Objective Study of Sensor Relevance for Automatic Cough Detection

Thomas Drugman, Jerome Urbain, Nathalie Bauwens, Ricardo Chessini, Carlos Valderrama,
Patrick Lebecque, and Thierry Dutoit

Abstract—The development of a system for the automatic, objective, and reliable detection of cough events is a need underlined by the medical literature for years. The benefit of such a tool is clear as it would allow the assessment of pathology severity in chronic cough diseases. Even though some approaches have recently reported solutions achieving this task with a relative success, there is still no standardization about the method to adopt or the sensors to use. The goal of this paper is to study objectively the performance of several sensors for cough detection: ECG, thermistor, chest belt, accelerometer, contact, and audio microphones. Experiments are carried out on a database of 32 healthy subjects producing, in a confined room and in three situations, voluntary cough at various volumes as well as other event categories which can possibly lead to some detection errors: background noise, forced expiration, throat clearing, speech, and laugh. The relevance of each sensor is evaluated at three stages: mutual information conveyed by the features, ability to discriminate at the frame level cough from these latter other sources of ambiguity, and ability to detect cough events. In this latter experiment, with both an averaged sensitivity and specificity of about 94.5%, the proposed approach is shown to clearly outperform the commercial Karmelsonix system which achieved a specificity of 95.3% and a sensitivity of 64.9%.

Index Terms—Audio processing, biomedical engineering, cough detection, cystic fibrosis, multimodal, neural network, sensor.

I. INTRODUCTION

COUGH is defined in physiology textbooks as a three-phase act but also, for clinical purposes, as an expiratory maneuvre against a closed glottis, which produces a characteristic sound. It is a protective mechanism of the proximal respiratory tract. It may be voluntary or reflex. Cough can appear sporadically with common illnesses (e.g., cold), but when it becomes chronic it can severely impair life quality. This symptom is the commonest reason for which people seek medical advice [1]. It concerns one-third of pulmonologist consultations [2]. The

assessment of cough severity is an essential tool in clinical use. It requires a combination of measures characterizing cough frequency, intensity, and its impact on quality of life.

The evaluation of cough severity is currently subjective: it is based on cough scores, diaries, visual analogue scales, and symptom questionnaires, which are completed either by the patient himself or a parent [3], [4]. However, it has been shown that subjective assessments correlate modestly with objective measures of cough frequency [5]. The medical literature underlines the lack of an objective and reliable tool to measure the severity of this symptom [6]. Such a tool is also necessary to test the effectiveness of cough treatments (e.g., antitussives) or novel therapies.

Cough events can be counted manually on audio recordings but this task remains extremely laborious and time consuming, and hence economically unfeasible. Recent technological advances have enabled the development of automated and ambulatory cough monitors. Existing systems use audio signal(s) alone (microphone and/or contact microphone) or in combination with other sensors such as accelerometers, pneumographic belt, electromyography electrodes, electrocardiography electrodes, induction plethysmography, and pulse oxymeter. However, as it is underlined by the European Respiratory Society Committee, there is currently neither standardized methods for recording cough nor adequately validated, commercially available, and clinically acceptable cough monitors [6].

Automatic cough monitors have to face several challenges. First, cough can be confused with surrounding noise and other parasitic patient sounds (throat clearing, laughing, forced expiration, etc.). Second, the acoustic properties of cough events vary with individuals and diseases, but also within individuals. Finally, it is desirable to monitor cough continuously for long periods (over 24 h) and in the patient's own environment.

A few cough monitors have been developed so far [7]. It started around the 1950s with simple audio recording systems enabling to manually spot the cough events. It is only recently that (semi- or fully-) automated cough recorders have been designed.

Some cough monitors use the audio signal only. In 2006, the Hull Automated Cough Counter (HACC) [8] was developed. It consists of a single audio signal fed into an artificial neural network (ANN) for detecting cough events. The number of cough components per event is not computed by the system, but can be manually determined. The system presents a sensitivity of 80% (ranging from 55% to 100% across the ten validation patients) and a specificity of 96%. The Leicester Cough Monitor (LCM) [9] also relies only on audio recordings. It enables

24-h ambulatory recordings. Hidden Markov models are used to presegment possible cough events. Some of these possible cough segments are then presented to a human expert in order to develop a statistical model tailored to the current recording. Finally, the full recordings are processed with the developed models, achieving overall sensitivity and specificity of 91% and 99%, respectively. Matos *et al.* [10] also use Hidden Markov models trained on audio features only, with an approach inspired by keyword spotting: models are developed to characterize cough events, but also to represent the set of all other possible events (named "fillers"). The two models compete to score new recordings and the most likely sequence of coughs and fillers is retained. This system achieves a sensitivity of 78%, ranging from 50% to 99% depending on the subjects, with false alarm rates ranging from 2 to 31/h. The performance is lower than the LCM but here the process is fully automatic. The Vitalojak system [11] uses only a contact microphone placed on the thorax for a semiautomated detection. A lapel microphone is added to the system for manual validation. This system has been validated in a 24-h ambulatory context on ten patients [12]. It reaches a sensitivity higher than 99% while compressing the amount of data to check manually with three possible levels, ranging from 65 to 23 min on average. In [13], we proposed an audio-only system based on ANNs which achieved a sensitivity and specificity of about 95% on voluntary cough from ten healthy subjects in various conditions.

Two commercial systems rely on multiple signals to detect cough. In 2005 appeared the Lifeshirt [14], consisting of several sensors integrated in a shirt worn by the user: electrocardiogram, induction plethysmography, 3-axis accelerometer, and a contact microphone placed on the throat. The device achieved sensitivity and specificity of 78.1% and 99.6%, respectively. It must be noted that the Lifeshirt is no longer available due to the liquidation of the company in 2009. In 2010, the Karmelsonix company launched the PulmoTrack-CC [15], which includes a piezoelectric belt for monitoring respiration, one lapel microphone and two contact microphones placed on the trachea and the thorax. The performance of this device is reported in [15] to reach a sensitivity of 91% and a specificity of 94% on voluntary cough. Although it has not been clinically validated, it is currently commercialized and will be used as a comparison point in our experiments.

In this paper, we present the first development steps toward a new, reliable, ambulatory cough monitor, following the recommendations of the European Respiratory Society. This paper focuses on the selection of sensors and the development of algorithms to provide increased recording capabilities, detection rates, and robustness (i.e., efficiency for all the subjects) compared to the existing systems. Clinical validation will be performed in the near future. This paper is organized as follows: the acquisition system is presented in Section II; the database recorded to develop and evaluate the cough monitor is presented in Section III; the developed framework for cough detection is detailed in Section IV; results of feature selection are described in Section V; Section VI focuses on the discrimination of cough with regard to other event categories (background, forced expiration, throat clearing, speech, and laugh); in Section VII finally
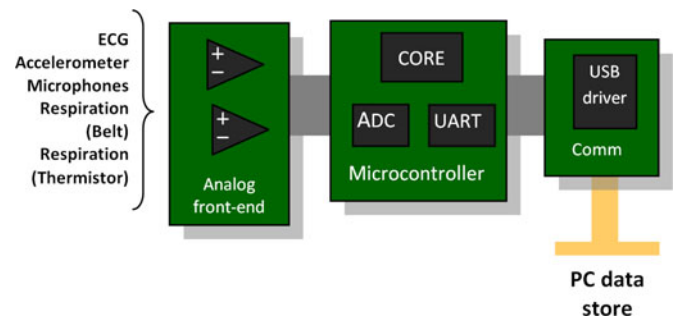

Fig. 1. Block diagram of the acquisition system.

a comparison between the proposed approach and the commercially available Karmelsonix system is carried out; finally Section VIII draws the conclusions of this study.

II. ACQUISITION SYSTEM

The acquisition system aims at capturing and storing data from sensors. Signals are not analyzed or processed along this step, which is performed offline (as explained in Section IV). The block diagram of the acquisition system is displayed in Fig. 1. It is implemented as a classical acquisition system by means of sensors, analog signal conditioning circuit (front-end), analog-to-digital conversion, communication, and storing functional blocks. The general specifications of the acquisition system are the following.

1) *Sensors*: Six sensors are used in this study: a) a disposable ECG monitoring electrodes $40 \times 36 \times 1$ mm; b) a SLP nasal thermocouple flow sensor from Medatec (specification not available) measuring the thermical flow at the patient's nose; c) a piezo respiratory effort belt; d) an ADXL103 analog dual axis accelerometer from *Analog Device* with ± 1.7 g/± 5 g/± 8 g of precision and ± 1 mg of resolution; e) a contact omnidirectional microphone placed on the throat, with 42 dB of signal-to-noise ratio and 40–12000 Hz of bandwidth, and with 2 mm of silicone cup; and f) a lapel audio electrets omnidirectional microphone featuring 60 dB of signal-to-noise ratio and 50–16000 Hz of bandwidth.

2) *Analog front end*: Each sensor has an amplifier circuit based on an AD620 instrumentation amplifier from *Analog Device* featuring 100 dB of common-mode rejection, 50 μV max input offset and 120 KHz of bandwidth. The analog circuits are supplied by a dedicated symmetric power of 5 V. Each sensor circuit was designed to adjust the signal between 0 and 2.6 V of amplitude and no more than 10 kHz of frequency as requirements.

3) *Analog-to-digital conversion*: This step is performed by an eight-channel and 12-bit ADC as part of ADuC841 microcontroller from *Analog Device*. For the two microphones, acquisition is made at 10 000 samples/s. For other sensors, since their signal (which we call *biosignals* in opposition to the acoustic signals captured by the microphones) have a much slower dynamics, the sampling rate is fixed to 1250 Hz.
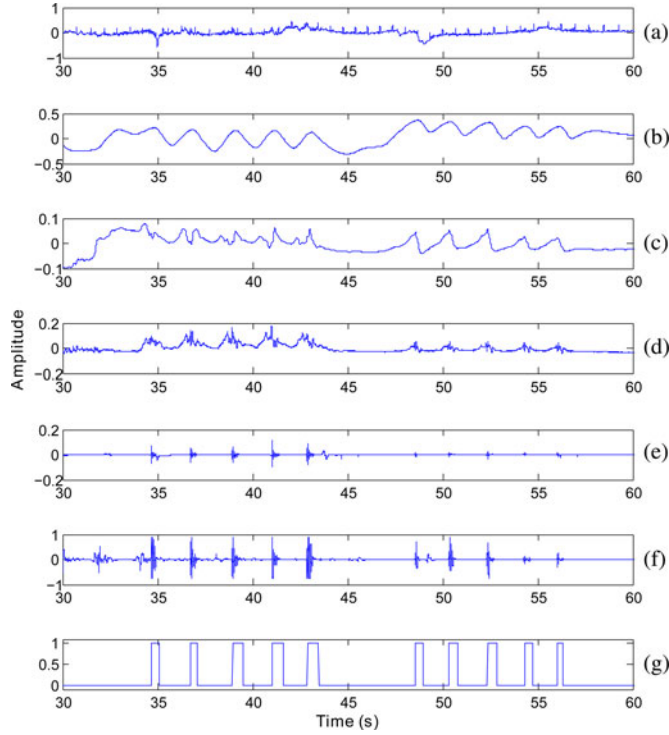
Fig. 2. Example of signals captured by the various sensors: (a) ECG, (b) thermistor, (c) chest belt, (d) accelerometer, (e) contact microphone, (f) audio microphone, and (g) manual cough event label.

TABLE I
STANDARDIZED PROTOCOL FOR DATA RECORDINGS

Sounds	Situations		
	Sitting down, quiet environment (n)	Sitting down, noisy environment (n)	Climbing/going down a stepladder (n)
High volume cough	5	5	5
Interm. vol. cough	5	5	5
Low vol. cough	5	5	5
Fit of coughing	3	3	3
Forced expiration	3	3	3
Throat clearing	5	5	5
Speaking	14	14	14
Laughing	3	3	3

periods. Finally the signals acquired by the contact and audio microphones are shown in panels (e) and (f), respectively. Although the contact microphone is more robust to ambient noise, it suffers from other acquisition perturbations, notably due to the fact that with such a sensor, it is difficult in some conditions (subject moving, etc) to maintain a good contact with the skin, leading to some degradation during the acquisition (albeit it is not the case in Fig. 2). The use of a contact microphone with better attachment might alleviate these robustness issues.

III. DATABASE

The study population was divided into two groups. The first set (A) included 22 healthy subjects (9 males, mean age $\pm$ SD: 22.8 ± 2.44, range: 20–28). The second set (B), consisting of ten additional healthy subjects (5 males, mean age $\pm$ SD: 23 ± 1.45, range: 22–26) was designed to compare our system to the commercially available KarmelSonix cough counter [15]. It is worth noting that these recordings were made across several sessions and in different rooms.

The aim of the database was to record various cough sounds but also some other sounds which are typically confused with cough. The participants followed a standardized protocol performed in three different situations, as detailed in Table I: 1) sitting down in a quiet environment; 2) sitting down in a noisy environment; and 3) climbing on/going down of a stepladder.

This protocol was inspired by the one used to develop and evaluate the PulmoTrack-CC [15], with the addition of low and intermediate volume coughs as these kinds of cough are more difficult to detect for automatic cough counters. All recordings have been precisely manually annotated by a trained observer. In total, the database contains 2338 coughs (among which 864 are from fits of coughing), 289 forced expirations, 479 throat clearings, 289 laughters, for a total duration of 237 min. Note that slight deviations were observed from the strict protocol, but that the manual annotation was made coherently.

IV. COUGH DETECTION FRAMEWORK

The general workflow for the automatic detection of cough used throughout this paper is displayed in Fig. 3. From the signals captured by a given sensor, or by several sensors in a

4) *Communication*: This step is based on a USB communication achieved by FT232R USB-to-serial component from FTDI, with intermediate serial UART communication from ADuC841 and a USB communication from a PC on 115200 bits/s of data rate.

5) *Storing*: A software in C++ manipulates the data from USB and stores it as *raw* files in the PC hard disk.

The key idea here is to be as exhaustive as possible by analyzing the performance of several sensors. Some of them, like the audio and contact microphones, the accelerometer or the chest belt, have been used in various existing cough monitors [11]. Some others (e.g., the nasal thermistor, or the ECG) are less common for detection, and their inclusion in this study is more exploratory by investigating their potential comparatively to more usual sensors.

Fig. 2 illustrates the signals recorded from these sensors by the acquisition system during a period of coughing (as indicated by the manual annotations in the last panel). Heartbeat pulses can be seen on the ECG recording [panel (a)]. This signal can exhibit some artifacts during cough events which can be turned into advantage for their detection [such as a bit before 35 s in panel (a)]. As observed in panels (b) and (c), the signals captured from the thermistor and the chest belt are mainly informative about the respiratory volume. They are both very slow-varying and show local maximum values around the cough instants, characterized by important expiratory flow and respiratory volume variations. The signal from the accelerometer [panel (d)] is clearly observed to present a much higher activity in the coughing neighborhood due to the abdominal contractions in such
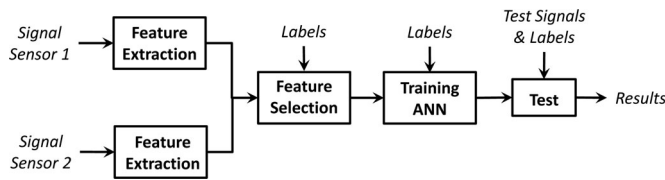
Fig. 3. Workflow used in this study for the automatic multimodal detection of cough.

multimodal approach, the first step aims at extracting a wide variety of features, as described in Section IV-A. Since this leads to a prohibitive number of features, a step of dimensionality reduction is necessary by selecting only the most relevant ones (see Section IV-B). This is here achieved based on some measures derived from the information theory. Finally, an ANN-based classifier is used for accurately modeling the feature distributions and for drawing the final cough detection decision (see Section IV-C).

A. Feature Extraction

Many features, both in the time and frequency domains, have been proposed in the literature to describe the characteristics of a signal. The key idea here is to extract the largest variety of features among which only the most relevant will be selected (see Section IV-B). In other words, we target at a comprehensive description of the signals from which data mining techniques will retain only the most relevant for our classification problem.

Features used in this study can be categorized into three classes.

1) *Features Describing the Waveform:* These descriptors are simply extracted from the signal waveform in the time domain. They consist of the signal amplitude at the analysis time as well as the signal slope computed at different scales ranging from 1.28 to 0.04 s. These features might be of interest for slow-varying signals, for example for the thermistor or the chest belt which capture respiratory characteristics. Unlike the next two categories, they are in the following extracted only for biosignals, and are, therefore, not considered for the two microphones for which they are not appropriate.

2) *Features Describing the Spectral Contents:* Several features characterizing the spectral contents have been proposed in [16]. For a comprehensive description of the magnitude spectrum, we extracted the three following types of descriptors: the Mel-Frequency Cepstral Coefficients (*MFCCs*, [17]) as widely used in automatic speech recognition, the specific *loudness* associated to each Bark band [18], and the *relative energy* in different frequency subbands linearly spaced by 500 Hz. In order to be as exhaustive as possible, we also used the *chroma* features [19] as widely done in music processing, although their potential might be limited for our issue.

Besides, several parameters describing the spectral shape are also employed, as proposed in [16]. The *spectral centroid* is defined as the barycenter of the amplitude spectrum. Similarly, the *spectral spread* is the dispersion

of the spectrum around its mean value. The *spectral decrease* is a perceptual measure quantifying the amount of decreasing of the spectral amplitude [16]. The *Spectral Variation* and *spectral flux* characterize the amount of variations of spectrum along time and are based on the normalized cross correlation between two successive amplitude spectra [16]. Finally, we also use the *energy* and *total loudness* which are mainly informative about the presence of activity in the signal.

3) *Features Measuring the Amount of Noise:* Quantifying the level of noise may be of interest for describing the cough sound for audio signals, or for detecting movements from accelerometer recordings. For this, several measures are extracted here. First, the *harmonic to noise ratio* (HNR) is calculated in the four frequency ranges [0–0.5 kHz], [0–1.5 kHz], [0–2.5 kHz], and [0–3.5 kHz]. The *spectral flatness* measures the noisiness/sinusoidality of a spectrum (or a part of it) in four frequency bands [16]. The *zero-crossing rate* quantifies the number of times the signal crosses the zero axis. It is expected that the greater the amount of noise, the higher the number of zero-crossings. The F_0 value and its related measure of periodicity based on the Summation of Residual Harmonics [20] are used as voicing measurements. As a last parameter quantifying the amount of noise in the audio signal, the *chirp group delay* is a phase-based measure proposed in [21] and shown to be suited for highlighting turbulences in the signal.

These features are extracted every 12 ms for all sensors so as to have synchronous streams. As mentioned previously, the first feature category was only extracted for biosignals, while the two other categories were used in all cases. For microphones, the window length was fixed to 30 ms, a standard value in audio processing, while two frame sizes were used for the biosignals: 500 ms and 1 s. Finally, we also added the first and second derivatives for each of these features in order to integrate the signal dynamics.

B. Feature Selection

A very large number of features have been extracted in Section IV-A: stud2 for the two microphones, and 369 for the biosignals (including the first and second derivatives). The goal of the feature selection algorithm is to retain the most relevant ones so as to alleviate the effect of the curse of dimensionality [22]. For this, we here make use of measures based on the mutual information. This allows to assess the intrinsic discrimination power of each feature separately, but also their possible complementarity or redundancy, and this independently of the subsequent classifier.

The feature selection algorithm used in this study has been proposed in [23]. It is a greedy method, which at each iteration chooses the feature conveying the greatest amount of new relevant information for the considered classification problem. This latter measure is estimated by considering the mutual information of this feature and its redundancy with the selected subset.
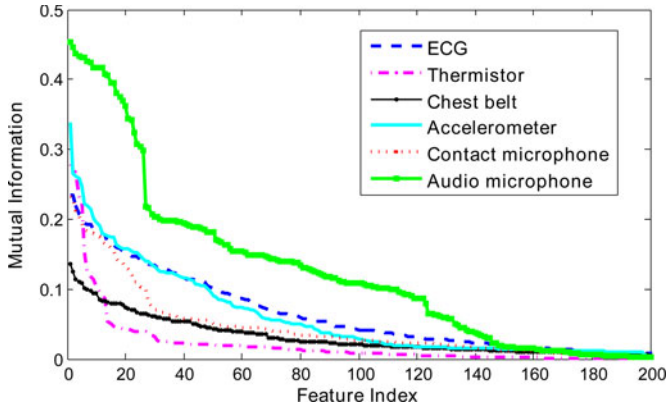
Fig. 4. Mutual information between a feature (considered independently) and the classes, for the 200 best features and for all sensors.

Fig. 5. Mutual information for the best feature (in blue), and two best features (considered jointly—in red) of each sensor.

In the rest of this paper, only 20 features are used for each configuration (i.e., for a particular signal or for a possible combination of several sensors). The relevance of these features is discussed throughout Section V.

C. Artificial Neural Network-Based Classification

For each of the issues tackled across experiments, a dedicated ANN has been trained. Our ANN implementation relies on the MATLAB Neural Network toolbox. Each ANN is made of a single hidden layer consisting of neurons (fixed to 32 neurons in this study) whose activation function is an hyperbolic tangent sigmoid transfer function. The output layer is a simple neuron with a logarithmic sigmoid function suited for a binary decision.

In order to provide contextual information to the ANNs, the feature vector at the considered analysis time is appended with its values 50 and 100 ms both in the past and in the future. Finally, note that when testing, the posterior probability of cough detection, provided by the ANN, is smoothed by a median filtering over a period of 50 ms so as to remove erroneous isolated decisions.

V. RESULTS OF FEATURE SELECTION

The various features for each sensor are first evaluated based on their mutual information with the classes, i.e., on their discrimination power for detecting cough. The computation of mutual information requires to estimate probability density functions of the features jointly with the classes. This is here carried out via a histogram approach [23]. The number of bins is set to 50 for each feature dimension, which results in a tradeoff between an adequately high number for an accurate estimation, while keeping sufficient samples per bin. Class labels correspond to the presence (or not) of a cough event. These measures are then normalized over the class entropy. This ensures a more intuitive interpretation with values ranging between 0 (the feature does not convey any relevant information) and 1 (the feature covers 100% of the classification information).

Mutual information measures are exhibited in Fig. 4 for the best 200 features *considered independently* for each sensor (therefore their redundancy is not taken into account). It
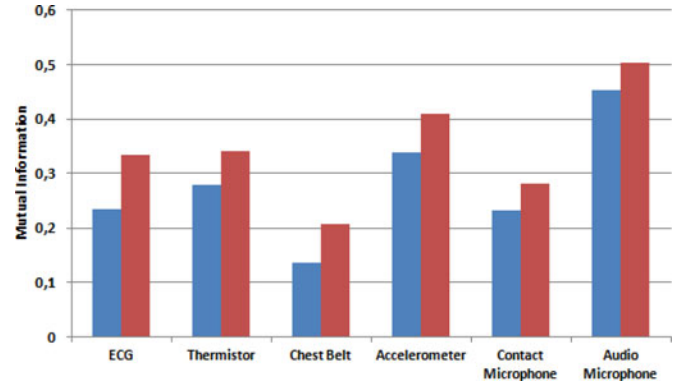
is clearly seen that the audio microphones provides the largest amount of relevant features. It is followed by the accelerometer and the ECG which seem to give a similar amount of information. It turns out for the thermistor that a limited number of features (about 15) bring an interesting amount of information, while its other features are poorly informative. Finally, the worst sensor is observed to be the chest belt.

Fig. 5 shows the mutual information for the best feature and the best two features considered jointly, i.e., taking their redundancy/synergy into account. This study of redundancy is only possible for two features since investigating more features would require to consider probability functions in spaces of high dimension, where they cannot be estimated accurately [22], [23]. With only two features, the best sensors are observed to be: the audio microphone (0.50), the accelerometer (0.41), the thermistor and the ECG (with both about 0.34), the contact microphone (0.26), and finally the chest belt turns out to be the worst sensor (0.20).

The most interesting features for each sensor, which were chosen by the feature selection algorithm, are 1) for the ECG, the specific loudness in the fourth and fifth Bark bands (from 204 to 399 Hz)) and their derivatives; 2) for the thermistor, the signal slopes at various scales (from 160 ms to 1.28 s) and the signal amplitude at the analysis time; 3) for the chest belt, the energy over a window of 500 ms, and the signal slopes at various scales (from 160 to 640 ms); 4) for the accelerometer, the specific loudness in the second and fourth Bark bands (from 13 to 297 Hz) and their derivatives, and the energy in a 1s-long window; 5) for the contact microphone, the specific loudness in the Bark bands 7, 11, 12, and 19 (respectively, [631–765], [1268–1479], [1479–1720], and [4386–5000] Hz); and 6) for the audio microphone, the total loudness and its derivative, as well as the first derivative of the specific loudness in the Bark bands 18 and 19 (from 3702 to 5000 Hz).

VI. DISCRIMINATING COUGH FROM OTHER EVENT CATEGORIES

Several event categories can be sources of ambiguity for the final cough detection issue, and the performance of a given sensor may vary across these categories. For example, it might turn
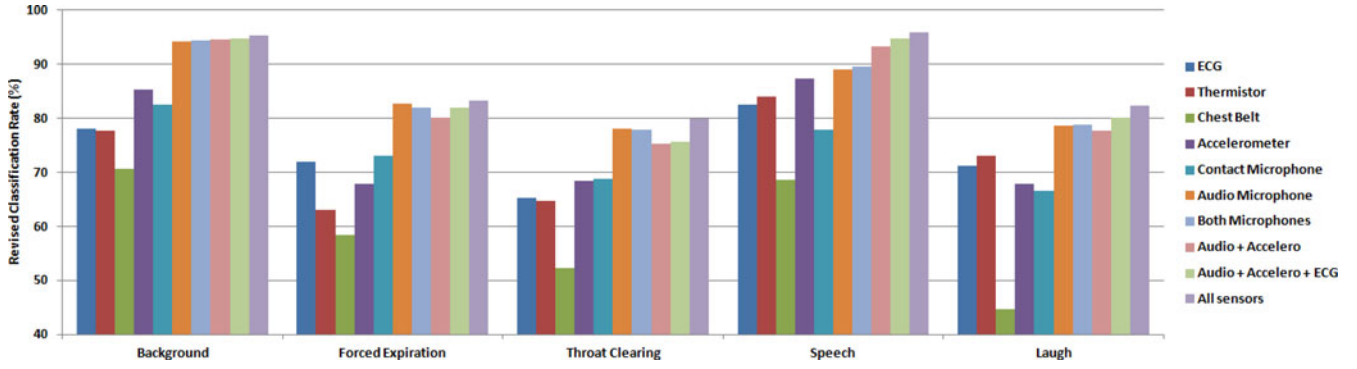
Fig. 6. For each of the problems discriminating cough with another event category (indicated on the horizontal axis), the RCR is given at the frame level using various sensors and their possible combinations.

out that a sensor discriminates well cough from laugh but has a poor performance for the differentiation with throat clearing. We, therefore, focus in this section on the detection of cough with regard to various other categories of events: background, forced expiration, throat clearing, speech, and laugh.

For this experiment, we only consider segments of the whole database (sets A and B) containing either cough or the considered category. For each one of these two categories, a specific ANN is trained. Performance is evaluated at the frame level using a leave-four–subjects-out cross-validation procedure. In other words, 28 out of the 32 subjects are used for the training and the four remaining are used for the test. This operation is repeated eight times so as to cover the whole database for the test.

Metrics we use at the frame level are the true positive rate (TPR) and false positive rate (FPR). By varying the threshold decision θ (applied on the posterior probability outputted by the ANN), a receiver operating characteristic (ROC) curve is obtained in the (FPR,TPR) plane [24]. As a single measure of performance at the frame level, we use the revised classification rate (RCR), defined as

$$\text{RCR} = 1 - \min_{\theta} \sqrt{(1 - \text{TPR}(\theta))^2 + \text{FPR}(\theta)^2}. \quad (1)$$

Indeed, an ideal classifier being characterized by a TPR = 1 and an FPR = 0, a single measure of performance is the revised error rate (RER = 1 − RCR) defined as the Euclidian distance from the top-left corner to the ROC curve. As a consequence, the higher RCR (the lower RER) the better the system. This criterion implies that an equal importance is given to both TPR and FPR criteria. Based on a medical advice, TPR or FPR could be emphasized by weighting its importance in (1).

Fig. 6 displays the RCR values obtained for the detection of cough with regard to each event category using the various sensors or their possible combination. The most ambiguous event classes turn out to be the throat clearing, forced expiration, and laugh, for which a rather similar performance is carried out. For these three classes, the best results (achieved with all sensors) reach a RCR varying between 80% and 83%. On the opposite, the distinction with the background and segments of speech leads to RCR values exceeding 95%.

Regarding the potential of the various sensors, it is clearly observed that the chest belt yields the worst results, and this

TABLE II
REVISED CLASSIFICATION RATE AT THE FRAME LEVEL USING VARIOUS SENSORS AND THEIR POSSIBLE COMBINATIONS FOR THE GENERAL COUGH DETECTION PROBLEM

Sensor	Revised Classification Rate (%)
ECG	72.68
Thermistor	76.34
Chest Belt	67.1
Accelerometer	83.94
Contact Microphone	79.57
Audio Microphone	93.17
Both Microphones	93.22
Audio + Accelero	93.27
Audio + Accelero + ECG	92.49
All sensors	92.9

across all categories. On the other side, the audio microphone outperforms all other sensors in all conditions. For the four remaining sensors (ECG, thermistor, accelerometer, and contact microphone), their relative ranking depends upon the considered event category. In this way, the accelerometer is noticed to be of high interest for the distinction with speech or with the background (for which abdominal movements are weak). The usefulness of the contact microphone is noticed for the discrimination with forced expirations and throat clearings. It is also worth noting that the thermistor is observed to be informative for the differentiation with laugh. This can be explained by the fact that laugh generally involves a nasal airflow [25], while the nose cavity is closed when coughing. Finally, the ECG is noticed to be ranked third or fourth sensor across all categories.

Inspecting results obtained with the combination of sensors, it is shown that the addition of other sensors to the audio microphone only leads to a minor improvement (and even sometimes to a slight reduction of performance), except for the distinction with speech and laugh. In these two latter cases, the complementarity mainly brought, respectively, by the accelerometer and the thermistor leads to an increase of RCR values of around 7% and 4%.

The RCRs obtained by each sensor and their combination for the general cough detection problem (where all ambiguous event categories are mixed) are exhibited in Table II. As expected, the audio microphone and chest belt are clearly, respectively, the best and worst sensors with 93.2% and 67.1%. Since the great
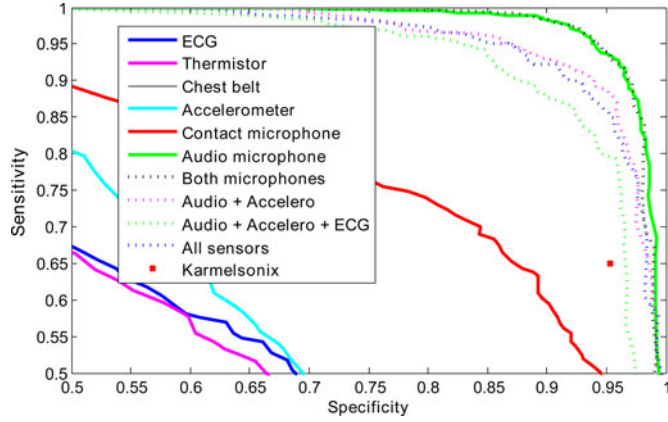
Fig. 7. Performance of cough detection in the specificity-sensitivity plane, using several sensors and their possible combinations, and comparison with the Karmelsonix system.

TABLE III
DETAIL OF THE PERFORMANCE ACHIEVED BY THE PROPOSED APPROACH USING ONLY THE AUDIO MICROPHONE AND WITH THE KARMELSONIX SYSTEM FOR THE TEN SUBJECTS OF SET B

Subj.	Proposed Audio-only		Karmelsonix	
	Sens.(%)	Spec.(%)	Sens.(%)	Spec.(%)
1	88.9	100	93.1	100
2	97.3	94.7	87.7	90.1
3	79.2	96.6	0.0	100
4	98.6	92.2	83.3	93.7
5	100	88.9	61.4	86.0
6	93.3	93.3	4.0	100
7	98.6	98.6	68.1	92.4
8	88.0	95.7	90.7	97.1
9	100	91.1	70.4	96.2
10	100	93.4	90.1	97.0
Avg.	94.39	94.45	64.9	95.3
STDV	7.0	3.4	34.8	4.7

majority of frames belong to the background and to speech (in a lesser extent), these results are mainly dominated by the performance for these two categories. It is, therefore, not surprising to find the accelerometer as second best sensor with about 84%, followed by the contact microphone (79.6%), the thermistor (76.3%), and the ECG (72.7%). Finally, it is emphasized again that the improvement brought by considering more sensors in complementarity with the audio microphone is negligible.

VII. COMPARISON WITH KARMELSONIX

The performance of the proposed approach is now compared to the Karmelsonix system, one of the few commercially available cough counters. Karmelsonix makes use of four sensors [15]: one audio microphone, two contact microphones (one on the trachea and one on the thorax), and a chest wall impedance belt. The acquisition system is also provided with an analysis software from which we here consider the cough detections.

In this experiment, training is performed on set A of the database (22 subjects) and the test is carried out on set B (10 subjects for which parallel recordings with Karmelsonix are available). The performance of the various sensors and their combination is assessed at the event level since Karmelsonix only detects cough event triggers. As standard metrics at the event level, we use the specificity and sensitivity measures.

In order to determine cough events with the proposed approach, our key idea was to focus only on the detection of the explosive phase of cough. Indeed, it is known that the intermediate phase of cough is very similar to a forced expiration [26]. Besides, the voiced phase does not appear in all cough events and could be confused with some parts of laughing or throat clearing. On the opposite, the explosive phase is charateristic of the beginning of any cough event and is, therefore, distinctive enough from other ambiguous event categories. Since explosive phases have not been manually annotated, we consider in the remainder of this paper that they dominate the first 60 ms of the cough events.

Fig. 7 shows the results, in the specificity-sensitivity plane by varying the decision threshold, of the proposed approach based

on each sensor separately or using their combination. The performance obtained by the Karmelsonix system is indicated as well. Note that the performance using the chest belt is so low that it is not represented in the plot range. It is also observed that considering the determination of the explosive phase for event detection modifies the relative ranking between the sensors, notably regarding the accelerometer and contact microphone. Nonetheless, the audio microphone still clearly remains the most effective sensor. Furthermore, our attempts to integrate a multimodal technique did not lead to any improvement compared to the audio-only approach. We, therefore, consider in the following that the proposed method is reduced to the use of the audio microphone. This approach is observed to outperform the Karmelsonix system which, albeit providing good specificity results (few false alarms), gives poor sensitivity results on average (about 65%, indicating a high number of missed detections). Interestingly, the proposed approach is observed to provide, for a comparable specificity, much better sensitivity capabilities. As working point for the proposed technique, we fixed in the following the decision threshold such that a compromise between misses and false alarms is found (i.e., in order to have comparable specificity and sensitivity performance).

The detection results specific to each subject are presented in Table III. It is noted that Karmelsonix suffers from a wide intersubject variability. Indeed, while subjects 1, 8, and 10 have a specificity higher than 90%, it turns out that the system misses almost all cough events for subjects 3 and 6. This drawback is the most important inconvenient of Karmelsonix. Note that this observation is in contradiction with the preliminary study carried out by Karmelsonix developers in [15] and where they report an overall sensitivity of 96%. On the opposite, while the proposed approach achieves comparable specificity results, its sensitivity never goes below 79% and reaches an average value of about 94.5%.

Finally, Table IV highlights the sources of false alarms for the proposed audio-based technique and the Karmelsonix system. Segments of speech and background noise are noticed to be well discriminated, with six false alarms for the proposed method

TABLE IV
REPARTITION OF COUGH EVENTS DETECTED BY THE PROPOSED AUDIO-BASED
APPROACH AND THE KARMELSONIX SYSTEM ACROSS ALL EVENT CATEGORIES

	Amount	# Cough detected with the proposed system	# Cough detected with Karmelsonix
High vol. cough	148	148	95
Int. vol. cough	150	137	95
Low vol. cough	155	148	86
Fit of coughing	267	251	192
Expiration	90	5	1
Throat Clearing	148	13	24
Speech	474 sec	1	0
Laugh	96	18	1
Background	3006 sec	5	2

and two for Karmelsonix, over a duration of 58 min. For the proposed approach, the most confusing classes are respectively laugh (18.7%), throat clearing (8.8%), and forced expiration (5.5%). Regarding the ambiguity for the Karmelsonix system, it turns out that throat clearing leads to the highest confusion (16.2%) while forced expiration and laugh are well discriminated. Albeit the proposed audio-based method leads to a higher number of false alarms (42 against 28 for Karmelsonix), it also produces much less missed cough events, providing on average a considerable improvement. Indeed, the Karmelsonix system is clearly seen to miss an important amount of cough events across all cough types: 28.1% for fits of coughing, around 36% for cough with high and intermediate volume, and 44.5% for cough with low volume. On the opposite, the proposed approach shows up a low miss rate across all cough types (ranging from 8.7% on cough with an intermediate volume, to 0% for cough with high volume).

VIII. CONCLUSION AND DISCUSSION

This paper aimed at providing an objective study on the relevance of several sensors to use in a cough detection system: ECG, thermistor, chest belt, accelerometer, contact, and audio mircophones. In a first experiment features extracted for each sensor were assessed through the mutual information they convey for the classification problem. Already at this stage it was highlighted that the audio microphone is the most promising sensor while the chest belt is the worst one. In a second time, we compared the ability of sensors for distinguish cough from several other event categories: background, forced expiration, throat clearing, speech, and laugh. Following an analysis at the frame level, it turned out that considering other sensors in addition to the audio microphone did not provide any substantial improvement, except regarding the discrimination of cough with speech (for which the accelerometer was observed to be particularly useful) and with laugh (where the thermistor showed an interesting performance). Regarding the most difficult categories to differentiate, it was seen that forced expiration, throat clearing (particularly, abundantly produced by subjects with cystic fibrosis) and laugh both led to the highest ambiguity with a misclassification rate of about 20% for all. Third and finally, cough event detection was carried out by recognizing the explosive phase of cough, and the proposed approach was compared to the Karmelsonix system, one of the few commercially available solutions. Two important conclusions could be drawn from this experiment. First, the audio microphone achieved the best results, even when compared to our attempts of multimodal approaches. This can be explained by the fact that the acoustic properties during the explosive phase of cough are sufficiently well differentiated and that considering other modalities does not bring any improvement for detecting cough event via this framework. From a practical point of view, this is a strong advantage since the cough monitor can be reduced to a simple audio microphone which is discreet and suited for an ambulatory 24-h usage. The second conclusion was that the performance yielded by the proposed approach reached around 94.5% of both specificity and sensitivity, clearly outpeforming the Karmelsonix system. Karmelsonix was observed to suffer from an important intersubject variability, with a sensitivity which could be very low for some subjects. On the opposite, this drawback was alleviated in our proposed audio-only method whose sensitivity varied between 79% and 100%, for a specificity ranging from 89% to 100%. Our future works encompass the validation of our prototype in patients with cystic fibrosis (and possibly other chronic cough diseases), in a 24-h ambulatory context, as well as the qualification of cough events to complement their frequency quantification.

REFERENCES

[1] S. M. Schappert and C. W. Burt, "Ambulatory care visits to physician offices, hospital, outpatient departments, and emergency departements: United States 2001–2002," *Vital Health Stat.*, vol. 13, pp. 1–66, 2006.

[2] L. P. McGarvey, L. G. Heaney, and J. McMahon, "A retrospective survey of diagnosis and management of patients presenting with chronic cough to a general chest clinic," *Int. J. Clin. Pract.*, vol. 52, no. 3, pp. 158–161, 1998.

[3] S. S. Birring, B. Prudon, A. J. Carr *et al.*, "Development of a symptom specific health status measure for patients with chronic cough: Leicester cough questionnaire (LCQ)," *Thorax*, vol. 58, pp. 339–343, 2003.

[4] P. A. Marsden, J. A. Smith, A. A. Kelsall *et al.*, "A comparison of objective and subjective measures of cough in asthma," *J. Allerg. Clin. Immunol.*, vol. 122, no. 5, pp. 903–907, 2008.

[5] S. Decalmer, D. Webster, A. Kelsall, K. McGuinness, A. Woodcock, and J. Smith, "Chronic cough : How do cough reflex sensitivity and subjective assessments correlate with objective cough counts during ambulatory monitoring?" *Thorax*, vol. 62, pp. 329–334, 2007.

[6] A. Morice, G. Fontana, M. Belvisi, S. Birring, K. Chung *et al.*, "ERS guidelines on the assessment of cough," *Eur. Respir. J.*, vol. 29, pp. 1256–1276, 2007.

[7] J. Smith, "Monitoring chronic cough: Current and future techniques," *Exper. Rev. Respirat. Med.*, vol. 4, no. 5, pp. 673–683, 2010.

[8] S. Barry, A. Dane, A. Morice, and A. Walmsley, "The automatic recognition and counting of cough," *Cough J.*, vol. 2, no. 8, 2006.

[9] S. Birring, T. Fleming, S. Matos *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, 2008.

[10] S. Matos, S.S. Birring, I. D. Pavord, and H. Evans, "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.

[11] J. Smith, "Cough: Assessment and equipment," *Buyers Guide Respirat. Care Products*, pp. 96–101, 2008.

[12] K. McGuinness, K. Holt, R. Dockry, and J. Smith, "Validation of the vitaloJAK 24 hour ambulatory cough monitor," *Thorax*, vol. 67, p. A131, 2012.

[13] T. Drugman, J. Urbain, and N. Bauwens *et al.*, "Audio and contact microphones for cough detection," presented at 13th Ann. Conf. Int. Speech Commun. Assoc., Portland, OR, USA, 2012

[14] M. Coyle, D. Keenan, L. Henderson, M. Watkins *et al.*, "Evaluation of an ambulatory system for the quantification of cough frequency in patients with chronic obstructive pulmonary disease," *Cough J.*, vol. 1, no. 3, 2005.

[15] E. Vizel, M. Yigla, Y. Goryachev, E. Dekel *et al.*, "Validation of an ambulatory cough detection and counting application using voluntary cough under different conditions," *Cough J.*, vol. 6, no. 3, 2010.

[16] G. Peeters, A large set of audio features for sound description (similarity and classification) in the CUIDADO project," 2003

[17] K. Tokuda, T. Kobayashi, T. Masuko, and S. Imai, "Mel generalized cepstral analysis—A unified approach to speech spectral estimation," in *Proc. Int. Conf. Spoken Language Process.*, 1994, vol. 3, pp. 1043–1046.

[18] B. Moore, B. Glasberg, and T. Baer, "A Model for the prediction of thresholds, loudness, and partial loudness," *JAES*, vol. 45, no. 4, pp. 224–240, 1997.

[19] G. Peeters, "Chroma-based estimation of musical key from audio-signal analysis," in *Proc. Int. Society Music Info. Retrieval Conf.*, 2006, pp. 115–120.

[20] T. Drugman and A. Alwan, "Joint robust voicing detection and pitch estimation based on residual harmonics," in *Proc. Int. Speech Commun. Assoc.*, Florence, Italy, 2011, pp. 1973–1976.

[21] T. Drugman, T. Dubuisson, and T. Dutoit, "Phase-based information for voice pathology detection," in *Int. Conf. Acoust. Speech Signal Process.*, 2011, pp. 4612–4615.

[22] R. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ, USA: Princeton Univ. Press, 1961.

[23] T. Drugman, M. Gurban, and J. P. Thiran, "Relevant feature selection for audio-visual speech recognition," in *Proc. IEEE Int. Worksh. Multimed. Signal Process.*, Oct. 2007, pp. 179–182.

[24] J. Hanley and B. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[25] J. Bachorowski, M. Smoski, and M. Owren, "The acoustic features of human laughter," *JASA*, vol. 110, no. 3, pp. 1581–1597, 2001.

[26] J. Korpas, J. Sadlonova, and M. Vrabec, "Analysis of the cough sound: An overview," *Pulmonary Pharmacol.*, vol. 9, pp. 261–268, 1996.

Authors photographs and biographies not available at the time of publication.