J Intell Robot Syst (2015) 80 (Suppl 1):S313–S330
DOI 10.1007/s10846-014-0168-9

Exploring Low Cost Laser Sensors to Identify Flying Insect Species

Evaluation of Machine Learning and Signal Processing Methods

Diego F. Silva · Vinícius M. A. Souza · Daniel P. W. Ellis · Eamonn J. Keogh · Gustavo E. A. P. A. Batista

Received: 28 February 2014 / Accepted: 1 December 2014 / Published online: 11 January 2015
© Springer Science+Business Media Dordrecht 2015

Abstract Insects have a close relationship with the humanity, in both positive and negative ways. Mosquito borne diseases kill millions of people and insect pests consume and destroy around US $40 billion worth of food each year. In contrast, insects pollinate at least two-thirds of all the food consumed in the world. In order to control populations of disease vectors and agricultural pests, researchers in entomology have developed numerous methods including chemical, biological and mechanical approaches. However, without the knowledge of the exact location of the insects, the use of these techniques becomes costly and inefficient. We are developing a novel sensor as a tool to control disease vectors and agricultural pests. This sensor, which is built from inexpensive commodity electronics, captures insect flight information using laser light and classifies the insects according to their species. The use of machine learning techniques allows the sensor to automatically identify the species without human intervention. Finally, the sensor can provide real-time estimates of insect species with virtually no time gap between the insect identification and the delivery of population estimates. In this paper, we present our solution to the most important challenge to make this sensor practical: the creation of an accurate classification system. We show that, with the correct combination of feature extraction and machine learning techniques, we can achieve an accuracy of almost 90 % in the task of identifying the correct insect species among nine species. Specifically, we show that we can achieve an accuracy of 95 % in the task of correctly recognizing if a given event was generated by a disease vector mosquito.

Keywords Classification · Feature extraction · Similarity search · Signal processing

D. F. Silva (✉) · V. M. A. Souza · G. E. A. P. A. Batista
Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Caixa Postal 668, São Carlos, SP, 13560-970, Brazil
e-mail: diegofsilva@icmc.usp.br

V. M. A. Souza
e-mail: vsouza@icmc.usp.br

G. E. A. P. A. Batista
e-mail: gbatista@icmc.usp.br

D. P. W. Ellis
Electrical Engineering, Columbia University, 500 W. 120th St., New York, NY 10027, USA
e-mail: dpwe@ee.columbia.edu

E. J. Keogh
Computer Science & Engineering Department, University of California, 900 University Ave, Riverside, CA 92521, USA
e-mail: eamonn@cs.ucr.edu

1 Introduction

Humans have always lived alongside insects, and insects impact our lives in many ways, both positive and negative. Mosquito borne diseases are a major problem across much of the world: It is estimated that dengue, a disease transmitted by mosquitoes of the genus *Aedes*, affects between 50 and 100 million people every year and is considered endemic in more than 100 countries [41]. Malaria, transmitted by mosquitoes of the genus *Anopheles*, affects around 6 % of the world's population and it is estimated that there are over 200 million cases per year and about 7 million lethal cases in the last decade [42]. In agriculture, insect pests consume and destroy around US $40 billion worth of food each year [26].

At the same time, insects pollinate at least two-thirds of all the food consumed in the world, with bees alone responsible for pollinating one-third of this total [4]. Furthermore, many species have been used as bioindicators of environmental quality, since their presence/absence, distribution, and density, define the quality of the ecosystem, especially in relation to contaminants in the air, soil, and water [14].

In order to control populations of disease vectors and agricultural pests, researchers in entomology have developed numerous methods of insect control [38]. Insects populations can be controlled using chemical methods, such as insecticides; biological methods, such as the release of male sterile individuals; and mechanical methods, for instance using insect traps. However, without the knowledge of the exact location of the insects, the use of these techniques becomes costly and inefficient.

Currently, information on the spatio-temporal distributions of insects is obtained with traps, usually adhesive, which are collected periodically and analyzed by experts who manually identify and count the insects. Although adhesive traps are very inexpensive, the tasks of distribution, collecting and analyzing these traps are labor intensive and therefore expensive in terms of human time. Adhesive traps also involve a large time lag between the moment the trap is installed in the field and the subsequent analysis by experts.

We are developing a novel sensor as a tool to control disease vectors and agricultural pests. This sensor captures insect flight information using laser light and classifies the insects according to their species. This sensor has several advantages when compared to the current techniques for estimating the distribution of insects. It uses commodity electronics and therefore is inexpensive. The use of machine learning techniques allows the sensor to automatically identify the species without human intervention. Finally, the sensor can provide real-time estimates of insect species with virtually no delay between the insect identification and the delivery of population estimates.

In this paper, we focus on the most significant challenge encountered in developing the sensor: the creation of an accurate classification system. An insect crossing the laser results in a brief perturbation in the signal. Such events last for tenths of a second and have a very simple structure, consequence of the wings movements. Nevertheless, we managed to successfully extract features containing adequate information for species identification using speech and audio analysis techniques.

We show that, with the correct combination of feature extraction and machine learning techniques, we can achieve an accuracy of almost 90 % in the task of identifying the correct insect species among nine species with data collected by the sensor. More importantly, we show that we can achieve an accuracy of 95 % in the task of correctly recognizing if a given event was generated by a disease vector mosquito.

This paper is an extended revision of [32, 33]. We provide a broader experimental evaluation that includes the classification based on similarity and the use of feature subset selection to analyze which feature extraction techniques provide the most relevant features for classification.

The rest of this paper is organized as follows. Section 2 discusses related work on automatic identification of insects. Section 3 describes the sensor used in this work, as well the data collection procedure. Section 4 describes the classification of time series by similarity and by feature extraction. Section 5 presents the results obtained by these approaches. Finally, we present our conclusions and directions for future work in Section 6.

2 Related Work

The idea of performing automatic classification of insects using acoustic devices dates back to 1945. Kahn et al. [12] used a microphone, a signal amplifier, a low-pass filter, and a recorder to register and

study the inaudible sounds produced by disease vector mosquitoes. They collected the sounds of four species: *Anopheles quadrimaculatus, Aedes aegypti, Aedes albopictus*, and *Culex pipiens*. It was necessary to establish an environment without external noise and under controlled conditions of temperature and humidity to collect the mosquitoes' sound. In this study, different sounds that could represent the insect behaviors were identified. Furthermore, the study showed that pitch can be used to distinguish males and females of the same species. This is possible because the sounds produced by male mosquitoes have a higher frequency than the sounds produced by female mosquitoes.

A few years later, Kahn & Offenhauser [13] mentioned that the fast evolution of electronic devices for sound recording would make the study of insect behavior using the sounds they produce easy, fast and accurate. However, after more than six decades, mechanical traps are still the most common technique for estimating the insect population in a given area.

While Kahn et al. worked with ideal conditions, microphones in non-controlled environments are highly vulnerable to external interference, such as the sounds produced by wind, cars, people, and other animals. Taking into account these difficulties, Moore et al. [21] proposed the use of an optical sensor based on the phototransistor previously presented by Unwin & Ellington [37]. The authors used the optical sensor to record the variation of the light caused by the passage of insects. Using this device, they performed an analysis of the wing-beat frequency of two species of the genus *Aedes* from both sexes. The automatic classification of species and sex was subsequently presented in [19].

Some years later, Moore [20] proposed an insect data collection system based on the previously proposed optical sensor connected to a computer running tools to process the obtained signal. Above the sensor, he positioned a transparent plastic jar with flying insects and above the jar a halogen lamp provided a light source. The main advantage of this sensor is that it is not susceptible to external interference. However, the use of a lamp as a light source may impact the behavior of insects that typically have activity periods influenced by daylight, known as circadian rhythms [35].

More recently, a research group from Universidade de São Paulo, University of California Riverside and ISCA Technologies proposed a new optical sensor to automatically identify flying insects [1]. The basic components of this sensor are a laser light source, an array of phototransistors, and a circuit board to filter and record the variation in the laser caused by the insects that cross the light plane.

Thanks to its coherence properties, a laser can be kept focused to a tight spot (which can be spread to a plane with a diffractor) and stay narrow over long distances. This allows the sensor to cover a large area without disturbing the surrounding environment. Machine learning techniques allow the sensor to automatically classify the data collected in real-time and opens a wide range of applications in vector and pest control. The next section provides more details about the sensor.

3 Laser Insect Sensor

The insect sensor under development is an invaluable tool to assist in the control of disease vectors and agricultural pests. In the next sections we provide an overview of the sensor, data collection procedure, and pre-processing.

3.1 Sensor Design

Figure 1 shows the general design of the sensor. It consists of a low-powered planar laser source pointed at an array of phototransistors. When a flying insect crosses the laser, its wings partially occlude the light, causing small light variations which are captured by the phototransistors. An electronic circuit board filters and amplifies the signal and a digital recorder captures the output.

The sensor signal is very similar to an audio signal captured by a microphone, even though the data are obtained optically. The data captured by the sensor consist of background noise with occasional events, resulting from brief moments in which an insect flies across the laser.

3.2 Collecting and Preprocessing Data

To evaluate and compare classification methods, we used laboratory data for which ground-truth labels were available. True class labels for each insect passage are necessary to assess the classification procedures. These data were collected in experimental

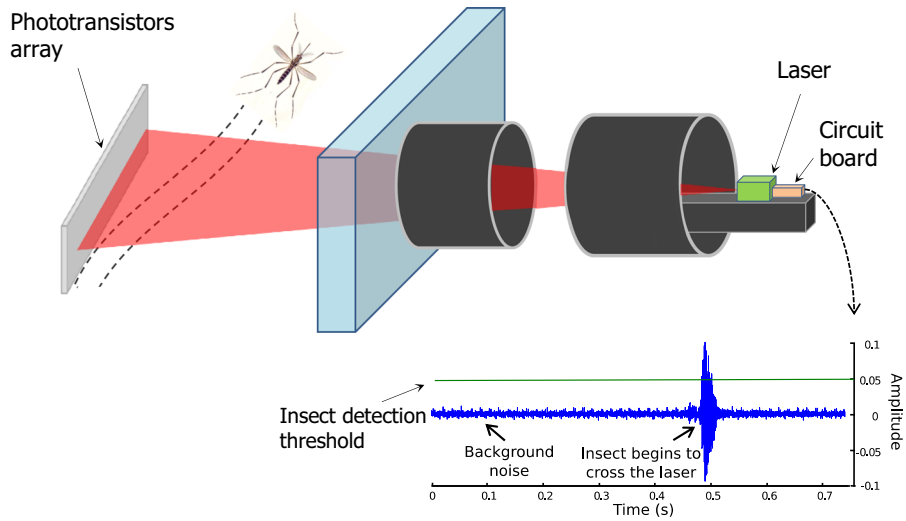


Fig. 1 The logical design of the sensor. A planar laser light is directed at an array of phototransistors. When an insect flies across the laser, a light variation is registered by the phototransistors as a time series

chambers ("insectaries") containing between 20 and 40 individuals of a single species, with an individual sensor attached to each chamber. Figure 2 shows some examples of insectaries used in our experiments.

After collecting the data, we preprocessed the recordings and detected the insect passages in the raw data. We designed a detector responsible for identifying the events of interest and separating them from background noise. The general idea of the detector is to move a sliding window across the raw data and calculate the spectrum of the signal inside the window. As most insects have wing-beat frequencies that range from 100 Hz to 1000 Hz, we used the maximum magnitude of the signal spectrum in this range as the detector confidence. All signals with magnitude above a user-specified threshold are considered an event generated by an insect. The high signal-to-noise ratio of the data collected by the sensor allows the user to specify low values for the threshold without the risk of false positives. Algorithm 1 shows a pseudo-code for the detector and Fig. 3 illustrates how the detector works.

Algorithm 1: Detector algorithm for wing-beat frequencies

Require: A sound data vector s
A window size w_s
A step size s_s
A minimum interest frequency $min_f = 100$ Hz
A maximum interest frequency $max_f = 1000$ Hz

```
s_z = Znormalize(s)
for i = 1 to length(s_z) - w_s step s_s do
    window = s_z(i : i + w_s)
    f = fft(window)
    d(i:i + w_s) = max(abs(f(min_f:max_f)))
end for
return d
```

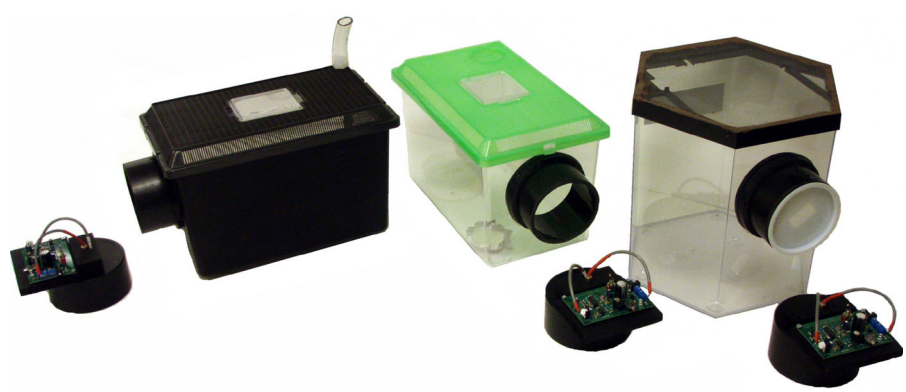


Fig. 2 Examples of experimental chambers for data collection

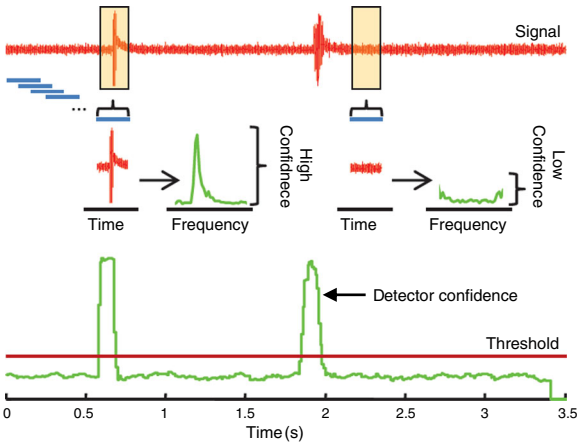


Fig. 3 Illustration of how the wing-beat detector works [2]

The detector outputs audio fragments which usually last for a few tenths of a second and have at least one insect passage. Due to the simplicity of our electronics, there is some noise mixed with the insect signals. We filtered most of the noise using a digital filter based on spectral subtraction, responsible for the removal of certain frequency ranges of signal [6]. This filter uses the spectrum of a window at the beginning of the signal as noise estimate. Then, a sliding window runs across the signal and, for each window, calculates the spectrum of the signal within the window. Finally, the filter reduces the spectral magnitude at each frequency by the corresponding noise estimate and calculates the inverse transform to reconstruct the signal without noise. Figure 4 shows an example of a filtered and segmented signal.

4 Computational Approaches for Automatic Insect Classification

Time series mining has attracted a huge amount of attention in the past few years. This is mainly due to the numerous application domains that generate temporal data, such as medicine, economics, and signal processing. Among all time series data mining tasks, our particular interest is in classification. In this task, an unknown time series must be associated to a label chosen from a finite set of labels.

In this section, we describe two strategies for classifying time series. We start presenting the basic concepts of time series classification by direct waveform similarity. Then, we discuss the strategy of classifying time series by feature extraction, focusing on the classification of audio signals.

4.1 Classification by Similarity

Studies have shown show that the simple nearest neighbor algorithm, often used in the classification of time series, can prove difficult to beat [7, 40]. This approach to time series classification assumes that similar series, as defined by a distance function calculated directly on the series values, are more likely to belong to the same class.

Of course, the success of this approach is wholly dependent on the choice of a distance measure. Many distance measures used for time series require a linear alignment of the series, meaning that the distance between the time series is represented by a distance function computed over pairs of observations located in the same position on the time axis. Table 1 presents twelve distance measures with such a characteristic, frequently known as non-elastic measures. A recent study compared these measures for classification of time series in different application domains [8].

An important problem with the distances in Table 1 is that many applications require a more flexible matching between observations, in which an observation of the time series x at the time t_x can be matched with an observation of the time series y at the time $t_y \neq t_x$. Dynamic Time Warping (DTW) is a basis for elastic distances that provide the shortest distance considering a non-linear alignment of observations according some constraints. Due to space restrictions, we recommend [30] for further details about DTW.

Although classification by direct sequence similarity is widely used in time series research, feature extraction allows the focus to be placed on particular signal properties, while discounting others, and facilitates the use of different machine learning approaches. For these reasons, this approach is extensively used in digital signal classification. The next section presents some concepts and techniques of feature extraction in time series domains.

4.2 Feature Extraction

The second classification strategy used in this paper is the use of standard Machine Learning approaches on a set of features extracted from the signals. This

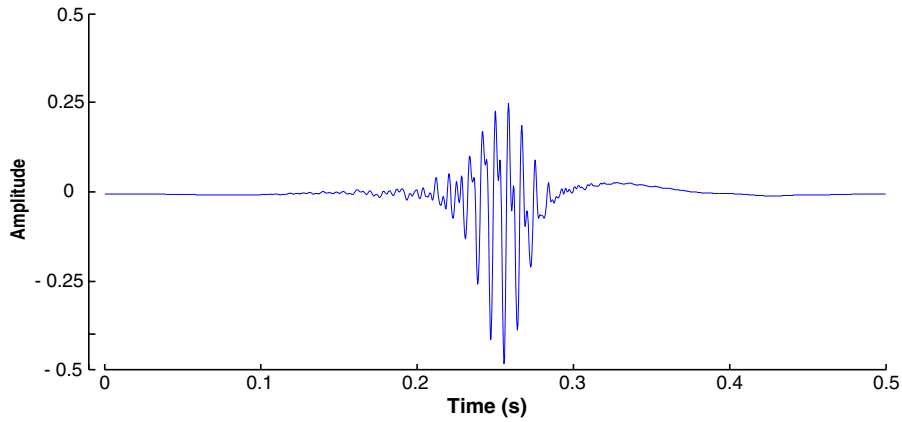


Fig. 4 Example of a segmented and filtered signal generated by an *Aedes aegypti* mosquito

Table 1 Distance measures to compare time series considering the linear (non-elastic) alignment of observations where x and y are time series with length n

Euclidean	$D_{euc}(x,y) = \sum_{k=1}^{n} \sqrt{(x_k - y_k)^2}$
Manhattan	$D_{man}(x,y) = \sum_{k=1}^{n} \lvert x_k - y_k \rvert$
Cosine	$D_{cos}(x,y) = 1 - \frac{\sum_{k=1}^{n} x_k y_k}{\sqrt{\sum_{k=1}^{n}(x_k)^2}\sqrt{\sum_{k=1}^{n}(y_k)^2}}$
Correlation	$D_{cor}(x,y) = 1 - \frac{n\sum_{k=1}^{n} x_i y_i - \sum_{k=1}^{n} x_i \sum_{k=1}^{n} y_i}{\sqrt{n\sum_{k=1}^{n} x_i^2 - (\sum_{k=1}^{n} x_i)^2}\sqrt{n\sum_{k=1}^{n} y_i^2 - (\sum_{k=1}^{n} y_i)^2}}$
Canberra	$D_{can}(x,y) = \sum_{k=1}^{n} \frac{\lvert x_k - y_k \rvert}{x_k + y_k}$
Chebyshev	$D_{che}(x,y) = \max_k \lvert x_k - y_k \rvert$
Jaccard	$D_{jac}(x,y) = \frac{\sum_{k=1}^{n}(x_k - y_k)^2}{\sum_{k=1}^{n} x_k^2 + \sum_{k=1}^{n} y_k^2 - \sum_{k=1}^{n} x_k y_k}$
Topsoe	$D_{top}(x,y) = \sum_{k=1}^{n} \left(x_k ln\left(\frac{2x_k}{x_k + y_k}\right) + y_k ln\left(\frac{2y_k}{x_k + y_k}\right)\right)$
Clark	$D_{clk}(x,y) = \sqrt{\sum_{k=1}^{n} \left(\frac{\lvert x_k - y_k \rvert}{x_k + y_k}\right)^2}$
Avg L_1 L_∞	$D_{avg}(x,y) = \frac{\sum_{k=1}^{n} \lvert x_k - y_k \rvert + \max_k \lvert x_k - y_k \rvert}{2}$
Squared χ^2	$D_{Sq\chi^2}(x,y) = \sum_{k=1}^{n} \frac{(x_k - y_k)^2}{x_k + y_k}$
Additive Symmetric χ^2	$D_{\chi^2 Ad}(x,y) = \sum_{k=1}^{n} \frac{(x_k - y_k)^2 (x_k + y_k)}{x_k y_k}$

strategy is commonly employed in research papers that perform classification of audio signals. We direct the interested reader to [5] for a detailed reference on audio processing. In the remainder of this section, we describe features that can be extracted from temporal, spectral and cepstral representations, as well as features based on linear prediction coefficients.

In the course of this section consider the following notation:

- A time series is represented by a vector x with length n;
- Each observation is represented by x_i, where $1 \leq i \leq n$;
- The frequency spectrum of the signal is represented by a vector Y;
- The length of vector Y is the number of different frequencies analyzed, N.

4.2.1 Feature Extraction in Temporal Representations

Features from temporal representations are useful to better understand the behavior of time series. For instance, the signal duration is perhaps the simplest feature that can be extracted from a time series. This information may be important in many domains of audio signal analysis. For instance, it can be used to distinguish violin notes caused by plucking (pizzicato) versus bowing.

Many temporal features aim to approximate the primitive features that describe a waveform, such as amplitude and period. For instance, the amplitude can be represented by the mean absolute amplitude of the signal, or the root mean square (RMS), defined by Eq. 1.

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{1}$$

An alternative way to estimate the amplitude is obtained by calculating the RMS without the square root. This feature is known as short-time energy (STE). Peak excursion is another feature easily derived from the waveform, defined as the difference between the maximum and minimum values of amplitude in the entire signal.

Another frequently-used class of features attempt to measure the concentration of signal energy in time, for instance, the temporal centroid (TC), Eq. 2, defines the temporal 'center of gravity' of the signal, providing an estimation of its energy location in time.

$$TC = \frac{\sum_{i=1}^{n} i x_i}{\sum_{i=1}^{n} x_i} \tag{2}$$

The period (or the wavelength) may be associated to the zero-crossing rate (ZCR), as defined by Eq. 3. The zero-crossing rate can also be used to estimate the level of noise in the signal. Larger ZCR values often indicate a high level of noise.

$$ZCR = \frac{1}{n-1}\sum_{i=2}^{n} |S(x_i) - S(x_{i-1})| \tag{3}$$

$$S(x_i) = \begin{cases} 1, & \text{if } x_i \geq 0 \\ 0, & \text{other case} \end{cases}$$

ZCR is commonly interpreted as an estimate of signal complexity. An alternative way to estimate the complexity of the signal is by measuring the variation of its level with time. Intuitively, the complexity estimate should have higher values when there are many peaks and valleys in the signal and lower values when the signal is more "well behaved". Equation 4 defines a complexity estimate measure [3].

$$CE = \sqrt{\sum_{i=1}^{n-1} (x_i - x_{i+1})^2} \tag{4}$$

Some descriptive statistics are also frequently used as features in the temporal representation. Two examples are variance and standard deviation. Moreover, higher-order measures such as skewness and kurtosis may be used to characterize the distribution of waveform values in the signal. Specifically, skewness is a measure of symmetry, where values close to zero mean symmetric, positive values mean a greater concentration in the beginning of the signal, and negative values mean that the energy is more concentrated at the end. Kurtosis measures the flatness of amplitude distributions, relative to the normal distribution. Low kurtosis values indicate a flat distribution. Both measures are commonly used as histogram descriptors, but may be used as time series features.

4.2.2 Feature Extraction in Spectral Representations

Frequently, certain properties of the signal are made explicit when the signal is represented in the frequency domain. For instance, in music applications, the fundamental frequency is closely related to the musical

note emitted. In the classification of flying insects, either by optical sensors or acoustic recordings, this characteristic often reflects the wing beat frequency of the insect.

Beyond the fundamental frequency, the spectrum of a signal also has harmonic components with (typically) smaller magnitudes multiples of the fundamental frequency. In a purely periodic signal that can be exactly represented as a sum of sine waves, this statement is easily verified. In complex signals, such analysis can experience difficulties. For instance, small perturbations can cause a slight displacement of the harmonic components. Therefore, feature extraction procedures frequently encompass a search to find the true positions of the harmonic components in the frequency spectrum.

Figure 5 shows an approach to perform this analysis based in a method proposed by Park [25]. This task, known as harmonic analysis, searches the peak magnitude at frequencies close to the theoretical harmonics. In other words, the method searches the frequency with highest magnitude in an area around the multiples of the fundamental frequency. The values found are of called estimated harmonics.

Harmonic analysis opens several possibilities for classification features. For instance, Eq. 5 defines the inharmonicity, which measures the average difference between the theoretical and the estimated harmonics.

$$Inharm = \sum_{i=1}^{N_{harm}} \frac{|F_i - i\,F_0|}{i\,F_0} \tag{5}$$

where, F_i is the position of the i-th estimated harmonic and N_{harm} is the number of analyzed components.

The *tristimulus* is another example of features from harmonic analysis. They are equivalent to the color features used by human vision [28]. The basic idea of tristimulus in human vision is that any color can be obtained by combining the primary colors. In relation to human hearing, any sound can be distinguished by characteristics related to the fundamental frequency and harmonics. These features are defined by Eqs. 6, 7 and 8, where H_k refers to the magnitude of the k-th harmonic, being the number 0 related to the fundamental frequency.

$$ts1 = \frac{H_0}{\sum_{k=0}^{N_{harm}} H_k} \tag{6}$$

$$ts2 = \frac{\sum_{k=1}^{3} H_k - 1}{\sum_{k=0}^{N_{harm}} H_k} \tag{7}$$

$$ts3 = \frac{\sum_{k=4}^{N_{harm}} H_k}{\sum_{k=0}^{N_{harm}} H_k} \tag{8}$$

Apart from harmonic analysis, exploring the variation of the spectrum along frequency can also provide useful information for analyzing audio signals. For instance, the spectral irregularity [16] (or spectral smoothness [18]) reveals the variability of neighboring frequencies in the spectrum. It considers the difference between the magnitudes of the current, previous, and next frequencies for each component. Formally, the spectral irregularity is defined by the Eq. 9.

$$SI = \sum_{i=2}^{N-1} \left(f_{si}(i) \frac{f_{si}(i-1) + f_{si}(i) + f_{si}(i+1)}{3} \right) \tag{9}$$

where $f_{si}(i) = 20 \log_{10}(Y_i)$

A variation of the Eq. 9 is also used as an estimate of variability between neighboring frequencies. This measure, named here as modified spectral irregularity, is described by Eq. 10.

$$SI_{modif} = \frac{\sum_{i=2}^{N}(Y_i - Y_{i-1})^2}{\sum_{i=2}^{N}(Y_i)^2} \tag{10}$$

The spectral flux is another measure to estimate the variability of magnitudes in the frequency spectrum. It estimates the rate of variation of magnitude values of frequency components. Equation 11 defines the spectral flux.

$$Flux = [\sum_{i=1}^{N-1}(|Y_i - Y_{i+1}|)^q]^{1/q} \tag{11}$$

where, q in an integer parameter, commonly set to 2 [39].

Other spectral features can be obtained by equations similar to those used for extracting features in the temporal representation. For example, the centroid is also a commonly used measure in the spectral domain. This measure represents the geometric center of the energy concentration of the frequency components. Intuitively, if there are components with high magnitudes at lower frequencies, the centroid should have a small value. If the energy is not concentrated in low frequencies, the value of the centroid will be higher.

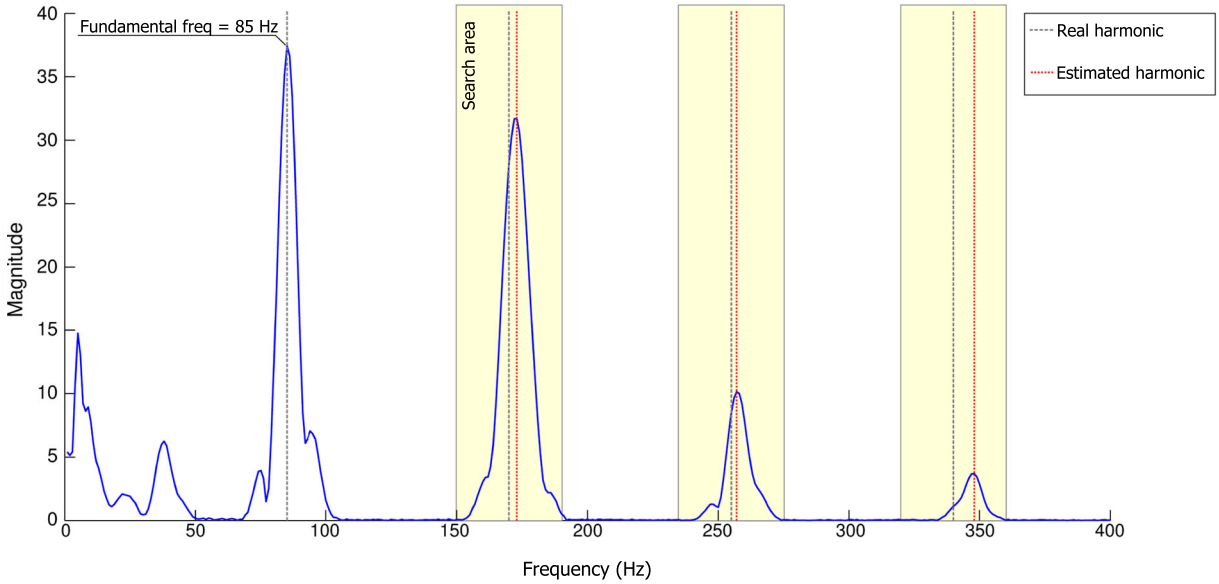


Fig. 5 Example of harmonic analysis procedure. In this example, the frequencies relative to the theoretical harmonics are 170 Hz, 255 Hz and 340 Hz, and the frequencies related to the estimated harmonics are 173 Hz, 257 Hz and 348 Hz

Similarly to the centroid, spectral roll-off measure estimates the concentration of magnitudes in the frequency ranges. This feature determines the frequency below which the energy spectrum contains some proportion (e.g., 85 %) of its total. Thus, the more energy is concentrated at low frequencies, the lower the value of the roll-off.

The shape of spectrum can also be estimated by measuring its flatness. This measure is defined by the ratio between geometric and arithmetic means of the magnitudes of the spectrum. Values close to 0 indicate approximately sinusoidal signals (all energy concentrated at a single frequency).

Other statistics, including variance, standard deviation, kurtosis and skewness, as well as average magnitude and energy, are also simply adapted to spectral representations.

4.2.3 Feature Extraction in Cepstral Representations

Cepstral coefficients are the most common features from the cepstral domain. Such coefficients are frequently represented in an acoustically defined scale created from a study by Stevens et al. [34]. This study relates the physical frequencies to the frequencies perceived by the human ear. This scale, called *mel*, is the basis for the mel-frequency cepstral coefficients (MFCC). MFCCs are popular features in various application domains, particularly speech and speaker recognition [43] as well as musical instruments classification [36]. Equation 12 defines the scale conversion from frequency (f) to mel (m).

$$m = 2595 log_{10}\left(1 + \frac{f}{700}\right) \tag{12}$$

In some applications, the assumption that a scale based on human auditory system is the most appropriate can be inappropriate. Therefore, other scales may be used. An example is the logarithmic scale. We can also perform the same operation without using any scale transformation.

In order to calculate the MFCC, we take the magnitudes of the frequency components using the mel scale and apply the discrete cosine transform (DCT) - i.e., using only cosine waves as components, widely used for data compression - over the logarithm of these values. The MFCCs are the amplitudes of the cepstrum generated by this operation.

In addition to cepstrum coefficients, cepstral representations can also be used to estimate the fundamental frequency [22]. On a linear frequency axis, a periodic signal will give rise to regularly-spaced harmonics, and the DCT will reflect this periodicity in a corresponding bin. Figure 6 shows an example of the fundamental frequency estimation in the cepstral

domain. In this case, the period found was 0.001563 s. Therefore, the frequency is $1/0.001563 = 639.79$ Hz.

4.2.4 Linear Prediction Features

Linear prediction (LP) is a technique used in many speech applications, such as recognition, compression and modeling [17, 29]. The idea behind linear prediction is to represent a signal as a linear combination of previously observed values, as in Eq. 13.

$$\hat{x}_k = \sum_{i=1}^{p} a_i x_{k-i} \tag{13}$$

where k is the time index and p is the order of LP, i.e., the number of employed LP coefficients. The a_i coefficients are calculated in order to minimize the prediction error using a covariance or auto-correlation method. This model turns out to be a good match to many real-world signals, including speech.

In a data transmission, it is only necessary to send a_l coefficients and the prediction error E_k, so that the amount of data sent can be much less than the original signal. For this, an analysis filter using a transmission function, which attempts to suppress frequencies with high magnitudes, compresses the signal. To receive this signal, a receive filter uses the inverse function of the transmission function, amplifying the attenuated frequencies [5].

Equation 13 can be rewritten in the frequency domain with a $z-$transform [23]. In this way, a short segment of speech is assumed to be generated as the output of an all-pole filter $H(z) = 1/A(z)$, where $A(z)$ is the inverse filter such that:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{14}$$

The Line Spectral Frequencies (LSF) representation, introduced by Itakura [11], is an alternative way to represent LP coefficients. In order to calculate LSF coefficients, the inverse polynomial filter is decomposed into two polynomials $P(z)$ and $Q(z)$:

$$P(z) = A(z) + z^{p+1}A(z^{-1}) \text{ and } Q(z) = A(z) - z^{p+1}A(z^{-1}) \tag{15}$$

where $P(z)$ is a symmetric polynomial and $Q(z)$ is an antisymmetric polynomial. The roots of $P(z)$ and $Q(z)$ determine the LSF coefficients.

LSF is well suited for quantization and interpolation [24]. Therefore LSF can represent the speech signal, mapping a large signal to a small number of coefficients, more efficiently than other LP representations.

5 Experimental Results

In this paper, we used the strategies of classification by direct similarity and feature extraction to automatically identify insect species. In this section, we present the results obtained using both approaches. First, we include a description of the dataset used in our experiments.

5.1 Dataset Description

Our dataset consists of four species of mosquitoes: *Aedes aegypti*, vector of filariasis, dengue fever, yellow fever, and West Nile virus; *Anopheles gambiae*, vector of malaria; *Culex quinquefasciatus*, vector of lymphatic filariasis; and *Culex tarsalis*, vector of St. Louis Encephalitis and Western Equine Encephalitis. Also included are three species of flies: *Drosophila melanogaster*; *Musca domestica* and *Psychodidae diptera*, as well as the beetle *Cotinis mutabilis* and the bee *Apis mellifera*.

The dataset has 18,115 instances. This relatively large number of instances allows us to split the dataset into training and test partitions. Such a procedure facilitates the direct comparison among classifiers and is less computational demanding than resampling methods. We performed a stratified division with 33 % of the examples in the training set and the remaining in the test set. The large test sets are important to reduce variance and to increase the confidence in the results. Table 2 presents the class distributions in the dataset.

5.2 Classification by Similarity on Signal Representations

In this section, we evaluate different distance measures applied to the spectrum and the cepstrum of the signals. Since the data represented in the time domain is high-dimensional, complex, and constituted of weak features, we do not evaluate this representation in our similarity experiments.

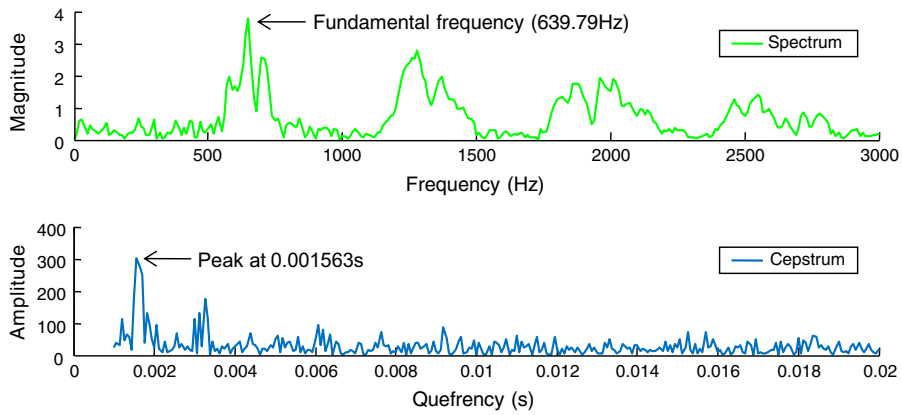


Fig. 6 Example of the spectrum *(top)* and the cepstrum *(bottom)* of a signal indicating the peak of highest amplitude in the cepstrum, representing the period related to the fundamental frequency

Table 3 presents the distance measures used in this experiment and the classification accuracy rates. We used DTW with a constraint band [31] fixed at 5 observations. In other words, we restricted the maximum distance between two observations matched in the non-linear alignment. This procedure avoids spurious matches, such as magnitude peaks in distant frequencies in the spectrum.

In general, we can see in Table 3 that similarity results in the spectrum outperform the cepstral domain. This is more evident by the fact that the accuracies presented a high variance in the cepstral domain. Among the best results, only the DTW measure has achieved a good result in the cepstrum when compared with those achieved in the spectrum. However, DTW is the most expensive distance among all the distance measures used in this work.

Table 2 Dataset class distribution by species

Species	# of Instances	Distribution (%)
Aedes aegypti	4,756	26.25
Anopheles gambiae	1,411	7.79
Apis mellifera	511	2.82
Cotinis mutabilis	172	0.95
Culex quinquefasciatus	3,137	17.32
Culex tarsalis	5,309	29.31
Drosophila melanogaster	777	4.29
Musca domestica	1,343	7.41
Psychodidae diptera	699	3.86
Total	18,115	100.00 %

5.3 Machine Learning Techniques on Extracted Features

In this section, we present the results obtained by different Machine Learning systems trained over features extracted from the signal. This approach uses the temporal, spectral, and cepstral representations of the signal presented in Section 4.

Throughout this section we use the designations *temporal features* and *spectral features* to refer to the feature vectors extracted from time and frequency domains, respectively. Table 4 lists the features that compose each of these vectors.

Table 3 Result of classification by similarity over the spectrum and the cepstrum

Distance	Accuracy (%)	
Measure	Spectrum	Cepstrum
Euclidean	76.14	78.66
Manhattan	80.09	67.24
Cosine	77.25	76.29
Correlation	76.60	75.34
Canberra	72.28	27.78
Chebyshev	71.20	74.71
Jaccard	77.26	78.73
Topsoe	81.54	73.50
Clark	75.59	26.79
Average L_1 L_∞	80.09	67.61
Squared χ^2	81.38	73.89
Additive Symmetric χ^2	81.01	47.99
DTW (constraint band = 5 observations)	81.04	80.34

Table 4 List of features that compose temporal and spectral feature vectors

Domain	Feature
Temporal	Mean amplitude, Root mean square, Short-time energy, Interval, Temporal centroid, Zero-crossing rate, Complexity estimate, Variance, Standard deviation, Skewness, Kurtosis, Duration
Spectral	Fundamental frequency, Inharmonicity, Tristimulus 1, Tristimulus 2, Tristimulus 3, Flux, Spectral centroid, Spectral irregularity, Modified spectral irregularity, Variance, Standard deviation, Skewness, Kurtosis, Mean magnitude, Energy, Roll-off, Flatness

Most learning algorithms have parameters that can significantly influence their performance. Therefore, in each experiment we performed a search in the parameter space seeking to maximize classification accuracy. Since the use of test data is restricted to the final classifier evaluation, we used 10-fold cross-validation on the training data to search the parameter values. For each possible combination of parameter values, the accuracy of the classifier was measured in the "internal" cross-validation test sets. We use the best combination of parameter values for a given learning algorithm as the final setting, and then use this combination to learn over the entire training set and evaluate the resulting classifier on the test set.

In the case of the Support Vector Machine classifier, we used grid search [10] to vary the parameters of the base algorithm and of the kernel. Given values of minimum, maximum and step size, we evaluate the cross-validation accuracy of each combination of parameters. However, this search is performed with only a coarse estimate, using 2-fold cross-validation. The search is then refined in regions with the best results. Table 5 describes the learning algorithms, as well as parameter ranges.

Table 6 presents the accuracy results different classifiers and feature sets. For reasons of readability, we omit results obtained by Naïve Bayes and J48 classifiers, since they achieved the worst results across all feature sets. Additionally, we only show the results for SVM RBF since SVM Poly had inferior results.

The best results were obtained by using MFCC, with LFC and LSF achieving slightly lower accuracy rates. The results obtained with temporal features and LPC were substantially lower than the other feature sets. The best single-classifier performance, 87.33 %, was obtained with the SVM RBF classifier with MFCC features. This is a respectable accuracy rate given the similarity among the signals generated from different classes and the simplicity of the signals in terms of length and structure.

Table 7 shows the confusion matrix obtained by the execution of the SVM RBF classifier trained on 40 MFCC values. This matrix indicates that the confusions mainly occur between the most similar species. For example, a relatively common error is related to the *Aedes aegypti* being classified as one of the other three species of mosquitoes.

The second experiment investigates combining the output of different classifiers. This experiment explores the possibility that classifiers trained with different feature sets can make independent errors, so that an ensemble of classifiers can exploit this diversity to improve classification. We explored three different strategies to combine the results of the base classifiers. The first and simplest is voting: each classifier votes for the predicted class and the class with the highest number of votes gives the final answer. In case of tie, the class with highest prior probability is chosen.

The other two strategies use sum and product functions over the output score of each classifier. One advantage of these strategies in relation to voting is that they consider the fact that classifiers can assign similar score values to two different classes when an object is close to the decision border. Therefore, the classification of borderline cases can potentially benefit from such forms of classifier combination.

We evaluated the hypothesis that the combination of different representations can provide enough diversity to improve the classification accuracy. We performed experiments with different combinations of feature sets using the same induction algorithm.

First, we checked if different frequency scales used to extract cepstral coefficients can be complementary. Thus, we created combinations of LFC, LLFC

Table 5 Learning algorithms with their respective parameter ranges

Algorithm	Parameters range (initial:step:final)
Decision Tree (J48 implementation)	Pruning factor P = 0.1:0.1:0.5
Gaussian Mixture Models (GMM)	Number of components N = 3:2:21
K-Nearest Neighbors (KNN)	Number of neighbors K = 1:2:25
Naïve Bayes (NB)	–
Random Forest (RF)	Number of trees N = 5:2:75
Support Vector Machine – Polynomial kernel (SVM Poly)	Complexity C / Polynomial Degree $C = 10^i, i = -7:1:5$ / D = 1:1:3
Support Vector Machine – RBF kernel (SVM RBF)	Complexity C / γ $C = 10^i, i = -7:1:5$ / $\gamma = 10^i, i = -4:1:0$

and MFCC. We also used LSF and spectral features in combination with MFCC, since they are the best known and most used cepstral features and achieved some of the best results in our first experiment, and LFC, which obtained competitive results in comparison to MFCC. In addition, we also evaluated the combination of all feature sets (LFC, LLFC, MFCC, LSF and spectral). Table 8 shows the results.

The combination of different feature sets improved the accuracy in several cases. In total, 31 (64.58 %) of the analyzed cases showed some improvement. The combination of all feature sets improves the accuracy over the base classifiers in all cases.

So far, we have used combinations of the classifiers outputs, trained by different feature sets. A related analysis is to build a dataset with all features and evaluate the performance of classifiers induced over it. Such dataset consists of 529 features, being 100 LFC, 100 LLFC, 100 MFCC, 100 LSF, 100 LPC, 12 temporal features and 17 spectral features.

Obviously, a dataset with such a large number of features obtained from a signal with simple structure is very likely to have redundant or irrelevant features. Therefore, we performed an additional experiment and created classifiers based on feature subsets. These subsets were obtained with two well-known feature selection algorithms, Correlation-based Feature Selection (CFS) [9] and Relief-F [15]. We have chosen these two algorithms because they present very different approaches to the problem. Relief-F is an algorithm focused in selecting the most relevant features, while CFS also takes into account the redundancy between the selected features. Another difference is that CFS evaluates an entire feature set while Relief-F evaluates the features independently. Although CFS tends to provide more consistent results in terms of redundancy and relevancy, Relief-F allows the evaluation of feature subsets of different sizes.

Since Relief-F provides ratings to each individual feature, we can create subsets of different values selecting the top k-ranked features. We selected 5 %, 10 %, 20 % and 30 % of total number of features (in absolute values 27, 53, 106 and 159 features). The CFS algorithm automatically selected 74 features. Table 9 presents the accuracy results for all features as well as the feature subsets selected by CFS and Relief-F. We show the results for KNN, SVM with RBF and Random Forest since these algorithms presented the best results in previous experiments.

The use of all features improved the performance for the Random Forest and SVM classifiers and decreased the performance for KNN. A likely reason for the performance decrease of KNN is that this algorithm depends on a distance function that tends to lose its discriminative power in higher dimensions. Regarding the feature selection algorithm, these methods seem to consistently improve the classification accuracy. In the case of CFS, it improved classification performance in all cases. For Relief-F, the classification improved when a reasonable amount of features

Table 6 Accuracy results per classifier and feature set with the corresponding parameter values

Feature Set	Algorithm	Selected Parameter Configuration	Accuracy (%)
LFC	KNN	#c= 75. $k = 7$	81.71
	RF	#c= 80. $T = 75$	83.49
	SVM RBF	#c= 95. $c = 10.\gamma = 1$	**86.93**
	GMM	#c= 100. $G = 9$	83.17
LLFC	KNN	#c= 15. $k = 7$	74.70
	RF	#c= 20. $T = 60$	76.30
	SVM RBF	#c= 70. $c = 10^4.\gamma = 0.01$	**79.05**
	GMM	#c= 20. $G = 17$	74.03
MFCC	KNN	#c= 30. $k = 5$	83.61
	RF	#c= 35. $T = 75$	85.39
	SVM RBF	#c= 40. $c = 10.\gamma = 1$	**87.33**
	GMM	#c= 45. $G = 13$	82.42
LPC	KNN	#c= 45. $k = 21$	56.18
	RF	#c= 65. $T = 75$	60.90
	SVM RBF	#c= 45. $c = 10^5.\gamma = 0.1$	**66.85**
	GMM	#c= 40. $G = 19$	54.15
LSF	KNN	#c= 95. $k = 5$	80.23
	RF	#c= 95. $T = 75$	84.25
	SVM RBF	#c= 100. $c = 10.\gamma = 1$	**84.97**
	GMM	#c= 75. $G = 17$	75.28
Temporal	KNN	$k = 11$	50.91
	RF	$T = 75$	60.13
	SVM RBF	$c = 10^5.\gamma = 0.1$	**60.62**
	GMM	$G = 19$	42.76
Spectral	KNN	$k = 5$	70.51
	RF	$T = 50$	**79.38**
	SVM RBF	$c = 10^5.\gamma = 0.1$	76.24
	GMM	$G = 21$	63.73

The best result in each feature set is marked in boldface

Table 7 Confusion matrix obtained by SVM RBF with 40 MFCC

Actual class	Predicted as								
	ae	ag	am	cm	cq	ct	dm	md	pd
Ae. aegypti (ae)	2890	67	0	0	69	125	0	2	0
An. gambiae (ag)	160	731	0	2	4	70	2	0	0
Ap. mellifera (am)	2	0	176	1	3	16	33	74	29
Co. mutabilis (cm)	2	0	9	100	0	0	3	6	1
Cu. quinquefasciatus (cq)	64	10	8	0	1893	78	11	7	5
Cu. tarsalis (ct)	110	51	8	1	91	3199	15	35	0
Dr. melanogaster (dm)	24	13	35	5	8	25	388	20	4
Mu. domestica (md)	0	0	64	1	1	16	8	776	34
Ps. diptera (pd)	0	0	26	0	0	1	8	62	395

Table 8 Results achieved by the combination of different feature sets with the same learning algorithm

Algorithm	Best Acc. (%)	Combined Feature Sets	Accuracy (%) Sum	Product	Voting
SVM RBF	87.33	LFC, LLFC, MFCC	**87.46**	87.27	**87.91**
		LFC, LSF, Spectral	86.83	86.44	87.09
		MFCC, LSF, Spectral	86.85	86.35	87.14
		All feature sets	**88.70**	**88.47**	**88.44**
KNN	83.61	LFC, LLFC, MFCC	**85.48**	**85.57**	**84.57**
		LFC, LSF, Spectral	**83.94**	83.56	82.46
		MFCC, LSF, Spectral	**84.82**	**84.45**	83.05
		All feature sets	**86.15**	**86.00**	**85.18**
GMM	83.17	LFC, LLFC, MFCC	**85.50**	**86.35**	**84.72**
		LFC, LSF, Spectral	83.17	**84.16**	81.49
		MFCC, LSF, Spectral	82.86	82.68	81.18
		All feature sets	**86.20**	**86.01**	**85.50**
RF	85.39	LFC, LLFC, MFCC	**86.69**	**86.93**	84.82
		LFC, LSF, Spectral	**86.50**	**86.36**	84.76
		MFCC, LSF, Spectral	**86.99**	**86.89**	**85.44**
		All feature sets	**87.83**	**87.97**	**86.14**

The boldface results represent an accuracy gain over the base classifier

is selected. In this case, 20 % (106) and 30 % (159) of the total.

Table 10 summarizes this analysis, showing the number of features selected from each feature extraction technique. MFCC and LSF have the highest number of selected coefficients. In particular, MFCC coefficients are selected in relatively large amounts in several settings. This seems to indicate that MFCC provides the most informative features.

5.4 Binary Classification

So far, we have evaluated our classifiers in a multiclass setting. However, many applications of the sensor will require a simpler binary-class setting. For instance, in public health and agriculture, frequently the main goal is to estimate the density of a disease vector or pest of interest, such as mosquitoes of genus *Aedes* or *Anopheles* in places that dengue or malaria are endemic, respectively. All other species are not of immediate interest and should be classified in a general negative class.

In this context, we analyzed the performance of classifiers that consider disease-carrying mosquitoes as positive class and other species as negative class. This setting leads to a considerable change in the classes' distribution. More specifically, we consider four scenarios in which each one of the following species is considered the positive class: *Aedes aegypti, Anopheles gambiae, Culex quinquefasciatus*

Table 9 Classification accuracy with all feature sets and feature selection techniques

Learning Algorithm	Best Feature Set	CFS	Relief-F 5 %	Relief-F 10 %	Relief-F 20 %	Relief-F 30 %	All Features
KNN	83.61	**86.19**	83.07	82.76	**83.85**	**85.23**	83.51
RF	85.39	**88.37**	**85.63**	**86.16**	**86.86**	**87.54**	**86.98**
SVM RBF	87.33	**88.78**	85.88	86.96	**87.38**	**89.55**	**89.14**

The boldface values are relative to those with better performance than the base classifier considering the best feature set for it

Table 10 Features selected by CFS and Relief-F algorithms

Feature Selection	Number of selected features							
Algorithm	LFC	LLFC	MFCC	LSF	LPC	Temporal	Spectral	Total
CFS	18/100	10/100	19/100	16/100	1/100	2/12	8/17	74/529
Relief-F 5 %	2/100	1/100	11/100	11/100	0/100	0/12	2/17	27/529
Relief-F 10 %	4/100	3/100	25/100	18/100	0/100	0/12	3/17	53/529
Relief-F 20 %	13/100	7/100	51/100	27/100	1/100	0/12	7/17	106/529
Relief-F 30 %	27/100	10/100	51/100	38/100	2/100	3/12	9/17	159/529

and *Culex tarsalis* while the remaining eight classes are considered negative. When *Aedes aegypti* is the positive class, the distribution is 26.25 %/73.75 %, when *Anopheles gambiae* is the positive class, the distribution is 7.79 %/92.21 %, when *Culex quinquefasciatus* is the positive class, the distribution is 17.32 %/82.68 % and, when *Culex tarsalis* is the positive class, the distribution is 29.31 %/70.69 %.

Since the class distributions change considerably in this experiment, we consider the area under the ROC curve (AUC) as an additional performance measure. AUC is not sensitive to changes in the operational conditions such as class distribution and misclassification costs [27].

We used 40 MFCC with a SVM-RBF classifier, since this configuration achieved the best result for a combination of classifier and feature extraction technique. Table 11 shows the accuracy and AUC for the four mosquito species considered as positive class.

5.5 Results Analysis

In this section we summarize, compare and make recommendations based on the experimental results obtained in this paper. We started with similarity classifiers on the spectrum and cepstrum. The spectrum frequently presented better classification performance with less variance among the different distances than cepstrum. Therefore, for similarity classification, we would recommend the use of spectrum for this application because of its superior classification performance, lower variance, and also because of its slightly lower computational cost.

Regarding the distances, Manhattan, Clark, Average, L_1 L_∞, Squared χ^2, Additive Symmetric χ^2 and DTW showed the best results. All but one distance measures have $O(n)$ time complexity. The exception is DTW, which is $O(n \times m)$, where m is the constraint band size. We use DTW with a very strict constraint band that diminishes the runtime differences between DTW and the other distances. Despite this, we recommend the use of the linear time measures due to their simplicity and faster running times.

For the feature extracting methods, MFCCs provide the best and most consistent results. The improvement of using SVM RBF with 40 MFCCs over the best similarity method was approximately 6 percentage points. The 87.33 % accuracy rate obtained by SVM is a respectable result that was difficult to beat even for more complex methods.

All improvements over SVM RBF trained with MFCCs obtained in other experiments were somewhat marginal and involved classifiers significantly more complex and computationally more expensive. The best ensemble method obtained 88.70 % (a 1.33 %

Table 11 Accuracy and area under ROC curve considering each of the disease vector mosquito species as positive class and the remaining as negative class

	Positive class			
	Ae. aegypti	*An. gambiae*	*Cu. quinquefasciatus*	*Cu. tarsalis*
Accuracy	95.00 %	96.93 %	96.91 %	94.41 %
AUC	93.20 %	86.10 %	94.20 %	93.00 %

percentage point improvement) and the combination of all feature sets provided the best accuracy of 89.14 % (a 1.81 % percentage point improvement). Although these more sophisticated methods tend to provide better accuracy rates, they involve considerably more computations to extract multiple feature sets and/or train and test multiple classifiers.

Given our results, we recommend MFCCs as the feature set that provides the best trade-off between classification performance and computational cost. MFCCs can be used by state-of-the-art machine learning classifiers to provide accurate classifiers.

6 Conclusion

In this paper we demonstrated the usefulness of signal processing features and similarity classification for an important application in public health and agriculture. We showed that MFCC features provide a good compromise between processing power necessary to obtain the features and classification accuracy. Such a compromise is very important in a scenario of embedded classification, since the sensor has limited capacity in terms of processing power and available memory. We obtained an accuracy of 87.33 % using a combination of 40 MFCC and SVM RBF, and 89.55 % using a more comprehensible set of features on a dataset with nine classes. In a binary setting, we obtained accuracy rates around 95 % and similar AUC values when four disease vector mosquito species are considered as a positive class.

We believe that our results support the application of the sensor in real world applications. There are several applications that require real-time estimation of spatio-temporal distributions of important insects. For instance, the sensor is a key component to create effective alarm systems for insect outbreaks. Better knowledge of the insect populations in a given area also allows the intelligent use of insect control techniques, such as insecticides. The idea is that such knowledge can be used to support more local application of the control techniques, reducing cost and increasing effectiveness. Finally, the sensor will be the heart of the next generation of insect traps that will capture only specific target species. Such traps will automatically recognize the insects captured and selectively trap only certain species, reducing the impact of the control device over the environment.

As future work, we plan to look at the sensor data as a stream data set, and adapt the techniques that present the most promising performance in this context. The sensor data stream has potential concept drifts caused by variations of environmental conditions, such as temperature, humidity, etc. Therefore, the classification techniques must adaptively learn from the data stream.

Acknowledgments This work was funded by São Paulo Research Foundation (FAPESP). Grants #2011/17698-5, #2012/50714-7 and #2013/26151-5.

References

1. Batista, G.E.A.P.A., Keogh, E.J., Mafra-Neto, A.: Sensors and software to allow computational entomology, an emerging application of data mining. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 761–764 (2011)
2. Batista, G.E.A.P.A., Hao, Y., Keogh, E.J., Mafra-Neto, A.: Towards automatic classification on flying insects using inexpensive sensors. In: IEEE International Conference on Machine Learning and Applications Workshops, vol. 1, pp. 364–369 (2011)
3. Batista, G.E.A.P.A., Keogh, E.J., Tataw, O.M., Souza, V.M.A.: CID: an efficient complexity-invariant distance for time series. Data Min. Knowl. Disc. **28**(3), 1–36 (2013)
4. Benedict, M.Q., Robinson, A.S.: The first releases of transgenic mosquitoes: an argument for the sterile insect technique. Trends Parasitol. **19**(8), 349–355 (2003)
5. Benesty, J., Sondhi, M.M., Huang, Y. (eds.): Springer Handbook of Speech Processing. Springer, Berlin (2008)
6. Boll, S.: Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. **27**(2), 113–120 (1979)
7. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.J.: Querying and mining of time series data: experimental comparison of representations and distance measures. Proceedings of the VLDB Endowment **1**(2), 1542–1552 (2008)
8. Giusti, R., Batista, G.E.A.P.A.: An empirical comparison of dissimilarity measures for time series classification. In: Brazilian Conference on Intelligent Systems, pp. 82–88 (2013)
9. Hall, M.A.: Correlation-based Feature Selection for Machine Learning. Ph.D. thesis, The University of Waikato (1999)
10. Hsu, C.W., Chang, C.C., Lin, C.J.: A Practical Guide to Support Vector Classification. Tech. rep., Department of Computer Science, National Taiwan University (2003)
11. Itakura, F.: Line spectrum representation of linear predictor coefficients of speech signals. J. Acoust. Soc. Am. **57**, S35 (1975)

12. Kahn, M.C., Celestin, W., Offenhauser, W. Jr.: Recording of sounds produced by certain disease-carrying mosquitoes. Science **101**(2622), 335–336 (1945)
13. Kahn, M.C., Offenhauser, W. Jr.: The identification of certain west African mosquitoes by sound. Am. J. Trop. Med. Hyg. **29**(5), 827–836 (1949)
14. Kevan, P.: Pollinators as bioindicators of the state of the environment: species, activity and diversity. Agric. Ecosyst. Environ. **74**(1-3), 373–393 (1999)
15. Kononenko, I., Šimec, E., Robnik-Šikonja, M.: Overcoming the myopia of inductive learning algorithms with RELIEFF. Appl. Intell. **7**(1), 39–55 (1997)
16. Krimphoff, J., McAdams, S., Winsberg, S.: Caractérisation du timbre des sons complexes. ii. analyses acoustiques et quantification psychophysique. Le Journal de Physique **4**(C5), C5–625 (1994)
17. Markel, J., Gray, A.: Linear Prediction of Speech, vol. 12. Springer-Verlag, New York (1976)
18. McAdams, S., Beauchamp, J.W., Meneguzzi, S.: Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. J. Acoust. Soc. Am. **105**, 882–897 (1999)
19. Moore, A.: Artificial neural network trained to identify mosquitoes in flight. J. Insect Behav. **4**, 391–396 (1991)
20. Moore, A.: Development of a data acquisition system for long-term outdoor recording of insect flight activity using a photosensor. In: Conference on Aerobiology and Biometeorology. American Meteorological Society (1998)
21. Moore, A., Miller, J.R., Tabashnik, B.E., Gage, S.H.: Automated identification of flying insects by analysis of wingbeat frequencies. J. Econ. Entomol. **79**(6), 1703–1706 (1986)
22. Noll, A.M.: Cepstrum pitch determination. J. Acoust. Soc. Am. **41**(2), 293–309 (1967)
23. Oppenheim, A., Schafer, R., Buck, J., et al.: Discrete-time signal processing, vol. 2. Prentice Hall, New Jersey (1989)
24. Paliwal, K., Kleijn, W.: Quantization of LPC parameters. In: Speech Coding and Synthesis, pp. 433–466 (1995)
25. Park, T.H.: Towards Automatic Musical Instrument Timbre Recognition. Ph.D. thesis, Princeton University (2004)
26. Pimentel, D.: Environmental and economic costs of the application of pesticides primarily in the United States. In: Integrated Pest Management: Innovation-Development Process, pp. 89–111 (2009)
27. Prati, R.C., Batista, G.E.A.P.A., Monard, M.C.: A survey on graphical methods for classification predictive performance evaluation. IEEE Trans. Knowl. Data Eng. **23**(1), 1601–1618 (2011)
28. Pollard, H., Jansson, E.: A tristimulus method for the specification of musical timbre. Acustica **51**(3), 162–171 (1982)
29. Rabiner, L., Schafer, R.: Digital Processing of Speech Signals, vol. 100. Prentice-Hall, Englewood Cliffs, NJ (1978)
30. Ratanamahatana, C.A., Keogh, E.J.: Making time-series classification more accurate using learned constraints. In: SIAM International Conference on Data Mining, pp. 11–22 (2004)
31. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics. Speech Sig. Process. **26**(1), 43–49 (1978)
32. Silva, D.F., Souza, V.M.A., Batista, G.E.A.P.A., Keogh, E.J., Ellis, D.P.W.: Applying machine learning and audio analysis techniques to insect recognition in intelligent traps. In: International Conference on Machine Learning and Applications, pp. 99–104 (2013)
33. Souza, V.M.A., Silva, D.F., Garcia, P.R., Batista, G.E.A.P.A.: Avaliação de classificadores para o reconhecimento automático de insetos (in English: Evaluation of classifiers for automatic insect recognition). In: Encontro Nacional de Inteligência Artificial e Computacional, pp. 1–12 (2013)
34. Stevens, S.S., Volkmann, J., Newman, E.B.: A scale for the measurement of the psychological magnitude pitch. J. Acoust. Soc. Am. **8**(3), 185–190 (1937)
35. Taylor, B., Jones, M.D.R.: The circadian rhythm of flight activity in the mosquito Aedes aegypti (L.): the phase-setting effects of light-on and light-off. J. Exp. Biol. **51**(1), 59–70 (1969)
36. Terasawa, H., Slaney, M., Berger, J.: The 13th Colors of Timbre, pp. 323–326 (2005)
37. Unwin, D.M., Ellington, C.P.: An optical tachometer for measurement of the wing-beat frequency of free-flying insects. J. Exp. Biol. **82**(1), 377–378 (1979)
38. Walker, K.: A Review of Control Methods for African Malaria Vector. Tech. Rep. 108, Bureau for Global Health (2002)
39. Wang, W., Yu, X., Wang, Y.H., Swaminathan, R.: Audio fingerprint based on spectral flux for audio retrieval. In: International Conference on Audio, Language and Image Processing. IEEE, pp. 1104–1107 (2012)
40. Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., Keogh, E.J.: Experimental comparison of representation methods and distance measures for time series data. Data Min. Knowl. Disc. **26**(2), 275–309 (2013)
41. W.H.O.: Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control. Tech. rep., World Health Organization (2009)
42. W.H.O.: The World Malaria report. Tech. rep., World Health Organization (2012)
43. Zhen, B., Wu, X., Liu, Z., Chi, H.: On the importance of components of the MFCC in speech and speaker recognition. In: Annual Conference of the International Speech Communication Association, pp. 487–490 (2000)