Detection of moving vehicles via dimensionality reduction

Alon Schclar Neta Rabin Amir Averbuch Valery Zheludev

School of Computer Science
Tel Aviv University, Tel Aviv 69978, Israel

Abstract

Automatic vehicle detection is a common task in security and surveillance systems. A microphone is placed in a designated area and a hardware/software system processes the sounds that are intercepted by the microphone to detect vehicles as they pass by in the designated area. We propose a novel algorithm for real-time automatic detection of vehicles. The proposed scheme uses dimensionality reduction and utilizes ideas from compressed sensing.

1 Introduction

Moving vehicles produce typical sounds that are mainly influenced by the engine vibrations and the friction between the tires and the road. Airplanes, helicopter, wind and speech create sounds that have different acoustic features when compared to vehicles. Similar types of vehicles produce similar sounds, however, it is not a trivial task to identify similar vehicles that travel in diverse speeds, in various distances from the recording device and on different types of roads (land, asphalt, etc). Our goal is to separate between vehicles and non-vehicles by analyzing their acoustic dynamic sounds.

Every sound emitting device can be characterized according to the acoustic features of the sounds it produces. We refer to these characteristic features as *acoustic signatures* and we use them in this paper to differentiate between vehicles and non-vehicles. Looking for these signatures is traditionally done by signal processing. Wavelet tools are well known for that. In our proposed scheme we use ideas that come from the field of compressed sensing [6, 5, 4]

to uncover dominating features of an unknown acoustic signal. The short-term dynamics of the acoustic signal is treated as a signal $x \in \mathbb{R}^m$. It is correlated with approximately $logM$ random vectors in $\mathbb{R}^m$. The outcome of this process is a set of features that is further processed to obtain the acoustic signature.

The algorithm is composed of two phases: an offline training phase and an online detection phase. In the training phase the data, which consists of vehicle and non-vehicle recordings, is analyzed and features that characterize it are extracted to produce acoustic signatures. These signatures are used during the online detection phase.

The proposed algorithm is generic and can be tailored to different tasks that have to separate between different clusters.

The rest of this paper is organized as follows: in section 2 we review previous works related to the problem at hand. The proposed algorithm is described in section 3. Experimental results are given in section 4. Future work and conclusions are given in section 5.

2 Related Work

Several papers have been published in the area of separation between vehicle and non-vehicle sounds. Most of them describe systems for a military context.

Choe *et al.* [9], extracted the acoustic features by using discrete wavelet transform. The feature vectors were compared with the reference vectors in the database using statistical pattern matching to determine the type of vehicle from where the signal originated. In [7], the discrete cosine transform was applied on the signals and used a time-varying autoregressive modeling approach for their analysis. Averbuch *et al.* [2], designed a system that is based on wavelet packets coefficients in order to discriminate between different types of vehicles. Classification and Regression Trees (CARTs) were used for classification of new unknown signals. In a later paper [1], Averbuch *et al.* used similar methods with multiscale local cosine transform applied on the frequency domain of the acoustic signal. The classifier that was used was based on "Parallel Coordinates" methodology. Wu *et al.* [8] used the "eigenfaces method" [13], which was originally used for human face recognition, to distinguish between different vehicle sound signatures. The data was sliced into frames - short series of time slices. Each frame was then transformed into the frequency domain. Classification was done by projecting new frames on the principal components that were calculated for a known training set. Munich [12] compared between several speech recognition techniques for classification of vehicle types. These methods were applied on short time Fourier transform of the vehicles' acoustic signatures.

3 The algorithm

The proposed acoustic classification algorithm consists of two phases:

1. A learning phase in which acoustic signatures (features) are extracted from sample recordings whose classification classes are known.

2. A classification phase that processes an acoustic signal in order to determine whether it is a vehicle or not according to the its acoustic signatures.

The classification phase does not process the entire signal in a batch manner but rather a short segment at a time. This fact along with the high efficiency of the algorithm renders the classification suitable for real-time applications.

3.1 Background

In the proposed algorithm, we make use of two dimensionality reduction schemes: principal component analysis (PCA) [14] and random projections [6, 5, 4]. In the following, we give a brief description of these methods. The input to both methods is a set of column vectors $\Gamma = \{x_i\}_{i=1}^n$ where $x_i \in \mathbb{R}^m$. Both methods embed the vectors into a lower dimensional space $\mathbb{R}^q$ where $q \ll m$ and their output is a set of column vectors in the lower dimensional space $\widetilde{\Gamma} = \{\widetilde{x}_i\}_{i=1}^n$, $\widetilde{x}_i \in \mathbb{R}^q$. We refer to the vectors in the set $\widetilde{\Gamma}$ as the *embedding* vectors.

3.1.1 Principal component analysis (PCA)

In PCA the vectors are centered around the origin in Γ. Let $\overline{\Gamma} = \{\overline{x}_i\}_{i=1}^n$ be the set Γ centered at the origin where $\overline{x}_i \triangleq x_i - \frac{1}{n}\sum_{j=1}^n x_j$. We define a $m \times n$ matrix X whose columns are comprised of the vectors $\{\overline{x}_i\}$

$$X = (\overline{x}_1 | \overline{x}_2 | \ldots | \overline{x}_n) .$$

The covariance matrix $C_{\overline{\Gamma}}$ of the set $\overline{\Gamma}$ in computed by :

$$C_{\overline{\Gamma}} = \frac{1}{n-1} X X^T .$$

$C_{\overline{\Gamma}}$ is a symmetric $m \times m$ matrix that captures the correlations between all possible pairs of measurements. We perform an eigen-decomposition of $C_{\overline{\Gamma}}$ to produce a set of eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and their corresponding eigenvectors $\varphi_1, \ldots, \varphi_n$. The eigenvector φ_1 gives the direction where the variance of the data is maximal. The magnitude of this variance is given by λ_1. In a similar way, the eigenvector φ_2 gives the direction of the second largest variance λ_2, and so forth.

The low dimensional vector $\widetilde{x}_i$, which is the embedding of x_i, is given by the projection of $\overline{x}_i$ onto the first q eigenvectors $\varphi_1, \ldots, \varphi_q$:

$$\widetilde{x}_i \triangleq \left(\left\langle \overline{x}_i, \varphi_1 \right\rangle, \left\langle \overline{x}_i, \varphi_2 \right\rangle, \ldots, \left\langle \overline{x}_i, \varphi_q \right\rangle\right)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product operator.

3.1.2 Random projections

The feasibility of dimensionality reduction using random projections was first theoretically proved by Johnson and Lindenstrauss [10]. Specifically, they showed that N points in N dimensional space can almost always be projected to a space of dimension $ClogN$ with control on the ratio of distances and the error (distortion). Bourgain [3] showed later that any metric space with N points can be embedded by a bi-Lipschitz map into an Euclidean space of $logN$ dimension with a bi-Lipschitz constant of $logN$. Various randomized versions of this theorem are used for protein mapping [11] and for the reconstruction of frequency-sparse signals [6, 5, 4].

In order to reduce the dimensionality of Γ using random projections, we first generate a random basis $\Upsilon = \{\rho_i\}_{i=1}^{q}$ where $\rho_i \in \mathbb{R}^m$ and $q \ll m$ is the dimension of the target reduced space. Two common choices for generating a random basis are:

1. The vectors $\{\rho_i\}_{i=1}^{q}$ are uniformly distributed on the q dimensional unit sphere.
2. The elements of the vectors $\{\rho_i\}_{i=1}^{q}$ are chosen from a Bernoulli $+1/-1$ distribution and the vectors are normalized so that $\|\rho_i\|_{l_2} = 1$ for $i = 1, \ldots, q$.

The embedding $\widetilde{x}_i$ of x_i into a low dimensional space is obtained by its projection onto the random vectors $\rho_1, \ldots, \rho_q$:

$$\widetilde{x}_i \triangleq \left(\left\langle x_i, \rho_1 \right\rangle, \left\langle x_i, \rho_2 \right\rangle, \ldots, \left\langle x_i, \rho_q \right\rangle\right).$$

3.2 The learning phase

The learning phase analyzes a test sample set of recordings $TSS = \{s_i\}_{i=1}^{\tau}$ whose classifications are known a-priori where τ is the number of samples in the training set. The signals do not necessarily have the same size. We regard a signal as a sequence of numbers composed of $|s_i|$ elements.

The learning phase consists of the following steps:

1. Decomposition of every signal s_i into overlapping segments $W_i = \left\{w_i^j\right\}$ that we refer to as *windows*.

2. Transformation of the windows to the frequency domain via the Fourier transform.

3. Dimensionality reduction via random projections: each window is projected onto a given number n_B of random generated bases [6, 5, 4]. These results are referred to as *random projections.*

4. Construction of *paths* $\{\pi_m\}$ from the random projections - a path contains the random projections of μ *consecutive* windows where μ is a given parameter.

5. Dimensionality reduction of the paths $\{\pi_m\}$ via principle component analysis (PCA).

The result of step 5 contains the acoustic signatures for the learned signals. In the following we describe in details each step.

Step *I*: Going to a higher dimension - decomposition into windows - Let s_i be a test sample signal. It is comprised of a sequence of numbers:

$$s_i = \left(s_i\left(0\right), \ldots, s_i\left(\left|s_i\right| - 1\right)\right)$$

where $s_i\left(t\right)$, $t = 0, \ldots, \left|s_i\right| - 1$, denotes the signal value at time t and $\left|s_i\right|$ is the length of the signal s_i.
Denote by l the size of the windows that the signal is to be decomposed into and by ξ the size of the overlapping (number of elements) between two consecutive windows. We define

$$d \triangleq l - \xi \tag{1}$$

to be the portion in each window that does not overlap with its consecutive window.
Each signal s_i is decomposed into overlapping windows $W_i = \left\{w_i^j\right\}$ where the j^{th} window is given by:

$$w_i^j \triangleq \left(s_i\left(\left(j-1\right)\cdot d\right), \ldots, s_i\left(\left(j-1\right)\cdot d + l - 1\right)\right)$$

and the number of windows is given by

$$\left|W_i\right| = \left\lfloor \frac{\left|s_i\right| - l}{d} \right\rfloor + 1 \tag{2}$$

where $\lfloor \cdot \rfloor$ denotes the floor operator.

Every signal $s_i, i = 1, \ldots, \tau$, is decomposed into windows and the results are concatenated to a single set:

$$\Omega = \bigcup_{i=1}^{\tau} W_i.$$

For notation convenience, we use a single index for w_i^j and the output of this step is denoted by:

$$\Omega = \{w_j\}_{j=1}^{n_w}$$

where the total number of windows resulting from the decomposition of all the signals is given by $n_w \triangleq |\Omega| = \sum_{i=1}^{\tau} |W_i| = \sum_{i=1}^{\tau} \left\lfloor \frac{|s_i|-l}{d} \right\rfloor + 1$.

Step *II*: Transformation into the frequency domain - We follow the assumption that the acoustic signature of a signal is found in its frequency domain. Therefore, we apply the Fast Fourier transform (FFT) to each window w_j.

Let $\hat{w}_j \triangleq (\hat{w}_j(0), \ldots, \hat{w}_j(l-1))$, $1 \leq j \leq n_w$, be the result from the application of the FFT to w_j i.e.

$$\hat{w}_j(n) = \sum_{k=0}^{l-1} w_j(k) e^{2\pi i k n/l}, \; n = 0, \ldots, l-1.$$

We are only interested in the magnitudes of the frequencies since they distinguish between two acoustic signatures. Hence, we calculate the frequency magnitudes (moduli)

$$|\hat{w}_j| \triangleq (|\hat{w}_j(0)|, \ldots, |\hat{w}_j(l-1)|)$$

where

$$|\hat{w}_j(n)| = \sqrt{\Re^2(\hat{w}_j(n)) + \Im^2(\hat{w}_j(n))}, \; n = 0, \ldots, l-1,$$

and $\Re(z)$, $\Im(z)$ denote the real and imaginary parts of a complex number z, respectively.

Since s consists of real numbers, we have $\hat{w}_j(n) = \overline{\hat{w}_j(l-n)}$ and consequently $|\hat{w}_j(n)| = |\hat{w}_j(l-n)|$ for $n = 0, \ldots, l-1$. Thus, we only need to retain the first

$$h \triangleq \left\lceil \frac{l}{2} \right\rceil \tag{3}$$

magnitudes from each window where $\lceil \cdot \rceil$ denotes the ceiling operator. Furthermore,

we reduce the dynamic range of $|\hat{w}_j|$, by taking the logarithm of the magnitudes

$$u_j \triangleq \left(\log\left(\left|\hat{w}_j\left(0\right)\right|+\epsilon\right),\ldots,\log\left(\left|\hat{w}_j\left(h\right)\right|+\epsilon\right)\right)$$

where ϵ is a small constant that is added to avoid taking the logarithm of zero-magnitude frequencies. We denote by $U \triangleq \{u_j\}_{i=1}^{n_w}$, $u_j \in \mathbb{R}^h$, the result of this step.

Step *III*: Dimensionality reduction via random projection - We generate n_B random bases

$$RB = \left\{B^i\right\}_{i=1}^{n_B} \tag{4}$$

where $B^i = \{\beta_k^i\}_{k=1}^r$, $\beta_k^i \in \mathbb{R}^h$, is the i^{th} basis and $r \ll h$ is the dimension of the reduced space.

The dimension of the set U is reduced by projecting it onto *every* basis in RB. Each projection onto a basis B^i produces a *single* embedding into a dimension-reduced space. We refer to the projection of U onto a random basis B^i as a *random projection* and we denote it by

$$\widetilde{U}^i \triangleq \left\{\widetilde{u}_j^i\right\}_{j=1}^{n_w}$$

where $\widetilde{u}_j^i \in \mathbb{R}^r$ and $\widetilde{u}_j^i\left(k\right)$ - the k^{th}element of the vector $\widetilde{u}_j^i$ - is obtained by

$$\widetilde{u}_j^i\left(k\right) \triangleq \left\langle \beta_k^i, u_j \right\rangle,\ k=1,\ldots,r.$$

We refer to a single projection $\widetilde{u}_j^i$ as a dimension-reduced-window (DRW) and denote the set of all random projections on RB by

$$\widetilde{U} = \left\{\widetilde{U}^i\right\}_{i=1}^{n_B}.$$

Step *IV*: Path construction - Given a random projection $\widetilde{U}^i \triangleq \left\{\widetilde{u}_j^i\right\}_{j=1}^{n_w}$ and a given length μ, we construct vectors in $\mathbb{R}^{rp}$ by concatenating sequences of μ *temporally consecutive* DRWs. We refer to these sequences as *paths*. This construction is done separately for each subset of DRWs according to the signal s they originated from. In the following, we formally describe this process.

For notation convenience, we temporarily remove the random basis index i from $\widetilde{U}^i$, $\widetilde{u}_j^i$ and describe the path construction for a single random projection $\widetilde{U} = \{\widetilde{u}_j\}_{j=1}^{n_w}$. First, $\widetilde{U}$ is separated into sets according to the signals the DRWs were derived from. Specifically, every vector $\widetilde{u}_j$ is labeled according to the class of its corresponding signal s_k. As mentioned above, the classifications of the signals that are analyzed during the

learning phase are known a-priory. We refer to this classification as the *label* of the signal and for a signal s_k we denote it by $L(s_k) \triangleq k$. We denote by $L(\widetilde{u}_j)$ the label of the DRW $\widetilde{u}_j$ and set its value to the label of the signal it originated from.
We separate $\widetilde{U}$ according to the labels of the DRWs

$$\widetilde{U} = \bigcup_{k=1}^{\tau} A^k$$

where $A^k \triangleq \left\{\alpha_n^k\right\}_{n=1}^{|W_k|} = \left\{\widetilde{u}_j | L(\widetilde{u}_j) = k\right\}$ is the set of DRWs that was derived from the signal s_k and the size of A^k is given by $\left|A^k\right| = |W_k| = \left\lfloor \frac{|s_k|-l}{d} \right\rfloor + 1$, where $|W_k|$ was defined in Eq. 2.
Next, we construct paths from every set A^k. We denote these paths by

$$p^k \triangleq \left\{\pi_m^k\right\}_{m=1}^{\left|A^k\right|-\mu+1}$$

where each path is given by

$$\pi_m^k \triangleq \left(\alpha_m^k, \ldots, \alpha_{m+\mu-1}^k\right),\ \pi_m^k \in \mathbb{R}^{rp},\ m = 1, \ldots, \left|A^k\right| - \mu + 1.$$

For convenience, we denote $\left|A^k\right| - \mu + 1$ by μ_k. This path construction is performed for every set A^k and for every random projection $\widetilde{U}^i$. We denote by P_i the paths constructed from the DRW of all the signals that were obtained by the random basis B^i:

$$P_i = \bigcup_{k=1}^{\tau} p^k = \left\{\pi_j^k\right\}_{j=1}^{|P_i|} \tag{5}$$

where $|P_i| = \sum_{k=1}^{\tau} \mu_k = \sum_{k=1}^{\tau} \left(\left|A^k\right| - \mu + 1\right)$.

Step *V*: Dimensionality reduction via PCA - Let P_i be the paths constructed from the DRW for all the signals by a random basis B^i as described in Eq. 5. First, the paths in P_i are shifted to be centered around the origin. We denote the center of gravity of P_i by

$$c_i = \frac{1}{|P_i|} \sum_{j=1}^{|P_i|} \pi_j. \tag{6}$$

We construct a new set of vectors $\left\{\overline{\pi}_m\right\}_{m=1}^{|P_i|}$ such that $\overline{\pi}_m \triangleq \pi_m - c_i$ for $m = 1, \ldots, |P_i|$. We apply principle component analysis (PCA) to $\left\{\overline{\pi}_m\right\}_{m=1}^{|P_i|}$ in order to reduce its dimensionality. Let

$$\left\{\varphi_i\right\}_{i=1}^{rp}, \varphi_i \in \mathbb{R}^{rp} \tag{7}$$

be the principle axes found by the PCA and their corresponding eigenvalues $\{\lambda_i\}_{i=1}^{rp}$ such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{rp}$. We embed the set $\{\overline{\pi}_m\}_{m=1}^{|P_i|}$ into a space of size q and denote the dimension-reduced paths (DRP) by

$$\{\widetilde{\pi}_m\}_{m=1}^{|P_i|}, \widetilde{\pi}_m \in \mathbb{R}^q \tag{8}$$

where the j^{th} coordinate of $\widetilde{\pi}_m$ is given by $\widetilde{\pi}_m(j) = \langle \widetilde{\pi}_m, \varphi_j \rangle$. We set the label of each DRP according to the label of its corresponding signal and denote it by $L(\widetilde{\pi}_m)$.
This step is performed for every set of paths P_i, $i = 1, \ldots, n_B$, that was produced in step *IV*.

Storing the relevant data for the classification phase - The classification phase needs the following data items from the learning phase and therefore they are stored offline at the end of the learning phase:

l	- size of the windows the signal was decomposed into (step *I*).
ξ	- size of the overlapping between windows (step *I*).
RB	- the n_B random bases that were generated in Step *III*.
r	- dimension of the random projected reduced space (step *III*).
μ	- the path length from step *IV*.
$\{c_i\}_{i=1}^{n_B}$	- the centers that were calculated in Eq. 6.
q	- the dimension of the PCA reduced space (step *V*).
$\{\varphi_i\}_{i=1}^{q}$	- the first q principal components obtained by the PCA in step *V*.
$\left\{\widetilde{P}_i\right\}_{i=1}^{n_B}$	- the dimension reduced vectors obtained by the application of PCA in Section 3.2 and their corresponding labels $L(\widetilde{\pi}_m)$.

3.3 The classification phase

The classification phase is done online. There is no need to wait for the entire signal to be received. In order to classify the signal *at time* t, the algorithm only needs the path that ends at time t, i.e., the μ consecutive overlapping windows of size l with overlapping ξ that immediately preceded time t. The values of μ and l are the same as the ones used in the learning phase. Hence, only the $\nu \triangleq l + d \cdot (\mu - 1)$ signal values that were received up to time t are required where d was defined in Eq. 1. Essentially, these samples constitute a path of μ overlapping windows of size l. The classification is performed only for values of t that obey $t = l + d \cdot i$ where $\mu - 1 \leq i \in \mathbb{N}$.

We denote by $\sigma_t = (\sigma(t - \nu + 1), \sigma(t - \nu + 2), \ldots, \sigma(t))$ the sequence of ν signal values that were received up to time t where $\sigma(x)$ is the signal value that was received at time x.

In order to classify σ_t, we employ steps that are similar to steps *I*-*V* (section 3.2) in the leaning process. Specifically, the classification of σ_t consists of the following steps:

1. Decomposition of σ_t into overlapping windows.
2. Application of the FFT.
3. Dimensionality reduction using random projections onto the bases that were generated in Step *III* (section 3.2).
4. Construction of a path from the output of the previous step.
5. Dimensionality reduction using the principal components that were calculated in step *V* (section 3.2).

Decomposition to windows - σ_t is decomposed to μ overlapping windows $\{\omega_i\}_{i=1}^{\mu}$ of size l with overlapping ξ - similar to step *I* as defined in section 3.2 - where

$$\omega_i = \left(\sigma\left(t-\nu+1+d\cdot(i-1)\right),\ldots,\sigma\left(t-\nu+d\cdot i\right)\right)\ ,\ \ i=1,\ldots,\mu.$$

FFT application - The FFT is applied to each window in $\{\omega_i\}_{i=1}^{\mu}$. We denote the result by $\{\widehat{\omega}_i\}_{i=1}^{\mu}$. In a similar manner to step *II* (section 3.2), we calculate

$$\upsilon_i \triangleq \left(\log\left(|\widehat{\omega}_i\left(0\right)|+\epsilon\right),\ldots,\log\left(|\widehat{\omega}_i\left(h\right)|+\epsilon\right)\right)\ ,\ \ i=1,\ldots,\mu$$

where h was defined in Eq. 3 and $\widehat{\omega}_i\left(j\right)$ denotes the j^{th} coordinate of the result of the FFT application to window ω_i. We denote the result of this step by $\{\upsilon_i\}_{i=1}^{\mu}$.

Dimensionality reduction using random projections - The dimensionality of $\{\upsilon_i\}_{i=1}^{\mu}$ is reduced by projecting it onto the random bases that were generated in step *III* (section 3.2). Specifically, let $RB=\{B^j\}_{j=1}^{n_B}$ be the random bases that were generated in step *III*. The set $\{\upsilon_i\}_{i=1}^{\mu}$ is projected onto every basis $B^j=\left\{\beta_k^j\right\}_{k=1}^{r}$. The projection on a single basis B^j produces a set of dimension reduced vectors $\left\{\widetilde{\upsilon}_i^j\right\}_{i=1}^{\mu}$ where the k^{th} coordinate of $\widetilde{\upsilon}_i^j$ is given by:

$$\widetilde{\upsilon}_i^j\left(k\right) \triangleq \left\langle \beta_k^j, \upsilon_i \right\rangle,\ k=1,\ldots,r.$$

Path construction - For every single basis B^j, the vectors $\left\{\widetilde{\upsilon}_i^j\right\}_{i=1}^{\mu}$ are concatenated into a path ζ^j. Thus, the output of this step is a set of n_B paths $\{\zeta^j\}_{j=1}^{n_B}$.

Dimensionality reduction via PCA and classification - The set of paths $\{\zeta^j\}_{j=1}^{n_B}$ is projected on the first q principal components $\varphi_1, \varphi_2, \ldots, \varphi_q$ that were calculated in step V (section 3.2) and defined in Eq. 7. This embeds the paths $\{\zeta^j\}_{j=1}^{n_B}$ in the same space of the vectors in Eq. 8. We denote these embeddings by $\left\{\widetilde{\zeta}^j\right\}_{j=1}^{n_B}$ where the k^{th}coordinate of $\widetilde{\zeta}^j$ is given by $\widetilde{\zeta}^j(k) = \left\langle \widetilde{\zeta}^j, \varphi_k \right\rangle$, $k = 1, \ldots, q$.

Classification - Let $\widetilde{P}_j = \{\widetilde{\pi}_m\}_{m=1}^{|P_j|}$ be the set of DRPs obtained from the set of learned signals via the random basis B^j (Eq. 8). For each embedding $\widetilde{\zeta}^j$, its δ nearest neighbors from the set $\widetilde{P}_j$ are found where δ is given as a parameter. This set is denoted by

$$N_\delta\left(\widetilde{\zeta}^j\right) \triangleq \left\{ \widetilde{\pi}_k | \, \widetilde{\pi}_k \in \widetilde{P}_j \, and \, \widetilde{\pi}_k \, is \, one \, of \, the \, \delta \, nearest \, neighbors \, of \, \widetilde{\zeta}^j \right\}_{k=1}^{\delta}$$

and the set of its corresponding labels is denoted by

$$L\left(N_\delta\left(\widetilde{\zeta}^j\right)\right) \triangleq \left\{ L\left(\widetilde{\pi}_k\right) | \, \widetilde{\pi}_k \in N_\delta\left(\widetilde{\zeta}^j\right) \right\}_{k=1}^{\delta}.$$

The set $L\left(N_\delta\left(\widetilde{\zeta}^j\right)\right)$ is found for every random basis B^j and the set of all labels is given by

$$\Lambda = \bigcup_{j=1}^{n_B} L\left(N_\delta\left(\widetilde{\zeta}^j\right)\right).$$

σ_t is classified by the label with the highest number of occurrences in Λ.

4 Experimental Results

For the learning phase, we used 42 recordings: 21 of them were recordings of vehicles that include cars, trucks and vans and the remaining part of the test sample set consisted of non-vehicle recordings such as airplanes, helicopters, silence, speech, etc. The recordings were sampled at 1000Hz. The following parameters were used for the learning and classification phases: $l = 512$, $\xi = 50\%$, $RB = 4$, $r = 30$, $\mu = 5$ and $q = 5$ (see section 3.2). These parameters were determined empirically.

The classification phase was tested on recordings that were taken in various road and wind conditions. For every recording that was used in the classification phase we show: (a) the plot of the original signal; (b) a graph showing the vehicle classification probability that the online classification algorithm produced; (c) the spectogram of the signal. The spectogram is obtained by calculating the windowed discrete-time Fourier transform of the signal using a sliding window. The spectogram plots as an image the magnitude of this function. Each

column in the image contains the Fourier magnitudes of a single local window. We include the spectograms since they provide a good way to notice vehicles. Specifically, vehicles (or any other object that emits a frequency coherent sound for this matter) are manifested as lines in the spectogram. These lines have slopes due to the Doppler effect. The slopes show whether the distance from the object to the microphone is increasing (negative slope - the perceived frequency by the microphone is decreasing) or decreasing (positive slope - the perceived frequency by the microphone is increasing).

Figure 1 contains the results of a recording in which a car and a truck at times $t = 30\,sec$ and $t = 50\,sec$, respectively. The sections between seconds 37 and 43 and between seconds 50 and 60 were part of the TSS in the learning phase. It can be noticed that during seconds 53 to 57 the truck is not detected. This is due to the fact that when the sound of the vehicle is too loud, the microphone is saturated and the resulting acoustic signature is similar to the that of white noise. The same phenomenon occurs when wind is present in the recording, however as we shall see later, our algorithm successfully handles the latter case.

Figure 2 displays the results of a recording that contains sounds emitted by a car. During the first 6 seconds the engine of the car is running and the car is in a standing position. The car starts to move at $t = 7\,sec$ and it is heard until $t = 16\,sec$.

Figure 3 shows the results of a recording that contains sounds emitted by a car (at around $t = 11\,sec$) and by another car (at around $t = 27\,sec$). A helicopter is heard starting from $t = 28\,sec$ until the end of the recording and a third car appears at $t = 70\,sec$ while the helicopter is still heard. The car at $t = 70\,sec$ is detected in spite of the helicopter sound in the background.

Figure 4 contains the results of a recording that contains speech. The classifier did not classify this recording as a vehicle.

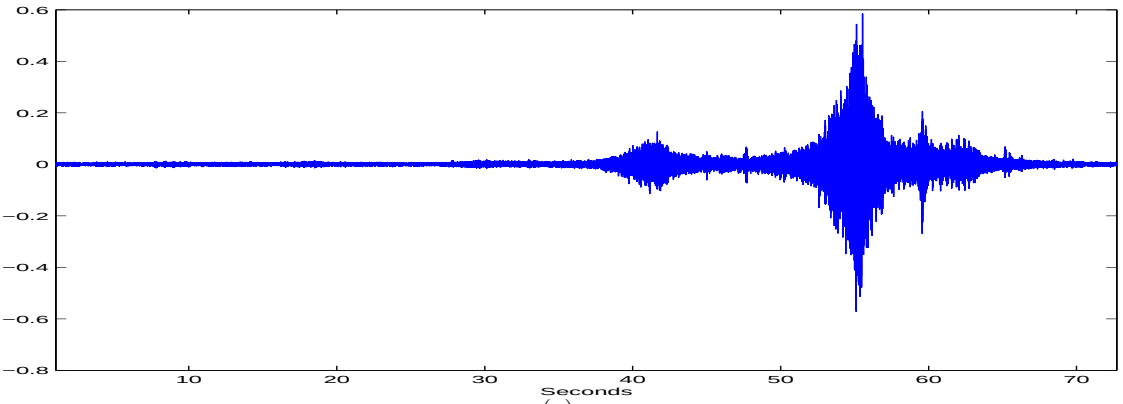
Figure 5 illustrates the results of a recording that contains sounds emitted by a van between $t = 0\,sec$ and $t = 13\,sec$.

Figure 6 shows results for a recording in which two helicopters are heard from $t = 15\,sec$ until $t = 100\,sec$ and two trucks are driven at high speed around $t = 110\,sec$ and around $t = 160\,sec$.

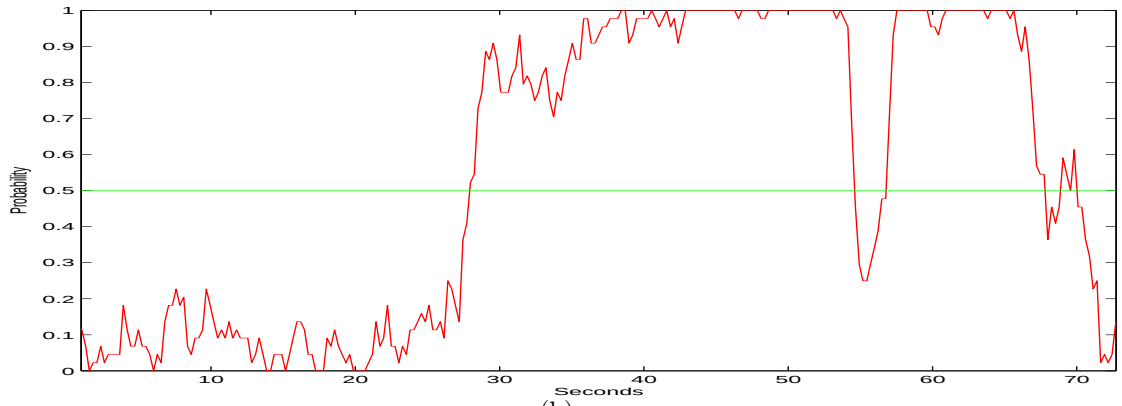
Figure 7 contains the results for a recording in which a heavy vehicle is heard twice: first, between $t = 20\,sec$ and $t = 50\,sec$ and second, between $t = 90\,sec$ and $t = 110\,sec$. During the last 30 seconds of the recording a plane is heard.

Figure 8 shows the results for a recording in which a truck passes near the recording device between $t = 80\,sec$ and $t = 120\,sec$. A plane is heard around $t = 180\,sec$. Strong wind is heard many times throughout the recording. The recordings in Figs. 2-8 were not part of the TSS that was used in the training phase.

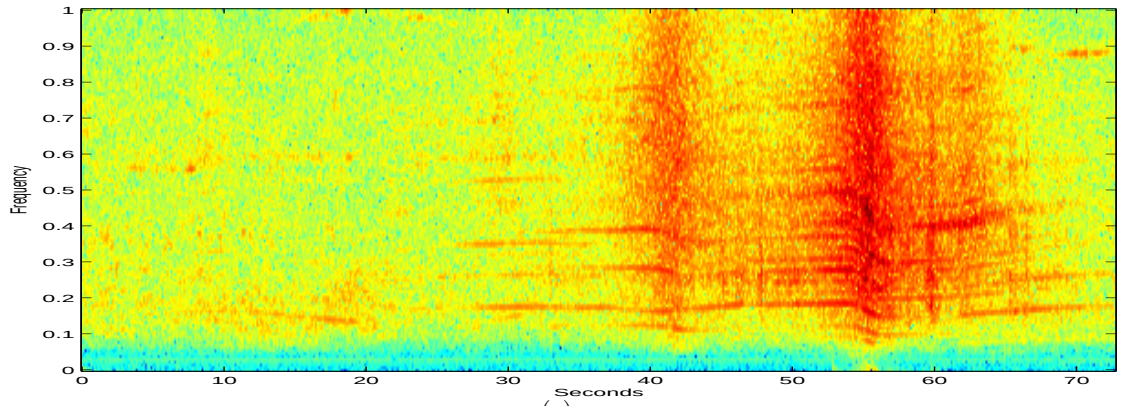
The algorithm exhibits very good results and is robust to background noise and wind. Furthermore, the algorithm successfully screens out background noises.



(a)

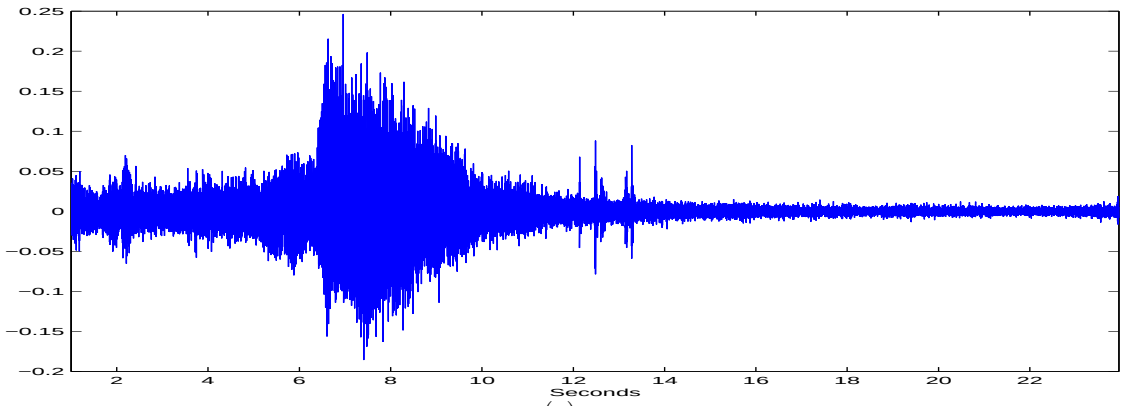


(b)

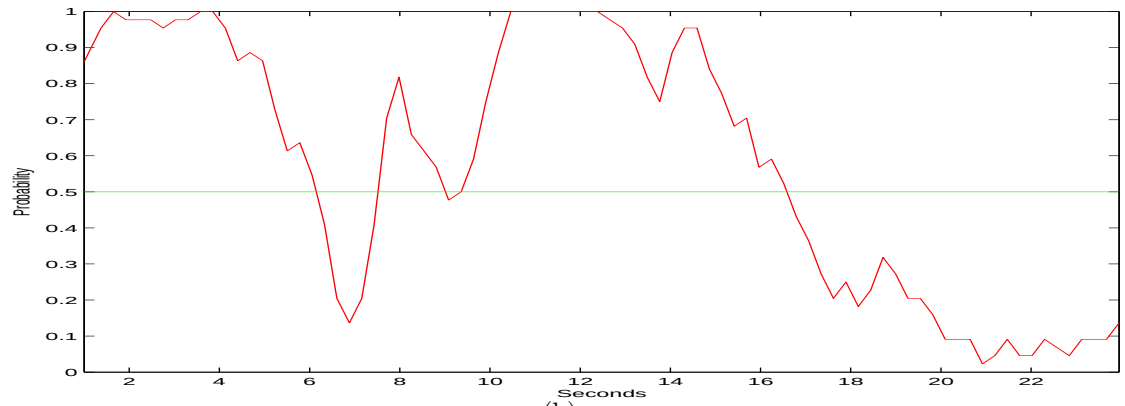


(c)

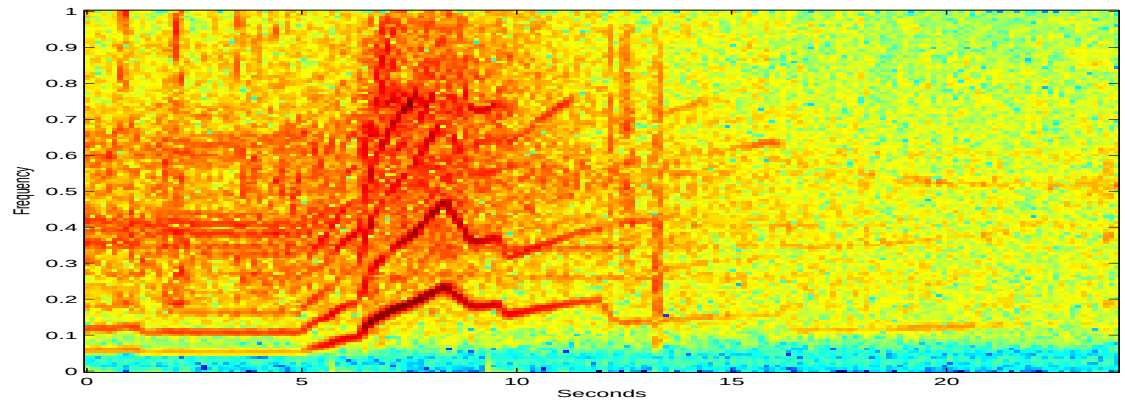
Figure 1: Results for a recording that contains sounds emitted by a car at $t = 30\,sec$ and by a truck $t = 50\,sec$. The car and truck sections of this recording were part of the TSS in the training phase. (a) The original recording; (b) The vehicle detection-probability; (c) The spectogram of the original signal.



(a)



(b)



(c)

Figure 2: Results for a recording that contains sounds emitted by a car. During the first 6 seconds the car is standing with its engine running. The car starts to move at $t = 7\,sec$ and it is heard until $t = 16\,sec$. (a) The original recording; (b) The vehicle detection-probability; (c) The spectogram of the original signal.

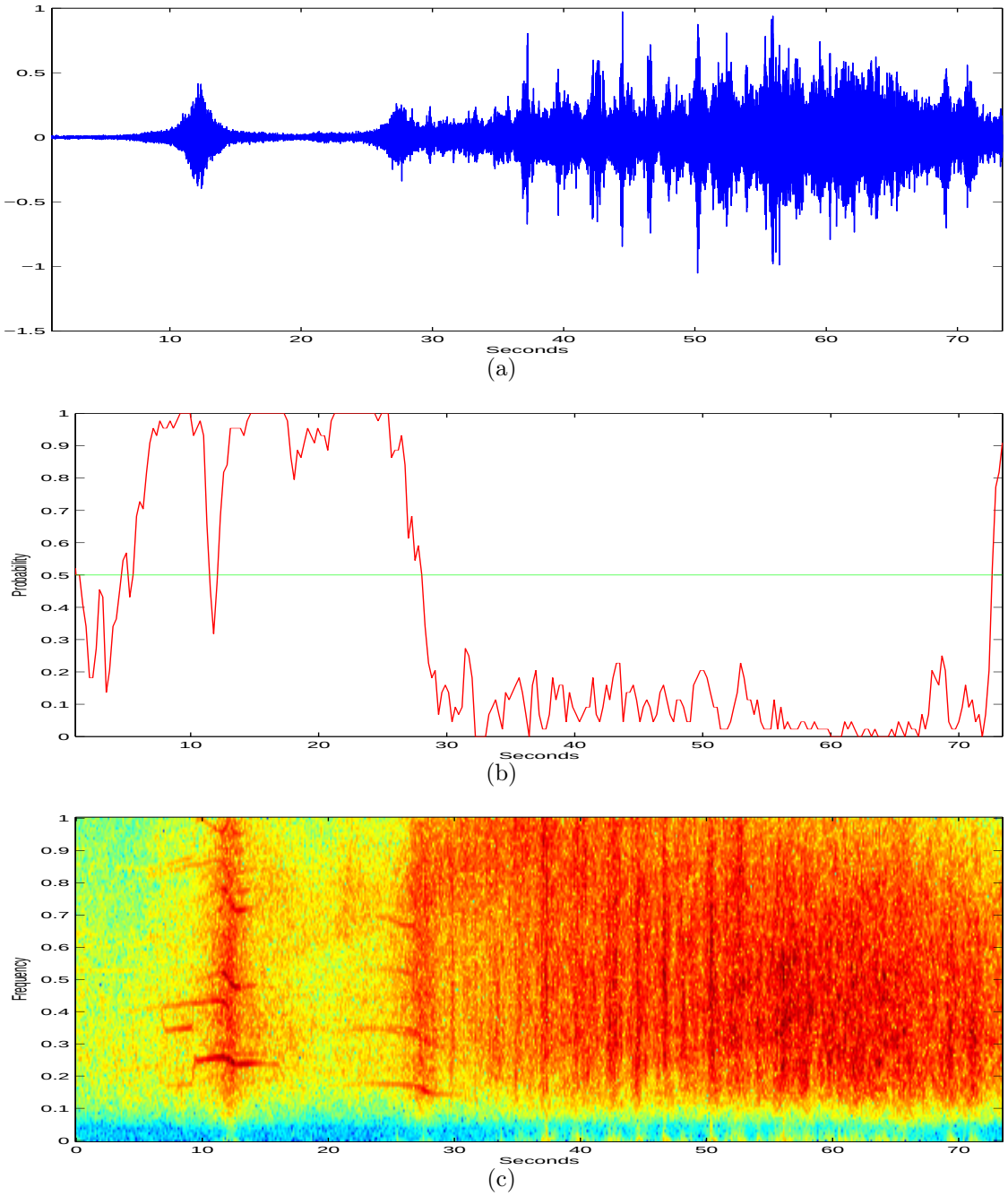


Figure 3: Results for a recording that contains sounds emitted by a car (at around $t = 11\,sec$) and by another car (at around $t = 27\,sec$). A helicopter is heard starting from $t = 28\,sec$ until the end of the recording and a third car appears at $t = 70\,sec$ while the helicopter is still heard. The car at $t = 70\,sec$ is detected in spite of the helicopter sound in the background. (a) The original recording; (b) The vehicle detection-probability; (c) The spectogram of the original signal.

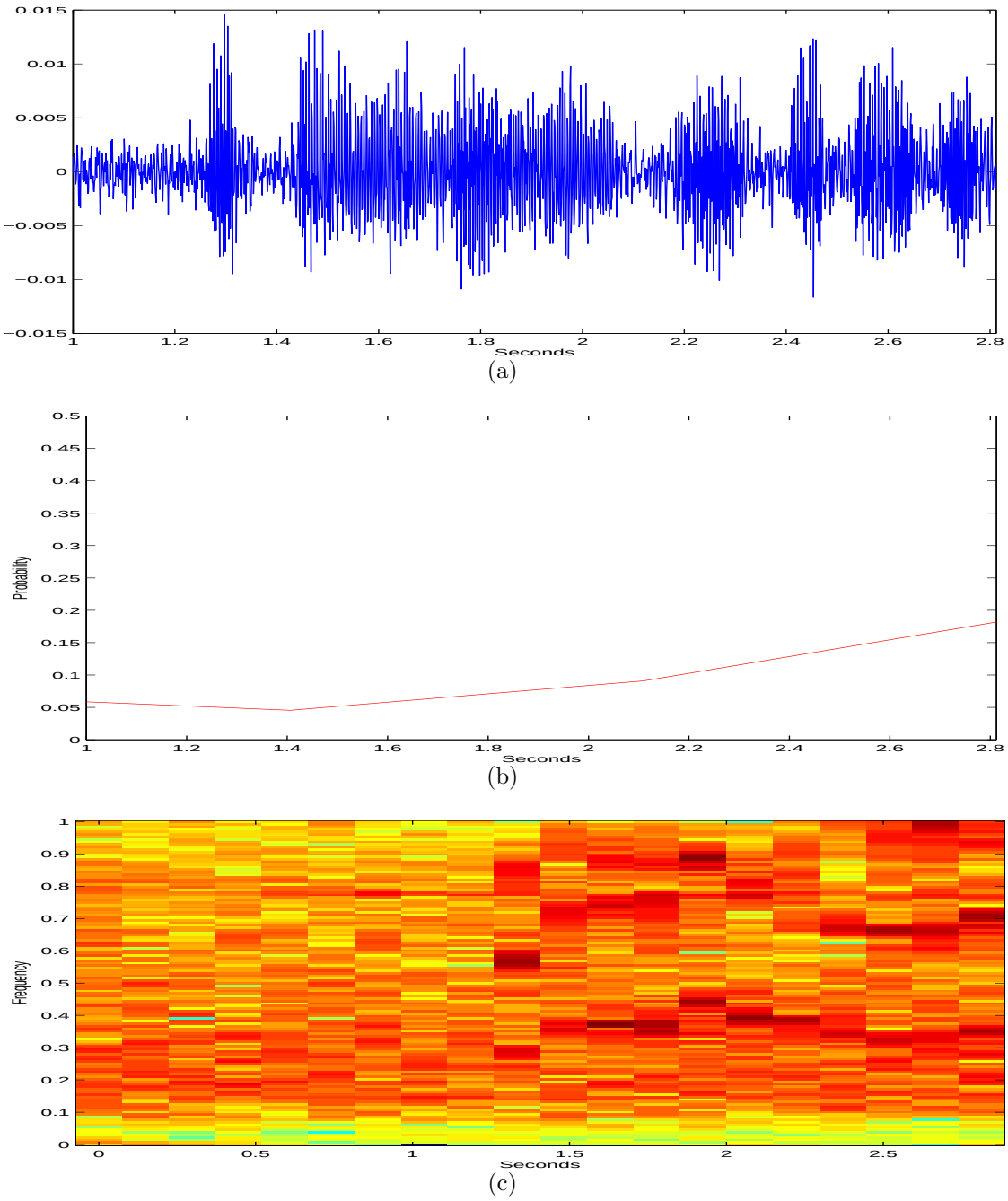


Figure 4: Results for a recording the contains speech sounds. The classifier did not classify this recording as a vehicle.(a) The original recording; (b) The vehicle detection-probability; (c) The spectogram of the original signal.

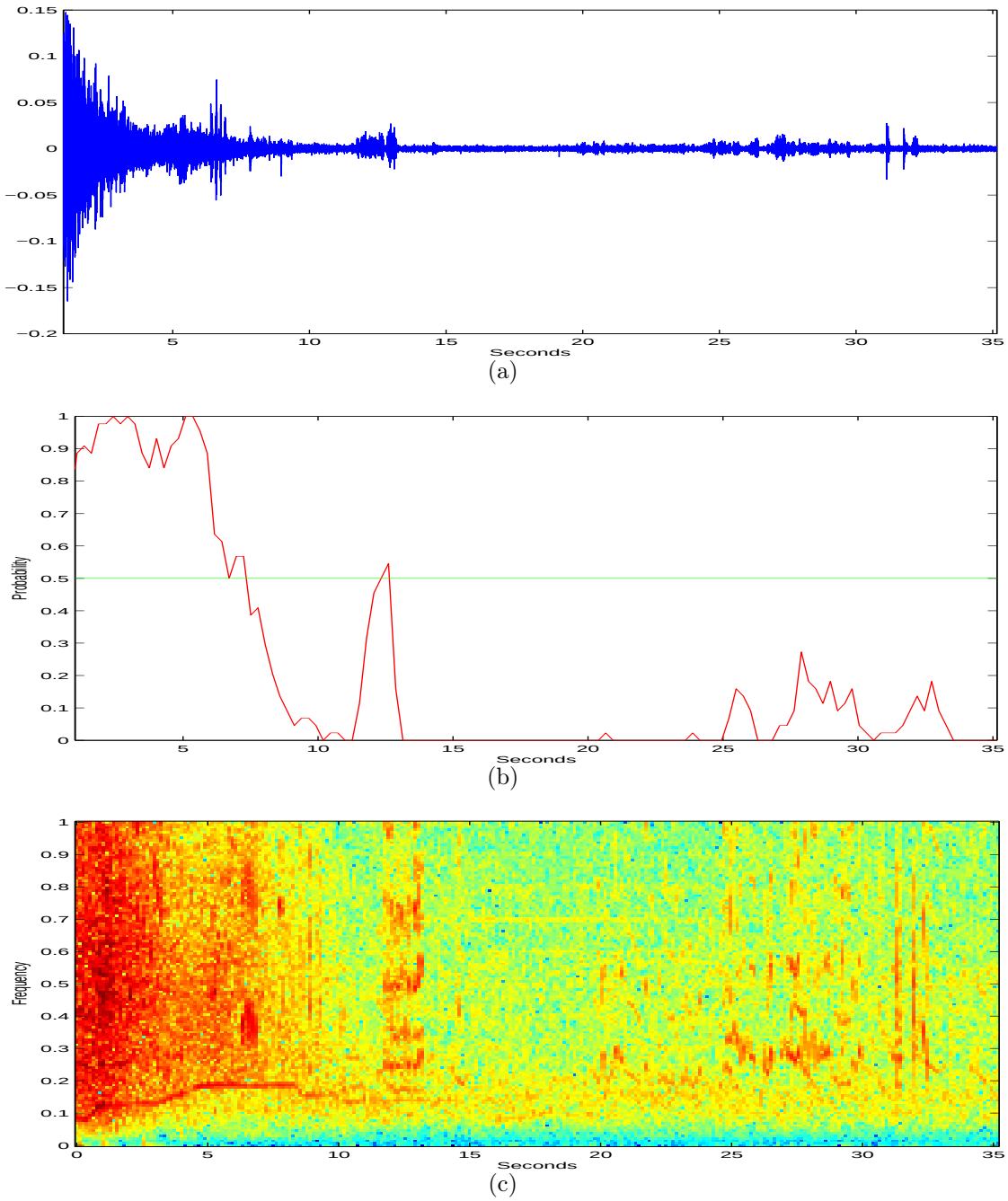


Figure 5: Results for a recording that contains sounds emitted by a van between $t = 0\,sec$ and $t = 13\,sec$. (a) The original recording; (b) The vehicle detection-probability; (c) The spectogram of the original signal.

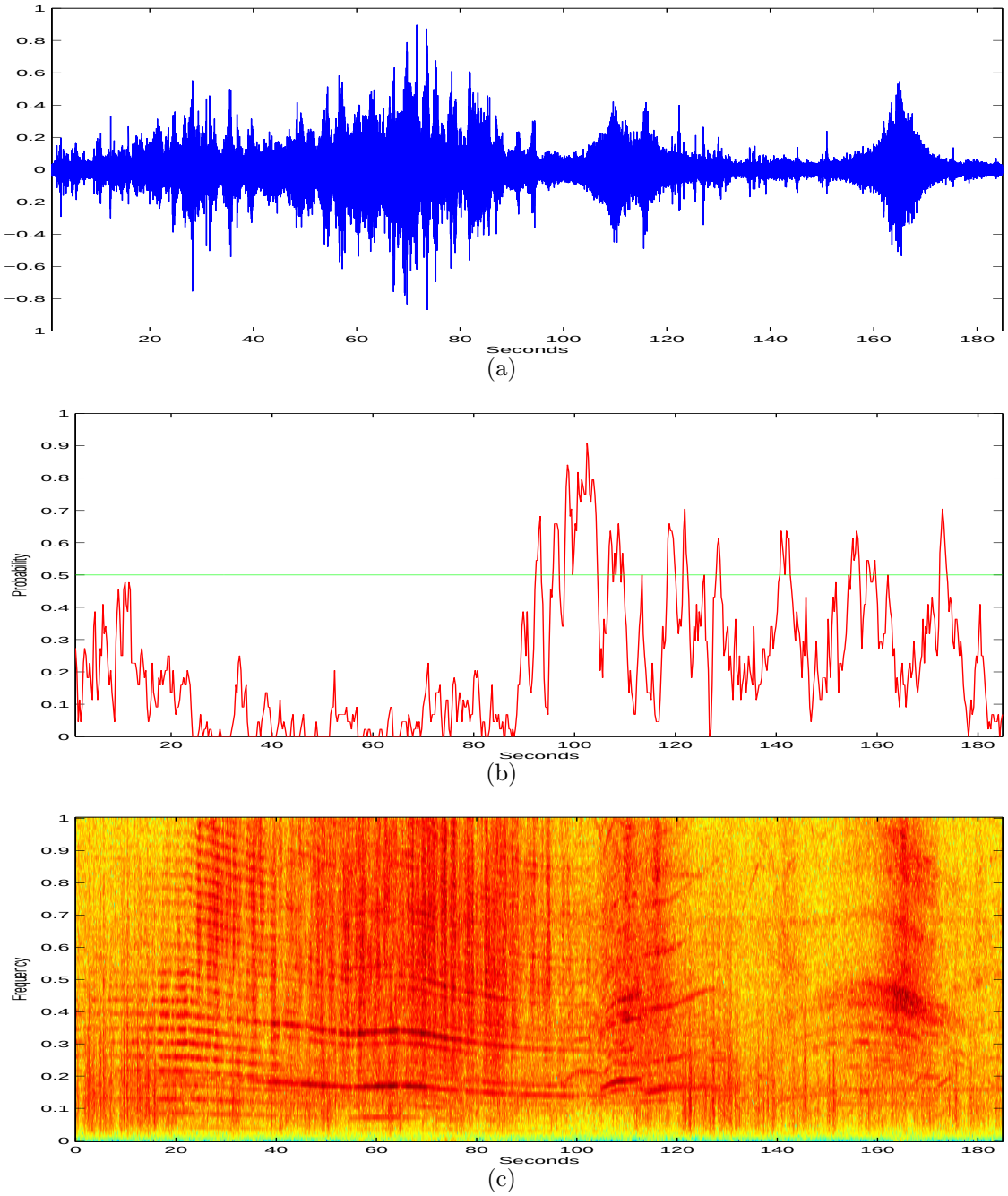
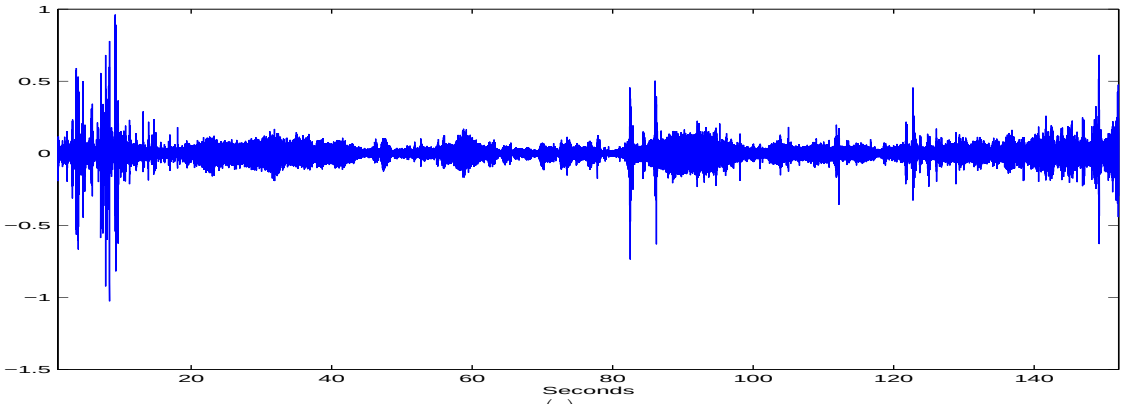
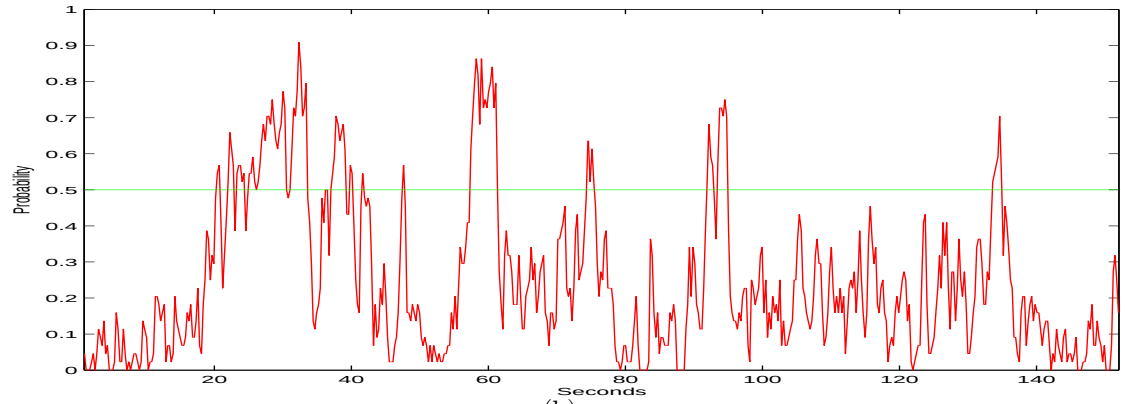


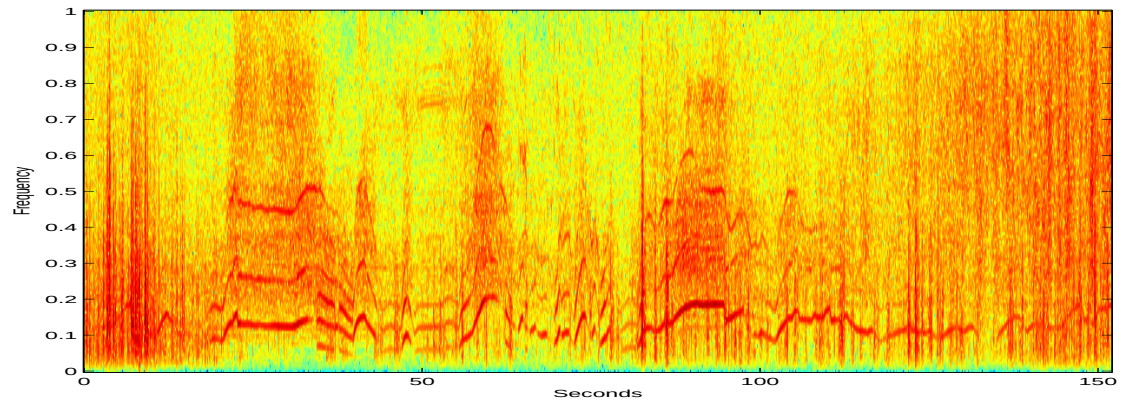
Figure 6: Results for a recording in which two helicopters are heard from $t = 15\,sec$ until $t = 100\,sec$ and two trucks are driven at high speed pass by around $t = 110\,sec$ and $t = 160\,sec$. (a) The original recording; (b) The vehicle detection-probability; (c) The spectogram of the original signal.



(a)

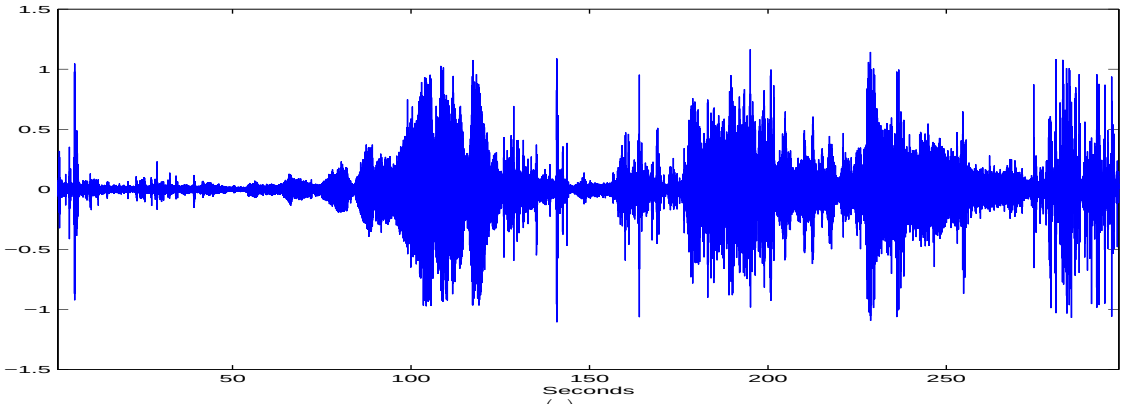


(b)

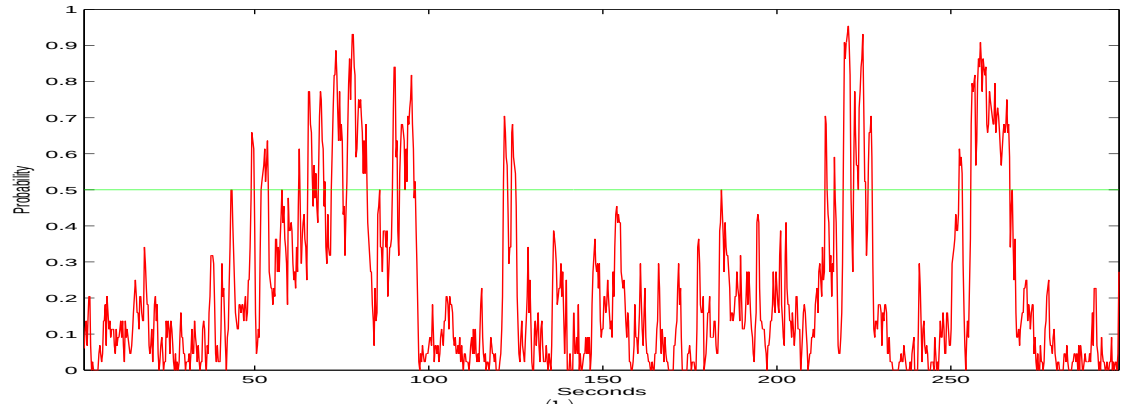


(c)

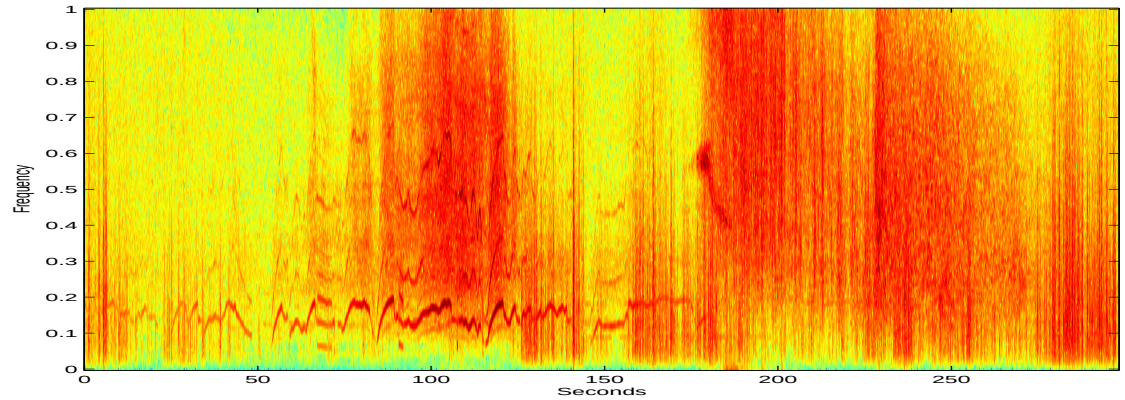
Figure 7: Results for a recording in which a heavy vehicle is heard twice: first between $t = 20\,sec$ and $t = 50\,sec$ and second between $t = 90\,sec$ and $t = 110\,sec$. During the last 30 seconds of the recording a plane is heard. (a) The original recording; (b) The vehicle detection-probability; (c) The spectogram of the original signal.



(a)



(b)



(c)

Figure 8: Results for a recording in which a truck passes near the recording device between $t = 80\,sec$ and $t = 120\,sec$. A plane is heard around $t = 180\,sec$. Strong wind is heard many times throughout the recording. (a) The original recording; (b) The vehicle detection-probability; (c) The spectogram of the original signal.

5 Conclusions and Future work

The promising results that are presented above motivate to explore both analytically and empirically the connection between the quality of the detection and the dimension of the reduced spaces. Naturally, a higher dimension will yield a better result. However, an analytic way to quantify this still remains an open problem. Moreover, one should take into consideration the quality of the recordings in the test sample set when exploring this open problem.

References

[1] V. Zheludev A. Averbuch, E. Hulata. Identification of acoustic signatures for vehicles via reduction of dimensionality. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(1):1–22, 2004.

[2] V. Zheludev A. Averbuch, E. Hulata and I. Kozlov. A wavelet packet algorithm for classification and detection of moving vehicles. *Multidimensional Systems and Signal Processing*, 12(1):9–31, 2001.

[3] J. Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Mathematics*, 52:46–52, 1985.

[4] Emmanuel Cand?s, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, February 2006.

[5] Y. Tsaig D. L. Donoho, I. Drori and J.L.Strack. Sparse solution of underdetermined linear equations by stagewise orthogonol machine pursuit. *Technical report No. 2006-2 Departement of Statistics, Stanford University*, April 2006.

[6] David Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[7] Kie B. Eom. Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models. *Multidimensional Systems and Signal Processing*, 10:357–378, 1999.

[8] M. Siegel H. Wu and P. Khosla. Vehicle sound signature recognition by frequency vector principal component analysis. *IEEE Transactions on Instrumentation and Measurement*, 48(5):1005–1008, October 1999.

[9] T. Meitzler G.R. Gerhart H.C. Choe, R.E Karlsen and D. Gorsich. Wavelet-based ground vehicle recognition using acoustic signals. *Proc of the SPIE, 2762:434-445*, 1996.

[10] W. B. Johnson and J. Lindenstrauss. Extensions of lipshitz mapping into hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[11] M. Linial, N. Linial, N. Tishby, and G. Yona. Global self-organization of all known protein sequences reveals inherent biological signatures. *Journal of Molecular Biology*, 268(2):539–556, May 1997.

[12] M.E. Munich. Bayesian subspace method for acoustic signature recognition of vehicles. *Proc. of the 12th European Signal Processing Conf. EUSIPCO*, 2004.

[13] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterzation of human faces. *J. Opt. Soc. Aner. Soc. A*, 4(1), March 1987.

[14] A. R. Webb. *Statistical Pattern Recognition, 2nd Edition*. John Wiley and Sons, October 2002.