2006.

[20] K. Yatani and K. N. Truong, "BodyScope: a wearable acoustic sensor for activity recognition," in *Proc. ACM Annu. Int. Conf. Ubiq. Comp.*, 2012, pp. 341-350.

[21] S. Ezgi and M. Sert, "Audio-based event detection in office live envrionments using optimized MFCC-SVM approach," in *Proc. IEEE Annu. Int. Conf. Semantic Computing*, 2015, pp. 475-480.

[22] T. Drugman, *et al.* (2012, Sep.). Audio and contact microphones for cough detection. *INTERSPEECH 13th Annu. Int. Conf. Speech Communication Association*. [Online]. Available: http://tcts.fpms.ac.be/publications/papers/2012/interspeech2012_cough_tdjurctd.pdf

[23] T. Drugman, "Using mutual information in supervised temporal event detection: application to cough detection," *Biomed Signal Process Control*, vol. 10, pp. 50-57, Mar. 2014.

[24] E. Agu, *et al.*, "The smartphone as a medical device: assessing enablers, benefits and challenges," *IEEE Int. Workshop of IoT-NC*, 2013, pp. 48-52.

[25] J. Monge-Álvarez, *et al.*, "Effect of importance sampling on robust segmentation of audio-cough events in noisy environments," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3740-3744

[26] Y. Sun, *et al.*, "Weighted spectral features based on local Hu moments for speech emotion recognition," *Biomed Signal Process Control*, vol. 18, pp. 80-90.

[27] M. A. Coyle, *et al.*, "Evaluation of an ambulatory system for the quantification of cough frequency in patients with chronic obstructive pulmonary disease," *Cough*, vol. 1, no. 3, Aug. 2005.

[28] S. J. Barry, *et al.*, "The automatic recognition and counting of cough," *Cough*, vol. 2, no. 8, Sep. 2006.

[29] S. S. Birring, *et al.*, "The Leicester cough monitor: preliminary validation of an automated cough detection system in chronic cough," *Eur Respir J*, vol. 31, no. 5, pp. 1013-1018, May. 2008.

[30] E. Vizel, *et al.*, "Validation of an ambulatory cough detection and counting application using voluntary cough under different conditions," *Cough*, vol. 6, no. 3, May. 2010.

[31] M. You *et al.*, "Novel feature extraction method for cough detection using NMF," *IET Signal Processing*, vol. 11, no. 5, p. 515-520, 2017.

[32] J-M. Liu *et al.*, "Cough detection using deep neural networks," in *IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, 2014, pp. 560-563.

[33] J. Amoh and K. Odame, "Deep Neural Networks for identifying cough sounds," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 5, pp. 1003-1011, Oct. 2016.

[34] K. Kosasish, *et al.*, "Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis," *IEEE Trans. Biomed. Eng.,* vol. 62, no. 4, pp. 1185-1194, Apr. 2015.

[35] J. MacAuslan, "Cough Analysis," U.S. Patent 20170055879 A1, Mar. 2, 2017.

[36] P. Casaseca-de-la-Higuera, *et al.*, "Effect of downsampling and compressive sensing on audio-based continuous cough monitoring," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2015, pp. 6231-6235.

[37] S. Haykin, *Communication systems*, John Wiley & Sons, New York, 2001.

[38] K. Tokuda, *et al.*, "Mel generalized cepstral analysis – a unified approach to speech spectral estimation," in *Proc. Annu. Int. Conf. on Spoken Language and Processing*, vol. 3, 1994, pp. 1043-1046.

[39] R. Mammone, *et al.*, "Robust speaker recognition: a feature-based approach," *IEEE Signal Process Mag,* vol. 13, no. 5, pp. 58-71, 1996.

[40] J. M. Liu, *et al.*, "Cough signal recognition with gammatone cepstral coefficients," in *Proc. IEEE Int. Conf. on Sig. and Img. Processing*, 2013, pp. 160-164.

[41] X. Valero and F. Alías, "Gammatone cepstral coefficients: biologically inspired features for non-speech audio classification," *IEEE T Multimedia,* vol. 14, no. 6, pp. 1684-1689, Dec. 2012.

[42] C. H. Lee, *et al.*, "Automatic music genre classification using modulation spectral contrast feature," *IEEE T Multimedia*, vol. 11, no. 4, pp. 670-682, May. 2009.

[43] H. G. Kim, *et al.*, "Audio classification based on MPEG-7 spectral basis representations," *IEEE T Circ Syst Vid,* vol. 14, no. 5, pp. 716-725, May. 2004.

[44] B. Gajic and K. K. Paliwal, "Robust speech recognition in noisy environments based on subband spectral centroid histograms," *IEEE T Speech Audi P,* vol. 14, no. 2, pp. 600-608, Mar. 2006.

[45] A. Ramalingam and S. Krishman, "Gaussian mixture modeling of short-time Fourier transform features for audio fingerprinting," *IEEE T Inf Foren Sec,* vol. 1, no. 4, Dec. 2006.

[46] P. R. B. Barbosa, *et al.*, "Spectral turbulence analysis of the signal-averaged electrocardiogram of the atrial activation as predictor of recurrence of idiopathic and persistent atrial fibrillation," *Int J Cardiol,* vol. 107, no. 3, pp. 307-316, Mar. 2006.

[47] X. Chen and P. J. Ramadge, "Music genre classification using multiscale scattering and sparse representations," in *Proc. IEEE Annu. Conf. on Information Sciencie and Systems,* 2013, pp. 1-6.

[48] M. Wisniewski and T. P. Zielinski, "Application of tonal index to pulmonary wheezes detection in asthma monitoring," in *Proc. Annu. European Conf. Signal Processing*, 2011, pp. 1544-1548.

[49] J. Poza, *et al.*, "Regional analysis of spontaneous MEG rhythms in patients with Alzheimer's disease using spectral entropies," *Ann Biomed Eng,* vol. 36, no. 1, pp. 141-152, Nov. 2007.

[50] S. Haykin, "Support Vector Machines," in *Neural networks: a comprehensive foundation*, 2nd ed., Upper Saddle River, NJ, 1998, pp. 340-365.

[51] R. O. Duda, P. E. Hart, D. G. Stork, "Nonparametric techniques", in Pattern classification, 2nd ed., Ed. Wiley, 2000, pp. 174-188.

[52] Theodoros Giannakopoulos and Aggelos Pikrakis. *Introduction to Audio Analysis: A MATLAB Approach* (1st ed.). Academic Press, 2014..

[53] 4uall, http://www.4uall.net/free-sound-effects/, latest visit: 15/11/2015.

[54] Universal soundbank, http://eng.universal-soundbank.com/, latest visit: 15/11/2015.

[55] B. Rosner, Fundamentals of Biostatistics. Pacific Grove, CA: Duxbury Thomson Learning, 2000.

[56] J. Dennis, *et al.*, "Spectrogram image feature for sound event classification in mismatched conditions," *IEEE Signal Process. Lett.*, vol. 18, no. 2, pp. 130-133, Feb. 2011.

[57] J. Martinek, *et al.*, "Distinction between voluntary cough sounds and speech in volunteers by spectral and complexity analysis," *J Physiol Pharmacol*, vol. 59, suppl. 6, pp. 433-440, Dec. 2008.

[58] P. G. Gibson, *et al.*, "Self-management education and regular practitioner review for adults with asthma," *Cochrane Database Syst Rev,* vol. 1, CD001117, Jul. 2002.