PATTERN RECOGNITION OF MIXED SIGNALS USING GUIDED UNDER-DETERMINED SOURCE SIGNAL SEPARATION

A Thesis presented to the Faculty of the Graduate School
University of Missouri

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

---

by

LUIS ALBERTO RIVERA ESTRADA

Dr. Guilherme N. DeSouza, Thesis Advisor

DECEMBER 2011

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

PATTERN RECOGNITION OF MIXED SIGNALS USING GUIDED UNDER-DETERMINED SOURCE SIGNAL SEPARATION

Presented by Luis Alberto Rivera Estrada,
a candidate for the degree of Master of Science,
and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Guilherme DeSouza, Associate Professor, Dept. of Electrical and Computer Engineering

---

Dr. Satish Nair, Professor, Dept. of Electrical and Computer Engineering

---

Dr. Chi-Ren Shyu, Professor, Dept. of Computer Science

To my family, my friends, and my country, Guatemala.

ACKNOWLEDGEMENTS

I would like to thank:

God, for all the blessings I have received throughout my life.

My parents Julio and Rosalinda, for their love and unconditional support; for being such great role models; and for everything they have taught me.

My family, particularly my sisters, aunts, uncles, cousins, nieces and brother in law, for always being there for me and for their affection.

My advisor, Professor Guilherme DeSouza, for his guidance and for everything he has allowed me to learn, and for his encouragement and support for me to continue with my studies.

Darren Gabbert and Dinal Andreasen, for their insights throughout the development of this research. Also, for lending me the EMG switches that were used for the tests.

All the test subjects and everyone else who helped me develop this work.

Fulbright and LASPAU Academic and Professional Programs for the Americas, for giving me the opportunity to pursue my Master's degree here at Mizzou.

Universidad Del Valle de Guatemala, its authorities, and my former teachers and colleagues. In particular, María Eugenia de Nieves, Irene Aguilar, Enrique Mencos and María Luisa de Boehm, for giving me the opportunity to work at Del Valle and for helping me obtain my Fulbright-LASPAU scholarship.

And my great friends Alejandro González, Axel Fuentes, José Velásquez, Miguel Guzmán, José Luis Mendizábal and Luis Raúl Díaz.

Contents

List of Figures

List of Tables

ABSTRACT

In typical problems involving pattern recognition, the challenge lies in selecting a good set of features and in devising a reliable algorithm to recognize any particular vector of such features. In general, even when noise is present in the feature vectors, a single pattern naturally emerges from the domain, and the goal becomes to find the class of patterns that most resemble the observed vector. In many other cases, however, various patterns emerge at the same time, and the complexity of the problem increases since the patterns must now be separated before they can be classified. In this research, we propose a new technique, called Guided Under-determined Source Signal Separation (GUSSS), for recognizing patterns in mixtures of signals. Our proposed method is not only capable of dealing with mixtures of signals, but it also requires a single measurement to do so. Traditional techniques for signal separation, such as Independent Component Analysis (ICA), require the number of measurements to equal the number of sources. In our method, however, only one single measurement is required despite the existence of multiple patterns mixed in the measurement. We tested our method using two different frameworks: classification of chemical compounds using Terahertz signatures; and recognition of hand gestures using surface Electromyographic signals (sEMG) in Robotic Assistive Technology. Our experimental results demonstrate that the proposed technique achieves very good results using a small number of features.

Chapter 1

Introduction

1.1 Problem Statement

The ability to detect specific patterns in a signal is beneficial in many different areas such as material detection and assistive technology. In material detection, being able to recognize signatures indicating the presence of explosives or illegal substances is very important. Much effort has been made to detect improvised explosive devices (IEDs) in the fields, as well as drugs or hazardous substances in airports and other public places. Terahertz technology has attracted the attention of researchers for many years, and it has seen many advances over the past decades. Special interest for the use of this technology is actually directed to the detection of illegal drugs, explosives and other hazardous materials. These materials exhibit characteristic signatures at terahertz wavelengths which may be used to identify them.

The human body generates various types of signals. For example, electromyographic (EMG) signals are electrical signals generated when there is muscle activity. These signals find applications in many areas such as rehabilitation, prosthesis and human-machine interaction. Systems reliant on them require various forms of machine learning algorithms for recognition of specific components or patterns. Those systems vary in terms of the signal detection methods, the feature selection and the classification algorithm used. However, in all those cases, the use of multiple sensors and complex analysis and classification algorithms are constant requirements.

In this work we develop a new technique for pattern recognition and signal separation named Guided Under-determined Source Signal Separation (GUSSS). Our technique is based on Independent Component Analysis (ICA), a popular method for Blind Source Signal Separation. Traditionally, ICA is used to separate a number N of independent components that are linearly mixed in at least N different ways. A simple example is having N microphones capturing sounds in a room where N sources are emitting sounds simultaneously. The captured sound signals can be modeled as linear combinations (mixtures) of the source signals and ICA can be used to separate N source signals using N mixtures.

1.2 Approach and Contributions

Our approach employs ICA to detect whether a particular source signal is present given one single mixture of source signals, even when the mixture contains many other unknown signals. In other words, we deal with an under-determined case where N could be greater than one, but we only have one sensed signal.

We tested our method on two different areas. The first one is material detection using Terahertz signatures. The method was tested using a public THz database, achieving high true positive and true negative percentages. The second area is assistive technology using surface electromyographic (sEMG) signals. We tested the method in a power wheelchair control system that relies on a single sEMG sensor. Compared to other approaches in the literature, the proposed technique relies on a much simpler classifier and uses a very small number of features to achieve reasonable results.

To summarize, the contributions of this research are as follows:

- a new pattern recognition technique using signal separation that deals with an extreme case of under-determination.
- a framework for material detection using Terahertz signatures, which could lead to the development of systems capable of detecting hazardous materials, such as RDX, TNT, drugs, etc.

- a framework for classifying electrical signals generated by muscle activity from hand gestures, eyebrows, etc., which can be used to interface with computers and control devices such as power wheelchairs, etc.

Chapter 2

Background and Related Work

In this chapter we present an overview on the topics and techniques related to this thesis. We begin with a classical problem known as the Cocktail Party Problem, which illustrates some of the ideas that we further develop later on. We then discuss Blind Source Signal Separation and Independent Component Analysis, both of which are key method in our approach.

2.1 The Cocktail Party Problem

The cocktail party problem is the task of hearing a sound of interest in an environment where many sources emit sounds concurrently. The sounds are added together generating the mixed signal that enters the ear (microphone, sensor). At different locations, individual sounds would have different intensities [1, 2].

Consider 2 speakers and 2 microphones, as illustrated in Figure 2.1. Each sensor captures a different fraction of the signals emitted by the sources. So, the sensed mixed signals can be expressed as linear combinations of the sources:

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) \end{aligned} \tag{2.1}$$

where $s_i(t)$ denotes the signals emitted by the i^{th} source, $x_j(t)$ denotes the mixed signal captured by the j^{th} sensor, and the a_{ij} are the mixing coefficients of the lin-

ear combination model, which depend on the distances of the microphones from the speakers.

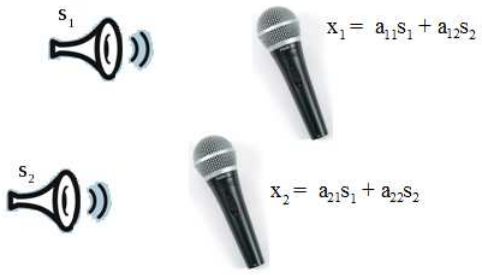


Figure 2.1: Two independent sources (speakers) emit sounds that mix in air and reach the microphones.

The important question becomes how to recover the original source signals $s_i(t)$ using the sensed signals $x_j(t)$. If there is a prior knowledge of the mixing coefficients, the linear system in eq. 2.1 can be solved by classical methods. However, it is rarely the case that information on the coefficients is available. The problem can then be addressed by estimating the coefficients a_{ij} using some information on the statistical properties of the signals $s_i(t)$. Specifically, if these source signals are statistically independent at each time instant, then it is possible to estimate the mixing coefficients [2, 3].

2.2 Blind Source Signal Separation using Independent Component Analysis (ICA)

Traditional Blind Source Signal Separation using ICA (BSSS-ICA) is a powerful technique for signal separation [3, 4]. It is assumed that sensed signals contain various statistically independent components. It is important to notice that each component actually originates from a different source. We present more details on statistical independence and how to measure its validity in section 2.3.

Mathematically, the goal of BSSS-ICA is to recover N source signals, $S = [s_1(t), \ldots, s_N(t)]^T$, which were linearly mixed, producing the observed signals $X = [x_1(t), \ldots, x_M(t)]^T$. A typical example would be N independent sounds emanating from different sources and being detected as mixed signals by M microphones spread over the space [5]. Figure 2.1 depicts this example for $N = 2$ sound sources $M = 2$ and microphones. Later, we explain how these same concepts can be applied to Terahertz signals and the detection of chemical compounds.

A very simple, artificial example is shown in Figure 2.2. On the left there are two mixed signals that could have been obtained from two sensors. On the right we see two separated, independent signals recovered using BSSS-ICA. It is not difficult to recognize the mixtures as having been produced by linear combinations of these two source signals.

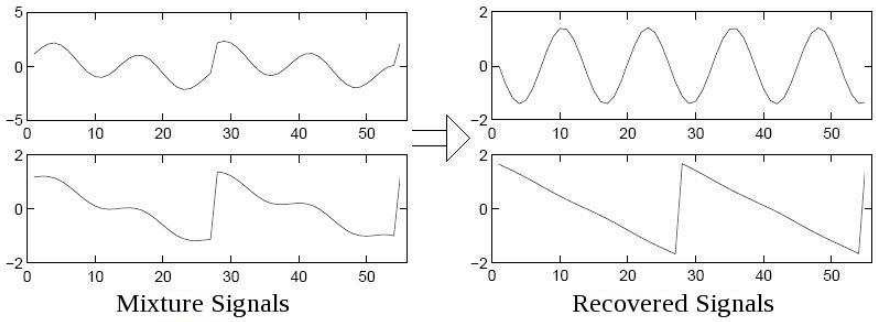


Figure 2.2: Out of two recorded mixture signals we obtain the original, independent signals which generated the mixtures in the first place.

2.2.1 Definition of ICA

A general definition of Independent Component Analysis (ICA) as stated in [3] is the following: ICA of the random vector X consists of estimating the generative model for the data: $X = AS$, where the components s_i in vector $S = [s_1, \ldots, s_N]^T$ are assumed independent and matrix A is a constant $M \times N$ matrix, usually called mixing matrix.

The model being identifiable can be assured if (1) all the independent components s_i, with the possible exception of one component, are non-Gaussian; (2) the num-

ber of observed linear mixtures M is at least as large as the number of independent components N, i.e $M \geq N$; and (3) the matrix A is full column rank [3].

Traditional ICA methods are able to separate the signals whenever $M \geq N$. (If $M > N$, the dimension of the observed vector X can always be reduced so that $M = N$). In those cases, the source signals and the observed signals can be related in a matrix form such as $X = AS$ – where A contains the coefficients of the linear combination of the sources. The methods can solve the (overdetermined) system of equations through the expression $S = A^{-1}X = WX$. The solution is found using a constrained optimization algorithm that maximizes the independence of the signals in S. We present some measures of independence and optimization algorithms in section 2.3.

2.2.2 Applications of ICA

As we have mentioned, blind source signal separation (BSSS) is a classical application of the ICA model. The typical example is the cocktail party problem described in section 2.1, but BSSS-ICA has also been used for separation of electroencephalographic (EEG) and magnetoencephalographic (MEG) data [2, 3, 5].

In financial data there are situations in which parallel time series are available, such as currency exchange rates or daily returns of stocks, that may have some common underlying factors. In [6], for example, ICA was used for decomposing parallel financial time series of weekly sales into basic factors. The cash flow of several stores belonging to the same retail chain was investigated, trying to find the fundamental factors common to all stores that affect the cash flow data. The effect of the actions taken at the individual stores and in its local environment could therefore be analyzed.

Noise reduction in natural images is another application discussed in [2]. A noise cleaning result is presented there, comparing the results of the Sparse Code Shrinkage method and classic wiener filtering to an ICA based noise filter.

ICA has also been used for feature extraction [3]. The columns of the mixing matrix A would represent features, and s_i would be the coefficient of the i^{th} feature in an observed data vector X.

Other areas where ICA can be applied include telecommunications, psychology and other social sciences. ICA could be considered in applications where projection pursuit and factor analysis are used.

2.3 Principles of ICA Estimation

In the previous sections we have mentioned the importance of the statistical independence of the components (i.e. source signals). In this section we present formal details on this key concept and the principles of an ICA estimation.

Let $y_1, \ldots, y_N$ be random variables. Let $p(y_1, \ldots, y_N)$ be the joint probability density function (pdf) and let $p_1(y_1), \ldots, p_N(y_N)$ be the marginal pdfs of $y_1, \ldots, y_N$, which can be obtained as

$$p_i(y_i) = \int p(y_1, \ldots, y_N)\, dy_1 \cdots dy_{i-1} dy_{i+1} \cdots dy_N$$

The variables $y_1, \ldots, y_N$ are said to be statistically independent if and only if the joint pdf is factorisable as the product of the N individual marginal pdfs, i.e.

$$p(y_1, \ldots, y_N) = p_1(y_1)\, p_2(y_2) \cdots p_N(y_N)$$

Two random variables y_i, y_j, $i \neq j$, are uncorrelated if their covariance is zero, i.e.

$$E\{y_i y_j\} - E\{y_i\}\, E\{y_j\} = 0$$

where $E\{\cdot\}$ is the expectation operator. It is well known that if two variables are independent, then they are uncorrelated. However, the inverse implication is not true. In general, uncorrelated variables are not necessarily independent. The equivalence between independence and uncorrelatedness holds for Gaussian random variables, though.

A restriction imposed in ICA is that independent components have non-Gaussian distributions. The reason is that the distribution of any orthogonal transformation of independent Gaussian variables y_i, y_j has the same distribution as the original y_i, y_j

variables. Therefore, the ICA model can only be estimated up to an orthogonal transformation. The mixing matrix A will not be identifiable for Gaussian independent components. It must be pointed out, however, that if only one of the independent components is indeed Gaussian, the ICA model can still be estimated [1–3].

2.3.1 Nongaussianity and Independence

According to the Central Limit Theorem, under certain conditions, the distribution of a sum of independent random variables tends toward a Gaussian distribution. In other words, the distribution of the sum of two independent variables is usually closer to a Gaussian distribution than any of the distributions of the original random variables.

In [2] it is shown that maximizing the nongaussianity leads to finding component as independent as possible. The basic ideas are the following. Consider a data vector X distributed according to the ICA model $X = AS$, i.e. it is a combination of independent components. Consider a component $y = w^T X = \sum_i w_i x_i$, where w is a vector that needs to be determined. A is the mixing matrix, so if w was one of the rows of A^{-1}, then y would equal one of the independent components in S. However, there is no prior knowledge of A, so there is a need for a good approximation.

Let $z = A^T w$. Then, $y = w^T X = w^T AS = z^T S$, a linear combination of the s_i. The coefficients or weights are the elements z_i of vector z. Due to the fact that the sum of independent random variables is more Gaussian than the original variables, the variable $y = z^T S$ is more Gaussian than any of the individual s_i. Conversely, y would be the least Gaussian if only one of the elements z_i was nonzero. This would be the case if y was in fact equal to one of the s_i. So, it would be desired to find a vector w that maximizes the nongaussianity of $w^T X$. This maximization would provide one of the independent components. In general, the optimization of nongaussianity in the N-dimensional space of vectors w results in $2N$ local maxima, two for each independent component. Those correspond to s_i and $-s_i$. This implies that the independent components can be estimated only up to a multiplicative sign.

We have seen that nongaussianity is key to ICA estimation. So, there is a need for a quantitative measure of the nongaussianity of a random variable. Next we present a

brief review of some of these measures. It is important to mention that if a random variable is centered (i.e. it has a zero mean) and has a unity variance, the calculations are simplified. Traditional ICA algorithms include pre-processing steps such as centering and whitening so that the ICA estimation becomes simpler and better conditioned [2, 3].

2.3.2 Measures of Nongaussianity

Kurtosis

The kurtosis of the random variable y is defined by

$$kurt\,(y) = \frac{E\left\{(y-\mu)^4\right\}}{\sigma^4}$$

where μ is the mean and σ^2 is the variance of y [7]. If y is a zero mean Gaussian, then $kurt\,(y) = 3$, since $E\left\{y^4\right\} = 3\left(E\left\{y^2\right\}\right)^2 = 3\left(\sigma^2+\mu^2\right)^2 = 3\left(\sigma^2\right)^2 = 3\sigma^4$.

Some authors define the kurtosis so that Gaussians have kurtosis equal to zero. In [2], kurtosis is defined as

$$kurt\,(y) = E\left\{y^4\right\} - 3\left(E\left\{y^2\right\}\right)^2$$

With this definition, it is clear that kurtosis is zero if y is a zero mean Gaussian. It is nonzero for most nongaussian random variables. It can be positive or negative. The absolute value and the square of the kurtosis are typically used as a nongaussianity measures. They have been used in ICA and related fields, mainly due to their simplicity: both computational and theoretical. In simple terms, given a data vector X, the ICA algorithm would search for the weight vectors w that maximize the absolute value of the kurtosis of $y = w^T X$. It is worth recalling that ICA algorithms pre-process the data vector X to center it. An important drawback for the use of kurtosis is its sensitivity to outliers and therefore, it is not a very robust measure of nongaussianity [2].

Negentropy

The concept of entropy is very important in information theory. For a random variable, the more unpredictable and unstructured (i.e. the more *random*) it is, the larger its entropy is. For a discrete random variable Y, entropy is defined as

$$H(Y) = -\sum_i P(Y = a_i) \, log\left[P(Y = a_i)\right]$$

where a_i are the possible values of Y. The generalization for continuous random variables is usually called differential entropy, and is defined as

$$H(y) = -\int p(y) \, log\left[p(y)\right] dy$$

where $p(y)$ is the density function of the random variable y. It turns out that a Gaussian variable has the largest entropy among all random variables of equal variance [8]. Therefore, it could be used as a measure of nongaussianity.

A modified version of the differential entropy is the negentropy. It is defined as

$$J(y) = H(y_{gauss}) - H(y)$$

where y_{gauss} is a Gaussian random variable of the same covariance matrix as y. Negentropy has the properties of being always non-negative, and being zero if and only if y has a Gaussian distribution. As a measure of nongaussianity, negentropy is a well justified by statistical theory. Nevertheless, the use of this measure is problematic in practice, since it is computationally difficult to calculate. Therefore, approximations of negentropy have to be used. A classical approximation of negentropy is

$$J(y) \approx \frac{1}{12} E\left\{y^3\right\}^2 + \frac{1}{48} kurt(y)^2$$

where y is assumed to be of zero mean and unit variance. This approximation has similar problems to those of kurtosis, particularly the lack of robustness [2].

A more robust approximation is

$$J(y) \approx \sum_{i=1}^{p} k_i \left[E\{G_i(y)\} - E\{G_i(\nu)\}\right]^2$$

with k_i positive constants, G_i non-quadratic functions, ν is a standardized Gaussian and y is assumed to be zero mean and unit variance also. This approximation allows the construction of a measure with the properties of being non-negative and equal to zero only for Gaussian distributions. If only one non-quadratic function G is used, then

$$J(y) \propto \left[E\{G(y)\} - E\{G(\nu)\}\right]^2$$

Robust approximations of negentropy are obtained by using non-quadratic functions that do not grow too fast. For example, the functions $G_1(u) = \frac{1}{a_1} log\left[cosh\left(a_1 u\right)\right]$, $1 \leq a_1 \leq 2$, and $G_2(u) = -exp\left(-\frac{u^2}{2}\right)$ have been used [2]. ICA algorithms have been developed using these approximations of negentropy as objective functions to be maximized. One such algorithm is the FastICA algorithm [9].

2.3.3 Mutual Information

Another measure of the dependence between random variables is mutual information. For the random variables $y_1, \ldots, y_N$, the mutual information I between them is defined as

$$I(y_1, \ldots, y_N) = \sum_{i=1}^{N} H(y_i) - H(\vec{y})$$

where $H(y_i)$ is the differential entropy of the individual random variables and $H(\vec{y}) = H(y_1, \ldots, y_N)$ is the "joint" differential entropy. I is non-negative and zero if and only if the variables are statistically independent [2].

ICA can be defined by the mutual information. Given the data vector X, the goal is to find an invertible transformation W that minimizes the mutual information of the transformed components s_i, in the model $S = WX$. It turns out that ICA estimation by minimizing mutual information is equivalent to maximizing the sum of nongaussianities of the estimates when those estimates are constrained to be uncorrelated [2]. This is

due to the following relationship between negentropy and mutual information:

$$I(y_1, \ldots, y_N) = C - \sum_{i=1}^{N} J(y_i)$$

where it is assumed that the y_i's are uncorrelated and of unit variance. C is a constant. It is worth mentioning that the uncorrelatedness constraint is not necessary, but it simplifies the computations. As it has been mentioned before, ICA algorithms include pre-processing steps of centering and whitening of the data, precisely to simplify the calculations.

2.4 Under-determined BSSS

In the previous sections we presented principles of ICA. As we have pointed out, BSSS-ICA is well suited for over and well determined cases ($M \geq N$). For the under-determined cases, that is, when the number of sensors is smaller than the number of independent sources ($M < N$), methods for signal separation have been proposed [10, 11] and referred to as Under-determined BSSS. However, these methods produce losses in the recovered (separated) signals which increase with the reduction of the number of sensors. This affects drastically the accuracy of any subsequent classification of the source signals.

2.5 Concluding Remarks

In this chapter we provided a quick overview of the concepts required for the presentation of the method proposed in this thesis, which we describe in the next chapter. The main concept discussed here was Independent Component Analysis (ICA). As it will be explained next, our method relies on this technique.

After the introduction of our method in chapter 3, we will present two frameworks for using the method for two different applications. We will leave the background and related work reviews on those areas for the corresponding chapters (5 and 6).

Chapter 3

Proposed Method

The main contribution of this thesis is the development of a method named Guided Under-determined Source Signal Separation (GUSSS). This method handles an extreme case of under-determination where the number of sensors is actually equal to one – i.e. $M = 1$. Unlike BSSS-ICA, where the source signals to be separated are unknown, in the proposed method, it is assumed that the source signals are one of a number of expected *signatures*, hence the term "guided" in GUSSS.

3.1 GUSSS and GUSSS Ratio

In order to explain the proposed method, let x_1 be a linear combination of N independent components. That is, x_1 represents a sensed signal from the single sensor of a given system. Such a system can be an Improvised Explosive Device (IED) detector trying to determine the presence of a hazardous substance, or a classifier of electrical signals generated by muscle activity. Next, let s_p be a particular known component, or signature, that the system is trying to identify within the observed signal x_1. Since the sensor captures not only s_p, but also various other components s_i, we should write:

$$\begin{aligned} x_1 &= c_1 s_1 + c_2 s_2 + \cdots + c_p s_p + \cdots + c_N s_N \\ &= c_p s_p + \sum_{i \neq p} c_i s_i \\ &= c_p s_p + \tilde{s} \end{aligned} \tag{3.1}$$

where c_i, $i = 1, \cdots, N$ are unknown mixing coefficients. It is assumed that $c_i \geq 0$. The expression stresses the fact that x_1 can be considered a linear combination of the signature of interest and a set $\tilde{s}$ of scaled signatures of other components. Since initially we are interested in separating or identifying only s_p from the observed signal, it is assumed that $\tilde{s}$ is independent from s_p. This assumption is an obvious consequence of the assumption that all N components are independent – i.e. if N components can be regarded as independent, any linear combination of $N-1$ components must also be independent of the remaining one. Moreover, the algorithm for GUSSS will successfully identify s_p within x_1 whenever $c_p \neq 0$. This could mean, for instance, that a chemical compound associated with the signature s_p is present in a scene. So, the question remaining becomes how to determine c_p.

As it has just been implied, two situations may arise: the desired signature is indeed present in the mixed signal x_1, or it is not. In order to distinguish between those two situations, the algorithm creates a second synthesized signal x_p by *injecting* a weighted copy of the particular signature s_p into the sensed signal x_1. That is:

$$x_p = w_1 x_1 + w_p s_p \tag{3.2}$$

where w_1 and w_p are arbitrarily chosen constants. Substituting eq. (3.1) in eq. (3.2), we obtain:

$$x_p = w_1 \left(c_p s_p + \tilde{s}\right) + w_p s_p = w_1 \tilde{s} + \left(w_1 c_p + w_p\right) s_p \tag{3.3}$$

which now allows us to write two equations instead of one. That is

$$\begin{aligned} x_1 &= \tilde{s} + c_p s_p \\ x_p &= w_1 \tilde{s} + k_p s_p \end{aligned} \tag{3.4}$$

where $k_p = w_1 c_p + w_p$. Finally, we can express these equations in matrix form as

$$X_p = AS$$

where

$$X_p = \begin{bmatrix} x_1^T \\ x_p^T \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & c_p \\ w_1 & k_p \end{bmatrix}$$

$$S = \begin{bmatrix} \tilde{s}^T \\ s_p^T \end{bmatrix}$$

The last step of the algorithm is to solve for S. Since we now have two independent components and two linear equations on s_p and $\tilde{s}$, we can apply any traditional ICA algorithm to separate s_p and $\tilde{s}$ from x_1 and x_p. Moreover, a sub product of the ICA algorithm is the mixing matrix A, and the coefficients of such matrix can be used to infer whether or not a particular signature was present in the originally sensed signal x_1. For example, if we consider the case when the particular signal s_p is not present in the mixture signal x_1, the mixing coefficient c_p should be in theory zero. On the other hand, if s_p is indeed present in the mixture x_1, that coefficient must be greater than zero.

In practice, mainly due to noise, the coefficient c_p is never zero. However, it should be very small whenever the particular signature is not present in x_1 and it should be large otherwise. We define the GUSSS ratio as:

$$r_p = \left| \frac{1}{c_p} \right| \tag{3.5}$$

Finally, while what constitutes a "large" or a "small" value for the coefficient c_p may not be obvious, it is clear that the derived GUSSS ratio can be used as a criterion for determining whether a particular signature is present or not in the sensed signal. It can become a feature characterizing sensed signals. Note that the use of the absolute value is due to the fact that the ICA can estimate the independent components up to a multiplicative sign, as explained in section 2.3.1.

3.2 Concluding Remarks

In this chapter we presented the GUSSS method and introduced the GUSSS ratio. In the next few chapters we will propose frameworks for using them in very different areas: material detection using Terahertz signatures, and assistive technology using muscle activity signals. But first, in chapter 4 we present some preliminary experiments we did on sound signals using the proposed method. For all our tests we used the FastICA algorithm developed by [9].

Chapter 4

Preliminary Tests Using Sounds

The Cocktail Party problem was one of the first motivations for the method proposed in this thesis. In section 4.1 we explain the experiments we designed to test the ability of GUSSS to determine the presence or not of a particular sound signature. As we show in section 4.2, the results obtained were very promising.

4.1 Description of the Experiments

We considered 9 different sound sources or signatures. Let s_1, s_2, ..., s_9 be those signatures, which are shown in Figure 4.1. Assume there is only $M = 1$ microphone (sensor) that can capture a sound signal. This signal is a combination or mixture of the 9 source signatures, and it can be modeled as:

$$x = c_1 s_1 + c_2 s_2 + \cdots + c_9 s_9 \tag{4.1}$$

The mixing coefficients c_i represent the intensities that each individual signature s_i has in the mixed signal, due to the different distances between the sources and the microphone. Note that if a particular signature, say s_p, contributed to the mixture x, the corresponding coefficient c_p is non-zero. If source p could not be detected when the mixture was obtained, then $c_p = 0$.

Our goal is: given a sensed mixture x, we want to determine if signature s_p was present or not within x. Since we have no knowledge of the mixing coefficients, but

we do have knowledge of the signature s_p, we can use the proposed GUSSS method to estimate c_p and thus the GUSSS ratio r_p. The GUSSS ratio can then be used to determine the presence or not of the signature in the mixture.

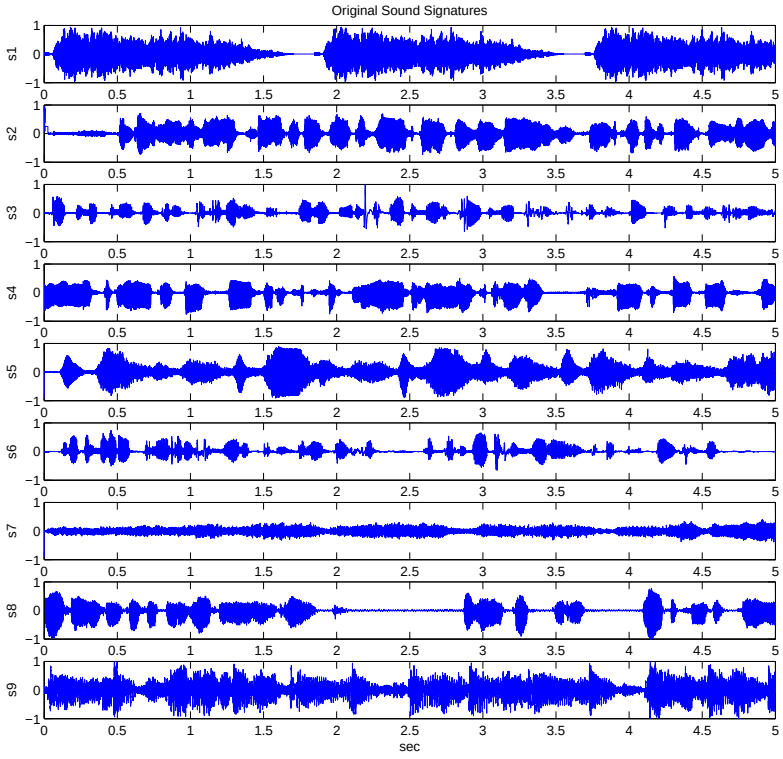


Figure 4.1: Nine sound signatures used for the preliminary tests of the proposed method. They correspond to: a police car's siren (s_1); a man speaking English (s_2); an old man speaking a foreign language (s_3); a woman speaking English (s_4); a band playing classical music (s_5); another man speaking a foreign language (s_6); an opera singer (s_7); a woman speaking a foreign language (s_8); a band playing pop music (s_9).

4.1.1 Experiments #1: No Noise

We ran tests for each one of the 9 signatures available. For example, for each p^{th} signature s_p as the target signature, i.e, the one we want to detect in a mixture, we created 1000 test mixtures using arbitrary linear combinations of all signatures, as in equation 4.2. Half of these test signals corresponded to mixtures containing the target

signature s_p. We refer to those 500 test signals as belonging to class "present". As we explained before, to guarantee that s_p is present in the mixture, the corresponding coefficient c_p in the linear combination has to be non-zero. So we employed a lower limit (ll) and an upper limit (ul) that satisfy $0 < ll \leq ul \leq 1$. Then, the coefficient c_p is randomly chosen from a uniform distribution defined over the interval [ll, ul]. The remaining coefficients were randomly chosen from a uniform distribution over the unit interval [0, 1] and then normalized so that $\sum_{i \neq p} c_i = 1 - c_p$. Note that ll can be interpreted as a minimum percentage of the particular signature s_p present in the mixtures.

The other half of the test mixtures belonged to class "not present", that is, those signals did not contain the target signature s_p. Therefore, for each one of them the corresponding coefficient c_p was set to zero. The other coefficients were once again randomly chosen from the unit interval and normalized as before.

For this first set of experiments we did not add any noise, as we will do in the next set of experiments. For all the test mixtures we applied the proposed method injecting the corresponding signature (section 3.1), and we calculated the GUSSS ratios and plotted them. We present the results in section 4.2.

4.1.2 Experiments #2: Adding Noise

As in the previous set of experiments, we created 1000 test mixtures for each of the signatures, divided into two classes (present and not present). This time we added noise to the mixtures:

$$x = c_1 s_1 + c_2 s_2 + \cdots + c_9 s_9 + \eta \tag{4.2}$$

Where η is Gaussian noise with zero mean and standard deviation σ. We tried different noise levels by changing σ. The mixing coefficients were randomly chosen as explained before. For the proposed method we used noisy versions of the signatures as well. That is, we injected

$$\check{s}_p = s_p + \xi \tag{4.3}$$

to create the synthesized signal (eq. 3.2) and applied the method described in section 3.1 to obtain the GUSSS ratios. We introduce the noise term ξ to simulate situations when the available signatures are corrupted, or when they have been obtained using noisy sensors. The term ξ is Gaussian noise with zero mean and standard deviation $\sigma/5$. We use this smaller noise level because we expect the mixtures to be noisier than the individual signatures available.

4.1.3 Experiments #3: Noise and Time Shifting

For the last set of experiments we created mixtures using shifted versions of the original signatures, plus noise. Let $\overleftrightarrow{s}_1$, $\overleftrightarrow{s}_2$, ..., $\overleftrightarrow{s}_9$ be horizontally shifted versions of the original sound signatures. This shifting of the signatures represent delays in the emission of the sounds by the sources or delays in the detection of those same sounds. We created test mixtures as:

$$x = c_1 \overleftrightarrow{s}_1 + c_2 \overleftrightarrow{s}_2 + \ldots + c_9 \overleftrightarrow{s}_9 + \eta \tag{4.4}$$

As before, η is Gaussian noise. The amount of shift in each signature was random, ranging from $-\tau_{max}$ to τ_{max}. In other words, each sound from a source is a delayed or advanced version of the original signature, and the amount of delay or advance can be as much as τ_{max}. The amount of shift for each test mixture was made different.

Once again, 1000 test mixtures were created for each of the signatures: 500 in the "present" class, and 500 in the "not-present" class. For calculating the GUSSS ratios we injected noisy, non-shifted versions of the original signatures.

4.2 Results

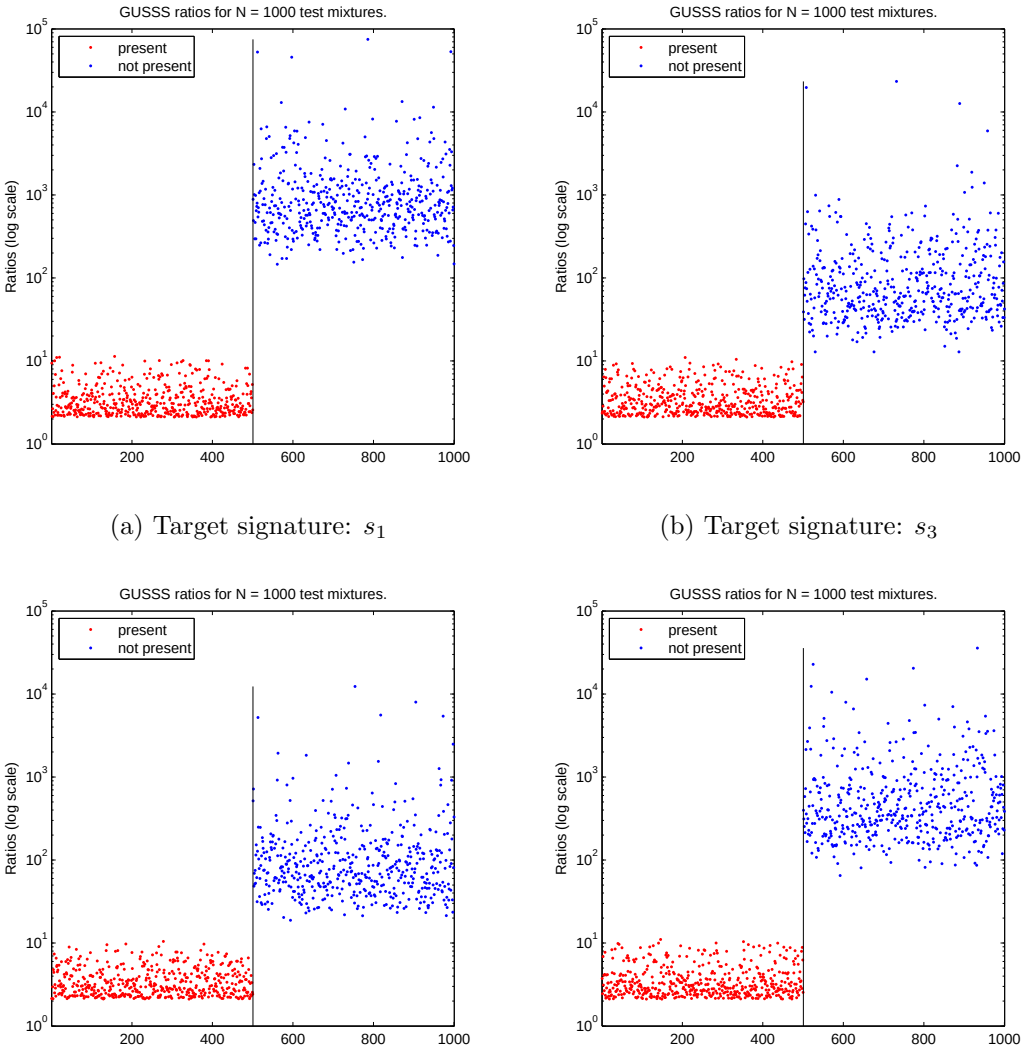
In this section we present the results obtained for the three experiments described above. Here, for each experiment we present plots of the GUSSS ratios obtained for only 4 out of the 9 signatures used. The complete set of plots can be found in Appendix A. For all the experiments we set the lower limit $ll = 0.1$ and the upper limit $ul = 0.9$. Those limits indicate that in the cases when the target signature is present in the

mixture, its contribution to the mixtures can be from 10% to 90%. We also show some example mixtures and the recovered signals we obtained using our method.

4.2.1 Results Experiments #1

Figure 4.2 (a) illustrates a set of GUSSS ratios from one of the experiments #1, specifically, the case where the target signature was s_1 (the police car's siren). The first 500 GUSSS ratios, shown in red, belong to the class "present", i.e., the class of mixtures for which signature s_1 was present in the scene. The remaining 500 ratios, shown in blue, belong to the class "not present". The ratios are shown in log scale. All of the following figures show similar plots as the one just described.

Note that in all the experiments shown here there is a significant difference between the ratios of class "present" and the ratios of class "not present". It would not be difficult to choose a threshold for separating the two classes, giving a perfect classification rate.



(a) Target signature: s_1

(b) Target signature: s_3

(c) Target signature: s_7

(d) Target signature: s_9

Figure 4.2: GUSSS ratios corresponding to 500 mixtures containing a target signature, and 500 mixtures without the target signature. The vertical line separates those two groups. Y-axis (GUSSS ratio axis) is shown in log scale.

Figure 4.3 illustrates an example mixture containing a target signature (s_4 in this particular case), the signature itself, and the recovered signature we obtain using our method. As explained before, no noise was added in this case. We note that the original signature and the recoverd one are almost identical.

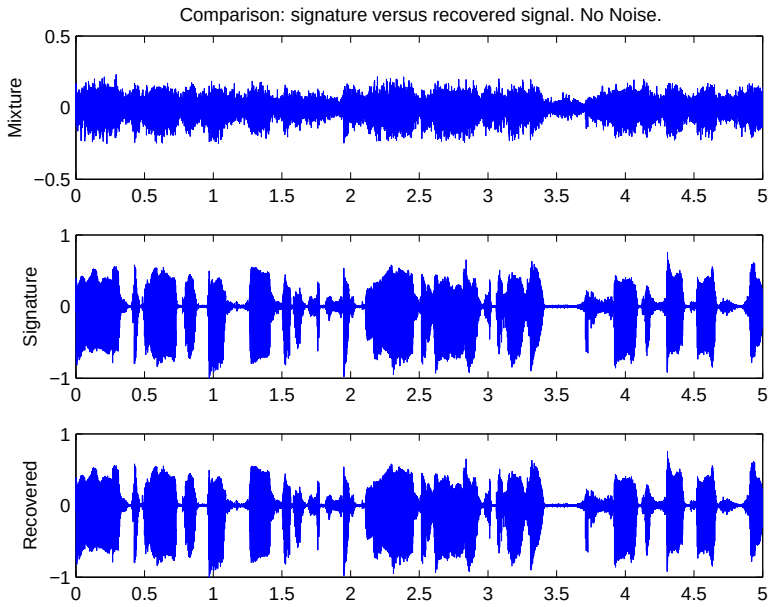
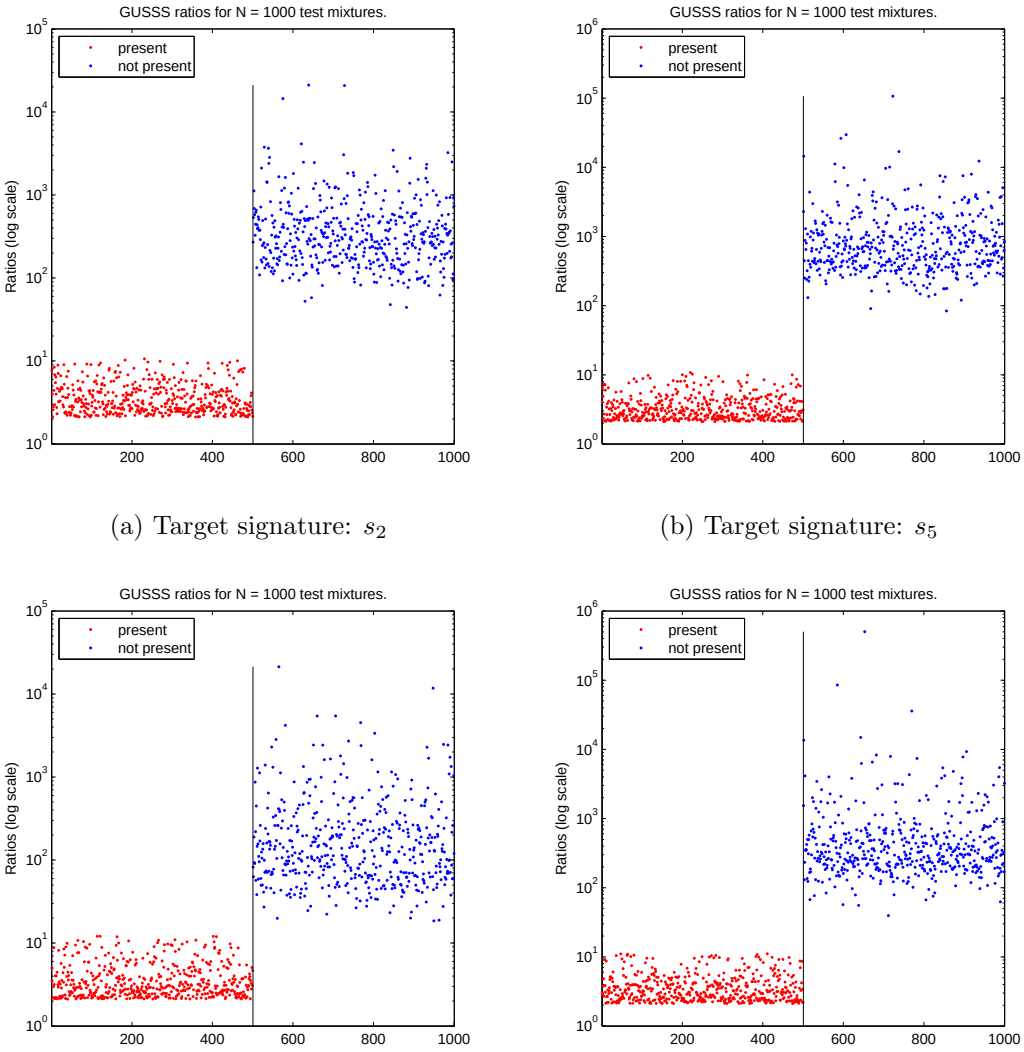


Figure 4.3: A mixture of signatures, the target signature s_4, and the recovered signal.

4.2.2 Results Experiments #2

For these experiments we used two different noise levels, $\sigma = 0.01$ and $\sigma = 0.1$. The amplitudes of the original sound signatures are constrained to the interval $[-1, 1]$. As we did for the experiments with no noise, we show the GUSSS ratios for 4 out of the 9 target signatures. The remaining cases using the remaining 5 signatures can be found in Appendix A.

Noise with $\sigma = 0.01$:



(a) Target signature: s_2

(b) Target signature: s_5

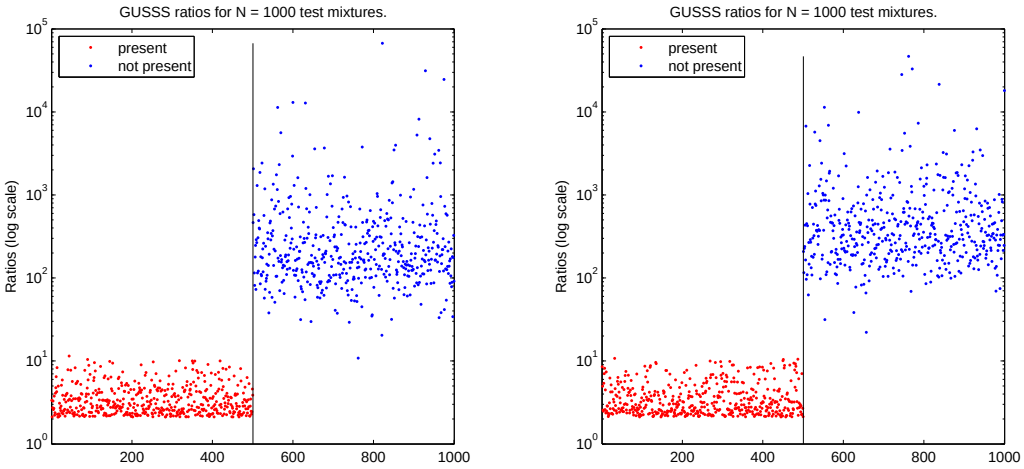
(c) Target signature: s_6

(d) Target signature: s_8

Figure 4.4: GUSSS ratios corresponding to 500 noisy mixtures containing a target signature, and 500 mixtures without the target signature. The vertical line separates those two groups. Y-axis (GUSSS ratio axis) is shown in log scale. Noise level: $\sigma = 0.01$

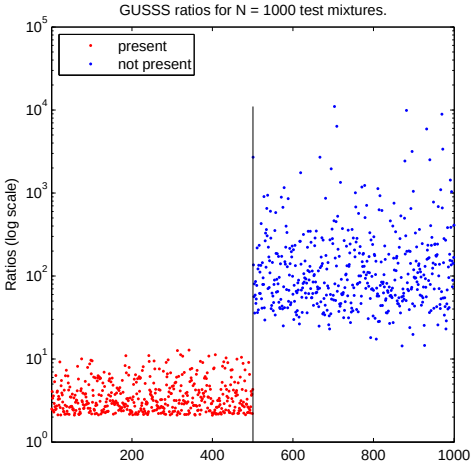
For these cases we would also get a 100% correct classification rate with any classifier, even a simple threshold based one.

Noise with $\sigma = 0.1$:

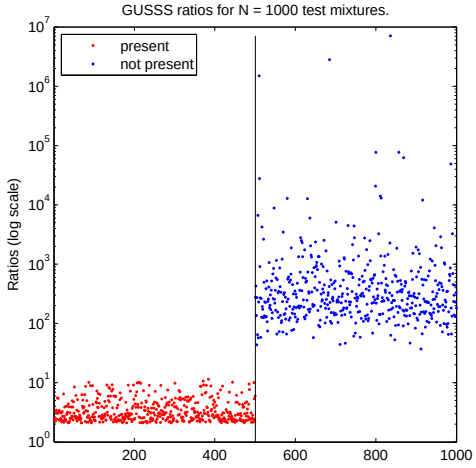


(a) Target signature: s_2

(b) Target signature: s_5



(c) Target signature: s_6



(d) Target signature: s_8

Figure 4.5: GUSSS ratios corresponding to 500 noisy mixtures containing a target signature, and 500 mixtures without the target signature. The vertical line separates those two groups. Y-axis (GUSSS ratio axis) is shown in log scale. Noise level: $\sigma = 0.1$

We note that with this higher noise level the two classes come closer together. In Appendix A it can be seen that there are some cases where the classes actually overlap. Nonetheless, even with a simple hard threshold approach, the overall correct classification rate is above 99%.

As we did before, we illustrate an example mixture containing the target signature s_4, the signature itself, and the recovered signature (Figure 4.6). This time we added noise with $\sigma = 0.1$. Note that the recovered signal is a noisy version of the original signature.

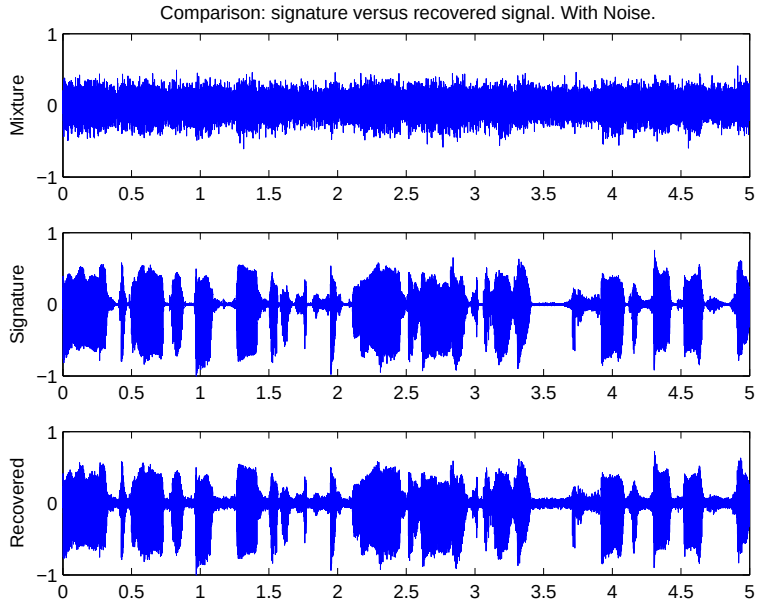
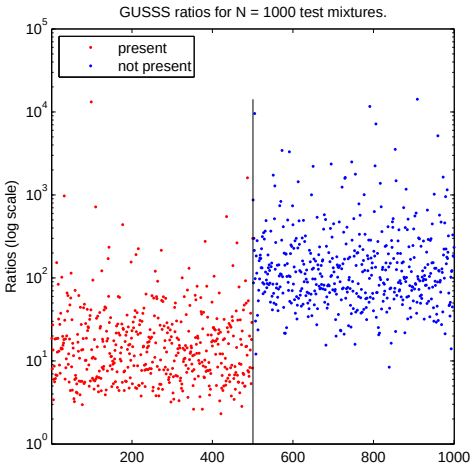


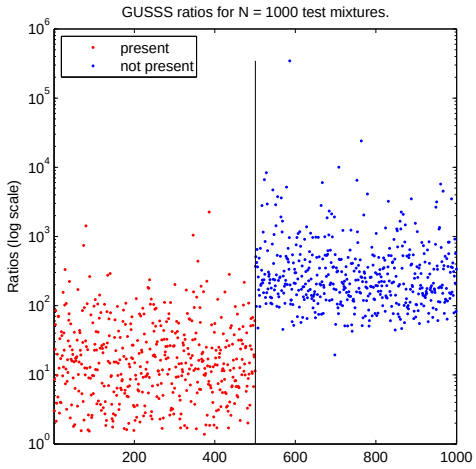
Figure 4.6: A mixture of signatures with noise level $\sigma = 0.1$, the target signature s_4, and the recovered signal.

4.2.3 Results Experiments #3

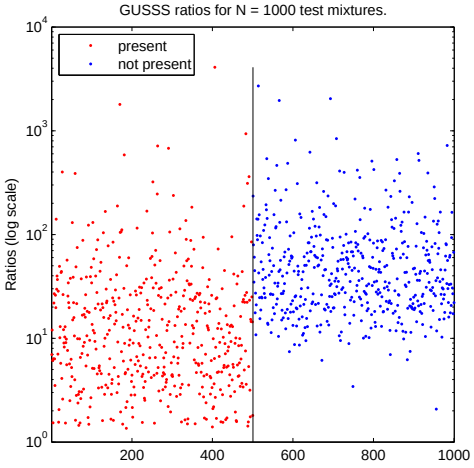
For these experiments we used a noise level of $\sigma = 0.05$ and a maximum time shift of $\tau_{max} = 0.01s$. The original sound signals last for about $5s$. Once again, we show here 4 of the 9 cases. The remaining cases are shown in Appendix A.



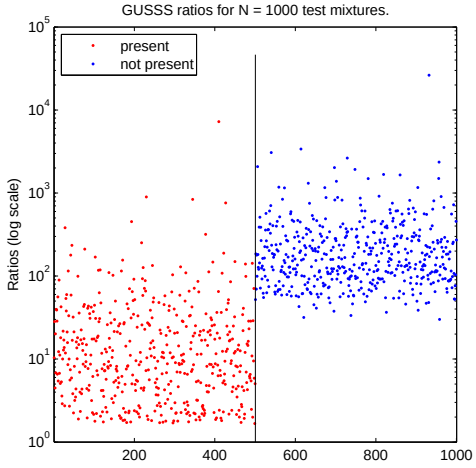
(a) Target signature: s_3



(b) Target signature: s_4



(c) Target signature: s_7



(d) Target signature: s_9

Figure 4.7: GUSSS ratios corresponding to 500 noisy mixtures with time shift containing a target signature, and 500 mixtures without the target signature. The vertical line separates those two groups. Y-axis (GUSSS ratio axis) is shown in log scale.

Once again, we illustrate an example mixture of shifted signatures (including the target signature s_4), the signature itself, and the recovered signature (Figure 4.8). Once again the recovered signal seems very good, although it is not as clean as in the case without noise (Figure 4.3).

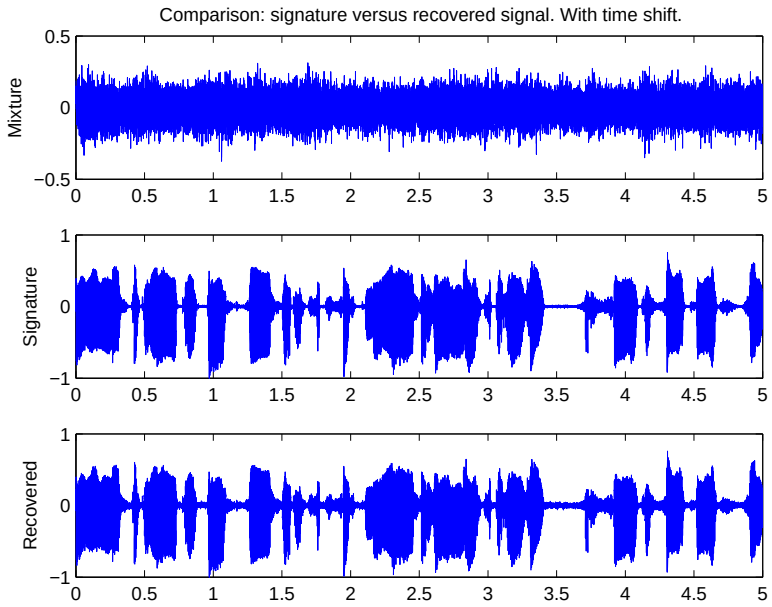


Figure 4.8: A mixture of shifted signatures, the target signature s_4, and the recovered signal.

These experiments show that horizontal (time) shifts have an important effect on the classification performance. Figure 4.7 (c), which corresponds to signature s_7, shows considerable overlapping of the two classes, which leads to a classification rate of around 75%. Still, with the exception of signature s_6, in all the other cases we achieve classification rates above 90% with simple classifiers (see Appendix A).

4.3 Concluding Remarks

The results presented in this chapter suggest that the GUSSS ratio could be used as a criterion to recognize when a signature of interest is present within a sensed signal. Classes "present" and "not present" where clearly separated in the cases without and with noise. In the cases with a higher noise level the classes seem to come closer together, but the effect was not considerable. In all these cases, a simple hard threshold would be sufficient to discriminate between the two classes. In the last set of experiments where the signals where time shifted, the two classes do overlap. However, the

classes are almost completely separable for most of the target signatures. It must be pointed out that it is important to sample the signals correctly, in order to reduce or eliminate delays.

In the next chapters we present frameworks for using GUSSS and the GUSSS ratio in two different areas such as material detection and assistive technology.

Chapter 5

Material Detection Using Terahertz Signals

Terahertz technology (THz) has been greatly developed over the past decade. The terahertz electromagnetic radiation lies between light and radio waves ($0.1 - 0.3$ to 10 THz) [12–14]. It has attracted a lot of attention, especially because of its potential in innovative sensing systems, ultra-fast wireless communication systems, devices for medical examinations and detection of hazardous materials [15]. The main advantages of this radiation are: the ability to penetrate many common barrier materials enabling hidden objects to be seen; adequate spatial resolution for imaging or localisation of threat objects due to short wavelengths; non-ionising properties; safe to use on people at modest intensities; etc. [16].

Research on terahertz technology has focused on sources and sensors, as well as guiding structures, quasi-optics, antennas, filters and submillimeter-wave materials [12]. This includes the development of semiconductor devices such as THz quantum cascade lasers, THz-range quantum well photodetectors and high-precision tunable continuous wave sources. Furthermore, efforts have been made to develop pulsed THz measurement systems, to model and measure atmospheric propagation and to establish frameworks to construct a materials database in the THz range, including standardization of the measurement protocol [15].

Many objects and chemicals including explosives, illegal drugs, proteins and other

biological materials have characteristic spectroscopic signatures at THz wavelengths which can be used to identify them [15, 16]. The work presented in [13] describes a method for the detection of explosives using THz pulsed spectroscopic imaging. In [17] the authors propose a THz-wave device for the nondestructive detection of illicit drugs and hazardous substances hidden in sealed envelopes.

In this chapter we present a brief review of terahertz technology, pattern recognition techniques that can be applied to Terahertz (THz) signatures, and the framework proposed and used to test GUSSS in this area. Such a framework could be used for detection of Improvised Explosive Devices (IED), for detection of hazardous or illegal substances in airports, etc.

5.1 Background on Terahertz Technology

The THz region in the electromagnetic spectrum is situated between infrared light and microwave radiation. Common techniques usually applied to neighbouring bands have not shown the same success for THz radiation [14]. So, for many years there has been this "undeveloped gap" between the electronics of the microwave and the optics of infrared [15]. Some decades ago THz technologies were mainly used in astronomy for searching far-infrared radiation, in laser fusion for the diagnostic of plasmas [18] and in chemistry for spectral characterization of rotational and vibrational resonances and thermal-emission lines of simple molecules [14].

Currently, one of the main uses of Terahertz technology is in spectroscopy [18], which allows investigating properties of materials as a function of frequency. Many methods have been developed for performing THz spectroscopy, including: Fourier transform spectroscopy (FTS); narrowband spectroscopy with tunable THZ source or detector; and THz time-domain spectroscopy (THz-TDS). The latter, a more recent technique, uses short pulses of broadband THz radiation typically generated using ultra-fast laser beams [14]. It has various advantages over FTS, including being able to calculate refractive indices and absorption coefficients from the phase and amplitude of the waveforms, and being able to use information directly from the time domain

waveforms [15].

The first demonstration of THz wave time-domain spectroscopy was in the late 1980's, and after that there has been a series of significant advances thanks to improved materials leading to more powerful THz sources and more sensitive detectors [14, 18]. The development of Quantum Cascade Lasers and Terahertz-Range Quantum Well Photodetectors are of great impact [15]. All these advances have provided new opportunities for understanding the THz frequency range, allowing major technical developments, which in turn have greatly extended the potential of THz systems. The use of THz radiation is now open to fields such as physics, chemistry, biology, materials science and medicine [12, 14, 18].

A major interest for THz technology is the detection of explosive devices such as improvised explosive devices (IED). The National Research Council has made recommendations regarding the application of this technology for explosive detection techniques and security screening. These recommendations are based on the fact that most explosives exhibit strong absorption and dispersion in the THz frequency range [18], and they exhibit characteristic spectroscopic signatures at THz wavelengths which can be used to identify these explosives [13, 16, 19].

Several studies have reported the spectra of energetic compounds like RDX, PETN, HMX and TNT, as well as commercial explosives based on those compounds [13,20–22]. Practical implementation of detection and security systems would need to operate in reflection mode rather than in transmission mode because of the high absorption coefficients of the explosives [13, 19].

There is also an interest to detect and identify other threat materials such as chemical and biological warfare compounds, and illegal substances like drugs of abuse [19]. THz signatures of a number of drugs were identified in [23], and a THz-wave device for nondestructive detection of illicit drugs and hazardous substances hidden in sealed envelopes was proposed in [17]. Detection of hidden objects depends on the transmission of radiation through barrier materials. Envelopes and most of barrier materials such as cloth, paper, cardboard and plastics are semi-transparent to THz radiation [19].

The interest in identifying materials by means of their THz signatures has led to

the development of spectra databases. These are valuable to enlarge the range of applications of terahertz spectroscopy by making it a popular measure of materials [15]. There are several open databases for THz spectroscopy, including the result of the European project "THz-Bridge: Tera-Hertz radiation in Biological Research, Investigation on Diagnostics and study of potential Genotoxic Effects" [24], and the public databases from Tera-photonics Laboratory, RIKEN Sendai and the National Institute of Information and Communications Technology (NICT) [25,26] – which were used in this research.

5.2 Pattern Recognition of THz Signatures

Most of the work done for identification and classification of materials using THz technology use individual samples of the materials under study. In those scenarios, a few pattern recognition and classification methods can be found in the literature. In [27] the authors proposed a system for identification and classification of four explosive and bio-chemical materials using THz spectroscopy. They used Principal Component Analysis (PCA) for feature dimensionality reduction, a minimum distance classifier and a neural network based classifier. In [28] the author discusses the Mahalanobis distance classifier, the Euclidean discrimination matrix and Support Vector Machines (SVM) and presents results of case studies for biomedical specimen identification using those methods.

Unlike previous methods, the approach presented here aims to detect materials even if they are mixed together with other materials. That is achieved by using the GUSSS technique developed in chapter 3, which can be used to separate the different signatures that may be present in a scene. In the following sections we discuss a framework proposed to achieve that purpose.

5.3 Proposed Framework for THz Technology

Imagine a scenario where a vehicle or person must navigate in a hostile environment with potential IEDs hidden underground. A ground penetrating radar can be successfully utilized as a counter measure, but as such, it is desirable that only one sensor be employed by the radar.

Now, imagine another scenario where the passengers luggage must be scanned in search of illegal substances or explosives [29], [30]. Once again, the use of a single sensor, even a portable one, can simplify immensely the job of security officers in airports, schools, train stations, etc.

That is precisely our goal here. We propose a framework that requires only one sensor. The main feature used for the learning and testing process is the GUSSS ratio we defined in 3. It is calculated for training signals in the system, and it is used to classify the testing signals, as we will explain in section 5.4. Next we describe two approaches used with the GUSSS ratio as feature for classification.

5.3.1 Hard Threshold Approaches

Using a hard threshold is a very simple classification method for two-class, one dimensional problems, such as the one with which we are dealing here. Figure 5.1 illustrates a set of GUSSS ratios from one of the experiments we ran in this work. It is similar to the figures presented in the previous chapter. The first 500 GUSSS ratios, shown in red, belong to the class "present", i.e., the class of sensed signals for which a certain material with signature s_p was present in the scene. Similarly, the remaining 500 ratios, shown in blue, belong to the class "not present". The ratios are shown in log scale. As the figure indicates, it is not difficult to choose a threshold for separating the two classes.

The first method for obtaining such a threshold is based on geometric means. We calculated the geometric mean of the GUSSS ratios of each class (present and not present), and then the threshold was simply the geometric mean of those two. Obviously, if the number ratios is the same for each class, then the threshold is simply

the geometric mean of all the ratios. We used geometric means instead of simple (arithmetic) means because that fits better the distribution of the ratios.

A second method used relied on modeling the distributions of the ratios. Based on preliminary observations, we determined that it was better to use the logarithm of the ratios to learn the threshold. That is, given samples from both classes of GUSSS ratios (signature present and not present), we calculated their logarithms. We assumed these logarithms follow normal distributions and for each set of ratios we calculated the sample mean and variance, so that we could find the intersections of the two distributions. Since the variances are likely to be different, there should be two intersection points. However, only one of them is of interest, namely the one closer to the midpoint between the two sample means. Figure 5.2 illustrates this idea. Note that the ratio's axis was made horizontal (i.e. rotated with respect to Figure 5.1) for visualization purposes. Now if we let ρ be the value of the intersection of interest and since the parameters used to calculate the intersection corresponded to the base-10 logarithms of the GUSSS ratios, the threshold can be obtained from: $thr = 10^{\rho}$.

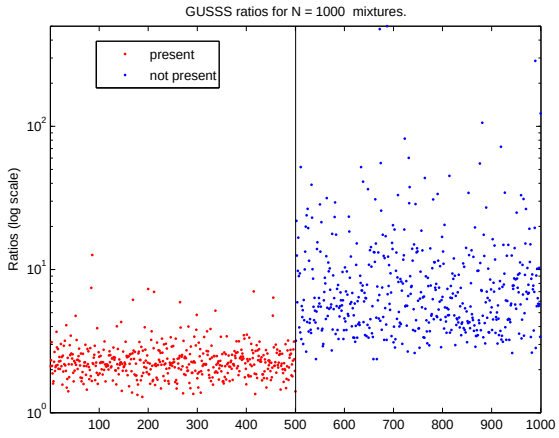


Figure 5.1: Ratios corresponding to 500 mixture signals containing a particular signature of interest, and 500 mixture signals without the particular signature. The vertical line separates those two groups. Y-axis (ratio axis) is in log scale.

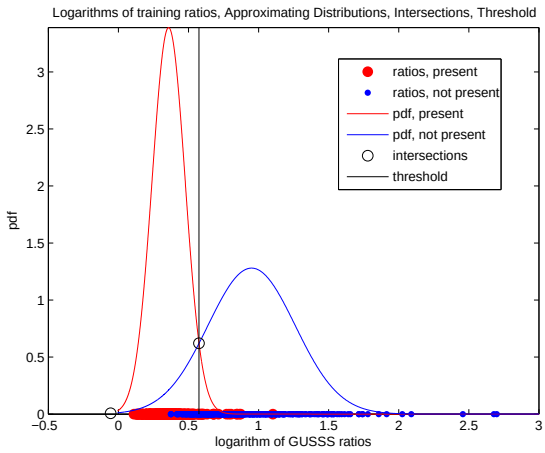


Figure 5.2: Obtaining the threshold approximating the logarithms of the GUSSS ratios via Gaussian distributions. The ratios are the same as those in Figure 5.1, but in this figure we plotted their base 10 logarithms, all on the same horizontal axis.

5.3.2 Support Vector Machine (SVM)

A third approach to train the system relied on a more sophisticated method than the calculation of thresholds. We used Support Vector Machine (SVM) with Gaussian Kernels to train and test the system. As before, we used the base-10 logarithm of the GUSSS ratios as input for the SVM algorithms. With this approach we don't actually learn a threshold, but we get a set of support vectors that are later used for classifying the testing ratios (their logarithms, to be exact).

SVM algorithms are convenient for two class problems. They rely on pre-processing the data to represent patterns in a high dimensional space, typically much higher than the original feature space. With an appropriate nonlinear mapping to a sufficient high dimension, data from two categories can always be separated by a hyperplane. SVM is therefore a machine learning method able to handle problems of nonlinear classification. It is based on the rule of structure risk minimum. Therefore, it becomes a decision machine as opposed to Bayesian learning methods that provide posterior probabilities [31].

5.4 System Implementation

In this section we discuss in more detail how the techniques presented earlier were used in the proposed system. Figure 5.3 shows a block diagram illustrating the various parts of the system and the flow of the process. We start with a collection of THz signatures, which in the experiments ran for this work come from public databases. In our approach we use time domain signals, so we need to transform the THz signatures from the databases (IFFT block in Figure 5.3). Time signatures were used for training the system, for creating test signals, and for obtaining the GUSSS ratios while testing. We explain the details of each part of the system in the following subsections.

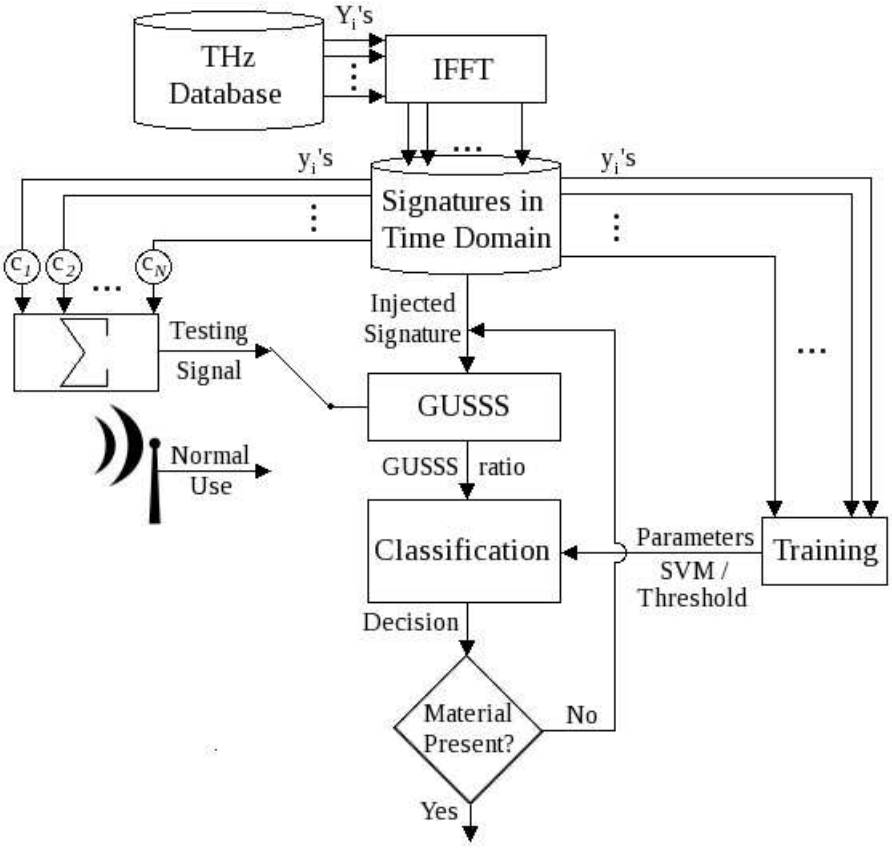
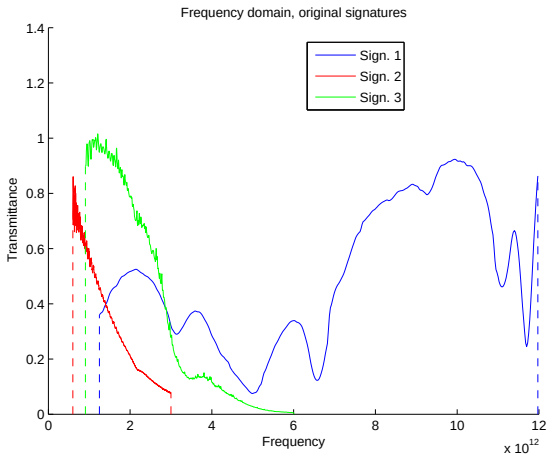


Figure 5.3: Proposed system.

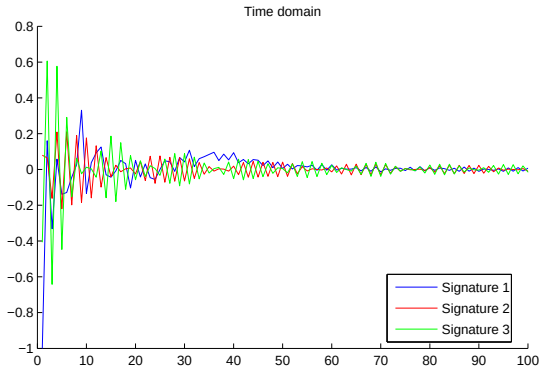
5.4.1 Database of THz Signatures

The experiments that we conducted rely on THz signatures obtained from two public databases [25, 26]. These signatures are originally in frequency domain, but we apply the methods and techniques described in section 5.3 to signals in time domain. Therefore, we apply the Inverse Fast Fourier Transform (IFFT) to the original signatures in frequency domain to obtain their time domain representations. This is a preliminary step of the system, as shown in Figure 5.3. Figure 5.4 illustrates three arbitrary signatures from the THz database both in frequency domain and in time domain.

Next we describe how we use the signatures to create the signals necessary to train and test the system.



(a) Original signatures



(b) Transformed signatures using IFFT

Figure 5.4: Arbitrary samples of a) Original THz signatures in frequency domain from the database; and b) derived time domain signatures.

5.4.2 Training the System

Let $Y_1, \ldots, Y_N$ be a set of THz signatures, and let one of them, say Y_p, be the signature corresponding to a particular material or chemical that we are interested in detecting in a sensed signal x. Let $y_1, \ldots, y_p, \ldots, y_N$ be the corresponding time domain signatures. These signatures are inputs for the training module shown in Figure 5.3. The output of this module are classification parameters later necessary for testing.

The training module uses the time signatures to create training mixed signals as the following linear combinations:

$$x = c_1 y_1 + c_2 y_2 + \cdots + c_p y_p + \cdots + c_N y_N \tag{5.1}$$

The mixing coefficients c_i represent the intensities that each individual signature y_i has in the mixed signals. In other words, the training signals simulate sensed signals we would get from a real THz sensor. So, we must create two classes of training signals. The first class corresponds to signals containing the particular signature y_p, i.e. class "present". To ensure that signature y_p is present in the mixture, the corresponding coefficient c_p has to be non-zero. As we did for the sound signal tests presented in chapter 4, we employed a lower limit (ll) and an upper limit (ul) that satisfy $0 < ll \leq ul \leq 1$. Then, the coefficient c_p is randomly chosen from a uniform distribution defined over the interval [ll, ul]. The remaining coefficients are randomly chosen from a uniform distribution over the unit interval [0, 1] and then normalized so that $\sum_{i \neq p} c_i = 1 - c_p$. As for the sound signals, ll can be interpreted as a minimum percentage of the particular signature y_p present in the mixtures.

The second class of training signals corresponds to "not present", that is, signals not containing the particular signature y_p. Therefore, the corresponding coefficient c_p is set to zero. The other coefficients are once again randomly chosen from the unit interval and normalized as before.

Next we apply the GUSSS algorithm described in section 3.1 to all training signals.

That is, in equation (3.2) we let $x_1 = x$ be the sensed signal and $s_p = y_p$ be the signature to be injected. The weights w_1 and w_p are set to 1 (we tried some other values, but there was no significant difference in the results). We obtain GUSSS ratios labeled "present" and "not present". Finally, we obtain the training parameter(s) that will be used later to classify the test mixture signals. When using the hard threshold approaches the parameters are simply the threshold. For the SVM approach, the training parameters are the support vectors (see sections 5.3.1 and 5.3.2).

5.4.3 Testing the System (for Classification)

As we mentioned earlier, the goal of the system is to detect the presence of specific materials among sensed mixtures. So, after training the system for the particular signatures that we want to investigate, we use the signatures for testing.

Given a test signal, we want to determine if a certain material (i.e. a certain signature) is present or not. As before, we assume that the test signal is a linear combination of arbitrary time signatures. As Figure 5.3 shows, one of the inputs of the GUSSS module is the test signal. The other input of this module is the signature of the desired material. In other words, if we want to determine if the particular material P_1 with signature y_{p_1} is present in the mixture signal, then we would inject a copy of y_{p_1} in the GUSSS module.

For the threshold-based approaches, the classification is achieved by direct comparison of the GUSSS ratio obtained for $y_{p_1}$and the learned threshold for y_{p_1}. For the SVM-based approach we calculate the logarithm of the testing ratio and provide it as input to an SVM classification function along with the learned SVM parameters. This function determines whether a given ratio's logarithm (and therefore the corresponding test mixture) belongs to class "present" or class "not present". Despite the decision reached by the system for material P_1, different materials P_i can be next tested for their presence or absence, too.

5.5 Concluding Remarks

In this chapter we propose a framework for material detection using the GUSSS method. As we have pointed out, we designed several experiments using real terahertz signatures of many materials, which we obtained from public databases. Further details on the experiments and the results obtained are presented in chapter 7. In the next chapter, we present another framework for using GUSSS, this time applied to robotic assistive technology.

Chapter 6

Robotic Assistive Technology

In this chapter we present a brief review of Robotic Assistive Technology, with special emphasis in the use of Electromyographic signals (EMG) for human-machine interaction (HCI). We discuss some pattern recognition techniques that can be applied to EMG signals, and then we present the proposed framework used to test GUSSS in this area. Specifically, we propose using sEMG signals generated by muscle activity to control a power wheelchair, but the same framework can be adapted to interface with computers, cell phones, etc. Our method allows using few sensors and simple classification schemes to detect and classify user movements such as hand movements, which can then be related to commands for the wheelchair.

6.1 Background on Assistive Technology

Robotic Assistive Technology (RAT) is a field that addresses the development of systems to assist people with different levels of impairments in carrying out routine activities [32]. People who have lost limbs or who have suffered strokes or spinal cord injuries, patients engaged in physical therapy, elder people and people with any kind of physical disabilities may benefit from advances in this field [32–34]. The range of assistive devices that the technological advances have made possible is very wide. Research in this field is multi-disciplinary and very active. Even though a lot of advances have been made, there is constant need of finding better, faster, more reliable, more

adaptable, user-friendlier and less expensive solutions.

In a more wider sense, Assistive Technology (AT) devices may be categorized under the following types [34]: *wearable devices*, which may seek to monitor a person's status or activities, or which may provide some active support; *ambient devices*, which should provide unobtrusive services; *audio*, *video, GPS* units; *robotics*, which may be in the form of robot suits for mobility aids, mobile robots like intelligent wheelchairs, or prosthetic limbs. AT devices may also be categorized based on generic applications [34]: *Event Driven*, which encompass systems that are only required to react in case of an event or accident (e.g. falls, fire, security breaches, etc.); *continuous*, which refer to systems acting to aid assisted people in carrying out activities throughout their daily lives (e.g. walking aids, communication, cognitive training, environmental control); *trend analysis*, including systems for general monitoring of patients (vital signs, medication, food intake, hygiene), and for location tracking/occupancy.

Stroke is a major cause of acquired adult disability. Motor impairment is perhaps the most serious consequence of a stroke. Therefore, the restoring the walking ability of these patients is the most important objective in post stroke rehabilitation [35]. There is a lot of work on developing and evaluating AT systems for stroke patients. In [35] a biofeedback cycling training system to improve gait symmetry was investigated. A motion controlled gait enhancing mobile Shoe for rehabilitation was designed and tested in [36]. There is also a lot of work focusing on upper limb rehabilitation of post stroke patients. For example, [37] presents a haptic robot for hand movement therapy, and [38] discusses effector force requirements to enable robotic systems to provide assisted exercise for those patients.

Physical rehabilitation for people who have suffered an injury or have had surgery is essential to recover normal function for daily activities. And assistive technology can help in that process. For instance, [39] presents a knee orthopaedic device to illustrate how robotic technology can improve the outcome in knee rehabilitation. Unfortunately, a lot of people do not have the possibility to fully recover. For instance, people with degenerative diseases like progressive muscle dystrophy, Multiple Sclerosis (MS), Alzheimer's and Parkinson's diseases; elder people with weakened muscles or

degenerative joint diseases; blind, deaf or amputees, etc. There is a need for devices and technologies to assist them. There are a lot of commercial devices available today, but research continues. To illustrate, in [40] an electric lifting chair meant to aid in standing up and sitting down is assessed; [41] presents a robot assisted rehabilitation protocol designed to treat cerebellar and motor symptoms in subjects with MS; [42] introduces an exoskeletal meal assistance system for progressive Muscle dystrophy patients; and [43] studies haptic and auditory interfaces that can help blind people to interact with computers.

Intelligent AT systems and devices rely on different kinds of signals or features coming from the environment or from the human users themselves. Temperature, velocity, acceleration, orientation, force, torque and many other quantities are detected and used. Audio and video signals are captured and analyzed, too. Some other signals such as electrical signals generated by brain, cardiac or muscle activity are also used for evaluating patients and for controlling AT devices. The method presented in this thesis was applied to those electrical signals, specifically the ones generated by muscles. In the next section we present a more specific background on them.

6.2 Background on EMG

Electromyographic signals collected at the surface of the skin (sEMG) have been used in many applications, including rehabilitation, prosthesis, computer interfacing, wheelchair control, etc. [44–46]. When it comes to rehabilitation, more specifically for power wheelchair control, EMG signals have been often used as on/off switches. In those cases, menu driven approaches [33], finite state machines [47], and combinations of multiple muscles and sensors [48] are common techniques employed to provide multiple dimensions in the operation to the interface.

Several sEMG-based systems that rely on more elaborated pattern recognition of EMG signatures have also been proposed, including for exoskeleton robotics [49], and they vary widely in terms of: the classification approach employed; the feature selection criteria; and the number of sensors used. But, again, they constantly require multiple

sensors and much investment on the feature extraction and classification algorithms. Our goal in this work is instead to present a much simpler and yet effective technique using a single EMG sensor, allowing for other muscles and sensors to be used in other interfaces or to add modalities of operation to the interface.

As we mentioned above, sEMG-based systems vary widely and in terms of the classification algorithm, Artificial Neural Networks (ANN) [50–52], Fuzzy Logic and Fuzzy Control systems [4,51], are possibly the most common methods used to classify muscle activity – i.e. classify motor unit action potentials trains (MUAPT). The ability to recognize MUAPT can be applied, for example, to hand gesture recognition, control of electro-mechanical prosthesis, computer mouse movement, etc. [46]. One such example can be found in [51], where an ANN was compared to a Fuzzy Inference System (FIS) for classification in a hand prosthesis control. In this work, the authors concluded that for their application the best performance was achieved using the FIS classifier, with an 83% accuracy.

In another work presented in [52], several techniques for classification were employed in order to identify hand gestures using sEMG signals extracted from the forearm of human subjects. The authors reported good performance using ANN, Random Forest (RF), 1-Nearest-Neighbor (1NN), Support Vector Machine (SVM), Decision Tree (DT) and Decision Tree with Boosting (DT/B) as some of the different classification techniques used. In that case, the ANN approach presented a better performance than the other methods.

In terms of feature selection, the features can be extracted from the *time* or the *time-frequency* domains [46, 50, 51]. These features typically include: number of Zero Crossings (ZC), Mean Absolute Value (MAV), Slope Sign Changes (SSC), coefficients of Auto-regressive models (AR) [50,51]; Absolute Maximum/Minimum, Maximum minus Minimum, Median Value (Med), Variance, Waveform Length (WL) [46]; coefficients of the Short Time Fourier Transform (STFT) [46]; Wavelets Transform (WT) [45, 46], etc.

Given the wide range of features and their large dimensionality, many systems also employ dimensionality reduction techniques to the set of features. In those cases,

Class Separability (CS), Principal Component Analysis (PCA), Analysis of Variance (ANOVA) or Multivariate ANOVA (MANOVA) are the techniques frequently used. In [51], for example, the authors developed a feature selection that employed CS and PCA for dimensionality reduction. In that system, as well as in [52] where ANOVA was the technique of choice, the main concern is always to reduce dimensionality without affecting the classification in a significant manner.

Finally, in terms of number of sensors used, as far as we know all systems developed to date have made use of two or more sEMG signals derived from multiple sensors. For example, in [51], the authors reported using only two differential sEMG electrodes placed on the forearm of the test subjects. As we mentioned earlier, their system used multiple features and a FIS+PCA classifier to achieve 83% accuracy. A better performance (93.3%) was obtained in [52], but with the cost of relying on more sensors – 5 to be more specific – and using ANN as the classification algorithm.

As it can be inferred from the literature review, the use of multiple electrodes and of sophisticated classification algorithms help coping with a major disadvantage of surface EMG: the occurrence of cross-talk from adjacent muscles [44]. It is exactly this cross-talk of MUAPTs that makes the use of a single sensor a quite challenging problem.

Several systems have been proposed to control wheelchairs using EMG signals. To illustrate, in [53], the authors developed a wheelchair controller for users with high-level spinal cord injury. They used Fuzzy Min-Max Neural Networks to classify forward, left and right movements, and rest. And they measured EMG signals from muscles in each side of the neck. In [48] the authors used EMG signals from the neck and the arm muscles to implement a "joystick-like" model to control a wheelchair. And in [47] the authors presented a hands-free control system based on EMG signals recorded from eyebrow muscle activity for directional control, and electro-oculography signals detected from eye movements for speed control.

Once again, systems like the one above either rely on very complex processing and classification algorithms, or use the EMG signals mainly as on/off switches. Moreover, they seek patterns based on the simple presence and the frequency of the activation.

In this thesis, we propose a system for operating a wheelchair that recognizes muscle movements derived from hand movements or gestures. In our framework, we use GUSSS on the *"cross-talked"* MUAPTs signals from a single sEMG electrode. Unlike other methods based on ICA [4], our method relies on a single sEMG source. Finally, our proposed method combined with a simple distance classifier was applied to the control of a power wheelchair using three different hand gestures.

6.3 Proposed Method for Using GUSSS on sEMG

As we just mentioned, we propose a wheelchair control system based on the recognition of hand gestures using GUSSS. The use of hand gestures was simply to illustrate the fact that any EMG pattern or signature derived from a natural and repetitive muscle activation can be employed by our system. Clearly, in the case of a person with severe impairment, another natural movement such as an eyebrow or eyelid could be used instead.

The proposed framework for our method is illustrated in Figure 6.1 and it consists of three parts: 1) signal detection and acquisition; 2) feature extraction and classification; 3) command transmission to the wheelchair and execution.

As the name implies, the first module of our framework is responsible for detecting and sampling the sEMG signal. The next module in the flow, as presented in Figure 6.1, extracts the features that are used for classification of the hand movements. These features are the Mean Absolute Value (MAV) and the GUSSS ratio, which we introduced in chapter 3. After the features have been extracted the classification takes place. In order to demonstrate the power of the proposed GUSSS ratio as a classification feature, we based our system on a simple distance classifier. A better classifier using more features should lead to better accuracy and classification. The last module in Figure 6.1 transmits a command to the wheelchair based on the output of the classifier.

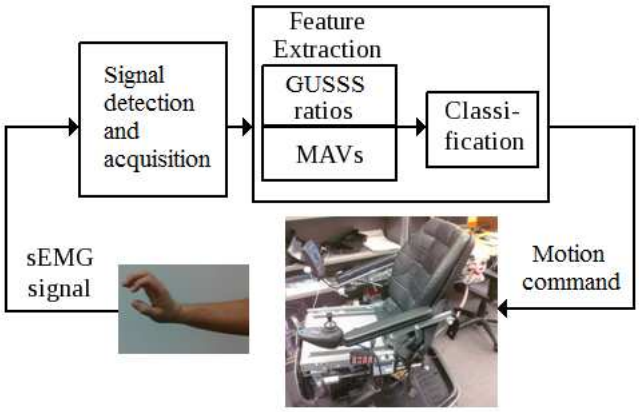


Figure 6.1: Framework of the proposed classification system.

6.3.1 Identifying Multiple Signatures in a Sensed Signal

In chapter 3 we explained how a particular signature can be identified or separated from a sensed signal, say x_1. Recall from equation 3.1 that this sensed signal is considered a linear combination of N independent components. In this case, the components are electrical signals (MUAPT) originating from different muscles [44]. In other words, as we mentioned earlier, the sensed signals are linear combinations of independent MUAPTs that become mixed due to cross-talk inside the subject's arm.

In order to identify the presence or not of all possible signatures, the framework employs an iterative method. That is, first, we assume that the system needs to identify N sEMG signatures corresponding to the N possible hand gestures (We will explain how to obtain the signatures in Chapter 8). Next, from the test signal x_1, we obtained N ratios by injecting iteratively the desired signature into x_1 – equations (3.2)-(3.5) . That is, we find

$$x_p = x_1 + s_p \qquad for\ p = 1\ to\ N$$

and once again, we apply the ICA algorithm to each

$$X_p = \begin{bmatrix} x_1^T \\ x_p^T \end{bmatrix} \qquad for\ p = 1\ to\ N$$

to obtain the GUSSS ratios r_1, r_2, ..., r_N. Finally, it should go without saying that if r_i is the smallest of the N ratios found by the GUSSS, it is likely that the sensed signal x_1 is the signature s_i, and thus, the hand gesture i is the one being sought.

6.3.2 Mean Absolute Value as a Classification Feature

We considered a second feature for the classifier: the Mean Absolute Value (MAV) of the signals. The MAV of a signal $x(t)$ is obtained by calculating the average of the absolute values of x at all instants t. If $x(t)$ is continuous in time, then

$$MAV = \frac{1}{T}\int_T |x(t)|\, dt$$

where T is the time interval for which $x(t)$ is defined. If the signal is discrete, then

$$MAV = \frac{1}{K}\sum_{k=1}^{K} |x(k)|$$

where K is the number of samples that constitute $x(k)$.

6.3.3 Classification Module

As we pointed out earlier, the goal of the GUSSS is to identify which signature is present in the observed sEMG signal x_1. In order to do so, the same signature must be injected to synthesize a secondary signal x_p. Our framework uses a training set of sEMG signals to learn those signatures and in chapter 8 we will explain two approaches used for this purpose. Here, we assume that the signatures are available.

Furthermore, from the training signals we also learn the average MAVs for the different signatures, i.e. for the different gestures or classes of gestures. In mathematical terms, let μ_1, μ_2, ..., μ_N be the average MAVs obtained from the training set and corresponding to N different gestures to be recognized. Let σ_1, σ_2, ..., σ_N the corresponding standard deviations. Given the input signal x_1, the algorithm calculates its MAV, m_1,

and based on this value, it computes the Mahalanobis distances to the average MAVs of the gestures. That is:

$$d_p = \frac{|m_1 - \mu_p|}{\sigma_p} \qquad for\ p = 1\ to\ N$$

It should be noted that if x_1 is the result of gesture i, it is likely that m_i is similar to μ_i. In that case, d_i would also be the smallest of the N distances above.

Using both the GUSSS ratios and the MAVs distances above, we can define the distance classifier. The only missing step is the normalization of both features. That is, we define the normalized GUSSS ratio and the normalized MAV distance as, respectively:

$$\bar{r}_p = \frac{r_p}{\sum_{j=1}^{N} r_j}$$

$$\bar{d}_p = \frac{d_p}{\sum_{j=1}^{N} d_j}$$

Those features are grouped in the feature vector:

$$\overrightarrow{v_p} = \begin{bmatrix} \bar{r}_p \\ \bar{d}_p \end{bmatrix}$$

for $p = 1, \ldots, N$, corresponding to each of the N gestures to be identified. The classification is obtained by assigning x_1 to that gesture (i.e. the class) for which the corresponding feature vector $\overrightarrow{v_i}$ is smallest. The reason for the normalization of the ratios and distances is, of course, to allow both features to have the same weight in the classification process.

6.4 Concluding Remarks

In this chapter we propose a framework for using GUSSS to classify sEMG signals. We developed an interface to control a real wheelchair using different hand gestures. Further details of the experiments and the results obtained are presented in chapter 8.

Chapter 7

Experimental Results in Material Detection Using Terahertz

In this section we present experimental results of the proposed system for material detection described in chapter 5. The THz signatures for our experiments were obtained from two public databases [25, 26] containing THz signals for hundreds of different materials or chemical compounds. We ran 18 tests, each divided into 11 different sub-cases representing different combinations of materials from the databases. For each of the 18 tests, we selected one material to be the target material – i.e. the one to be detected. We used its signature and the signatures of 7 additional materials to train the system and obtain the training parameters. Then, we created testing mixtures using the same 8 THz signatures plus 2 additional ones. We introduce those two new THz signatures to simulate situations when the target material is mixed with materials for which no training had been provided – i.e. these 2 extra materials had never been seen by the system before the tests.

After selecting the compounds for each of the 18 tests, we created 11 different sub-cases representing different potential concentrations of the target compound in the total mixture. We did that by randomly selecting the concentration of that target material between a certain lower limit, ll, and an upper limit, ul. Finally, for each of the 11 sub-cases in each test, we created 1000 mixture signals for training and another 1000 for testing. Half of the mixtures corresponded to cases when the target material

is present, and the other half to cases when the material is not present. As explained in section 5.4.2, the mixtures were created as linear combinations (eq. 4.2) of all 10 THz signatures in the test.

In section 5.3, we explained two ways of getting hard thresholds, and we discussed SVM. For all the tests we tried the thresholds and the SVM approaches for training the system and classifying testing mixture signals. The results obtained were similar, but the best overall performance was obtained using the SVM approach. Next I present the results from three of the tests. The complete set of tests is presented in Appendix B.

7.1 Test #1: Manzeb

Our first test consisted of detecting Manzeb among 9 other compounds (7 of which had been used during training and 2 had never been seen before this test). Table 7.1 shows the results obtained for this Test #1. The Table presents the results in the form of True Positive (TP) and True Negative (TN) percentages, as well as total correct classification percentages (CC). These results were obtained using the SVM approach.

The reader should notice that the classification performance is excellent for this compound. Only the cases with the lowest lower limit (ll) values resulted in less than perfect classification. The reason for that is mainly due to the distinct THz signature associated with this compound. Figure 7.1 shows the original THz signatures obtained from the databases for this test (left) and the transformed time domain signatures (right) used by our system. The first two plots (top) correspond to the Manzeb, while the remaining plots correspond to the other materials used in this experiment.

Another way to understand the excellent results of this test is provided by Figure 7.2, which illustrates the behavior of the GUSSS ratios for randomly selected values of the lower limit ll, ranging from 0.1 to 0.8. The figure depicts in red and blue the distributions of the logarithm of the GUSSS ratios corresponding respectively to the cases when Manzeb is present and when it is not present. For this specific plot, we set the upper limit (ul) equal to the lower limit (ll) in order to force the coefficient c_p

to be equal to that specific ll. Other than that, the training signals were created as described in section 5.4.2, using the THz signatures mentioned in table 7.1. From the figure, we can notice that the GUSSS ratios for the cases when the Manzeb is not present remain approximately within the same range, independent of ll – i.e. the averages of the logarithm of those GUSSS ratios are approximately constant throughout the plot. However, for the cases when Manzeb is present in the mixtures, the GUSSS ratios are smaller, and their values decrease further with the increase in the value of ll – i.e. with increases in the concentration of Manzeb in the mixture. Finally, the figure shows that the separation between the set of "not present" cases and the set of "present" cases increases as ll increases, which again explains the classification performance shown in Table 7.1.

As the discussion above already indicated, the distinct characteristics of a THz signature vis-a-vis the other signatures in the mixture leads to almost perfect performance by our proposed method. Also, the percentage of the target material in the mixture (given by the lower ll and upper ul limits) and the separation of the two classes "present" and "not-present" are also important factors in the performance of our system. In the next two cases, we present the good performance achieved by our system despite the relaxation of these two conditions.

Table 7.1: Target material: Manzeb. Training materials: ADP, Albumin, DAST, H_2S, HgS, Lumiflavin, MgO. Additional materials for testing: Glucose, SiN. All values are percentages.

[ll, ul]	TP	TN	CC
[0.10, 0.40]	99	99	99
[0.10, 0.75]	99	99	99
[0.20, 0.40]	100	100	100
[0.20, 0.75]	100	100	100
[0.30, 0.40]	100	100	100
[0.30, 0.75]	100	100	100
[0.50, 0.75]	100	100	100
[0.50, 0.95]	100	100	100
[0.70, 0.75]	100	100	100
[0.70, 0.95]	100	100	100
[0.80, 1.00]	100	100	100
Total	99	99	99

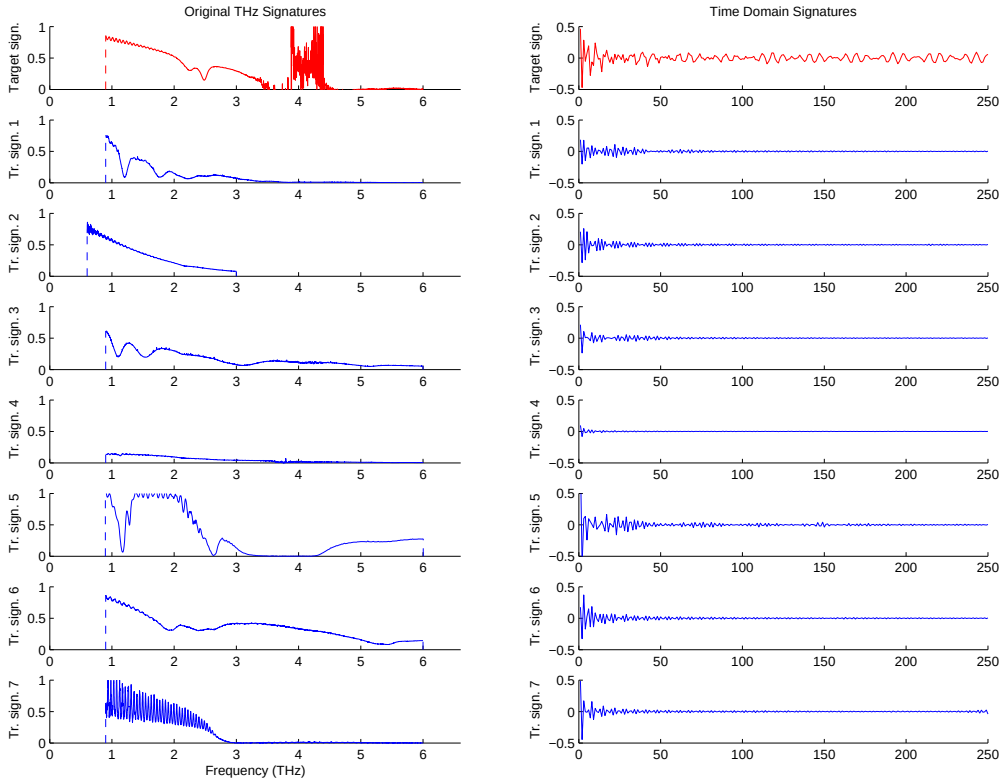


Figure 7.1: In red, the signature of Manzeb (target material). In blue, the additional signatures used for training. On the left: Frequency domain. On the right: Time domain.

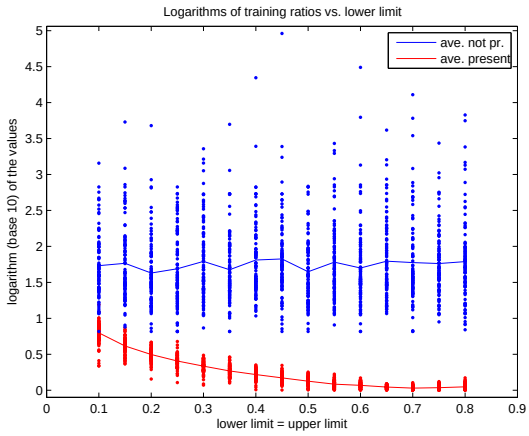


Figure 7.2: Distribution of the logarithm of GUSSS ratios vs. lower limit (ll). Shown are the GUSSS ratios for cases when the target material is present (red) and not present (blue). The upper limits (ul) were set equal to the ll. Also shown by the continuous lines are the averages of the logarithm of the GUSSS ratios. Target material: Manzeb.

7.2 Test #2: Talc

The second of the 18 tests performed used Talc as the target material. Table 7.2 shows the result obtained for this test, which was carried out using all three approaches for the classifier: that is, the SVM and the two threshold based methods. Overall, the best performance was once again achieved using the SVM-based classifier.

The results obtained for Test #2 are still very good, but not quite as good as in Test #1. Figure 7.3 shows the THz signatures and the corresponding time domain signatures for all materials used in this test. As before, the THz signature of the Talc in time domain seems mostly different from the signatures of the other materials, but it retains some similar traces of these signatures. Those similarities should explain the test results being slightly worse than the test using Manzeb.

Similar to the previous test, Figure 7.4 illustrates the logarithm of GUSSS ratios versus various ll values. However, in this test, one can notice that the GUSSS ratios of the two classes, present and not present, are closer when compared to the test using Manzeb. Still, for the "present" class, the averages of the logarithm of the GUSSS

ratios decrease as ll increase, leading to a better classification as the percentage of Talc in the mixture increases. In fact, according to this plot, the classification should be close to perfect for ll values greater than 0.5, which is exactly the case in Table 7.2.

Table 7.2: Target material: Talc. Training materials: Baking Soda, Caffeine, Cellulose, Chalk, Coffee Ground, Equal, O_2Si. Additional materials for testing: Ovalbumin, Sodium Chloride. All values are percentages.

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	66	83	75
[0.10, 0.75]	69	88	79
[0.20, 0.40]	77	82	80
[0.20, 0.75]	79	89	84
[0.30, 0.40]	82	81	82
[0.30, 0.75]	85	92	89
[0.50, 0.75]	94	95	95
[0.50, 0.95]	95	97	96
[0.70, 0.75]	98	99	98
[0.70, 0.95]	98	99	99
[0.80, 1.00]	99	99	99
Total	86	91	89

(a) SVM approach

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	69	71	70
[0.10, 0.75]	75	82	79
[0.20, 0.40]	76	76	76
[0.20, 0.75]	88	84	86
[0.30, 0.40]	81	78	80
[0.30, 0.75]	89	83	86
[0.50, 0.75]	97	84	90
[0.50, 0.95]	98	94	96
[0.70, 0.75]	99	89	94
[0.70, 0.95]	98	95	97
[0.80, 1.00]	100	97	98
Total	88	85	87

(b) Threshold approach, Geometric Mean

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	85	48	66
[0.10, 0.75]	89	63	76
[0.20, 0.40]	87	60	73
[0.20, 0.75]	91	74	82
[0.30, 0.40]	89	66	77
[0.30, 0.75]	94	80	87
[0.50, 0.75]	97	91	94
[0.50, 0.95]	95	95	95
[0.70, 0.75]	99	98	98
[0.70, 0.95]	99	99	99
[0.80, 1.00]	99	99	99
Total	93	79	86

(c) Threshold approach, Intersection of Gaussians

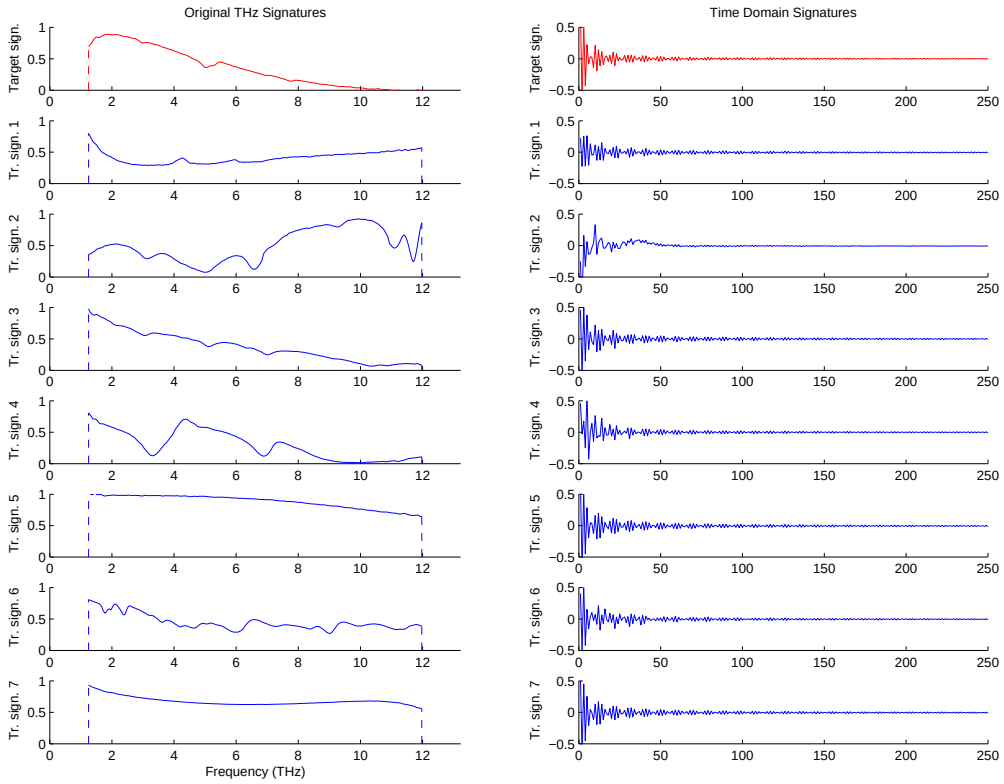


Figure 7.3: In red, the signature of Talc (target material). In blue, the additional signatures used for training. On the left: Frequency domain. On the right: Time domain.

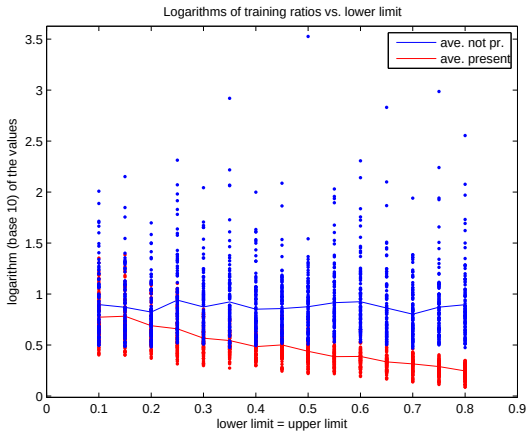


Figure 7.4: Distribution of the logarithm of GUSSS ratios vs. lower limit (ll). Shown are the GUSSS ratios for cases when the target material is present (red) and not present (blue). The upper limits (ul) were set equal to the ll. Also shown by the continuous lines are the averages of the logarithm of the GUSSS ratios. Target material: Talc.

7.3 Test #3: Baking Soda

Lastly, we now present in table 7.3 the results for the third of the 18 tests performed. The material detected this time is Baking Soda. Once again, we show the results for the threshold-based as well as the SVM-based approaches for the classifier. The classification performance for this test was the lowest among all tests. As before, Figure 7.5 shows the THz signatures in frequency and time domain for all the materials used in this test. However, as it can be observed from Figure 7.5, this time the time signatures seem very similar to each other, which explains the results of this last test as being the worst among all tests.

It is important to mention here that despite this test presented the worst results, such results are still very reasonable, especially for ll values higher than 0.5. Besides, our classifier is a very simple classifier when compared to typical system in the literature, and the performance of our system could be greatly enhanced by the inclusion of more features or the use of an even better classifier. Our goal in this work, however, was to show the excellent performance of our GUSSS method, even for simple classifiers.

As before, we show a plot of the logarithm of the GUSSS ratios versus ll in Figure 7.6. Here, it should be pointed out the much greater overlap of the two classes, "present" and "not-present", when compared with the previous two tests. This is consistent with the classification performance obtained.

Table 7.3: Target material: Baking Soda. Training materials: Caffeine, Cellulose, Coffee Ground, O_2Si, Ovalbumin, Sodium Chloride, Talc. Additional materials for testing: Equal, Trehalose. All values are percentages.

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	58	80	69
[0.10, 0.75]	67	85	76
[0.20, 0.40]	62	84	73
[0.20, 0.75]	65	87	76
[0.30, 0.40]	71	83	77
[0.30, 0.75]	75	92	83
[0.50, 0.75]	81	90	85
[0.50, 0.95]	80	88	84
[0.70, 0.75]	81	93	87
[0.70, 0.95]	81	87	84
[0.80, 1.00]	77	89	83
Total	73	87	80

(a) SVM approach

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	48	87	68
[0.10, 0.75]	60	93	77
[0.20, 0.40]	52	91	72
[0.20, 0.75]	65	95	80
[0.30, 0.40]	59	91	75
[0.30, 0.75]	72	95	84
[0.50, 0.75]	81	95	88
[0.50, 0.95]	80	97	88
[0.70, 0.75]	81	95	88
[0.70, 0.95]	80	95	87
[0.80, 1.00]	76	94	85
Total	69	94	81

(b) Threshold approach, Geometric Mean

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	67	72	70
[0.10, 0.75]	72	87	80
[0.20, 0.40]	71	74	72
[0.20, 0.75]	76	87	81
[0.30, 0.40]	71	86	78
[0.30, 0.75]	74	91	82
[0.50, 0.75]	80	96	88
[0.50, 0.95]	83	91	87
[0.70, 0.75]	80	96	88
[0.70, 0.95]	84	91	87
[0.80, 1.00]	88	88	88
Total	77	87	82

(c) Threshold approach, Intersection of Gaussians

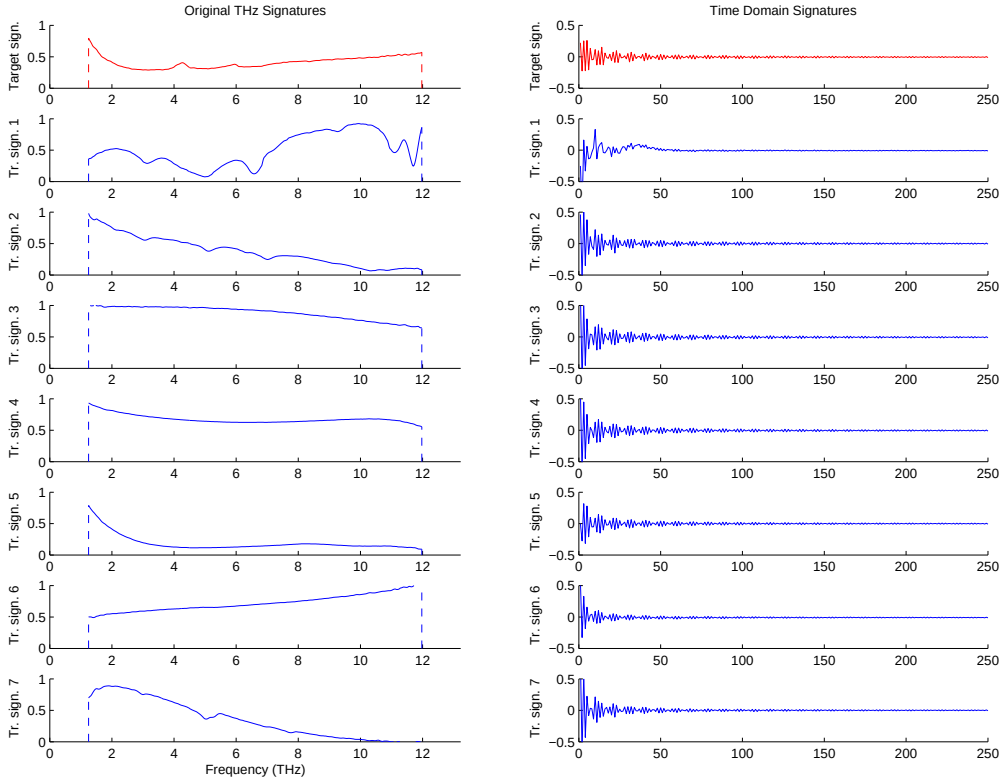


Figure 7.5: In red, the signature of Baking Soda (target material). In blue, the additional signatures used for training. On the left: Frequency domain. On the right: Time domain.

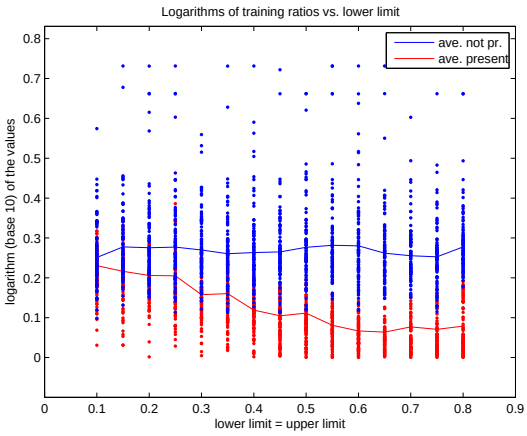


Figure 7.6: Distribution of the logarithm of GUSSS ratios vs. lower limit (ll). Shown are the GUSSS ratios for cases when the target material is present (red) and not present (blue). The upper limits (ul) were set equal to the ll. Also shown by the continuous lines are the averages of the logarithm of the GUSSS ratios. Target material: Baking Soda.

7.4 Overall Results and Discussion

As we mentioned before, we ran a total of 18 tests. Test #1 above corresponds to the material for which we got the best performance (Manzeb). Test #2 corresponds to a material with an average performance (Talc), while the third test shows the worst overall performance (Baking Soda) among the 18 tests. As mentioned before, we do not present the results for the remaining 15 tests in this chapter, but in Appendix B. Nevertheless, we do present the average performance of all 18 tests in Table 7.4, next.

Table 7.4: These results are the average percentages of all the experiments (18 materials).

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	74	80	77
[0.10, 0.75]	81	87	84
[0.20, 0.40]	82	83	83
[0.20, 0.75]	86	89	88
[0.30, 0.40]	86	87	86
[0.30, 0.75]	92	92	92
[0.50, 0.75]	95	94	94
[0.50, 0.95]	94	94	94
[0.70, 0.75]	95	96	95
[0.70, 0.95]	94	95	94
[0.80, 1.00]	94	95	94
Total	89	90	89

(a) SVM approach

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	79	74	76
[0.10, 0.75]	84	83	83
[0.20, 0.40]	85	78	81
[0.20, 0.75]	90	84	87
[0.30, 0.40]	89	80	85
[0.30, 0.75]	93	87	90
[0.50, 0.75]	96	88	92
[0.50, 0.95]	95	90	92
[0.70, 0.75]	97	88	92
[0.70, 0.95]	96	89	92
[0.80, 1.00]	95	89	92
Total	91	84	88

(b) Threshold approach, Geometric Mean

$[ll, ul]$	TP	TN	CC
[0.10, 0.40]	86	61	73
[0.10, 0.75]	88	75	82
[0.20, 0.40]	90	70	80
[0.20, 0.75]	92	81	86
[0.30, 0.40]	92	77	84
[0.30, 0.75]	93	86	90
[0.50, 0.75]	95	91	93
[0.50, 0.95]	95	90	93
[0.70, 0.75]	95	92	94
[0.70, 0.95]	95	90	93
[0.80, 1.00]	96	91	93
Total	92	82	87

(c) Threshold approach, Intersection of Gaussians

In general, the threshold-based approaches led to a slightly higher index of True Positives (TP) – and thus, to lower False Negatives (FN) – when compared to the SVM-based approach. This result is quite desirable, especially for applications involving explosive detection (e.g. airport security inspection, IED detection, etc.). However,

the improvement in TP when compared to the SVM-based classifier came with a much bigger price in terms of True Negatives (TN) and the complemented False Positives (FP). So, the overall higher index of correct classification (CC) achieved with the SVM-based classifier would justify its use for most applications where better correct classification would be desired.

Another important observation to make is that the results show that in general the performance increases with higher values of the lower and upper limits, ll and ul. This makes perfect sense since those limits determine the concentration of the target material in the mixture being tested. In an extremely ideal case, when that target material is the only one present, the coefficient c_p of equation 4.2 would be 1 and the GUSSS ratio would be 1. In general, if the value of ll is high (close to 1), then the coefficient c_p would also be close to 1 and the GUSSS ratio would be relatively low. On the other hand, when the target material is not present in the mixture, the coefficient c_p becomes close to 0 and the GUSSS ratio tends to infinity. Similarly, if the value of ll is low (close to 0), then c_p would also be low for a given mixture, and thus, the corresponding GUSSS ratio would be large. Based on the previous analysis, it was expected that the best classification performance resulted with the higher values of ll and ul, which was confirmed by the experiments.

The mixing coefficients c_i used to create the testing mixture signals were randomly chosen. They represent how much of the i^{th} material is present in a mixture. By examining those coefficients we gained some insights on the results obtained. Specifically, we want to better understand the miss classifications. On the one hand we have the cases when the material that we wanted to identify was present in a mixture signal, but the system determined that that signal came from a mixture not including the material sought for. These false negative (FN) cases may be due to lower concentrations of the particular material.

On the other hand we have the false positive (FP) cases. In a FP case the system classifies a testing signal as containing the material under investigation, when in fact it was not present. This cases may be due to strong presence of materials with similar time signatures (to the particular material being investigated) in a mixture. As shown

in figures 7.3 and 7.5, there are materials with similar time signatures, which may be confused in a certain degree. So, in order to confirm this hypothesis we examined the mixing coefficients c_i, $i \neq p$, which determine how much of each material different than the material investigated is present in a mixture (since the material investigated is not present, $c_p = 0$). We found out that the larger coefficients of an FP test signal corresponded to those materials which appear to have more similar signatures to the signature of the material of interest. For instance, in the case of material Talc (see Figure 7.3, right hand), Tr. signatures 3, 5 and 7 seem very similar to the signature of Talc (in red). In general, those signatures had the largest coefficients in the testing signals that ended up being FP cases. In contrast, for the true negative (TN) cases (those when a test mixture is correctly classified as not containing the material of interest) the larger coefficients corresponded to the time signatures more different than that of the material of interest, whereas the "similar" signatures usually had the lower coefficients.

7.5 Concluding Remarks

The results presented here and in Appendix B show a lot of promise for our method in the area of material detection. In chapters 9 and 10 we present general conclusions on the methods and experiments that we conducted for this work, and we comment on improvements and further tests to be conducted in the future.

Chapter 8

Experimental Results with sEMG Signals

In this section, we explain how we applied the proposed classification framework described in chapter 6. For these experiments we used three different hand gestures each time, i.e, we tested our method for classifying $N = 3$ possible movements. The sEMG signals of interest are those generated in the transition from a rest position of the hand to the actual gesture. All the sEMG signals were obtained using a Tinkertron EMG switch [54]. This device consists of circuitry for detection and amplification of sEMG signals. Since our framework relies on a single sEMG source, we placed a pair of differential electrodes on the extensor carpi radialis muscle along the subject's forearm. The use of this muscle has been previously reported [50, 51]. We choose this muscle for convenience and easy access. A reference (ground) electrode was also placed on the wrist of the opposite arm.

Before implementing the wheelchair control illustrated in Figure 6.1, we first did offline testings of the method. We described the tests and the results in section 8.1. Then, in section 8.2 we present the results obtained for the actual wheelchair. We conclude this chapter with a discussion of the results, the main problems we had, and possible improvements for the framework.

8.1 Offline Testing Using National Instruments Digitizer

The main goal of the experiments was to test our proposed GUSSS method for classification of sEMG signals. Having an interface to control a wheelchair (or any other robot or system) implies additional challenges, mostly technical ones. Therefore, the first tests we ran did not involve the wheelchair. We collected sEMG signals offline and under more controlled conditions. We processed and analyzed those signals to get classification rates, leaving out the last part of the framework presented in section 6.3 (command transmission to the wheelchair and execution). A simplified framework for the method is illustrated in Figure 8.1 and it consists of three parts: 1) signal acquisition and pre-processing; 2) feature extraction; and 3) classification. The features used and the classification module implemented are as described in chapter 6, section 6.3.

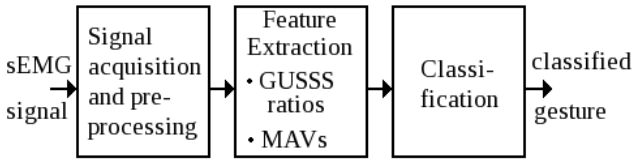


Figure 8.1: Simplified classification system.

For these offline experiments, the signals from the Tinkertron were sampled using a National Instruments digitizer. A typical signal recorded looks like the one in Figure 8.2. We need to process the signal before we can use it. First of all, in order to reduce undesired noise coming from the power lines, we implemented a digital filter to remove the 60 Hz component. Then, we need to extract the actual sEMG signal generated when the subject performs the hand movement. Figure 8.2 shows an initial period of time where there is only noise. Then comes the main part of the signal, the one of interest. Finally, there is another period of time with noise. The pre-processing module shown in Figure 8.1 consists of detecting where the main signal is so it can be

extracted and then used in the next module. The details of the extraction algorithm are presented in Appendix C.

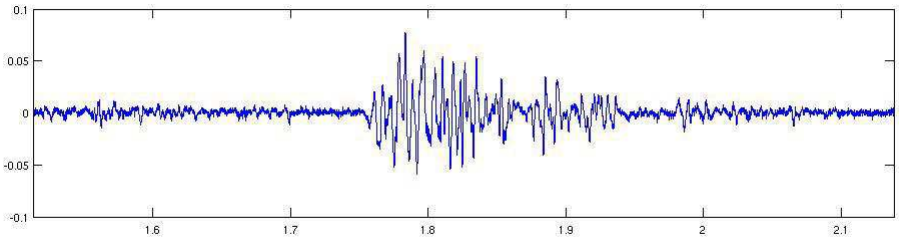


Figure 8.2: Typical sEMG signal captured using a National Instruments Digitizer.

8.1.1 Training: Obtaining the Signature Signals and Other Parameters

As we mentioned earlier, three different hand gestures were considered in our experiments. As we have also explained in previous chapters, the proposed GUSSS technique requires a signature associated to each of the gestures to be recognized. Before we present the performance of our framework, we explain how we learn the signatures.

Signatures by Averaging Given a training set with $3 \times J$ samples – i.e. J samples from each hand gesture, we did an averaging of the training signals grouped per hand gesture. That is, each of the J samples belonging to the same gesture were averaged creating a single time signal $s_p(t) = \frac{1}{J}\sum_{i=1}^{J} y_{pi}(t)$ for $p = 1, 2, 3$.

Other Training Parameters As explained in chapter 6, section 6.3, the classification module also needs the average MAVs and the standard deviations. All of them are obtained by analyzing the set of training sample signals for each of the gestures. Once the signatures, the mean average MAVs and the corresponding standard deviations are learnt, an additional set of testing signals can be analyzed.

It is important to mention that the training process is subject dependent. The signatures, the average MAVs and the corresponding standard deviations would need to be specifically calculated for each individual.

8.1.2 Experimental Results

Figure 8.3 illustrates the 3 hand gestures considered in the experiments. We collected sEMG signals from three individuals. We used a subset of the signals to obtain the signatures and the other parameters for each of the gestures. We then classified the remaining sEMG signals collected. Tables 8.1, 8.2, and 8.3 present the confusion matrices for the classification results.

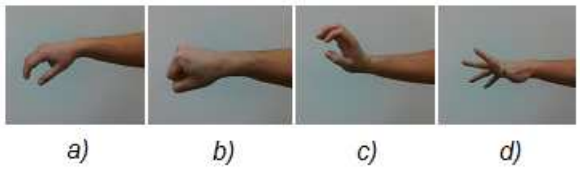


Figure 8.3: The hand gestures considered: a) Relax or resting position; b) "clench"; c) "up" ; d) "finger tapping".

Table 8.1: Confusion matrix for classification, test subject 1.

		Assigned gestures		
		clench	up	tapping
Real gestures	clench	35	5	0
	up	3	32	5
	tapping	1	8	31
Correct classification: 81.7%				

Table 8.2: Confusion matrix for classification, test subject 2.

		Assigned gestures		
		clench	up	tapping
Real gestures	clench	33	3	4
	up	3	30	7
	tapping	8	1	31
Correct classification: 78.3%				

Table 8.3: Confusion matrix for classification, test subject 3.

		Assigned gestures		
		clench	up	tapping
Real gestures	clench	29	2	9
	up	8	32	0
	tapping	4	0	36
Correct classification: 80.8%				

8.2 Controlling a Wheelchair

After testing our method offline, we developed a program to control a wheelchair using the sEMG signals. Now, instead of using the National Instruments digitizer, we used a TS-7250 embedded device from Technologic Systems to sample the signals from the Tinkertron EMG switch. This is part of the first software module in Figure 6.1 – Signal Detection and Acquisition. This module is constantly monitoring the sEMG signals, waiting for their levels to cross a certain threshold. Once the threshold is detected, the program stores the signal and then transmits the signal to the second software module of the system for the purpose of feature extraction and classification. This second module then sends the commands to the wheelchair, which then sends the control signals to the wheelchair motors to actually move forward, turn or stop it. It should be mentioned here that the two modules described above were implemented in a client-server fashion and each can run on the same computer or not. In this experiment,

the client – i.e. the Signal Detection and Acquisition module – ran on the TS-7250 as we just mentioned, while the server – i.e. the Feature Extraction and Classification module – ran on a desktop computer. The original programs for the offline testing (and for the sound and THz tests) were created using Matlab. For the wheelchair control we needed to translate the programs to C. We experienced problems with the real time functions for the TS-7250. Therefore, we ended up not using real time functions of the embedded device.

As in the offline testings, three different hand gestures were considered in the experiments (Figure 8.3). As we have also explained, the proposed GUSSS technique requires a signature associated to each of the gestures to be recognized. Before the subjects can use the wheelchair they need to go through a training process. The system allows the user to repeat certain gestures a number of times and it associates each gesture to any of the possible motions of the chair. The training signals obtained are analyzed and processed to create the signatures, which are then stored in the server. The signatures are obtained as described in subsection 8.1.1.

8.2.1 Experimental Results

Once again, three test subjects were asked to go through the training process. For each subject, ten signals per movement were acquired in order to train the system. For these experiments, the "clench" gesture (Figure 8.3 b) was assigned to the stop command, the "up" gesture (Figure 8.3 c) was assigned to the forward command and the "finger tapping" (Figure 8.3 d) was assigned to the turn command. After the training, the test subjects were asked to perform a total of 150 movements, 50 for each command. Tables 8.4, 8.5 and 8.6 show the confusion matrices for the classification results of these experiments.

It is important to mention that the results presented below encompass all aspects of the real time control of the wheelchair. In other words, the system must capture the signal, localize the time window of muscle activity, extract the features and classify the gesture correctly in order to be considered a successful classification. Therefore, the accuracy of the classification calculated and presented on the tables can be affected by

errors in any of these steps.

Table 8.4: Confusion matrix for classification, test subject 1.

		Executed command		
		stop	forward	turn
Actual hand movement	clench	43	7	0
	up	7	37	6
	tapping	0	19	31
Correct classification: 74%				

Table 8.5: Confusion matrix for classification, test subject 2.

		Executed command		
		stop	forward	turn
Actual hand movement	clench	41	4	5
	up	10	28	12
	tapping	4	13	33
Correct classification: 68%				

Table 8.6: Confusion matrix for classification, test subject 3.

		Executed command		
		stop	forward	turn
Actual hand movement	clench	29	6	15
	up	1	32	17
	tapping	2	5	43
Correct classification: 69%				

8.3 Discussion and Concluding Remarks

The proposed classification framework was applied to a wheelchair control system based on detecting specific hand gestures, using a single sEMG source. The classification

accuracy obtained in these experiments was lower than the accuracy obtained offline – that is, using the National Instrument acquisition equipment. In that case, the accuracy was around 80%. We assume that the lower performance was caused by hardware limitations of the AD converter on-board the TS-7250 embedded device and the algorithm used to automatically locate the window containing the EMG signature for classification. However, since both of these conditions are required for a real-life application of our proposed GUSSS method to the wheelchair control, we thought necessary to report the tests above.

Despite the problems with the embedded device, the results achieved here were still quite reasonable if we take into account the use of fewer sensors, fewer features, and a very straightforward classification algorithm. Other systems found in the literature use multiple sEMG sources and elaborated algorithms for classification. The training process that we used for this experiment was also very simple to be carried out – another advantage of our method over other methods in the literature. A future improvement of the system would be to upgrade the embedded device and take advantage of real time features.

Chapter 9

Conclusion

This work presented a new method for pattern recognition, signal detection and separation called Guided Under-determined Source Signal Separation (GUSSS). The method relies on Independent Component Analysis and it is useful when there is a need of separating individual components from mixtures captured by only one sensor. Our method differs from Under-determined Blind Source Signal Separation in that we use prior knowledge of particular source signals (signatures) that might be present within the sensed mixture signals. By incorporating that prior information we go from an under-determined problem, where we have fewer sensed signals than components, to a well determined problem, where we have two equations and two components that we want to estimate. This allows the use of traditional ICA algorithms to estimate the independent components and the mixing coefficients. Using those coefficients we introduced the GUSSS ratio, which can be used to determine if a particular signature is present or not in a sensed signal, or as a feature in a classification problem.

We first tested our proposed method with sound signals and we were able to detect the individual source signals within mixtures as they would be captured by a microphone. The results were very promising, even with noisy signals. After the sound tests we applied our method in two very different areas: material detection using terahertz signatures and assistive technology using sEMG signals.

To apply GUSSS to THz technology we proposed a framework for detecting the presence of particular chemicals or materials in a mixture. We used THz signatures

of various materials obtained from public databases for our simulations. Our method presented very good results for detecting a variety of such materials under various assumptions on the percentage of the target material in the scene. For these experiments, we used three very simple classification schemes and only a single feature. Our goal was to demonstrate the superior performance of the method when it relies solely on the GUSSS technique.

Unlike other methods that emphasize the detection of individual compounds, the results obtained demonstrated that the proposed approach allows the detection of materials within mixtures. The results also support the claim that the introduction of GUSSS presents a very powerful tool for the separation and identification of patterns in signals. We expect that our proposed approach will lead to the development of systems capable of detecting hazardous materials, such as RDX, TNT, drugs, etc., enhancing and simplifying law enforcement, anti-terrorism security, etc.

For the area of assistive technology we focused on classifying sEMG signals, which can be used for controlling devices like a power wheelchair. In order to use GUSSS we proposed a framework to detect and classify user movements such as hand movements, which were then be related to commands for a wheelchair. Our method allowed using few sensors and a simple classification scheme. The performance of our method with sEMG signals was not as impressive as with the sound and THz signals. However, the results were reasonable given the technical limitations and the simplicity of the feature vectors and the classifier used.

Overall, this work shows that the GUSSS method has the potential to be a powerful method for pattern recognition of mixed signals.

Chapter 10

Future Work

Future work in the area of Terahertz technology would consist of applying our method to THz signatures of RDX and other hazardous materials. Despite having been tested with real THz signatures from the public databases, the proposed framework should also be tested in real world scenarios. Furthermore, a future implementation of the proposed framework should benefit from the use of a more diverse set of features and a more sophisticated classifier.

In the area of assistive technology, our method could be tested on other muscles. For instance, people with severe disabilities may only be able to move face muscles like eyebrows. So, the proposed framework could be used to learn and use patterns of the eyebrow movements to control the wheelchair. Another area of investigation is to define better ways of obtaining the signatures for the sEMG signals. The learning process of those signatures could be enhanced by incorporating information coming from more muscles. The inclusion of more features and the use of a better classification scheme should lead to better performances, as for the THz framework. This should allow discriminating more gestures or movements.

The method could also be applied to other types of signals, for example, Electroencephalographic (EEG) or Electrocardiographic (EKG) signals. EEG signals, for instance, are widely used for clinical assessment and neurological studies and research. But collecting those signals requires a lot of sensors. The use of GUSSS may allow the reduction of the number of sensors, making the process of data collection easier for the

patients and doctors/researchers.

Future work should not be limited to the areas tackled in this work. The proposed method could be applied, for example, to image processing and computer vision. GUSSS could be used to detect specific patterns in an image or could be added to background subtraction algorithms. In essence, GUSSS could potentially be used in many applications where the concept of "signature" would fit.

In addition to extending the use of the method to other areas, work should be done to improve and extend the method itself. One main aspect would be to generalize it to any under-determined case, that is, when the number of sensors is still smaller than the number of independent sources ($M < N$), but that number is not necessarily equal to one. Moreover, the ICA algorithm used could be revised. Specifically, the optimization methodology could be modified to somehow include the prior information available (i.e., the signatures). The iterative methods for finding the independent components and the corresponding mixing coefficients need some initial values. Just like the signatures are injected to create the artificial signal (chapter 3) and thus transform the under-determined system into a determined system, it could be possible to produce a better initialization for the optimization algorithm. This could help improving the results, or at least could reduce computation time.

Finally, further tests should be conducted to have quantitative measures of the success/failure of the GUSSS method and the proposed frameworks, in addition to the qualitative results obtained and presented in this work. A major goal would be determining the theoretical conditions under which the method would be successful, and what its limitations are.

Bibliography

[1] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Computation*, vol. 17, pp. 1875–1902, Sep. 2005.

[2] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[3] A. Hyvaerinen, "Survey on independent component analysis," *Neural Comput. Sur.*, vol. 2, pp. 94–128, 1999.

[4] Y. Guangying, "Study of myoelectric prostheses hand based on independent component analysis and fuzzy controller," in *Eighth International Conference on Electronic Measurement and Instruments*, Aug. 2007, pp. 1–174–1–178.

[5] T. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T. Lee, and T. J. Sejnowski, "Imaging brain dynamics using independent component analysis," in *Proceedings of the IEEE*, vol. 89, no. 7, Jul. 2002, pp. 1107–1122.

[6] E. Oja, K. Kiviluoto, and S. Malaroiu, "Independent component analysis for financial time series," in *The IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, Oct. 2000, pp. 111–116.

[7] K. P. Balanda and H. L. MacGillivray, "Kurtosis a critical review," *The American Statistician*, vol. 42, no. 2, pp. 111–119, May 1988.

[8] D. E. Lake, "Renyi entropy measures of heart rate gaussianity," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 21–27, Jan. 2006.

[9] A. Hyvarinen and E. Oja, "Fastica," neural Networks Research Centre, Helsinki University of Technology, Finland. [Online]. Available: http://research.ics.tkk.fi/ica/fastica/

[10] Y. Li, S. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 423–437, Feb. 2006.

[11] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 87, pp. 1833–1847, Mar. 2007.

[12] P. H. Siegel, "Terahertz technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 910–928, Mar. 2002.

[13] Y. C. Shen, T. Lo, P. F. Taday, B. E. Cole, W. R. Tribe, and M. C. Kemp, "Detection and identification of explosives using terahertz pulsed spectroscopic imaging," *Applied Physics Letters*, vol. 86, p. 241116, 2005.

[14] B. Ferguson and X. Zhang, "Materials for terahertz science and technology," *Nature Materials*, vol. 1, pp. 26–33, Sep. 2002.

[15] I. Hosako, N. Sekine, M. Patrashin, S. Saito, K. Fukunaga, Y. Kasai, P. Baron, T. Seta, J. Mendrok, S. Ochiai, and H. Yasuda, "At the dawn of a new era in terahertz technology," *Proceedings of the IEEE*, vol. 95, no. 8, pp. 1611–1623, Aug. 2007.

[16] M. C. Kemp, "Detecting hidden objects: Security imaging using millimetre-waves and terahertz," in *IEEE Conference on Advanced Video and Signal Based Surveillance*, Sept. 2007, pp. 7–9.

[17] Y. Sasaki, M. Yamashita, A. Dobroiu, T. Shibuya, C. Otani, and K. Kawase, "Noninvasive detection of concealed powders using terahertz wave scattering," in *Infrared and Millimeter Waves and 13th International Conference on Terahertz Electronics*, vol. 2, Sept. 2005, pp. 648–649.

[18] A. Redo-Sanchez and X. Zhang, "Terahertz science and technology trends," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 14, no. 2, pp. 260–269, Mar./Apr. 2008.

[19] M. Kemp, "Millimetre wave and terahertz technology for the detection of concealed threats - a review," in *Proc. SPIE*, vol. 6402, 2006, pp. 1–19.

[20] K. Yamamoto, M. Yamaguchi, F. Miyamaru, M. Tani, M. Hangyo, T. Ikeda, A. Matsushita, K. Koide, M. Tatsuno, and Y. Minami, "Non-invasive inspection of c-4 explosive in mails by terahertz time-domain spectroscopy," *Japanese Journal of Applied Physics*, vol. 43, no. 3B, pp. L414–L417, 2004.

[21] F. Huang, B. Schulkin, H. Altan, J. Federici, D. Gary, R. Barat, D. Zimdars, M. Chen, and D. B. Tanner, "Terahertz study of 1,3,5-trinitro-s-triazine by time-domain and fourier transform infrared spectroscopy," *Applied Physics Letters*, vol. 85, no. 23, pp. 5535–5537, 2004.

[22] J. Chen, H. Zhao, G. J. Bastiaans, and X.-C. Zhang, "Absorption coefficients of selected explosives and related compounds in the range of 0.1-2.8 thz," *Optics Express*, vol. 15, no. 19, pp. 12 060–12 067, Sep. 2007.

[23] K. Kawase, Y. Ogawa, and Y. Watanabe, "Non-destructive terahertz imaging of illicit drugs using spectral fingerprints," *Optics Express*, vol. 11, no. 20, pp. 2549–2554, Oct. 2003.

[24] "Thz-bridge," tera-phtonics Laboratory, RIKEN Sendai. [Online]. Available: http://www.frascati.enea.it/thz-bridge/

[25] "Terahertz database." [Online]. Available: http://thzdb.org

[26] "Thz database web," tera-phtonics Laboratory, RIKEN Sendai. [Online]. Available: http://www.riken.jp/THzdatabase/

[27] H. Zhong, A. Redo-Sanchez, and X.-C. Zhang, "Identification and classification of chemicals using terahertz reflective spectroscopic focal plane imaging system," *Optics Express*, vol. 14, no. 20, pp. 9130–9141, Oct. 2006.

[28] X. Yin, "Pattern recognition and tomographic reconstruction with terahertz signals for applications in biomedical engineering," 2008, ph.D. dissertation, Department of Electrical and Electronic Engineering, University of Adelaide.

[29] D. Zimdars, J. White, G. Stuk, A. Chernovsky, G. Fichter, and S. L. Williamson, "Time domain terahertz detection of concealed threats in luggage and personnel," in *Proc. SPIE*, vol. 6212, Apr. 2006.

[30] R. S. Doyle, B. N. Lyons, A. H. Lettington, T. McEnroe, J. Walshe, J. McNaboe, P. Curtin, and S. Bleszynski, "Stand-off detection of hidden threat objects on personnel at checkpoints and in public areas using active millimetre-wave imaging," in *Proc. SPIE*, vol. 5619, Oct. 2004, pp. 90–97.

[31] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification, 2nd Ed.* Wiley Interscience, 2001.

[32] M. Aquilano, C. Salatino, and M. C. Carrozza, "Assistive technology: a new approach to evaluation," in *10th IEEE International Conference on Rehabilitation Robotics*, Jun. 2007, pp. 809–819.

[33] D. S. Andreasen and D. L. Gabbert, "Electromyographic switch navigation of power wheelchairs," in *Annual conference of the Rehabilitation Engineering and Assistive Technology Society of North America*, Jun. 2006.

[34] A. O'Brien and R. M. Ruairi, "Survey of assistive technology devices and applications for aging in place," in *Second International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services*, Sept. 2009, pp. 7–12.

[35] E. Ambrosini, S. Ferrante, A. Pedrocchi, G. Ferrigno, E. Guanziroli, and F. Molteni, "A novel biofeedback cycling training to improve gait symmetry in stroke patients: a case series study," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 1030–1035, eTH Zurich, Switzerland.

[36] I. Handzic, E. V. Vasudevan, and K. B. Reed, "Motion controlled gait enhancing mobile shoe for rehabilitation," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 583–588, eTH Zurich, Switzerland.

[37] M. A. Baniasad, f. Farahmand, and N. N. Ansari, "Wrist-robohab: A robot for treatment and evaluation of brain injury patients," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 1095–1099, eTH Zurich, Switzerland.

[38] A. E. Jackson, P. R. Culmer, M. C. Levesley, J. A. Cozens, S. G. Makower, and B. B. Bhakta, "Effector force requirements to enable robotic systems to provide assisted exercise in people with upper limb impairment after stroke," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 444–449, eTH Zurich, Switzerland.

[39] A. Koller-Hodac, D. Leonardo, S. Walpen, and D. Felder, "Knee orthopaedic device, how robotic technology can improve outcome in knee rehabilitation," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 186–191, eTH Zurich, Switzerland.

[40] J. Bae and I. Moon, "Biomechanical assessment of electric lifting chair for persons with disability," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 505–509, eTH Zurich, Switzerland.

[41] A. Basteris, A. D. Luca, V. Sanguineti, C. Solaro, M. Mueller, I. Carpinella, D. Cattaneo, R. Bertoni, and M. Ferrarin, "A tailored exercise of manipulation of virtual tools to treat upper limb impairment in multiple sclerosis," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 1112–1116, eTH Zurich, Switzerland.

[42] Y. Hasegawa and S. Oura, "Exoskeletal meal assistance system (emas ii) for progressive muscle dystrophy patient," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 725–730, eTH Zurich, Switzerland.

[43] V. Jaijongrak, I. Kumazawa, and S. Thiemjarus, "A haptic and auditory assistive user interface: Helping the blinds on their computer operations," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 154–159, eTH Zurich, Switzerland.

[44] J. G. Webster, Ed., *Encyclopedia of Medical Devices and Instrumentation. Electromyography.* John Wiley, 2006.

[45] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of emg signal analysis: detection, processing, classification and applications," *Biol. Proced. Online*, vol. 1, no. 8, pp. 11–35, March 2006.

[46] M. R. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "Advances in electromyogram signal classification to improve the quality of life for the disabled and aged people," *Journal of Computer Science*, vol. 7, no. 6, pp. 706–715, 2010.

[47] C. S. L. Tsui, P. Jia, J. Q. Gan, H. Hu, and K. Yuan, "Emg based hands free wheelchair control with eog attention shift detection," in *Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics*, Dec. 2007, pp. 1266–1271.

[48] K. Choi, M. Sato, and Y. Koike, "A new, human-centered wheelchair system controlled by the emg signal," in *International Joint Conference on Neural Networks*, Jul. 2006, pp. 4664–4671.

[49] Y. Wang, X. Zhang, J. Zhao, and C. He, "Pattern recognition of electromyography applied to exoskeleton robot," in *3rd International Congress on Image and Signal Processing (CISP)*, Oct. 2010, pp. 3802–3805.

[50] M. Khezri and M. Jahed, "Real-time intelligent pattern recognition algorithm for surface emg signals," *BioMedical Engineering OnLine*, 2007.

[51] ——, "A novel approach to recognize hand movements via semg patterns," in *29th Annual International Conference of the IEEE EMBS*, Aug. 2007, pp. 4907–4910.

[52] G. Shuman, "Using forearm electromyograms to classify hand gestures," in *IEEE International conference on Bioinformatics and Biomedicine*, 2009, pp. 261–264.

[53] J. Han, Z. Zenn, D. Kim, H. Lee, and J. Kim, "Human-machine interface for wheelchair control with emg and its evaluation," in *Proceedings of the 25th Annual International Conference of the IEEE EMBS*, Sep. 2003, pp. 1602–1605.

[54] Tinkertron EMG switch. [Online]. Available: http://www.tinkertron.com

Appendix A

Complete Set of Results of the Sound Experiments

In this Appendix we show the complete set of results obtained applying the GUSSS method to the sound signals, as described in Chapter 4.

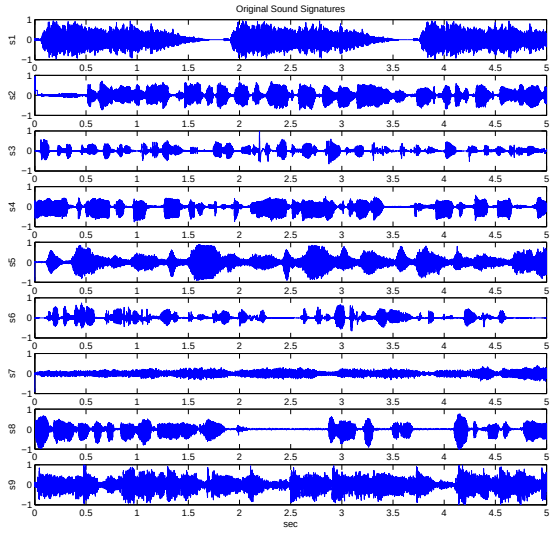
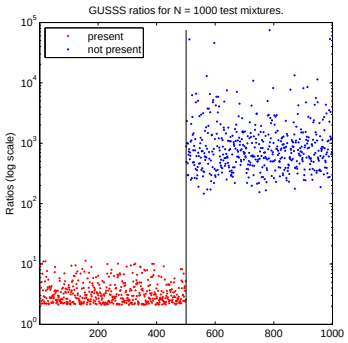
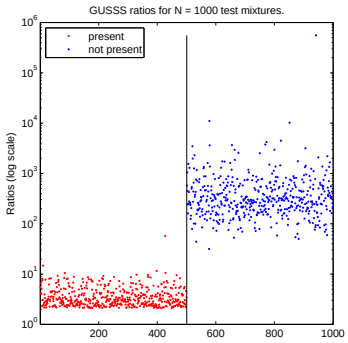


Figure A.1: Nine sound signatures used for the preliminary GUSSS tests described in Chapter 4. They correspond to: a police car's siren (s_1); a man speaking English (s_2); an old man speaking a foreign language (s_3); a woman speaking English (s_4); a band playing classical music (s_5); another man speaking a foreign language (s_6); an opera singer (s_7); a woman speaking a foreign language (s_8); a band playing pop music (s_9).

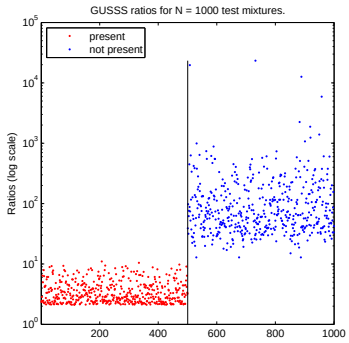
A.1 Experiments #1: No Noise



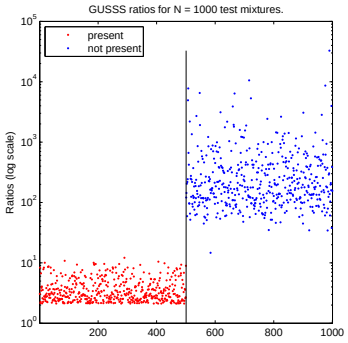
(a) Target signature: s_1



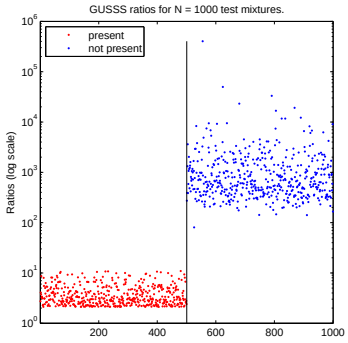
(b) Target signature: s_2



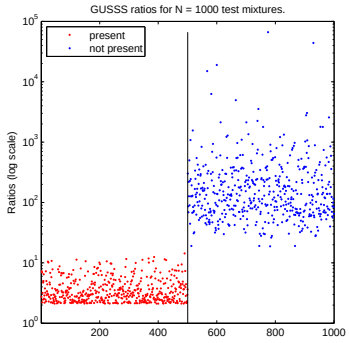
(c) Target signature: s_3



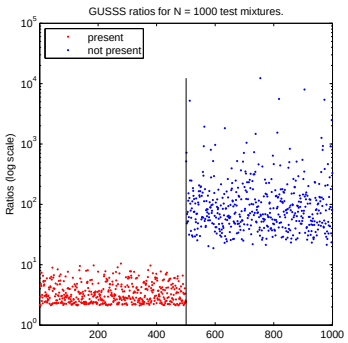
(d) Target signature: s_4



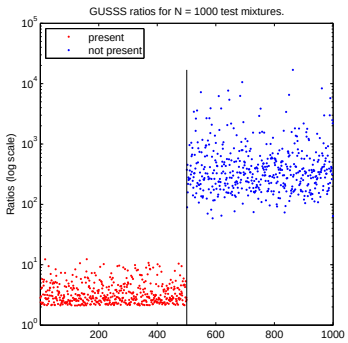
(e) Target signature: s_5



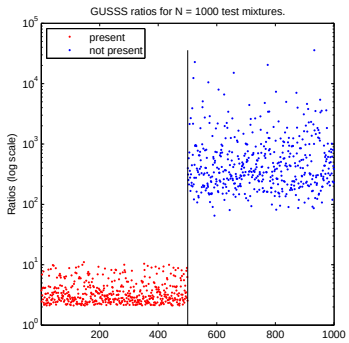
(f) Target signature: s_6



(g) Target signature: s_7



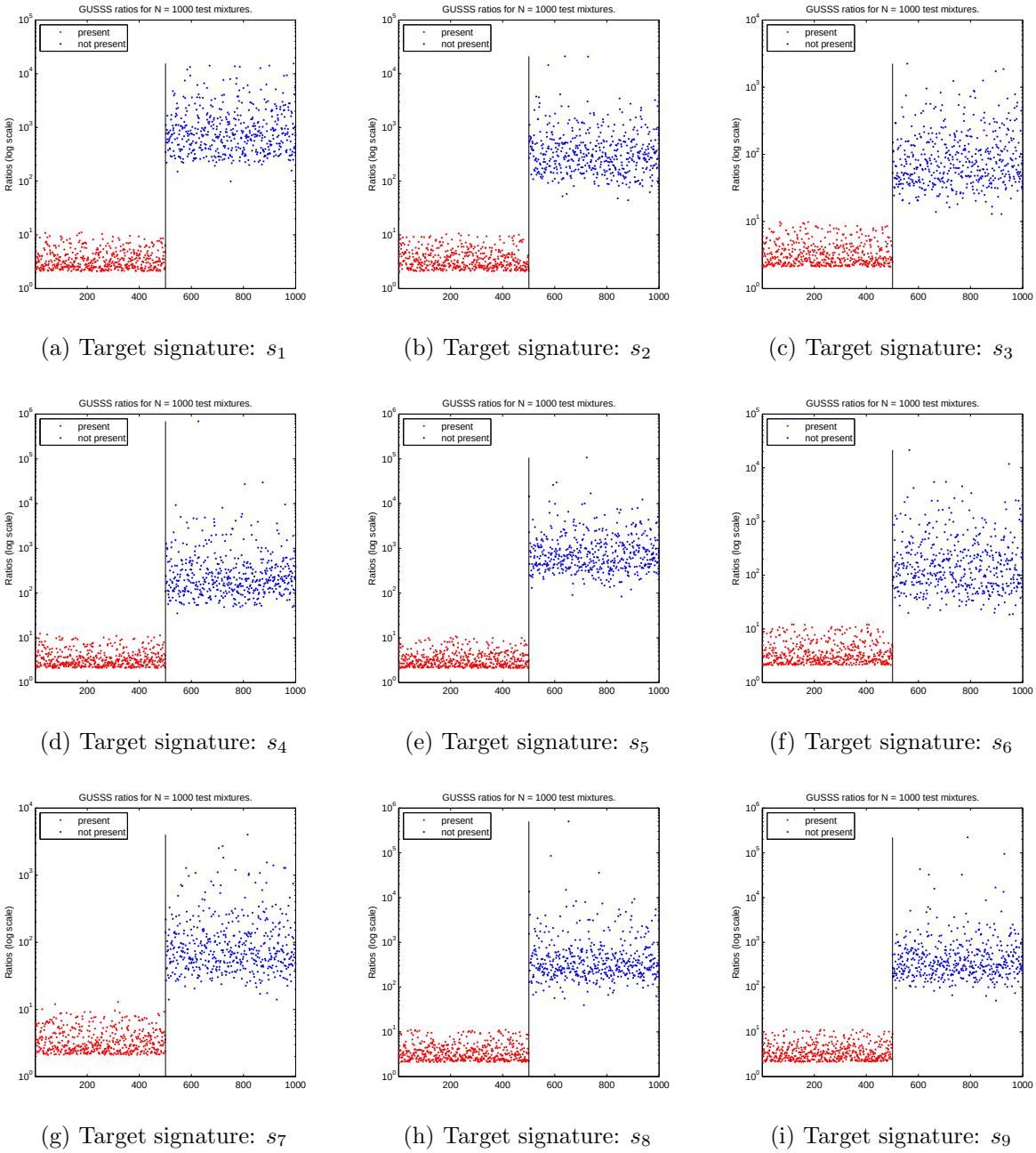
(h) Target signature: s_8



(i) Target signature: s_9

Figure A.2: GUSSS ratios corresponding to 500 mixtures containing a target signature, and 500 mixtures without the target signature. The vertical line separates those two groups. Y-axis (GUSSS ratio axis) is shown in log scale.

A.2 Experiments #2: With Noise

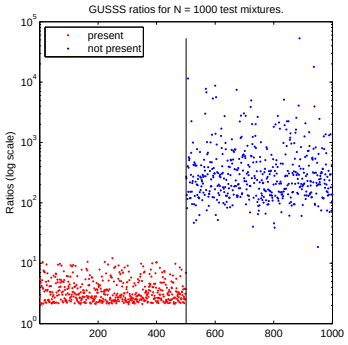


(a) Target signature: s_1 (b) Target signature: s_2 (c) Target signature: s_3

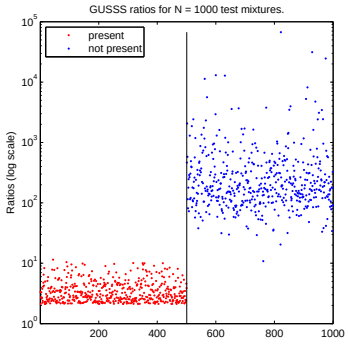
(d) Target signature: s_4 (e) Target signature: s_5 (f) Target signature: s_6

(g) Target signature: s_7 (h) Target signature: s_8 (i) Target signature: s_9

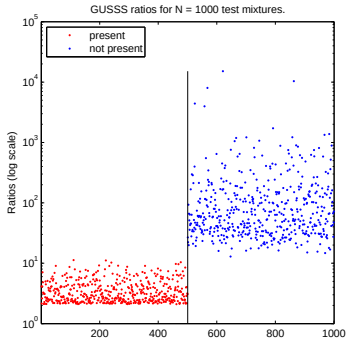
Figure A.3: GUSSS ratios corresponding to 500 noisy mixtures containing a target signature, and 500 mixtures without the target signature. The vertical line separates those two groups. Y-axis (GUSSS ratio axis) is shown in log scale. Noise level: $\sigma = 0.01$



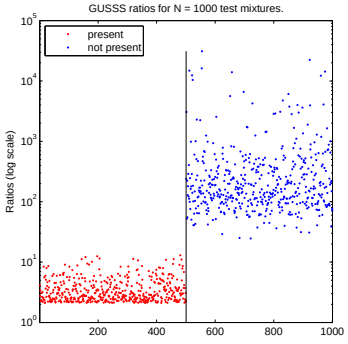
(a) Target signature: s_1



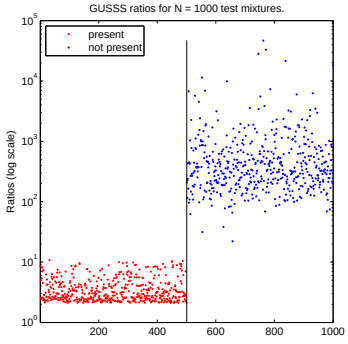
(b) Target signature: s_2



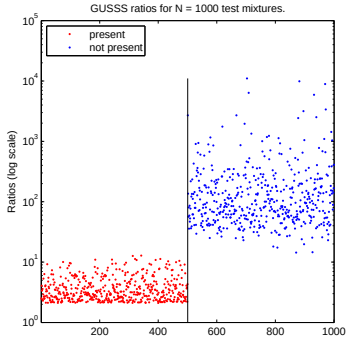
(c) Target signature: s_3



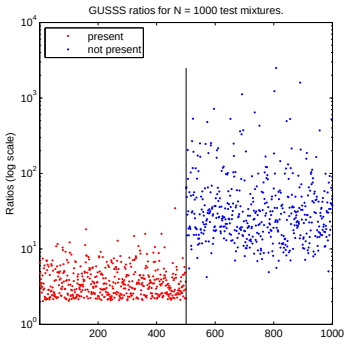
(d) Target signature: s_4



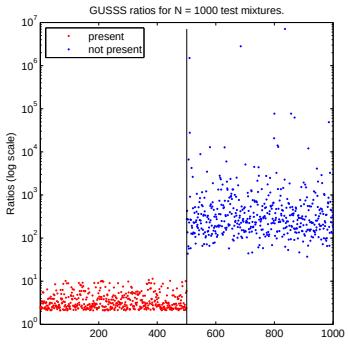
(e) Target signature: s_5



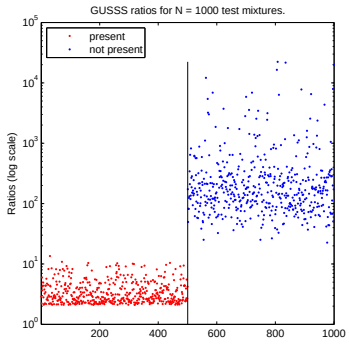
(f) Target signature: s_6



(g) Target signature: s_7



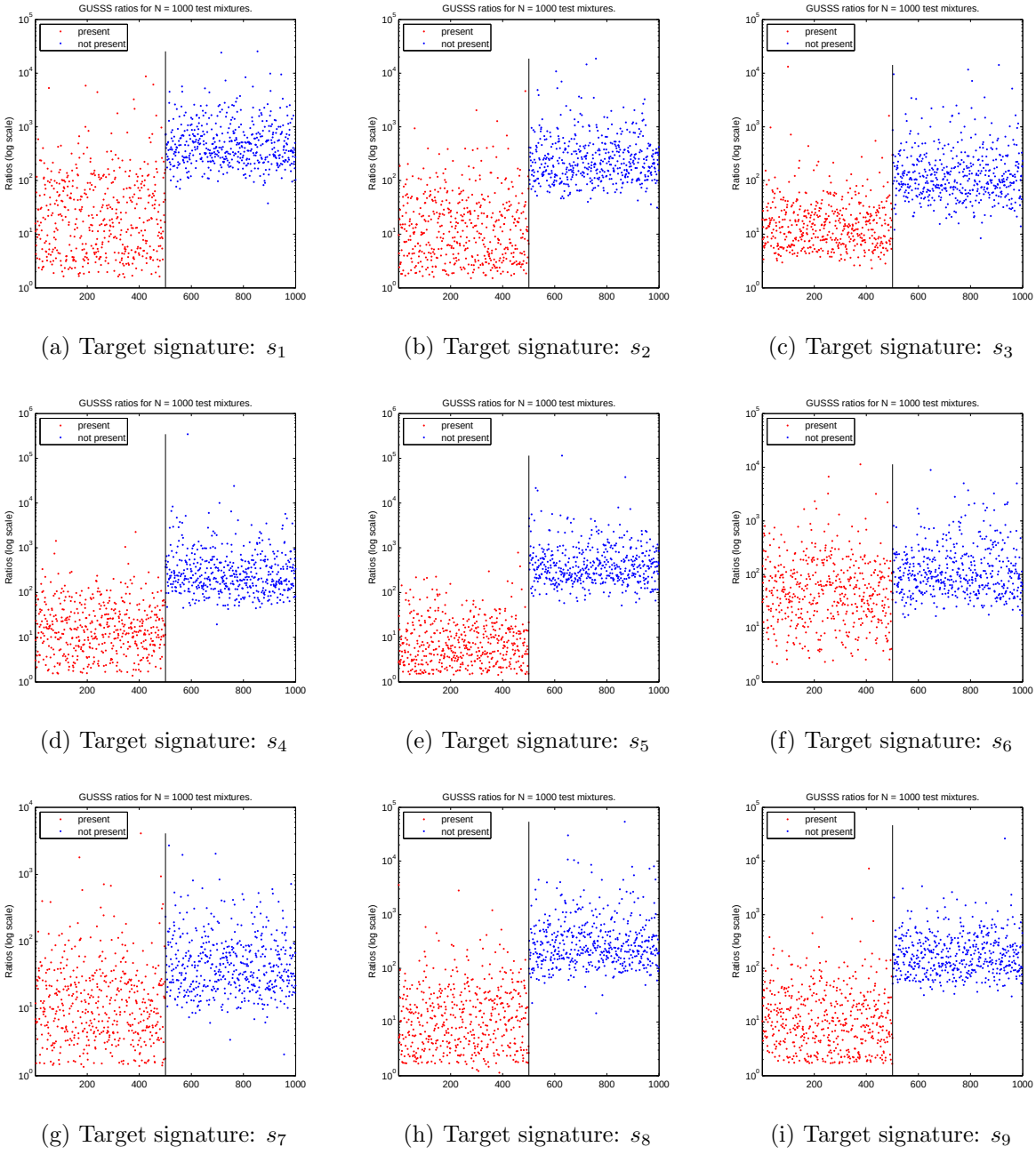
(h) Target signature: s_8



(i) Target signature: s_9

Figure A.4: GUSSS ratios corresponding to 500 noisy mixtures containing a target signature, and 500 mixtures without the target signature. The vertical line separates those two groups. Y-axis (GUSSS ratio axis) is shown in log scale. Noise level: $\sigma = 0.1$

A.3 Experiments #3: With Noise and Time Shift



(a) Target signature: s_1 (b) Target signature: s_2 (c) Target signature: s_3

(d) Target signature: s_4 (e) Target signature: s_5 (f) Target signature: s_6

(g) Target signature: s_7 (h) Target signature: s_8 (i) Target signature: s_9

Figure A.5: GUSSS ratios corresponding to 500 noisy mixtures with time shift containing a target signature, and 500 mixtures without the target signature. The vertical line separates those two groups. Y-axis (GUSSS ratio axis) is shown in log scale.

Appendix B

Complete Set of Results of the THz Experiments

In this Appendix we present the results of all 18 THz tests. For each material tested we show the tables with true positive (TP), true negative (TN) and Correct Classification (CC) percentages, for the three approaches explained in chapters 5 and 7 (threshold using Geometric Means, threshold using intersection of Gaussians and SVM).

Table B.1: Material Manzeb

Detected Material:	**Manzeb**
Additional Training Materials:	ADP, Albumin, DAST, H2S, HgS, Lumiflavin, MgO
Additional Testing Materials:	Glucose, SiN

Threshold using Geometric Means			
[ll, ul]	TP	TN	CC
[0.10, 0.40]	99.6	98.8	99.2
[0.10, 0.75]	99.6	99.6	99.6
[0.20, 0.40]	100.0	99.0	99.5
[0.20, 0.75]	100.0	99.2	99.6
[0.30, 0.40]	100.0	98.8	99.4
[0.30, 0.75]	100.0	99.6	99.8
[0.50, 0.75]	100.0	99.6	99.8
[0.50, 0.95]	100.0	99.0	99.5
[0.70, 0.75]	100.0	99.8	99.9
[0.70, 0.95]	100.0	100.0	100.0
[0.80, 1.00]	100.0	99.6	99.8
Total:	99.9	99.4	99.6

Threshold using of int. of Gaussians			
[ll, ul]	TP	TN	CC
[0.10, 0.40]	99.4	99.8	99.6
[0.10, 0.75]	99.6	99.4	99.5
[0.20, 0.40]	99.6	100.0	99.8
[0.20, 0.75]	99.4	100.0	99.7
[0.30, 0.40]	99.8	100.0	99.9
[0.30, 0.75]	98.6	100.0	99.3
[0.50, 0.75]	100.0	100.0	100.0
[0.50, 0.95]	99.0	100.0	99.5
[0.70, 0.75]	100.0	100.0	100.0
[0.70, 0.95]	100.0	100.0	100.0
[0.80, 1.00]	100.0	100.0	100.0
Total:	99.6	99.9	99.8

SVM Approach			
[ll, ul]	TP	TN	CC
[0.10, 0.40]	99.4	99.8	99.6
[0.10, 0.75]	99.4	99.6	99.5
[0.20, 0.40]	100.0	100.0	100.0
[0.20, 0.75]	100.0	100.0	100.0
[0.30, 0.40]	100.0	100.0	100.0
[0.30, 0.75]	100.0	100.0	100.0
[0.50, 0.75]	100.0	100.0	100.0
[0.50, 0.95]	100.0	100.0	100.0
[0.70, 0.75]	100.0	100.0	100.0
[0.70, 0.95]	100.0	100.0	100.0
[0.80, 1.00]	100.0	100.0	100.0
Total:	99.9	99.9	99.9

Table B.2: Material Talc

Detected Material:	**Talc**											
Additional Training Materials:	Baking Soda, Caffeine, Cellulose, Chalk, Coffee Ground, Equal, O2Si											
Additional Testing Materials:	Ovalbumin, Sodium Chloride											
Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach				
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	
[0.10, 0.40]	68.8	70.8	69.8	[0.10, 0.40]	85.2	47.6	66.4	[0.10, 0.40]	66.0	83.4	74.7	
[0.10, 0.75]	75.4	82.4	78.9	[0.10, 0.75]	89.2	62.6	75.9	[0.10, 0.75]	69.4	87.8	78.6	
[0.20, 0.40]	76.0	76.4	76.2	[0.20, 0.40]	86.6	59.8	73.2	[0.20, 0.40]	77.0	82.0	79.5	
[0.20, 0.75]	87.8	84.2	86.0	[0.20, 0.75]	91.2	73.6	82.4	[0.20, 0.75]	78.8	89.0	83.9	
[0.30, 0.40]	81.0	78.0	79.5	[0.30, 0.40]	89.2	65.6	77.4	[0.30, 0.40]	82.2	80.8	81.5	
[0.30, 0.75]	89.2	83.0	86.1	[0.30, 0.75]	93.6	80.2	86.9	[0.30, 0.75]	84.8	92.4	88.6	
[0.50, 0.75]	97.0	83.6	90.3	[0.50, 0.75]	97.4	90.8	94.1	[0.50, 0.75]	94.0	95.2	94.6	
[0.50, 0.95]	97.8	94.4	96.1	[0.50, 0.95]	95.2	94.6	94.9	[0.50, 0.95]	94.6	96.8	95.7	
[0.70, 0.75]	99.0	88.6	93.8	[0.70, 0.75]	98.6	97.6	98.1	[0.70, 0.75]	97.8	98.6	98.2	
[0.70, 0.95]	98.4	95.4	96.9	[0.70, 0.95]	99.2	98.6	98.9	[0.70, 0.95]	98.4	98.8	98.6	
[0.80, 1.00]	100.0	96.6	98.3	[0.80, 1.00]	99.6	99.8	99.7	[0.80, 1.00]	99.8	99.2	99.5	
Total:	88.2	84.9	86.5	Total:	93.2	79.2	86.2	Total:	85.7	91.3	88.5	

Table B.3: Material Baking Soda

Detected Material:	**Baking Soda**											
Additional Training Materials:	Caffeine, Cellulose, Coffee Ground, O2Si, Ovalbumin, Sodium Chloride, Talc											
Additional Testing Materials:	Equal, Trehalose											
Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach				
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	
[0.10, 0.40]	48.4	87.0	67.7	[0.10, 0.40]	67.0	72.2	69.6	[0.10, 0.40]	58.4	79.8	69.1	
[0.10, 0.75]	60.2	92.8	76.5	[0.10, 0.75]	72.4	87.4	79.9	[0.10, 0.75]	67.4	85.0	76.2	
[0.20, 0.40]	51.6	91.4	71.5	[0.20, 0.40]	70.6	74.2	72.4	[0.20, 0.40]	62.0	84.4	73.2	
[0.20, 0.75]	65.2	95.4	80.3	[0.20, 0.75]	75.6	86.8	81.2	[0.20, 0.75]	65.4	86.8	76.1	
[0.30, 0.40]	58.6	91.0	74.8	[0.30, 0.40]	71.0	85.8	78.4	[0.30, 0.40]	70.8	83.0	76.9	
[0.30, 0.75]	72.2	95.4	83.8	[0.30, 0.75]	73.8	91.0	82.4	[0.30, 0.75]	75.4	92.0	83.7	
[0.50, 0.75]	80.6	94.6	87.6	[0.50, 0.75]	80.4	96.4	88.4	[0.50, 0.75]	81.0	89.6	85.3	
[0.50, 0.95]	79.8	96.6	88.2	[0.50, 0.95]	83.4	91.0	87.2	[0.50, 0.95]	79.8	87.8	83.8	
[0.70, 0.75]	81.2	95.2	88.2	[0.70, 0.75]	80.4	96.2	88.3	[0.70, 0.75]	81.0	92.6	86.8	
[0.70, 0.95]	79.8	95.0	87.4	[0.70, 0.95]	83.6	90.8	87.2	[0.70, 0.95]	80.6	87.2	83.9	
[0.80, 1.00]	76.2	93.8	85.0	[0.80, 1.00]	88.4	88.4	88.4	[0.80, 1.00]	76.8	88.6	82.7	
Total:	68.5	93.5	81.0	Total:	77.0	87.3	82.1	Total:	72.6	87.0	79.8	

Table B.4: Material ADP

Detected Material:	**ADP**											
Additional Training Materials:	Talc, Albumin, DAST, Glucose, H2S, Lumiflavin, NAC											
Additional Testing Materials:	O2Si, MgO											
Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach				
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	
[0.10, 0.40]	86.4	91.6	89.0	[0.10, 0.40]	92.4	61.4	76.9	[0.10, 0.40]	83.0	90.6	86.8	
[0.10, 0.75]	88.4	92.0	90.2	[0.10, 0.75]	93.6	87.8	90.7	[0.10, 0.75]	89.0	96.2	92.6	
[0.20, 0.40]	94.0	92.0	93.0	[0.20, 0.40]	95.2	79.4	87.3	[0.20, 0.40]	91.4	92.8	92.1	
[0.20, 0.75]	95.8	93.6	94.7	[0.20, 0.75]	96.2	91.6	93.9	[0.20, 0.75]	94.4	92.4	93.4	
[0.30, 0.40]	93.8	93.2	93.5	[0.30, 0.40]	94.6	90.4	92.5	[0.30, 0.40]	93.2	93.8	93.5	
[0.30, 0.75]	97.2	94.0	95.6	[0.30, 0.75]	96.2	95.2	95.7	[0.30, 0.75]	94.8	94.0	94.4	
[0.50, 0.75]	96.6	95.6	96.1	[0.50, 0.75]	96.0	94.4	95.2	[0.50, 0.75]	95.4	95.0	95.2	
[0.50, 0.95]	93.4	93.2	93.3	[0.50, 0.95]	92.0	96.6	94.3	[0.50, 0.95]	92.2	95.6	93.9	
[0.70, 0.75]	94.0	93.6	93.8	[0.70, 0.75]	92.2	95.6	93.9	[0.70, 0.75]	87.2	97.0	92.1	
[0.70, 0.95]	89.2	92.6	90.9	[0.70, 0.95]	88.4	90.8	89.6	[0.70, 0.95]	86.0	95.0	90.5	
[0.80, 1.00]	85.2	93.4	89.3	[0.80, 1.00]	87.8	89.8	88.8	[0.80, 1.00]	82.8	97.2	90.0	
Total:	92.2	93.2	92.7	Total:	93.1	88.5	90.8	Total:	89.9	94.5	92.2	

Table B.5: Material Glucose

Detected Material:	Glucose		
Additional Training Materials:	Baking Soda, Cellulose, O2Si, Ovalbumin, Sodium Chloride, Talc, NAC		
Additional Testing Materials:	Coffee Ground, HgS		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	96.4	74.0	85.2	[0.10, 0.40]	93.8	82.0	87.9	[0.10, 0.40]	88.0	88.8	88.4
[0.10, 0.75]	98.0	82.6	90.3	[0.10, 0.75]	87.8	94.2	91.0	[0.10, 0.75]	86.8	96.8	91.8
[0.20, 0.40]	99.0	80.0	89.5	[0.20, 0.40]	93.0	94.6	93.8	[0.20, 0.40]	95.4	91.6	93.5
[0.20, 0.75]	99.4	88.2	93.8	[0.20, 0.75]	93.6	98.0	95.8	[0.20, 0.75]	93.8	94.8	94.3
[0.30, 0.40]	99.2	80.8	90.0	[0.30, 0.40]	94.8	97.4	96.1	[0.30, 0.40]	98.0	97.6	97.8
[0.30, 0.75]	99.6	87.0	93.3	[0.30, 0.75]	97.2	98.8	98.0	[0.30, 0.75]	97.4	98.6	98.0
[0.50, 0.75]	100.0	91.2	95.6	[0.50, 0.75]	89.4	100.0	94.7	[0.50, 0.75]	98.8	100.0	99.4
[0.50, 0.95]	99.8	93.6	96.7	[0.50, 0.95]	93.0	100.0	96.5	[0.50, 0.95]	98.6	99.8	99.2
[0.70, 0.75]	100.0	90.8	95.4	[0.70, 0.75]	87.6	100.0	93.8	[0.70, 0.75]	99.0	100.0	99.5
[0.70, 0.95]	100.0	87.0	93.5	[0.70, 0.95]	95.0	100.0	97.5	[0.70, 0.95]	98.8	100.0	99.4
[0.80, 1.00]	100.0	87.6	93.8	[0.80, 1.00]	96.6	100.0	98.3	[0.80, 1.00]	99.6	99.8	99.7
Total:	99.2	85.7	92.5	Total:	92.9	96.8	94.9	Total:	95.8	97.1	96.5

Table B.6: Material DAST

Detected Material:	DAST		
Additional Training Materials:	Caffeine, Equal, ADP, Albumin, H2S, Lumiflavin, NAC		
Additional Testing Materials:	Sodium Chloride, Glucose		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	79.6	62.6	71.1	[0.10, 0.40]	89.0	33.6	61.3	[0.10, 0.40]	70.2	79.2	74.7
[0.10, 0.75]	77.2	79.6	78.4	[0.10, 0.75]	88.2	45.4	66.8	[0.10, 0.75]	73.8	87.6	80.7
[0.20, 0.40]	83.0	71.6	77.3	[0.20, 0.40]	89.2	41.4	65.3	[0.20, 0.40]	74.4	84.2	79.3
[0.20, 0.75]	85.8	82.0	83.9	[0.20, 0.75]	87.2	57.4	72.3	[0.20, 0.75]	81.6	87.0	84.3
[0.30, 0.40]	86.6	73.8	80.2	[0.30, 0.40]	90.8	51.0	70.9	[0.30, 0.40]	82.4	85.6	84.0
[0.30, 0.75]	82.4	85.4	83.9	[0.30, 0.75]	87.4	63.8	75.6	[0.30, 0.75]	82.6	89.6	86.1
[0.50, 0.75]	84.8	83.2	84.0	[0.50, 0.75]	85.2	64.4	74.8	[0.50, 0.75]	86.0	93.6	89.8
[0.50, 0.95]	80.0	85.6	82.8	[0.50, 0.95]	87.2	47.8	67.5	[0.50, 0.95]	76.6	89.8	83.2
[0.70, 0.75]	81.8	87.0	84.4	[0.70, 0.75]	84.2	45.2	64.7	[0.70, 0.75]	78.4	95.0	86.7
[0.70, 0.95]	74.2	82.4	78.3	[0.70, 0.95]	87.8	34.6	61.2	[0.70, 0.95]	75.2	91.2	83.2
[0.80, 1.00]	78.8	80.2	79.5	[0.80, 1.00]	87.6	50.8	64.2	[0.80, 1.00]	74.6	86.0	80.3
Total:	81.3	79.4	80.3	Total:	87.6	48.7	67.7	Total:	77.8	88.1	82.9

Table B.7: Material Lumiflavin

Detected Material:	Lumiflavin		
Additional Training Materials:	Caffeine, O2Si, Albumin, DAST, Glucose, MgO, SiN		
Additional Testing Materials:	Cellulose, Coffee Ground		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	56.6	79.4	68.0	[0.10, 0.40]	79.6	49.0	64.3	[0.10, 0.40]	44.2	92.2	68.2
[0.10, 0.75]	75.6	85.6	80.6	[0.10, 0.75]	87.2	64.0	75.6	[0.10, 0.75]	67.8	94.6	81.2
[0.20, 0.40]	65.2	85.6	75.4	[0.20, 0.40]	81.0	55.0	68.0	[0.20, 0.40]	56.0	95.4	75.7
[0.20, 0.75]	79.4	87.2	83.3	[0.20, 0.75]	86.8	72.0	79.4	[0.20, 0.75]	70.2	96.2	83.2
[0.30, 0.40]	68.4	86.4	77.4	[0.30, 0.40]	85.8	57.2	71.5	[0.30, 0.40]	58.2	97.8	78.0
[0.30, 0.75]	86.4	85.4	85.9	[0.30, 0.75]	89.0	75.0	82.0	[0.30, 0.75]	83.8	97.6	90.7
[0.50, 0.75]	94.2	86.8	90.5	[0.50, 0.75]	94.8	84.4	89.6	[0.50, 0.75]	87.2	96.2	91.7
[0.50, 0.95]	93.6	83.0	88.3	[0.50, 0.95]	96.2	84.4	90.3	[0.50, 0.95]	91.2	93.2	92.2
[0.70, 0.75]	98.6	82.8	90.7	[0.70, 0.75]	97.0	86.4	91.7	[0.70, 0.75]	94.0	97.2	95.6
[0.70, 0.95]	98.4	83.2	90.8	[0.70, 0.95]	98.6	81.6	90.1	[0.70, 0.95]	91.8	92.8	92.3
[0.80, 1.00]	97.4	84.2	90.8	[0.80, 1.00]	98.0	83.0	90.5	[0.80, 1.00]	94.2	92.8	93.5
Total:	83.1	84.5	83.8	Total:	90.4	72.0	81.2	Total:	76.2	95.1	85.7

Table B.8: Material Cellulose

Detected Material:	**Cellulose**		
Additional Training Materials:	Caffeine, Coffee Ground, Equal, O2Si, Ovalbumin, Sodium Chloride, Talc		
Additional Testing Materials:	Chalk, NAC		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	58.6	53.8	56.2	[0.10, 0.40]	77.8	26.8	52.3	[0.10, 0.40]	57.4	33.8	45.6
[0.10, 0.75]	63.0	68.0	65.5	[0.10, 0.75]	86.4	32.0	59.2	[0.10, 0.75]	61.4	54.4	57.9
[0.20, 0.40]	60.4	55.6	58.0	[0.20, 0.40]	81.2	30.4	55.8	[0.20, 0.40]	67.8	23.6	45.7
[0.20, 0.75]	72.2	56.2	64.2	[0.20, 0.75]	84.8	42.0	63.4	[0.20, 0.75]	64.8	55.2	60.0
[0.30, 0.40]	67.2	57.4	62.3	[0.30, 0.40]	82.4	32.8	57.6	[0.30, 0.40]	57.2	52.2	54.7
[0.30, 0.75]	77.0	67.0	72.0	[0.30, 0.75]	92.4	40.0	66.2	[0.30, 0.75]	79.8	57.4	68.6
[0.50, 0.75]	86.4	68.8	77.6	[0.50, 0.75]	95.4	49.6	72.5	[0.50, 0.75]	81.6	63.4	72.5
[0.50, 0.95]	87.8	72.0	79.9	[0.50, 0.95]	94.4	57.6	76.0	[0.50, 0.95]	90.8	69.8	80.3
[0.70, 0.75]	95.4	73.2	84.3	[0.70, 0.75]	96.6	68.0	82.3	[0.70, 0.75]	95.6	69.4	82.5
[0.70, 0.95]	96.4	75.2	85.8	[0.70, 0.95]	97.4	74.6	86.0	[0.70, 0.95]	95.4	76.6	86.0
[0.80, 1.00]	97.6	78.4	88.0	[0.80, 1.00]	96.8	81.0	88.9	[0.80, 1.00]	95.8	80.4	88.1
Total:	78.4	66.0	72.2	Total:	89.6	48.6	69.1	Total:	77.1	57.8	67.4

Table B.9: Material MgO

Detected Material:	**MgO**		
Additional Training Materials:	ADP, Albumin, DAST, Glucose, H2S, HgS, NAC		
Additional Testing Materials:	Lumiflavin, Manzeb		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	80.4	77.0	78.7	[0.10, 0.40]	85.8	64.0	74.9	[0.10, 0.40]	79.0	77.8	78.4
[0.10, 0.75]	86.4	89.4	87.9	[0.10, 0.75]	89.2	76.0	82.6	[0.10, 0.75]	86.4	82.8	84.6
[0.20, 0.40]	86.4	81.8	84.1	[0.20, 0.40]	89.8	73.6	81.7	[0.20, 0.40]	87.2	81.8	84.5
[0.20, 0.75]	89.4	89.0	89.2	[0.20, 0.75]	94.4	80.6	87.5	[0.20, 0.75]	92.4	87.0	89.7
[0.30, 0.40]	91.0	85.4	88.2	[0.30, 0.40]	94.4	80.2	87.3	[0.30, 0.40]	92.0	85.6	88.8
[0.30, 0.75]	93.6	92.8	93.2	[0.30, 0.75]	96.4	92.4	94.4	[0.30, 0.75]	96.8	93.8	95.3
[0.50, 0.75]	98.6	92.4	95.5	[0.50, 0.75]	98.2	98.8	98.5	[0.50, 0.75]	98.6	97.2	97.9
[0.50, 0.95]	99.0	92.2	95.6	[0.50, 0.95]	94.4	98.8	96.6	[0.50, 0.95]	97.0	96.8	96.9
[0.70, 0.75]	99.4	89.4	94.4	[0.70, 0.75]	97.0	98.0	97.5	[0.70, 0.75]	98.0	98.6	98.3
[0.70, 0.95]	97.8	93.2	95.5	[0.70, 0.95]	96.6	99.0	97.8	[0.70, 0.95]	96.8	95.6	96.2
[0.80, 1.00]	98.0	94.6	96.3	[0.80, 1.00]	94.4	99.0	96.7	[0.80, 1.00]	95.2	97.8	96.5
Total:	92.7	88.8	90.8	Total:	93.7	87.3	90.5	Total:	92.7	90.4	91.6

Table B.10: Material HgS

Detected Material:	**HgS**		
Additional Training Materials:	Caffeine, Cellulose, Chalk, Coffee Ground, Equal, O2Si, Talc		
Additional Testing Materials:	Sodium Chloride, DAST		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	78.8	77.0	77.9	[0.10, 0.40]	90.4	68.4	79.4	[0.10, 0.40]	75.4	81.8	78.6
[0.10, 0.75]	85.0	89.0	87.0	[0.10, 0.75]	88.6	79.8	84.2	[0.10, 0.75]	84.6	92.0	88.3
[0.20, 0.40]	90.8	78.4	84.6	[0.20, 0.40]	96.8	79.8	88.3	[0.20, 0.40]	85.4	83.6	84.5
[0.20, 0.75]	90.6	94.2	92.4	[0.20, 0.75]	96.0	81.2	88.6	[0.20, 0.75]	90.8	93.0	91.9
[0.30, 0.40]	97.2	83.0	90.1	[0.30, 0.40]	97.0	85.4	91.2	[0.30, 0.40]	94.6	89.0	91.8
[0.30, 0.75]	97.2	93.4	95.3	[0.30, 0.75]	98.8	93.6	96.2	[0.30, 0.75]	96.2	96.2	96.2
[0.50, 0.75]	99.8	96.8	98.3	[0.50, 0.75]	99.4	99.2	99.3	[0.50, 0.75]	99.6	99.0	99.3
[0.50, 0.95]	99.6	98.6	99.1	[0.50, 0.95]	99.6	99.8	99.7	[0.50, 0.95]	99.8	99.6	99.7
[0.70, 0.75]	100.0	97.4	98.7	[0.70, 0.75]	99.8	100.0	99.9	[0.70, 0.75]	99.6	100.0	99.8
[0.70, 0.95]	100.0	98.6	99.3	[0.70, 0.95]	99.8	100.0	99.9	[0.70, 0.95]	100.0	100.0	100.0
[0.80, 1.00]	100.0	98.2	99.1	[0.80, 1.00]	99.8	100.0	99.9	[0.80, 1.00]	100.0	100.0	100.0
Total:	94.5	91.3	92.9	Total:	96.9	89.7	93.3	Total:	93.3	94.0	93.6

Table B.11: Material Caffeine

Detected Material:	**Caffeine**		
Additional Training Materials:	Baking Soda, Cellulose, Chalk, Coffee Ground, Equal, Sodium Chloride, DAST		
Additional Testing Materials:	Ovalbumin, Albumin		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	72.8	89.6	81.2	[0.10, 0.40]	77.4	85.0	81.2	[0.10, 0.40]	73.6	83.6	78.6
[0.10, 0.75]	84.4	91.0	87.7	[0.10, 0.75]	84.6	85.2	84.9	[0.10, 0.75]	86.2	90.2	88.2
[0.20, 0.40]	89.6	85.6	87.6	[0.20, 0.40]	85.6	89.2	87.4	[0.20, 0.40]	88.2	87.0	87.6
[0.20, 0.75]	91.8	88.6	90.2	[0.20, 0.75]	90.0	89.4	89.7	[0.20, 0.75]	93.8	91.8	92.8
[0.30, 0.40]	96.8	90.4	93.6	[0.30, 0.40]	92.4	90.8	91.6	[0.30, 0.40]	95.4	91.4	93.4
[0.30, 0.75]	98.4	89.8	94.1	[0.30, 0.75]	95.4	93.6	94.5	[0.30, 0.75]	95.8	90.4	93.1
[0.50, 0.75]	97.4	91.6	94.5	[0.50, 0.75]	94.8	91.2	93.0	[0.50, 0.75]	97.2	92.6	94.9
[0.50, 0.95]	94.4	87.8	91.1	[0.50, 0.95]	95.4	83.8	89.6	[0.50, 0.95]	94.0	88.8	91.4
[0.70, 0.75]	95.4	88.8	92.1	[0.70, 0.75]	96.2	86.8	91.5	[0.70, 0.75]	96.0	88.8	92.4
[0.70, 0.95]	93.4	83.6	88.5	[0.70, 0.95]	93.8	72.2	83.0	[0.70, 0.95]	93.6	82.4	88.0
[0.80, 1.00]	92.6	82.4	87.5	[0.80, 1.00]	94.6	73.2	83.9	[0.80, 1.00]	91.0	82.2	86.6
Total:	91.5	88.1	89.8	Total:	90.9	85.5	88.2	Total:	91.3	88.1	89.7

Table B.12: Material Chalk

Detected Material:	**Chalk**		
Additional Training Materials:	Cellulose, Coffee Ground, Equal, O2Si, Ovalbumin, Sodium Chloride, Talc		
Additional Testing Materials:	Baking Soda, Caffeine		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	77.4	90.6	84.0	[0.10, 0.40]	81.0	79.4	80.2	[0.10, 0.40]	75.2	90.6	82.9
[0.10, 0.75]	84.4	95.4	89.9	[0.10, 0.75]	87.4	92.8	90.1	[0.10, 0.75]	86.8	95.4	91.1
[0.20, 0.40]	89.2	91.4	90.3	[0.20, 0.40]	90.0	89.4	89.7	[0.20, 0.40]	89.2	93.0	91.1
[0.20, 0.75]	88.2	97.8	93.0	[0.20, 0.75]	91.0	96.0	93.5	[0.20, 0.75]	90.4	97.0	93.7
[0.30, 0.40]	93.6	94.0	93.8	[0.30, 0.40]	90.8	96.6	93.7	[0.30, 0.40]	91.2	96.6	93.9
[0.30, 0.75]	95.4	98.0	96.7	[0.30, 0.75]	96.8	99.0	97.9	[0.30, 0.75]	95.6	98.4	97.0
[0.50, 0.75]	98.8	98.0	98.4	[0.50, 0.75]	96.2	100.0	98.1	[0.50, 0.75]	98.6	99.6	99.1
[0.50, 0.95]	99.0	99.6	99.3	[0.50, 0.95]	98.2	99.2	98.7	[0.50, 0.95]	99.0	99.0	99.0
[0.70, 0.75]	99.0	98.0	98.5	[0.70, 0.75]	97.8	100.0	98.9	[0.70, 0.75]	99.0	99.6	99.3
[0.70, 0.95]	98.8	99.6	99.2	[0.70, 0.95]	98.6	99.8	99.2	[0.70, 0.95]	99.4	100.0	99.7
[0.80, 1.00]	98.8	99.6	99.2	[0.80, 1.00]	99.6	99.6	99.6	[0.80, 1.00]	99.8	99.8	99.8
Total:	93.0	96.5	94.8	Total:	93.4	95.6	94.5	Total:	93.1	97.2	95.1

Table B.13: Material Sodium Chloride

Detected Material:	**Sodium Chloride**		
Additional Training Materials:	Baking Soda, Caffeine, Cellulose, Coffee Ground, Equal, Ovalbumin, ADP		
Additional Testing Materials:	Chalk, O2Si		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	83.6	46.4	65.0	[0.10, 0.40]	93.8	25.8	59.8	[0.10, 0.40]	81.4	53.0	67.2
[0.10, 0.75]	88.0	51.6	69.8	[0.10, 0.75]	91.6	44.6	68.1	[0.10, 0.75]	83.2	62.0	72.6
[0.20, 0.40]	87.2	53.4	70.3	[0.20, 0.40]	95.2	34.2	64.7	[0.20, 0.40]	79.2	58.2	68.7
[0.20, 0.75]	93.2	54.4	73.8	[0.20, 0.75]	94.8	48.2	71.5	[0.20, 0.75]	87.8	68.8	78.3
[0.30, 0.40]	90.8	50.6	70.7	[0.30, 0.40]	95.4	39.6	67.5	[0.30, 0.40]	88.4	57.6	73.0
[0.30, 0.75]	96.6	59.0	77.8	[0.30, 0.75]	95.4	63.0	79.2	[0.30, 0.75]	91.6	70.2	80.9
[0.50, 0.75]	98.0	57.6	77.8	[0.50, 0.75]	96.4	79.2	87.8	[0.50, 0.75]	95.4	78.8	87.1
[0.50, 0.95]	97.0	59.4	78.2	[0.50, 0.95]	94.0	77.0	85.5	[0.50, 0.95]	91.6	83.8	87.7
[0.70, 0.75]	97.4	62.2	79.8	[0.70, 0.75]	95.4	82.2	88.8	[0.70, 0.75]	91.4	91.4	91.4
[0.70, 0.95]	97.0	68.2	82.6	[0.70, 0.95]	93.8	78.6	86.2	[0.70, 0.95]	88.0	84.8	86.4
[0.80, 1.00]	95.2	60.4	77.8	[0.80, 1.00]	92.4	68.4	80.4	[0.80, 1.00]	87.8	84.0	85.9
Total:	93.1	56.7	74.9	Total:	94.4	58.3	76.3	Total:	87.8	72.1	79.9

Table B.14: Material NAC

Detected Material:	NAC										
Additional Training Materials:	Caffeine, Cellulose, Chalk, O2Si, Ovalbumin, Sodium Chloride, Talc										
Additional Testing Materials:	Coffee Ground, Equal										
Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	88.0	87.0	87.5	[0.10, 0.40]	89.8	82.2	86.0	[0.10, 0.40]	81.6	95.8	88.7
[0.10, 0.75]	88.4	96.8	92.6	[0.10, 0.75]	89.6	96.2	92.9	[0.10, 0.75]	90.2	97.2	93.7
[0.20, 0.40]	95.4	92.2	93.8	[0.20, 0.40]	95.6	94.2	94.9	[0.20, 0.40]	92.0	98.2	95.1
[0.20, 0.75]	96.4	95.8	96.1	[0.20, 0.75]	97.2	96.8	97.0	[0.20, 0.75]	94.2	99.0	96.6
[0.30, 0.40]	97.2	92.2	94.7	[0.30, 0.40]	97.0	97.0	97.0	[0.30, 0.40]	94.8	98.0	96.4
[0.30, 0.75]	98.4	97.6	98.0	[0.30, 0.75]	96.6	99.0	97.8	[0.30, 0.75]	98.2	99.2	98.7
[0.50, 0.75]	100.0	97.8	98.9	[0.50, 0.75]	98.8	100.0	99.4	[0.50, 0.75]	99.6	99.4	99.5
[0.50, 0.95]	100.0	99.0	99.5	[0.50, 0.95]	97.8	100.0	98.9	[0.50, 0.95]	99.6	99.8	99.7
[0.70, 0.75]	100.0	98.0	99.0	[0.70, 0.75]	99.0	100.0	99.5	[0.70, 0.75]	100.0	100.0	100.0
[0.70, 0.95]	100.0	98.4	99.2	[0.70, 0.95]	98.8	100.0	99.4	[0.70, 0.95]	100.0	100.0	100.0
[0.80, 1.00]	100.0	99.0	99.5	[0.80, 1.00]	99.0	100.0	99.5	[0.80, 1.00]	100.0	100.0	100.0
Total:	96.7	95.8	96.3	Total:	96.3	96.9	96.6	Total:	95.5	98.8	97.1

Table B.15: Material SiN

Detected Material:	SiN										
Additional Training Materials:	Caffeine, Chalk, Coffee Ground, Equal, Ovalbumin, DAST, NAC										
Additional Testing Materials:	O2Si, Sodium Chloride										
Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	88.0	65.4	76.7	[0.10, 0.40]	90.4	58.2	74.3	[0.10, 0.40]	80.0	71.0	75.5
[0.10, 0.75]	89.4	74.6	82.0	[0.10, 0.75]	88.6	77.6	83.1	[0.10, 0.75]	80.2	86.4	83.3
[0.20, 0.40]	92.2	66.2	79.2	[0.20, 0.40]	91.6	68.8	80.2	[0.20, 0.40]	80.0	81.8	80.9
[0.20, 0.75]	95.2	76.2	85.7	[0.20, 0.75]	94.6	83.0	88.8	[0.20, 0.75]	85.6	91.6	88.6
[0.30, 0.40]	98.8	66.8	82.8	[0.30, 0.40]	96.2	75.6	85.9	[0.30, 0.40]	87.4	88.8	88.1
[0.30, 0.75]	97.4	80.0	88.7	[0.30, 0.75]	91.8	91.2	91.5	[0.30, 0.75]	93.0	92.6	92.8
[0.50, 0.75]	98.8	81.8	90.3	[0.50, 0.75]	97.2	98.2	97.7	[0.50, 0.75]	97.0	97.8	97.4
[0.50, 0.95]	99.2	81.8	90.5	[0.50, 0.95]	97.8	98.4	98.1	[0.50, 0.95]	98.4	98.0	98.2
[0.70, 0.75]	99.0	78.4	88.7	[0.70, 0.75]	97.4	99.4	98.4	[0.70, 0.75]	97.0	99.2	98.1
[0.70, 0.95]	99.4	79.8	89.6	[0.70, 0.95]	97.2	100.0	98.6	[0.70, 0.95]	98.6	99.6	99.1
[0.80, 1.00]	98.2	80.0	89.1	[0.80, 1.00]	98.2	100.0	99.1	[0.80, 1.00]	98.0	98.2	98.1
Total:	96.0	75.5	85.8	Total:	94.6	86.4	90.5	Total:	90.5	91.4	90.9

Table B.16: Material Albumin

Detected Material:	Albumin										
Additional Training Materials:	Coffee Ground, Equal, O2Si, Sodium Chloride, H2S, MgO, SiN										
Additional Testing Materials:	Talc, HgS										
Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	67.2	67.8	67.5	[0.10, 0.40]	72.4	46.8	59.6	[0.10, 0.40]	53.2	92.6	72.9
[0.10, 0.75]	76.8	88.2	82.5	[0.10, 0.75]	79.6	83.6	81.6	[0.10, 0.75]	72.4	95.8	84.1
[0.20, 0.40]	77.0	79.6	78.3	[0.20, 0.40]	88.6	58.6	73.6	[0.20, 0.40]	67.0	95.8	81.4
[0.20, 0.75]	91.0	89.0	90.0	[0.20, 0.75]	85.0	94.4	89.7	[0.20, 0.75]	86.4	96.8	91.6
[0.30, 0.40]	88.0	83.4	85.7	[0.30, 0.40]	89.4	84.2	86.8	[0.30, 0.40]	81.4	93.4	87.4
[0.30, 0.75]	96.2	96.2	96.2	[0.30, 0.75]	91.4	97.8	94.6	[0.30, 0.75]	92.0	98.4	95.2
[0.50, 0.75]	99.4	97.2	98.3	[0.50, 0.75]	97.0	100.0	98.5	[0.50, 0.75]	99.2	99.8	99.5
[0.50, 0.95]	99.6	98.4	99.0	[0.50, 0.95]	94.0	100.0	97.0	[0.50, 0.95]	98.4	100.0	99.2
[0.70, 0.75]	100.0	98.4	99.2	[0.70, 0.75]	99.0	100.0	99.5	[0.70, 0.75]	99.6	100.0	99.8
[0.70, 0.95]	100.0	97.0	98.5	[0.70, 0.95]	96.0	100.0	98.0	[0.70, 0.95]	100.0	100.0	100.0
[0.80, 1.00]	100.0	98.8	99.4	[0.80, 1.00]	99.6	100.0	99.8	[0.80, 1.00]	100.0	100.0	100.0
Total:	90.5	90.4	90.4	Total:	90.2	87.8	89.0	Total:	86.3	97.5	91.9

Table B.17: Material Ovalbumin

Detected Material:	Ovalbumin		
Additional Training Materials:	Baking Soda, Caffeine, Chalk, Coffee Ground, O2Si, Sodium Chloride, Talc		
Additional Testing Materials:	Cellulose, Equal		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	90.4	52.8	71.6	[0.10, 0.40]	91.6	56.2	73.9	[0.10, 0.40]	89.6	64.0	76.8
[0.10, 0.75]	90.0	71.8	80.9	[0.10, 0.75]	93.8	73.4	83.6	[0.10, 0.75]	91.2	76.4	83.8
[0.20, 0.40]	97.6	57.6	77.6	[0.20, 0.40]	96.8	67.6	82.2	[0.20, 0.40]	92.6	74.6	83.6
[0.20, 0.75]	96.8	70.2	83.5	[0.20, 0.75]	94.8	82.0	88.4	[0.20, 0.75]	94.2	80.0	87.1
[0.30, 0.40]	98.6	63.6	81.1	[0.30, 0.40]	99.0	76.0	87.5	[0.30, 0.40]	98.6	72.2	85.4
[0.30, 0.75]	98.2	80.2	89.2	[0.30, 0.75]	96.4	91.8	94.1	[0.30, 0.75]	98.6	89.8	94.2
[0.50, 0.75]	99.8	85.0	92.4	[0.50, 0.75]	98.6	98.6	98.6	[0.50, 0.75]	99.2	97.4	98.3
[0.50, 0.95]	100.0	85.6	92.8	[0.50, 0.95]	99.0	99.4	99.2	[0.50, 0.95]	99.6	96.6	98.1
[0.70, 0.75]	99.8	84.2	92.0	[0.70, 0.75]	99.0	99.8	99.4	[0.70, 0.75]	99.8	99.0	99.4
[0.70, 0.95]	99.8	93.4	96.6	[0.70, 0.95]	99.6	100.0	99.8	[0.70, 0.95]	99.8	99.4	99.6
[0.80, 1.00]	100.0	97.6	98.8	[0.80, 1.00]	99.6	99.8	99.7	[0.80, 1.00]	99.8	100.0	99.9
Total:	97.4	76.5	87.0	Total:	97.1	85.9	91.5	Total:	96.6	86.3	91.5

Table B.18: Material Trehalose

Detected Material:	Trehalose		
Additional Training Materials:	Coffee Ground, Equal, O2Si, Sodium Chloride, Talc, ADP, Glucose		
Additional Testing Materials:	Caffeine, DAST		

Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	94.6	52.0	73.3	[0.10, 0.40]	91.2	54.2	72.7	[0.10, 0.40]	81.6	87.2	84.4
[0.10, 0.75]	93.8	57.6	75.7	[0.10, 0.75]	91.8	67.4	79.6	[0.10, 0.75]	86.6	92.2	89.4
[0.20, 0.40]	95.4	63.0	79.2	[0.20, 0.40]	93.6	68.6	81.1	[0.20, 0.40]	90.6	87.4	89.0
[0.20, 0.75]	94.4	72.4	83.4	[0.20, 0.75]	94.4	79.8	87.1	[0.20, 0.75]	89.2	94.2	91.7
[0.30, 0.40]	94.6	70.0	82.3	[0.30, 0.40]	93.8	76.6	85.2	[0.30, 0.40]	90.0	92.8	91.4
[0.30, 0.75]	95.8	77.6	86.7	[0.30, 0.75]	92.6	90.4	91.5	[0.30, 0.75]	92.2	96.6	94.4
[0.50, 0.75]	96.0	74.0	85.0	[0.50, 0.75]	92.4	98.6	95.5	[0.50, 0.75]	93.8	100.0	96.9
[0.50, 0.95]	96.6	80.8	88.7	[0.50, 0.95]	94.0	99.4	96.7	[0.50, 0.95]	93.2	100.0	96.6
[0.70, 0.75]	97.0	76.4	86.7	[0.70, 0.75]	94.2	100.0	97.1	[0.70, 0.75]	92.8	99.8	96.3
[0.70, 0.95]	97.8	73.4	85.6	[0.70, 0.95]	92.6	100.0	96.3	[0.70, 0.95]	94.4	99.4	96.9
[0.80, 1.00]	99.0	80.2	89.6	[0.80, 1.00]	94.8	100.0	97.4	[0.80, 1.00]	95.8	99.2	97.5
Total:	95.9	70.7	83.3	Total:	93.2	85.0	89.1	Total:	90.9	95.3	93.1

Table B.19: Overall Averages

OVERALL AVERAGES											
Threshold using Geometric Means				Threshold using int. of Gaussians				SVM Approach			
[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC	[ll, ul]	TP	TN	CC
[0.10, 0.40]	78.6	73.5	76.1	[0.10, 0.40]	86.0	60.7	73.4	[0.10, 0.40]	74.3	80.3	77.3
[0.10, 0.75]	83.6	82.7	83.1	[0.10, 0.75]	88.3	75.0	81.6	[0.10, 0.75]	81.3	87.4	84.3
[0.20, 0.40]	85.0	77.8	81.4	[0.20, 0.40]	90.0	69.9	80.0	[0.20, 0.40]	82.0	83.1	82.5
[0.20, 0.75]	89.6	84.1	86.8	[0.20, 0.75]	91.5	80.7	86.1	[0.20, 0.75]	86.3	88.9	87.6
[0.30, 0.40]	89.0	79.9	84.5	[0.30, 0.40]	91.9	76.8	84.3	[0.30, 0.40]	86.4	86.5	86.4
[0.30, 0.75]	92.8	86.7	89.8	[0.30, 0.75]	93.3	86.4	89.9	[0.30, 0.75]	91.6	91.5	91.6
[0.50, 0.75]	95.9	87.5	91.7	[0.50, 0.75]	94.9	91.3	93.1	[0.50, 0.75]	94.6	94.1	94.4
[0.50, 0.95]	95.4	88.9	92.1	[0.50, 0.95]	94.7	90.4	92.6	[0.50, 0.95]	94.1	94.2	94.2
[0.70, 0.75]	96.5	87.9	92.2	[0.70, 0.75]	95.1	92.0	93.5	[0.70, 0.75]	94.8	95.9	95.3
[0.70, 0.95]	95.6	88.7	92.1	[0.70, 0.95]	95.4	90.0	92.7	[0.70, 0.95]	94.3	94.6	94.4
[0.80, 1.00]	95.4	89.1	92.3	[0.80, 1.00]	95.9	90.7	93.0	[0.80, 1.00]	93.9	94.7	94.3
Total:	90.7	84.3	87.5	Total:	92.4	82.2	87.3	Total:	88.5	90.1	89.3

Appendix C

Extraction Algorithm for sEMG Signals

In this appendix we describe the extraction algorithm implemented for detecting and extracting sEMG signals. As it was pointed out in chapter 8, before applying the GUSSS method we first need to get the main sEMG signal out of a recorded signal. Figure C.1 illustrates the problem of a recorded signal and the goal of the algorithm. Note there is an initial time period when there is only noise. At one point, the muscle activity generates the actual sEMG signal of interest. Then, when the muscle relaxes, the sEMG signal vanishes and we record noise once more. We want to detect the muscle activity and extract the corresponding signal.

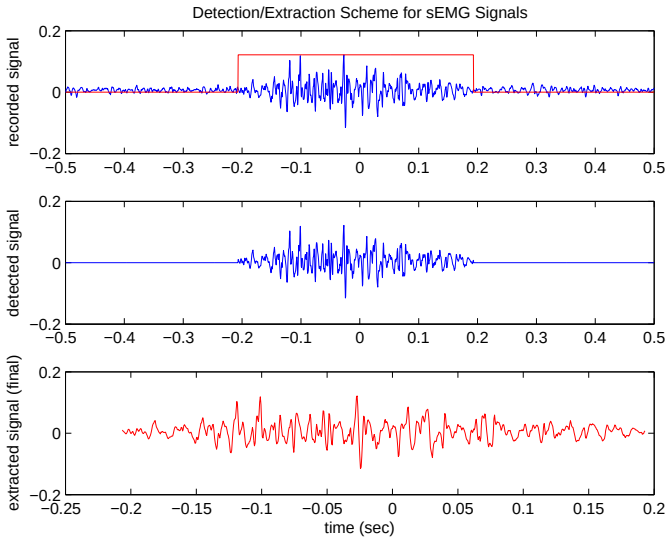


Figure C.1: Complete recorded signal and extracted sEMG signal.

The first signal is the recorded signal, consisting of 1 second worth of data. Note that the actual sEMG signal starts after approximately 200 ms, and lasts for about 400 ms. The second signal shows how the signal values outside the detection or activation time window are set to zero. The third signal is the extracted signal. It consists of 400 ms worth of data (this time window is a parameter of the algorithm).

C.1 Extraction Algorithm

Given a recorded signal $y = y(t)$, we first obtain the rectified signal $r = |y(t)|$. Figure C.2 illustrates both y and r.

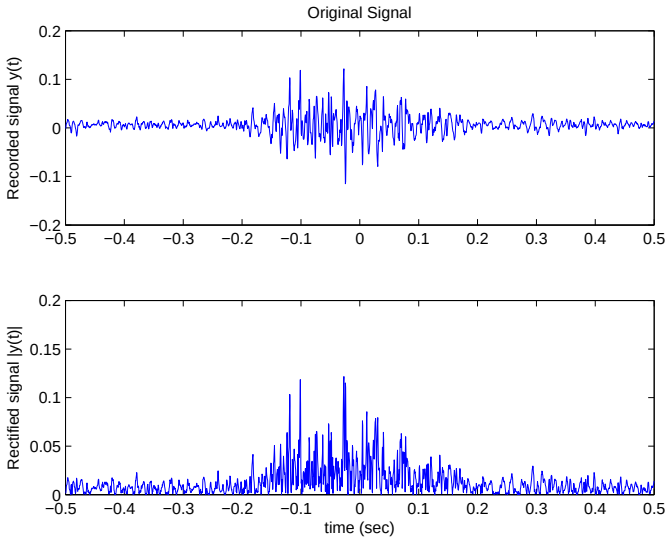


Figure C.2: Recorded signal and its rectified version.

We then compare r to a threshold value θ. This value is related to the noise level detected over a period of time when no sEMG signal is present. In essence, we record noise over a short period of time. Then, we calculate the maximum value of that noise, n_{max}. The threshold is then calculated as $\theta = kn_{max}$, where $k > 1$ is a constant. The top plot in Figure C.3 shows the rectified signal r and the threshold. The bottom plot, called *detection* plot, shows the places where the rectified signal surpasses the threshold level. This detection plot consists of zeros where $r \leq \theta$, and ones where $r > \theta$. It can be seen that there are gaps, because of the oscillating nature of the signal. There are also isolated peaks greater than the threshold, which are most likely due to noise. These are to be removed.

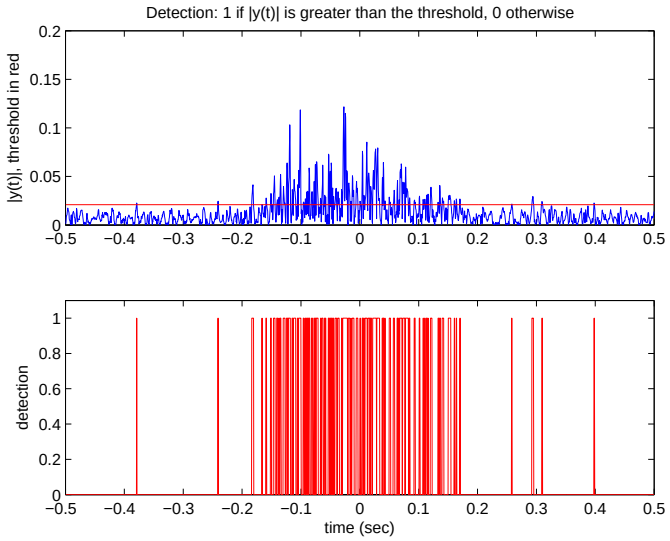


Figure C.3: Rectified signal and detection points.

The next step is to determine the *active* time interval(s). A certain density of ones in the detection values is required to consider a particular point (in y) to be part of the actual sEMG signal. Figure C.4 shows the recorded signal and the active intervals (so far) on top of it. The active intervals are there where there is a high density of ones, over a certain minimum time interval. Note that the isolated peaks at the beginning and the end of the bottom plot of figure C.3 were not considered for the active intervals.

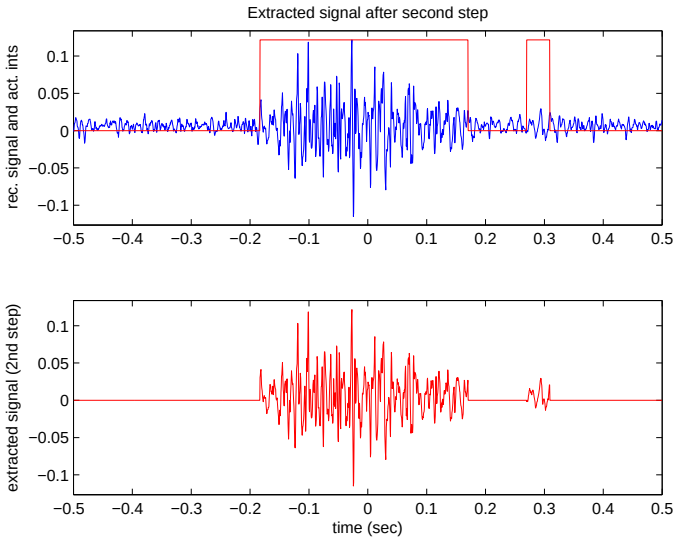


Figure C.4: Active Time Intervals.

Finally, we use the active intervals to extract the sEMG signal. Note in the figure above the small active interval on the right. In cases like that we consider them if they are close enough to the main active interval, or if they are long enough not to be regarded as noise. Otherwise, those small active intervals are ignored. Figure C.5 shows the final result of the algorithm. The small active interval was disregarded in this case. Note also that the final active interval was a little extended on both sides so to have a smoother extracted sEMG signal. With this we avoid sudden step-like beginnings of the extracted signals, or suddenly interrupted ones, as can be observed in the bottom plot in Figure C.4.

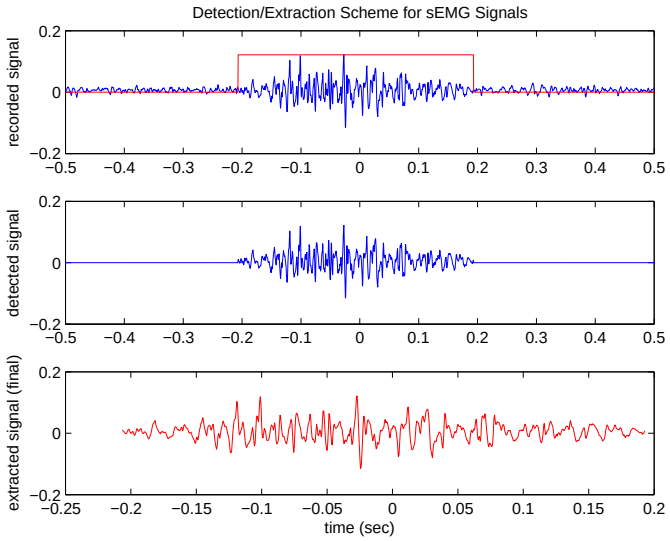


Figure C.5: Final extracted sEMG signal.