Using Mutual Information in Supervised Temporal Event Detection: Application to Cough Detection

Thomas Drugman

TCTS Lab, University of Mons, 31 Boulevard Dolez, 7000 Mons, Belgium
Email: thomas.drugman@umons.ac.be, Tel: +32 65 374749

Abstract

A large number of biomedical and surveillance applications target at identifying specific events from sensor recordings. These events can be defined as rare and relevant occurrences with a limited duration. When possible, human annotation is available and developed techniques generally adopt the standard recognition approach in which a statistical model is built for the event and non-event classes. However, the goal is not to detect the event in its complete length precisely, but rather to identify the presence of an event, which leads to an inconsistency in the standard framework. This paper proposes an approach in which labels and features are modified so that they are suited for time event detection. The technique consists of an iterative process made of two steps: finding the most discriminant segment inside each event, and synchronizing features. Both steps are performed using a mutual information-based criterion. Experiments are conducted in the context of audio-based automatic cough detection. Results show that the proposed method enhances the process of feature selection, and significantly increases the event detection capabilities compared to the baseline, providing an absolute reduction of the revised event error rate between 4 and 8%. Thanks to these improvements, the audio-only cough detection algorithm outperforms a commercial system using 4 sensors, with an absolute gain of 26% in terms of sensitivity, while preserving the same specificity performance.

Keywords: Event Detection, Biomedical Engineering, Cough Detection, Mutual Information, Supervised Learning

1. Introduction

This paper addresses the problem of automatic event detection from time series. A proper method for event detection is of interest in various biomedical, surveillance and signal-based applications involving a sensor-based monitoring of any phenomenon. These applications encompass the characterization of molecular events [1], cough detection [2], monitoring of biomedical measures (sleep apnea [3], muscle activity [4], etc.), seismic event detection [5], [6], anomaly detection [7], meteorological changes, traffic regulation [8], or event detection in social streams [9]. This paper proposes a new method of event detection based on mutual information in the context of supervised learning. The problem positioning is more precisely presented in Section 1.1. Since the validity of the proposed approach is illustrated in the frame of audio-based cough detection, the background on this issue is given in Section 1.2. The structure of the paper is finally described in Section 1.3.

1.1. Temporal Event Detection

The detection of events from time series data is a problem which gained interest from the research community [10], [11], [12]. In most cases, studies refer to the issue of unsupervised event detection, in which the underlying phenomenon is ill-understood, making human annotation impossible [10],[13]. In such a context, the goal is to identify the time points at which the system behavior change occurs. This is referred to as the *change-point detection* problem [10]. This is typically achieved by considering probability distributions from data in the past and present intervals, and by inspecting whether these two distributions are significantly different [12]. Semi-supervised learning has also been addressed in [14] for the detection of rare and unexpected events. This method can be applied when collecting a sufficient amount of labeled training data for supervised learning is practically infeasible (e.g. because manually annotating such a large corpus would be too time-consuming, and consequently too expensive).

On the other hand, there is a large number of applications for which human annotation is available and which target at identifying specific events from sensor recordings [2], [5], [8]. Events can then be defined as rare and relevant occurrences, generally with a limited duration. For such an issue, vectors of features characterizing the signal are extracted at a constant sampling rate. More precisely, the signal is windowed in so-called *frames* where a short-term analysis is performed [15]. For each frame, a set of characteristics (also called features, measurements or attributes) are extracted. The whole signal is analyzed by shifting the frames by a constant delay, consecutive frames possibly overlapping [15]. Based on these sequences of frames, developed techniques generally adopt the standard approach in which a statistical model is built for each class to be identified (presence or not of an event) [16]. This approach nonetheless suffers from a main drawback: models built at learning stage are based on the whole event duration. However, the goal is not to detect the event in its complete length precisely, but rather to identify the presence of an event, i.e. to detect an event trigger.

This induces a dramatic change from the typical training formalism: instead of building models so as to minimize the error rate at the *frame level*, the supervised learning has to focus on the detection ability at the *event level*. An extreme case to illustrate this concept would be a classification system which identifies correctly only one single frame among the several contained in each event. In the conventional framework, this would be characterized by low performance since the majority of the frames contained in the event are missed, although it leads to a perfect discrimination at the event level since the event has been properly detected.

In parallel, measures derived from the Information Theory [17] have been extensively used in machine learning. Among others, the usefulness of Mutual Information (MI) for selecting the most relevant features in a given classification task has been proven [18]. This efficiency is nonetheless also impaired if the traditional formalism aiming at detecting events in their whole duration is considered. Indeed, the relevance of a feature at the frame level does not necessarily imply its relevance at the event level, and vice versa.

The goal of this paper is precisely to investigate how MI can be used to alleviate the aforementioned drawbacks by localizing the relevant regions of interest in each event and by synchronizing features. Some concepts of the proposed approach and all our experimental results will be illustrated in the context of a particular application: audio-based cough detection.

1.2. Automatic Cough Detection

Cough is the commonest reason for which patients seek medical advice to the general practitioner (around 20% of consultations for children below 4 years old), the paediatrician and the pneumologist (for whom chronic cough represents one third of consultations). The impact of cough, notably chronic coughing, on life quality can be important [19].

In order to evaluate the cough severity, a subjective assessment is possible by making use of cough diaries, quality-of-life questionnaires or relying on a visual analog scale [20]. However, it has been shown that the subjective perception of cough is only slightly correlated with objective measurements of its severity [21]. Medical literature on this topic therefore underlines the lack of a tool allowing the automatic, objective and reliable quantification of this symptom [19]. This latter step is notably required prior to any correct evaluation of possible treatments.

Some approaches have been recently proposed to address the automatic detection of cough [2]. These systems generally couple various sensors to the audio signal [2]: accelerometer, chest impedance belt, contact microphone, ECG, respiratory inductance plethysmography etc. Although reported results are encouraging, there is currently neither standardized methods nor adequately validated, commercially available and clinically acceptable cough monitors [19], [2]. Besides, following the patient in ambulatory and 24h-long conditions (while preserving his daily habits) remains an open problem. As a result, cough quantification in the majority of hospitals is still nowadays performed by a tedious task of manual counting from audio recordings, or for validation by comparison using simultaneous video recordings.

For respiratory physiologists, cough is three-phase expulsive motor act characterized by an inspiratory effort, followed by a forced expiratory effort against a closed glottis and then by opening of the glottis and rapid expiratory airflow [19]. As shown in Figure 1, the acoustics of the cough sound is manifested by three phases, where the last one is optional [22]: an explosive phase, an intermediate period whose characteristics are similar to a forced expiration, and a voiced phase. At this point, it can then be understood that even for the detection of short events like cough: i) it might not make sense to try to detect the cough event in its complete duration, ii) as the signal properties vary across the duration of an event, the segments where features are particularly discriminative may not coincide. For example, some features might be relevant for detecting the explosive phase, while others would characterize the voiced phase. This might be particularly true when features arise from different sensors which might not be synchronous. This paper aims at addressing both of these issues.

1.3. Structure of the paper

This paper is structured as follows. Section 2 describes the proposed approach based on information localization inside events, and feature syn-

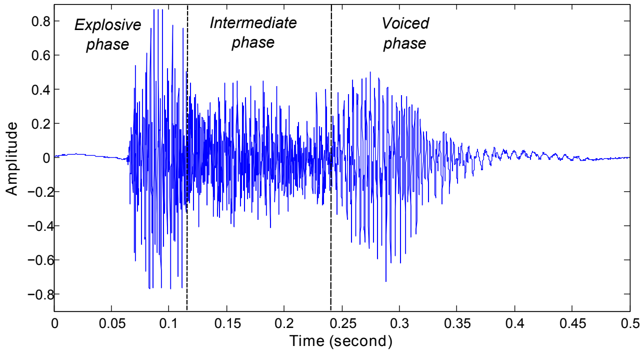

Figure 1: *Waveform of a typical cough sound with three phases.*

chronization. The experimental protocol is detailed in Section 3. Results of our evaluation are reported in Section 4 and the paper is concluded in Section 5.

2. Proposed Approach

The general workflow of the proposed approach is presented in Figure 2. The method starts with a sequence of feature vectors and with the initial event labels (resulting from the manual annotation). The algorithm consists of an iterative process aiming at localizing the relevant information inside the events, and at synchronizing features with each others and with labels. The motivation behind these steps is the following. First of all, the relevant segments of the events, i.e the portions of events which are the most distinguishable from other classes, are only a partial component of the whole event duration and have to be located. Secondly, features extracted from the time signal may characterize different aspects of this latter signal, which may occur at different instants. Besides, in some applications, features might even arise from various sensors, which strengthens this issue. For these reasons, features have to be synchronized such that their relevant segments emerge at the same time, which is expected to enhance the event discrimination capabilities of the classifier.

2.1. Feature Synchronization

As aforementioned, the period where a feature is particularly discriminative may not perfectly coincide with the class label indicating the presence of an event. Therefore, each feature must be synchronized with the class labels

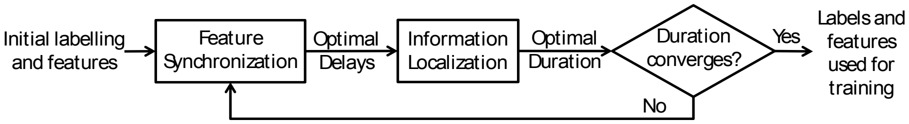

Figure 2: *Workflow of the proposed approach.*

by applying a certain delay. Denoting C_i the class labels at iteration i, the sequence of the j^{th} feature, noted X_j, is synchronized with a delay d_{ij} such that:

$$d_{ij} = \arg\max_{d} I(X_j(d); C_i), \tag{1}$$

where $X_j(d)$ represents the feature sequence X_j on which a delay of d frames has been applied (with regard to the initial features). The mutual information $I(X_j; C_i)$ between X_j and the classes C_i can be computed as [17]:

$$I(X_j; C_i) = \sum_{x_j} \sum_{c_i} p(x_j, c_i) \log_2 \frac{p(x_j, c_i)}{p(x_j)p(c_i)} \tag{2}$$

and can be viewed as the amount of information that feature X_j conveys about the considered classification problem, i.e. the individual discrimination power of this feature alone.

The goal of this step is then to find the optimal delays to apply to each feature so as to maximize its discrimination power. Figure 3 shows the evolution of the normalized MI as a function of the applied delay for a particular feature used in our cough detection application. In this case, MI is normalized by division with the entropy of the classes, such that it is bounded to 1 for a perfect classification [17], [23]. In the illustration of Figure 3, it can be observed that shifting the feature sequence by 3 frames in the past gives a normalized MI reaching 0.21, providing a clear improvement compared to the initial case where a value of only 0.14 is obtained.

2.2. Information Localization

The key idea of the information localization step is to find the optimal duration of the relevant segments inside the events such that the discrimination abilities of the feature set are maximized. As in Section 2.1, we would like this step to be independent of any classifier and to rely only on measures derived from the Information Theory. Unfortunately, computing MI from

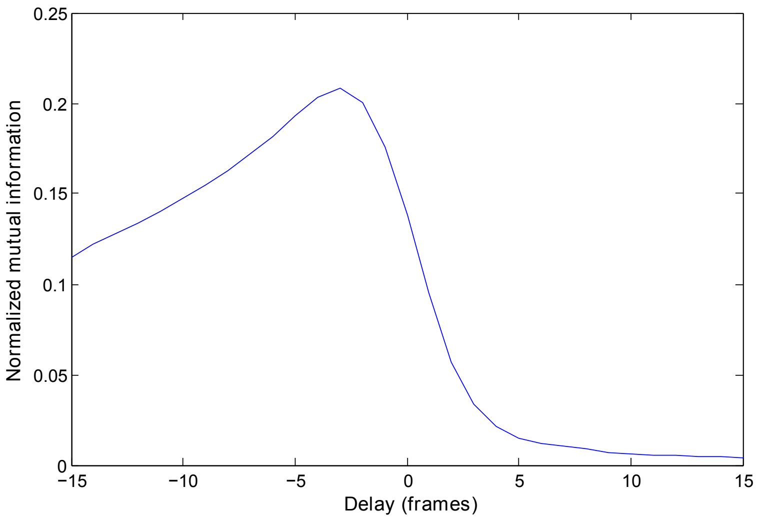

Figure 3: *Illustration of feature synchronization for a particular feature used in our cough detection application.*

data requires the estimation of probability densities, which cannot be accurately done in high dimensions. This is why the great majority of MI-based methods use measures based on up to three variables (two features plus the class label).

Therefore a MI-based assessment of the relevance of a feature set whose cardinality is higher than 2 is impossible to achieve accurately in practice, as it would require a prohibitive amount of data. For this reason, several strategies (mainly of feature selection) have been proposed to deal with the issue of redundancy management, i.e to estimate the redundancy and the amount of new relevant information of a given feature with an existing feature subset [18], [23].

In this paper, we use as feature selection method the following algorithm which is known [23] to provide among the best feature selection results. Let us denote F=$\{X_1$,X_2,...,$X_N\}$ the initial set of N features, and S_k the selected subset (with $S_k \subseteq F$) of k features at step k. The method is a greedy algorihm which starts from an empty set and which selects at each step k the feature Y_k maximizing:

$$Y_k = \arg \max_{X_p \in F \setminus S_{k-1}} [I(X_p; C) - \max_{Y_q \in S_{k-1}} I(X_p; Y_q; C)], \tag{3}$$

where $I(X_p; C)$ is the relevant information brought by feature X_p separately (i.e independently of any other feature), and where $I(X_p;Y_q;C)$ is the redundancy of relevant information between features X_p and Y_q and can be developed as:

$$I(X_p;Y_q;C) = \sum_{x_p}\sum_{y_q}\sum_{c} p(x_p,y_q,c)\cdot \log_2 \frac{p(x_p,y_q)p(x_p,c)p(y_q,c)}{p(x_p,y_q,c)p(x_p)p(y_q)p(c)} \quad (4)$$

In other words, the algorithm considers that the redundancy between X_p and the selected subset S_{k-1} is dominated by the most redundant feature in it.

In a similar spirit, we can consider in the following that, working with a set S_M of M features, the Relevant Information $RI(X_p,C,S_M)$ brought by feature X_p (contained in S_M) with regard to other selected frames can be expressed as:

$$RI(X_p,C,S_M) = I(X_p;C) - \max_{Y_m \in S_M \setminus \{X_p\}} I(X_p;Y_m;C), \quad (5)$$

In this context, information localization inside the events is made at iteration i by fixing the duration of the class labels C_i such that the discrimination of these events is optimal:

$$C_i = \arg\max_{C} \sum_{m=1}^{M} RI(X_m,C,S_M). \quad (6)$$

In other words, the idea is to set the duration of the new labels at iteration i in a way such that the discrimination abilities of the selected feature set S_M are optimized. Finally, it is worth noticing that the sum involved in Equation 6 does not make sense in the absolute (as this sum might excess $H(C)$) but allows a comparative evaluation between various feature sets (of the same cardinality), or between class labels as it is the case here.

2.3. Iterative Process

As depicted in Figure 2, starting from the initial non-synchronized features and manually-annotated class labels, an iterative process is used by repeating the steps of feature synchronization and information localization as explained here above. This is done untill convergence is reached for the label duration. According to our experiments, this was always achieved in less than 3 iterations. No divergence or oscillation between two or more possible solutions were observed. Note that common schemes used in non-linear optimization to prevent these issues can be applied here straightforwardly.

3. Experimental Protocol

We here detail the protocol used throughout our experiments led in the context of audio-based cough detection. The database is first described in Section 3.1. The recognition framework is presented in Section 3.2. First a large variety of audio descriptors is extracted (Section 3.2.1), among which only the most relevant will be selected as explained in Section 3.2.2. Finally Section 3.2.3 provides details about the methodology used for classification and assessment. Methods compared in our results are summarized in Section 3.3, and metrics employed in our evaluation are introduced in Section 3.4.

3.1. Database

The study population was divided into two groups. The first set (A) included 22 healthy subjects (9 Male, mean age±SD: 22.8 ± 2.44 ,range: $20-28$). The second set (B), consisting of 10 additional healthy subjects (5 Male, mean age±SD: 23 ± 1.45, range: $22-26$) was designed to compare our system to the commercially available KarmelSonix cough counter [24]. 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 1: (a) sitting down in a quiet environment, (b) sitting down in a noisy environment and (c) climbing on/going down of a stepladder.

This protocol was inspired by the one used to develop and evaluate the Karmelsonix system [24], with the addition of coughs at low and intermediate pulmonary volume 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 minutes. Note that slight deviations were observed from the strict protocol, but that the manual annotation was made coherently.

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

Table 1: *Standardized protocol for data recorings*

3.2. Details of implementation

3.2.1. Feature Extraction

The key idea here is to extract the largest variety of audio features among which only the most relevant will be selected. These features are extracted every 12 ms on a 30 ms-long frames and can be divided into two categories: features describing the spectral contents and measures of noise. We also added the first and second derivatives for each of these features in order to integrate the sound dynamics.

Several features characterizing the spectral shape have been proposed in [25]. For a comprehensive description of the magnitude spectrum, we used the widely-used Mel-Frequency Cepstral Coefficients (*MFCCs*), the *loudness* associated to each Bark band [25] and the *relative energy* in different frequency subbands. Besides, several parameters describing the spectral shape are also employed. 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 [25]. Finally, the *Spectral Variation* and *Spectral Flux* characterize the amount of variations of spectrum along time and are based on the normalized crosscorrelation between two successive amplitude spectra [25]. Besides, we also use the energy and total loudness which are informative mainly about the presence of audio activity.

Quantifying the level of noise in the signal is of interest for describing the cough sound. For this purpose, several measures are here extracted. First, the *Harmonic to Noise Ratio* (HNR) is calculated in four frequency ranges. The *Spectral Flatness* measures the noisiness/sinusoidality of a spectrum (or a part of it) in four frequency bands [25]. The *Zero-Crossing Rate* quantifies the number of times the signal crosses the zero axis. It is expected that the greater the amount of high-frequency 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 [26] 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 [27] for highlighting turbulences during glottal production.

3.2.2. Feature Selection

A total number of 222 features (including the first and second derivatives) has been extracted in Section 3.2.1. 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 [28]. The algorithm of feature selection we use throughout the rest of the paper is the one briefly described in Section 2.2 and whose details can be found in [23]. Probability density functions involved in the calculation of MI measures are estimated by a histogram approach. The number of bins is set to 50 for each feature dimension, which results in a trade-off between an adequately high number for an accurate estimation, while keeping sufficient samples per bin.

3.2.3. Classification

For each of the issues tackled across experiments, a dedicated Artificial Neural Network (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 16 neurons in this work, as it gave in [29] a good compromise between high performance and rather low complexity) 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 ms 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.

Except for the comparison with the Karmelsonix system (Section 4.2) for which training is performed on set A and testing on set B, a leave-four-subjects-out cross-validation approach is adopted in which models are trained on 28 out of the 32 subjects are used for training, and test is performed on the four remaining. This operation is repeated 8 times so as to cover the whole database for testing, and results are averaged across them.

3.3. Methods Compared

Four methods will be compared in the following, depending upon the steps involved in the proposed approach (see Section 2):

- *Local=0, Synchro=0 (baseline)*: denotes the traditional method in which none of the proposed steps is achieved. In other words, this is the classical framework where the initial features and labels are used, and it is considered as a baseline in the following.

- *Local=0, Synchro=1*: is the technique where features are synchronized according to the original labels.

- *Local=1, Synchro=0*: performs the step of information localization using the initial features.

- *Local=1, Synchro=1*: is the proposed approach described in Section 2 for which the complete iterative process has been carried out.

In addition to an assessment of these four methods, a part of the results (Section 4.2) will also be devoted to a comparison with the commercial Karmelsonix system [24].

3.4. Metrics

Metrics we use at the event level are the standard specificity and sensitivity measures [30]. Specificity is the complement of the so-called *"false positive rate"*, defined as the proportion of false alarms. A false alarm is an event which is incorrectly identified: in our case, this means that the algorithm detects a cough event when there is actually none. Similarly, sensitivity is the complement of the so-called *"false negative rate"*, defined as the proportion of misses. A miss is an event which is incorrectly rejected: in our case, that implies that the algorithm does not detect anything where there is actually a cough event. By varying the decision threshold θ (applied on the posterior probability outputted by the ANN), a Receiver Operating Characteristic (ROC) curve is obtained in the specificity-sensitivity plane [30]. Two measures are then employed to characterize the performance of the ROC curve. The first one is the well-known Area Under Curve (AUC), which reaches a value of 1 for a perfect classifier. As a second single measure summarizing the ROC curve, we defined the Revised Event Error Rate (REER) as:

$$REER = \min_{\theta} \sqrt{(1 - sens.(\theta))^2 + (1 - spec.(\theta))^2}, \qquad (7)$$

and which also benefits from a straightforward interpretation: REER is the Euclidean distance in the ROC curve plane from the ideal working point characterized by values of 1 for both specificity and sensitivity. This criterion implies that an equal importance is given to both specificity and sensitivity criteria. Based on a medical advice, one of these aspects could be emphasized by weighting its importance in Equation 7. Finally, the Revised Event Classification Rate (RECR) is defined as the complement of the REER (i.e $RECR = 1 - REER$). As a consequence, the higher AUC and RECR (the lower REER), the better the system performance.

4. Results

This section presents the results of our experiments. Section 4.1 first investigates the influence of the proposed approach in a mutual information-based feature selection scheme. Section 4.2 then highlights the classification abilities of the proposed method in comparison with a commercial system: Karmelsonix [24]. Section 4.3 finally further explores the classification performance of the proposed technique, in terms of error rate as a function of the number of selected features.

4.1. Impact on Feature Selection

As mentioned in the introduction, traditional state-of-the-art approaches have been designed for frame classification, not for event detection. In a first experiment, our goal is to show that applying the proposed method (mainly through its information localization step) makes the MI-based measures more suited for feature selection. For this, we computed for each feature separately *i)* its normalized MI with the classes, which is supposed to be an image of its discrimination ability, and *ii)* the area under the ROC curve using only this feature in a ANN classifier, which is a performance measure after classification. Ideally, both measures should be highly correlated such that MI can be reliably used for feature selection. Figure 4 shows the results we obtained with the proposed method for our 222 audio features used for cough detection.

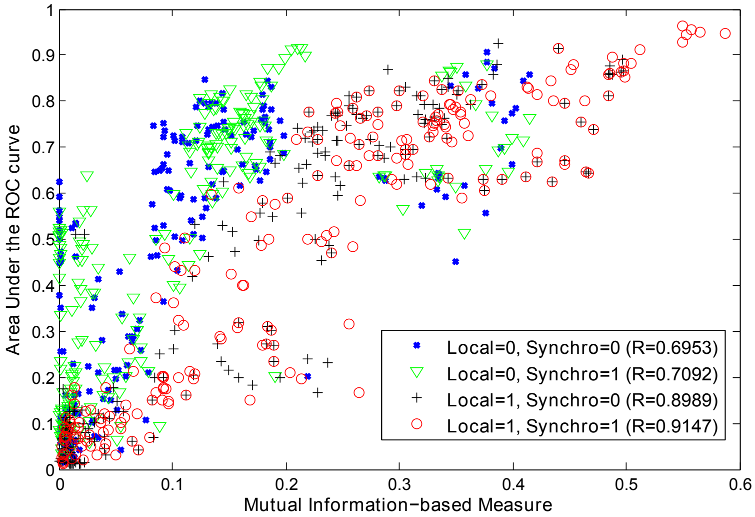

Figure 4: *Illustration, for the four compared techniques, of the significance of mutual information for feature selection. The baseline method is Local=0, Synchro=0.*

Three main problems can be noticed with the baseline approach (Local=0, Synchro=0). First, some features have very low MI values (close to 0) while their classification ability is rather good. Secondly, features having a comparable classification performance can lead to MI values dramatically different. This can be particularly observed for the two groups with MI values ranging between [0.1-0.2] and [0.3-0.4], and which have comparable classification performance. Thirdly, as a consequence, MI values are only poorly correlated with the classification performance measures, reaching a coefficient of Pearson correlation of only 0.6953. Therefore applying the baseline approach in the context of event classification makes the resulting MI measures inappropriate for efficient feature selection.

It is seen that the effect of feature synchronization on this issue is rather negligible, while the improvement brought by information localization is clear, leading to a correlation coefficient of 0.8989. The complete technique (Local=1, Synchro=1) still slightly enhances this trend. In this latter case, the three drawbacks observed with the traditional approach have been solved and the choice of features with high MI values ensures high classification abilities, leading to a proper feature selection.

4.2. Comparison with the Karmelsonix system

An example of ROC curves in the specificity-sensitivity plane, obtained using the four compared techniques with 20 features, is given in Figure 5. These results were obtained by training the four methods on set A of the database, and testing on set B, such that a comparison with the commercial Karmelsonix system is possible. First, it is clearly observed that the proposed approach outperforms the standard baseline (*Local=0, Synchro=0*), reaching higher values of both specificity and sensitivity. It can be observed that a non-negligible part of this improvement is brought by a single application of the information localization step. Further feature synchronization using the resulting new labels, and carrying out the iterative process allow to still increase the performance of cough detection. Another important conclusion is that albeit Karmelsonix provides an interesting specificity (95.3%), its sensitivity performance is rather poor (64.9%). The proposed approach is interestingly shown to lead to a dramatic improvement, yielding a sensitivity of about 91% for comparable specificity capabilities, or equivalently a corresponding absolute gain of 26%.

To give an idea, these results can be compared to other similar studies, although they were not obtained on the same database. In [31], the HACC system based on the analysis of audio recordings achieved a specificity of 96% and a sensitivity of 80%. In [32], it was reported that the commercialized LifeShirt system (using a microphone, a respiratory inductance plethysmography, and an accelerometer) gave a specificity and sensitivity of respectively 99.6% and 78.1%. Finally, the Leicester Cough Monitor was found in [33] to have a specificity and sensitivity of respectively 99% and 91%.

4.3. Impact on Classification

The benefit of applying the proposed method is evident by simple visual inspection of the ROC curves (as those exhibited in Figure 5). This is obviously reflected through the AUC and RECR values extracted from these

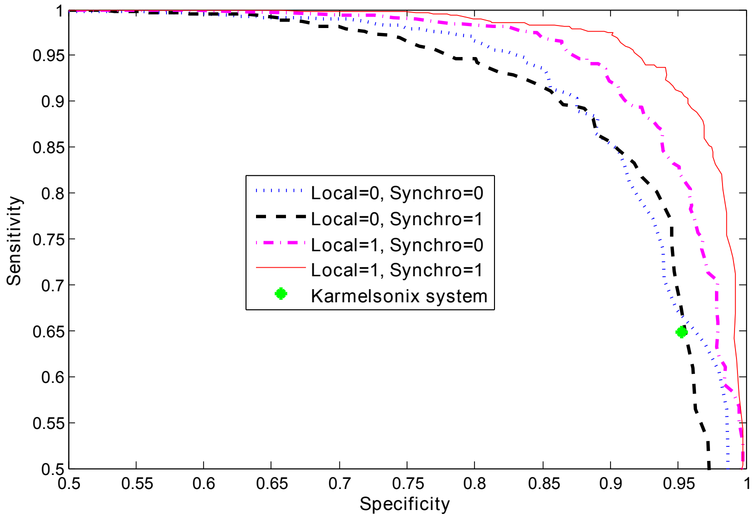

Figure 5: *ROC curves in the specificity-sensitivity plane for the four compared techniques using 20 features. The performance of the commercialized Karmelsonix system is also indicated.*

latter curves. As an exammple, Table 2 gives the classification rates achieved using 20 features. Figure 6 further shows, for the four compared techniques, how RECR varies with the number of selected features. Several conclusions can be drawn from this table and plot. First of all, it turns out that the proposed approach outperforms all other methods across all configurations. The gain with regard to the baseline (Local=0, Synchro=0) is clear, with an absolute gain of RECR varying between 4 and 8% (depending on the number of features). Secondly, the observation made in Section 4.1 about the fact that the proposed technique makes the MI-based measures more suited for feature selection is here also corroborated. Indeed the advantage of our method, although existent across all conditions, is also well emphasized for a low number of features.

Finally, it is worth noting that a key aspect of the method is the step of information localization. Focusing the detection on these specific segments is shown to significantly increase the performance of the system. On the opposite, it turns out that feature synchronization while keeping the original labels does not bring anything, and even deteriorates the results. Nevertheless, it is seen that this step makes sense in the whole iterative process (*Local=1, Synchro=0*), as it yields a slight but consistent enhancement over the *Local=1, Synchro=0* technique.

Method	L=0, S=0	L=0, S=1	L=1, S=0	L=1, S=1
RECR	83.07%	81.36%	88.90%	89.86%

Table 2: *Revised event clasiffication rate for the 4 compared techniques using 20 features. L and S respectively stand for* Local *and* Synchro*.*

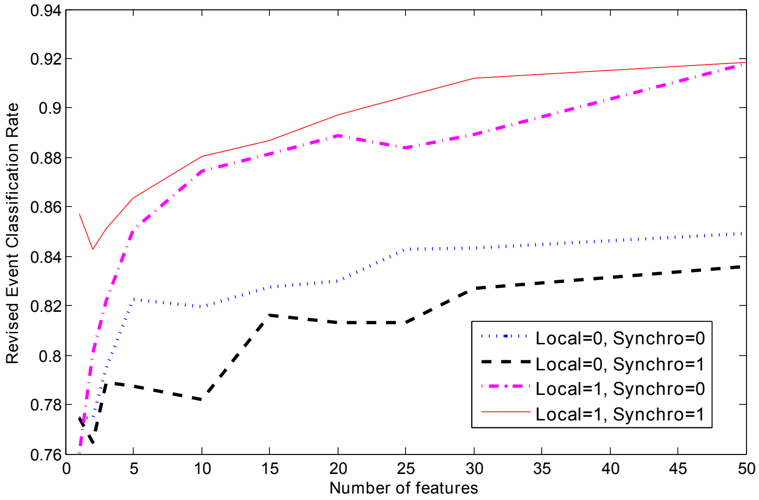

Figure 6: *Evolution, for the four compared methods, of RECR as a function of the number of selected features.*

5. Conclusion

This paper focused on a modification of the traditional recognition framework specifically devoted to the event detection issue, in the context of supervised learning. The proposed method consists of an iterative process made of two steps: *i)* information localization which identifies the most relevant segments inside each event, and *ii)* feature synchronization which ensures that the discrimination abilities of each feature emerge at the same time, even though they describe different aspects of the signal, or arise from various sensors. The proposed technique is assessed for a particular concrete application: audio-based cough detection. In a first experiment, it is shown that, compared to the baseline, it allows MI-based measures to be more suited for feature selection. These latter measures are indeed observed to be much more correlated with classification results, than what is obtained with the traditional baseline approach. This is reflected by a Pearson correlation coefficient increasing from around 0.7 to 0.9. The impact on feature selection is clear, and event detection with the proposed technique has been significantly improved independently of the number of selected features. As an illustration, working with 50 features, the revised event detection rate increased from about 85% to 92%. This enhancement was noticed to be mainly due to the information localization step. In a last experiment, the resulting audio-only cough detection method was compared to the commercialized Karmelsonix system which relies on four sensors. Applying the proposed algorithm was shown to clearly outperform Karmelsonix, as it led to a dramatic augmentation of sensitivity from 65% to 91%, keeping equal specificity performance (95%).

The potential applicability of the proposed approach covers any system targeting temporal event detection: monitoring of biomedical measures, seismic event detection, anomaly detection, meteorological changes, traffic regulation, etc. As perspectives of this study, our future works encompass: *i)* the clinical validation of the proposed technique in a 24-hour ambulatory cough counter for patients suffering from cystic fibrosis, *ii)* applying the proposed approach in an audio-based surveillance system for the automatic detection of abnormal events.

Acknowledgment

Thomas Drugman is supported by the Fonds National de la Recherche Scientifique (FNRS). The project is supported by the Walloon Region (Grant WIST 3 COMPTOUX # 1017071).

References

[1] Sarafraz, F., Eales, J., Mohammadi, R., Dickerson, J., Robertson, D., Nenadic, G., “Biomedical event detection using rules, conditional random fields and parse tree distances”, Proc. BioNLP09, pp. 115-118, 2009.

[2] Smith, J., “Cough: Assessment and Equipment”, The Buyers Guide to Respiratory Care Products, pp. 96-101, 2008.

[3] Gil, E., Maria Vergara, J., Laguna, P., “Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children”, Biomedical Signal Processing and Control, vol. 3, pp. 267-277, 2008.

[4] Beck, T., Von Tscharner, V., Housh, T., Cramer, J., Weir, J., Malek, M., Mielke, M., "Time/frequency events of surface mechanomyographic signals resolved by nonlinearly scaled wavelets", Biomedical Signal Processing and Control, vol. 3, pp. 255-266, 2008.

[5] Cansi, Y., "An automatic seismic event processing for detection and location: The P.M.C.C. Method", Geophysical Research Letters, vol. 22, pp. 1021-1024, 1995.

[6] Saragiotis, C., Hagjileondiadis, L., Rekanos, I., Panas, S., "Automatic p phase picking using maximum kurtosis and k-statistics criteria", IEEE Trans. Geosci. Remote Sensing Letters, pp. 147 - 151, 2004.

[7] Chandola, V., Banerjee, A., Kumar, V., "Anomaly detection: A survey", ACM Computing Surveys, vol. 41, art. 15, 2009.

[8] Xiaokun, L., Porikli, F.M., "A hidden Markov model framework for traffic event detection using video features", Proc. ICIP04, vol. 5, pp. 2901-2904, 2004.

[9] Sakaki, T., Okazaki, M., Matsuo, Y., "Earthquake shakes Twitter users: real-time event detection by social sensors", Proc. WWW10, pp. 851-860, 2010.

[10] Guralnik, V., Srivastava, J., "Event detection from time series data", Proc. KDD, pp. 33-42, 1999.

[11] Hunter, J., McIntosh, N., "Knowledge-Based Event Detection in Complex Time Series Data", Proc. AIMDM, pp. 271-280, 1999.

[12] Kawahara, Y., Sugiyama, M., "Change-Point Detection in Time-Series Data by Direct Density-Ratio Estimation", Proc. SDM, pp. 389-400, 2009.

[13] Tsien, C., "Event discovery in medical time-series data", Proc AMIA Symp., pp. 858-862, 2000.

[14] Zhang, D., Gatica-Perez, D., Bengio, S., Mccowan, I., "Semi-supervised adapted hmms for unusual event detection", Proc. IEEE CVPR, pp. 611-618, 2005.

[15] Quatieri, T., "Discrete-time speech signal processing", Prentice-Hall, 2002.

[16] Fukunaga, K., "Introduction to Statistical Pattern Recognition", 2nd Edition, Academic Press, 1990.

[17] Cover, T., Thomas, J., "Elements of Information Theory", Wiley Series in Telecommunications, New York, 1991.

[18] Battiti, R., "Using mutual information for selecting features in supervised neural networking", IEEE Transactions on Neural Networks, vol. 5, pp.537-550, 1994.

[19] Morice, A., Fontana, G., Belvisi, M., Birring, S., Chung, K., et al., "ERS guidelines on the assessment of cough", European Respiratory Journal, vol. 29, pp. 1256-1276, 2007.

[20] Kelsall, A., Decalmer, S., Webster, D., Brown, N., McGuinness, K., Woodcock, A., Smith, J., "How to quantify coughing: correlations with quality of life in chronic cough", Eur Respir Journal, 32, pp. 175179, 2008.

[21] Decalmer, S., Webster, D., Kelsall, A., McGuinness, K., Woodcock, A., Smith, J., "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.

[22] Korpas, J., Sadlonova, J., Vrabec, M., "Analysis of the Cough Sound: an Overview", Pulmonary Pharmacology, vol. 9, pp. 261-268, 1996.

[23] Drugman, T., Gurban, M., Thiran, JP., "Relevant Feature Selection for Audio-Visual Speech Recognition", IEEE International Workshop on Multimedia Signal Processing, pp. 179-182, 2007.

[24] Vizel, E., Yigla, M., Goryachev, Y., Dekel, E., et al., "Validation of an ambulatory cough detection and counting application using voluntary cough under different conditions", Cough Journal, 6:3, 2010.

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

[26] Drugman, T., Alwan, A., “Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics”, Proc. Interspeech, Florence, Italy, pp. 1973-1976, 2011.

[27] Drugman, T., Dubuisson, T., Dutoit, T., “Phase-based Information for Voice Pathology Detection”, Int. Conf. on Acoustics, Speech and Signal Processing, pp. 4612-4615, 2011.

[28] Bellman, R., “Adaptive Control Processes: a Guided Tour”, Princeton University Press, Princeton, NJ, 1961.

[29] Drugman, T., Urbain, J., Dutoit, T., “Assessment of Audio Features for Automatic Cough Detection”, 19th European Signal Processing Conference, pp. 1289-1293, 2011.

[30] Hanley, J., McNeil, B., “The meaning and use of the area under a receiver operating characteristic (ROC) curve”, Radiology, 143(1), pp. 29-36, 1982.

[31] Barry, S., Dane, A., Morice, A., Walmsley, A., “The automatic recognition and counting of cough”, Cough Journal, 2:8, 2006.

[32] Coyle, M., Keenan, D., Henderson, L., Watkins, M., et al., “Evaluation of an ambulatory system for the quantification of cough frequency in patients with chronic obstructive pulmonary disease”, Cough Journal, 1:3, 2005.

[33] Birring, S., Matos, S., Evans, D., Pavord, I., “Leicester Cough Monitor: a novel automated cough detection system”, Am J Resp Crit Care Med, 175:A379, 2007.