The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: A Statistical Validation of the Individuality of Guns Using 3D Images of Bullets

Author(s): Benjamin Bachrach

Document No.: 213674

Date Received: March 2006

Award Number: 1997-LB-VX-0008

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

A Statistical Validation of the Individuality of Guns Using 3D Images of Bullets

Grant Number: 97-LB-VX-0008

Final Report

This project was supported under award number 97-LB-VX-0008 from the National Institute of Justice Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Justice.

1. Table of Contents

2. Abstract

Weapon identification, its procedures and methodologies, have been developed over the past 100 years. These procedures are routinely used by firearms examiners and are the basis of their testimony in court. As currently practiced, these procedures involve a firearms examiner looking at the surface of bullets and attempting to determine whether they were fired by the same gun. In reaching such conclusions, the firearms examiner relies mostly on his/her training and judgment, making current matching procedures mostly subjective. The development of DNA identification techniques and the level of accuracy achievable in the estimation of error rates associated with DNA identification has raised the expectations of the quantitative precision that may be achieved in forensic analysis. Furthermore, recent Supreme Court decisions such as Daubert and Kumho are making it increasingly necessary to further formalize the presentation of scientific evidence in court. The subjective nature of current identification criteria, together with the inability of existing matching methodologies to estimate the probability of error associated with identification may pose a serious problem for the use of firearms evidence in court.

The present study was conducted by Intelligent Automation Inc. (IAI) under the support of the National Institute of Justice (NIJ). The first objective of this study was to improve on the state of the art of automated ballistic analysis systems, and to make such advances available to the law enforcement community. The second objective of this project was to develop and validate methodologies for ballistic identification, including the estimation of the probability-of-error in the identification process. Automated ballistic analysis systems are specifically designed for the objective comparison of large numbers of samples, making them an ideal instrument for the development of objective performance bounds. The development of such procedures reinforces the scientific foundations of ballistic evidence to be presented in court.

The scope of this study was considerable. The barrels used in this study were selected to span the spectrum of weapons commonly found in crime scenes. Over the duration of this study, more than 2800 bullets were fired by over 100 barrels of 9 different brands. The bullets fired by these barrels were retrieved and compared, and the results have been statistically analyzed. The effect of a variety of factors such as barrel manufacturing quality, bullet brand, barrel wear, number of control bullets, etc. have been analyzed and quantified.

3. Executive Summary

3.1 Introduction

In August 1997 the National Institute of Justice (NIJ) awarded Intelligent Automation Inc. (IAI) Grant Number: 97-LB-VX-0008 under the application title "Ballistics Matching Using 3D Images of Bullets and Cartridge Cases." The main purpose of this grant was to determine whether the use of three-dimensional data (3D data) of the surface of a bullet or cartridge case could be exploited to improve the performance of existing image-based automated ballistic comparison systems. In addition to the support provided by the National Institute of Justice towards the development of a 3D-based ballistic analysis system, IAI was also supported by a grant from the National Science Foundation (Grant Number DMI-9801361). Thanks to the support of these two agencies, IAI developed SCICLOPS™, a complete 3D-based ballistics identification system (see Figure 1). SCICLOPS™ made its public debut at the 30th Conference of the Association of Firearms and Tool mark Examiners (AFTE), in July 1999 in Williamsburg, Virginia. The debut of the system generated considerable interest because of its innovative approach to ballistic identification.

SCICLOPS™ was the first fully functional, fully automated, 3D-based ballistic analysis system ever developed. SCICLOPS™ incorporates the ability to automatically acquire, process and compare bullets in pristine condition (the first implementation of the system was not suitable for dealing with damaged bullets). The performance and originality of this system was so impressive, that Forensic Technology Incorporated (FTI), a world leader in the development and commercialization of automated ballistic analysis systems immediately expressed considerable interest in establishing a collaboration between IAI and FTI with the purpose of developing a 3D-based automated ballistic analysis system to be commercialized in the period of a few years. FTI's interest in 3D technology originated from the fact that FTI's analysis system, the Integrated Ballistic Analysis System (IBIS) operates using two-dimensional data (2D) data, and 3D offered the potential to improve their system's performance. In January 2000 FTI became the only manufacturer of automated ballistic analysis systems to be commercialized in the US. In May 2000 FTI and IAI reached an agreement to develop a commercial system which would integrate both 2D and 3D surface data. A prototype of such system was unveiled to the firearms examiner's community at the main exhibit floor of AFTE 2003 as part of the FTI booth (see Figure 2). Together with the unveiling of the prototype, a formal presentation on the new technology was given to the entire body of firearms examiners. By the end of 2004, FTI began the commercialization of BulletTRAX™-3D, a 3D based ballistic analysis system developed as a result of the collaboration between IAI and FTI (see Figure 3). BulletTRAX™-3D has received high praise from the firearms examiner's community, and brings the benefits in performance of topographical analysis of firearms evidence to the law enforcement

Figure 1: SCICLOPS™

community.

In parallel to these events, in September 2003, NIJ awarded IAI an extension of Grant Number: 97-LB-VX-0008 under the application title "A Statistical Validation of the Individuality of Guns Using 3D Images of Bullets." The purpose of this extension was to develop objective methodologies for ballistic identification, and to validate these methods through statistical means; including the estimation of the probability-of-error in the identification process. Furthermore, as part of this project, the effect of variables such as manufacture quality, ammunition and barrel wear on the probability of error was to be evaluated. The central motivation for the additional funding was to make use of the already developed 3D-based automated ballistic analysis system together with sound statistical techniques to further cement the scientific validity of the premises of firearms identification.

Figure 2: 3D-based Ballistic Analysis System Prototype Unveiled at AFTE 2003 as Part of FTI's Booth.

3.2 Scope of the Study

The scope of the present validation study is unprecedented in the arena of firearms examination. Over the three years of its duration, a new upgraded ballistic analysis platform capable of handling both pristine and damaged bullets was developed and manufactured. The barrels used in this study were selected to span (as best as possible) the spectrum of weapons commonly found in crime scenes. More than 2800 bullets fired by over 100 barrels of 9 different brands were collected over the duration of this study. A Mikrosil cast of each of the barrels involved in this evaluation was created prior to their firing, so as to preserve their original condition. Each of the bullets fired as part of this study was manually engraved prior to their being fired (so that the chance of "mixing up" the bullets would be minimized), fired into a water tank, and manually retrieved from these tanks. This process took place over a period of more than two years, in dozens of visits to volunteer firearms examiner's facilities who made their water tanks and time available for this purpose. Three different organizations assisted us in this

Figure 3: BulletTRAX™-3D

project: Washington State Police (thanks to the support of Evan Thompson), the Federal Bureau of Investigation Forensic Laboratory in Quantico VA, (thanks to the support of Paul Tangren and other FBI personnel), and Baltimore County Police (thanks to the support of Michael Thomas and Mark Ensor). We are extremely grateful to these firearms examiners who were willing to volunteer their time to this effort. This project would not have been possible without them.

Each of the bullets fired by these barrels were retrieved and compared using the 3D-based system developed for this purpose, and the results have been statistically analyzed. The volume of data available for analysis was so great, that it was necessary to develop significant amounts of software for the purpose of analyzing this data. Both the barrels and bullets collected as part of this effort have been preserved for future studies and analysis. Such provision has already proved to be of value. A portion of the test bullets created as part of this study are going to be used as part of the NIJ sponsored study “Assessing the Feasibility, Accuracy, and Technical Capability of a National Ballistic Database.” This study is being conducted by the National Academy of Sciences (NAS). Intelligent Automation Inc. participated in this study in collaboration with the National Institute of Standards and Technology (NIST) with members of NIST’s Office of Law Enforcement Standards (OLES), Manufacturing Engineering Laboratory (MEL), and Information Technology Laboratory (ITL). The Principal Investigator of this study, Dr. Benjamin Bachrach as well as Dr. Dan Xiang of IAI have played an active role in the NAS study as well.

3.3 Findings Summary

This study was structured into three main components. The first component dealt with the effect of barrel wear. The second component of the study dealt with the development of methodologies to address two main issues: a) the evaluation of the degree of individuality of barrels by looking at the bullets fired through them, and b) the estimation of the probability of error in bullet-to-barrel classifications. Both these two components of the study were analyzed using bullets in pristine condition. The third component of the study focused on the degree to which the conclusions of the previous sections can be applied to damaged bullets.

Barrel Wear Study:

As part of this study, two types of wear effect were considered. The first type of barrel wear was the wear which takes place over the first few shots of the life of the barrel. In this context, the questions of interest were: Are the features transferred to the very first bullets fired by a barrel any different than those fired later? The second type of barrel wear under consideration was the wear which takes place between any two bullets fired by a given barrel. In this context, the question of interest was: is there a difference between comparing the n^{th} and $(n+1)^{th}$ bullet fired by a barrel as opposed to the n^{th} and $(n+100)^{th}$ bullet fired by a barrel? In other words, does the proximity between firings make a difference? Although some wear effect could be detected in both of these scenarios, these effects do not appear to be significant enough to prevent correct identification.

It was decided to consider the question of barrel wear at the beginning of the study in order to prevent potential wear effects from contaminating other aspects of the study. As a result of the Barrel Wear Study it became clear that any concern about possible wear effects could be avoided by test firing the barrels approximately 5 – 10 times before continuing with the remaining parts of the study. In order to guarantee that no wear effects would permeate the remainder of the study, all barrels were fired 30 times before we began to collect sample bullets for the next two phases of the study.

Individuality and Classification for Pristine Bullets Study:

This portion of the study (also referred to as the “Pristine Bullet Study”) was by far the lengthiest both in terms of the number of bullets involved and the amount of analysis performed. Methodologies for the assessment of barrel individuality and bullet-to-gun identification were developed, and empirically tested. For this portion of the study, the questions of interest were: What procedures can be employed to validate that the features transferred between barrel and bullet are repeatable? Can it be shown that these features indeed repeat between firings? The results of this section were fairly consistent with the data obtained from the Barrel Wear Study. In other words, in the vast majority of cases, barrels brands which displayed good individuality as part of the Barrel Wear Study also displayed good individuality in this portion of the study. The second component of this portion of the study was the bullet-to-gun classification study. The questions of interest in this portion of the study were: Given an evidence bullet and a number of control bullets, what procedures can be employed to correctly identify the evidence bullet as matching or not matching the control bullets? What is the probability of error associated with these procedures? Which variables affect the probability of error? Not surprisingly, those barrel brands which displayed good individuality characteristics also displayed good classification characteristics.

Although there can be significant variations of performance (in terms of probability of error) for barrels of the same brand, most barrels of a given brand performed in a similar manner. This leads us to conclude that the barrel manufacture is the most dominant factor in both the individuality and classification performance of the bullets fired by it. For example, bullets fired by Ruger, Beretta and Smith & Wesson barrels could be identified with very low probability of error. Bullets fired by Taurus and Browning could be identified, but with somewhat larger probabilities of error. Finally, the ability of the system to identify bullets fired by HiPoint or SIG Sauer barrels was very limited. Bullets fired by Glock were virtually impossible to analyze, due to the peculiar rifling characteristics of the barrel (polygonal rifling). This was not a surprise; Glock barrels are known for requiring a different approach for identification. Time constraints prevented us from implementing the software necessary for this approach.

Next to the barrel manufacture, the ammunition brand (manufacture) seems to play the most dominant role in classification performance. As part of this study we used two brands of ammunition; Winchester and Remington. The classification performance using Winchester ammunition was in the vast majority of cases better than that achieved with Remington ammunition. Although a systematic study of the manufacture parameters which affect the classification performance is beyond the scope of this study, we conjecture that the dominant parameters affecting the manner in which features are consistently transferred between barrel and bullets are the bore machining accuracy and consistency (specially bore diameter) and the surface finish of the interior of the barrel. In the case of the bullet, we believe the dominant parameters to be the bullet diameter, and the hardness of the “jacket.”

Both barrel and bullet manufacture parameters are beyond the control of the firearms examiner. A parameter which is within his/her control is the number of control bullets test fired to perform the identification. The number of control bullets used in the identification can play a significant role in the classification probability of error. As the number of control bullets is increased, the probability of identification error decreases. However, the classification benefits associated with the increase of control bullets quickly reaches a point of diminishing returns.

Damaged Bullets Study:

The purpose of the Damaged Bullets Study was to test whether the conclusions and performance achieved in the case of pristine bullets could be extended to the case of damaged bullets. This is an important question because most bullets found in a crime scene will show some degree of damage. The acquisition challenges presented by a damaged bullet are mainly due to the fact that the bullet is no longer cylindrical, while the acquisition hardware was optimized for cylindrical targets. As can be expected, there is some significant deterioration of the classification performance when damaged bullets were used in place of pristine bullets. However, it is difficult to determine at this point to what extent the deterioration in performance was due to the deformation of the bullets, and to what extent it was due to the fact that these bullets are significantly more difficult to acquire in the current system. This additional degree of difficulty may have caused poorer data to be acquired, degrading the performance.

At the end of the day, those barrels for which the performance was good with pristine bullets still performed well with damaged bullets. Although the identification of damaged bullets is more challenging than that of pristine bullets, the statistical evidence indicates that it is possible to link a damaged bullet to the barrel which fired it.

4. Technical Report

4.1 Introduction

The ability to determine whether an evidence bullet was fired by a suspect gun can be extremely important in many criminal cases. Such identification is possible because the striations found on the surface of fired bullets are imprinted on them by microscopic imperfections found in the gun's barrel. The interior of a gun's barrel (see Figure 4) is manufactured with "grooves" whose purpose is to force the rotation of the bullet as it leaves the barrel of the weapon. By forcing the bullet to rotate during flight, a gyroscopic effect similar to that imparted by a quarterback on a football is generated, improving the accuracy and range of the projectile. These grooves (and their counterpart, "lands") in turn imprint groove and land impressions on the surface of the bullet (see Figure 4). Together with these impressions, imperfections on the barrel surface are incidentally transferred to the bullet's surface (striations). Because these imperfections are randomly generated during the barrel's manufacture, no two barrels have the same pattern of imperfections. These patterns of imperfections, therefore, amount to a "signature" that each barrel imprints on each of the bullets fired through it.

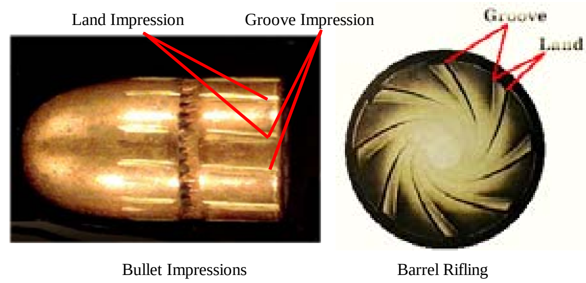


Figure 4: Generation of Striations on Bullets

One of the main tasks of the firearms examiner is to determine whether the patterns of imperfections found on two bullets were generated by the same barrel (gun). Until recently, the comparison of such patterns could only be made manually; i.e., by a firearms examiner inspecting a pair of bullets under a comparison microscope. The comparison microscope is an optical instrument which allows the examiner to manipulate and "line up" images of two bullets in an attempt to identify coinciding striations. The left side of Figure 5 shows a common such comparison microscope. The right side shows a typical image of a pair of matching land impressions as seen through the microscope.

Over the last ten years, computer aided ballistic analysis systems have been developed as an important tool to aid the task of the firearms examiner. These systems consist of an imaging device (often a microscope equipped with a digital camera), a database to store such images, and the software necessary to process and compare these images. Currently, one such automated system has a prominent place in United States forensic laboratories, namely, the Integrated Ballistics Identification System (IBIS). The data acquired by this system is a 2D image or a photograph of the bullet's surface. Figure 6 shows such an image of a land impression. Notice the similarity between this image and the comparison microscope image shown on the right side of Figure 5 (the image shown in Figure 5 is taken at a higher magnification than the one in Figure 6). Automated ballistic analysis systems can perform tasks ranging from preliminary classifications of bullets, to ranking a database of bullets against a questioned bullet by degree of similarity. Moreover, computers can perform these tasks in a fraction of the time it would take a firearms examiner. However, existing automated ballistic analysis systems cannot

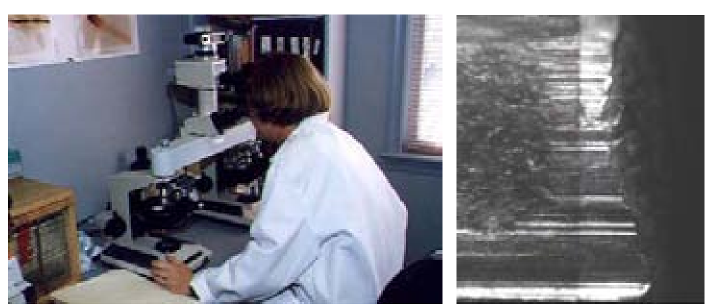


Figure 5: Comparison Microscope and Comparison Image

estimate the probability that a gun other than the suspect gun could have imprinted the pattern of striations found on the evidence bullet.

Figure 6: Digitized Image of Land Impression

Regardless of the of tool used by the firearms examiner (manual or automated), the determination of whether two bullets were fired by the same gun is currently made based on a firearms examiner looking at the surface of the bullets and reaching the conclusion - based mostly on his/her judgment - that they were (or were not) fired by the same gun. The **subjective** nature of current identification criteria, together with the inability of existing matching methodologies to estimate the probability of error associated with a match may in the future pose a serious problem for the use of firearms evidence in court.

In order to guarantee the future admissibility of ballistic evidence in court, it is necessary to develop objective standardized procedures to determine whether a given evidence bullet was fired by a suspect gun. These procedures should be founded on well-established scientific principles, and should therefore be verifiable, repeatable, and their probability of error quantifiable. Automated ballistic analysis systems are ideal tools to guarantee the objectiveness of the evaluation. The purpose of this study is to develop such procedures and to **validate** them through extensive statistical testing.

As part of the original effort of this grant, it was shown that the application of 3D methodologies for ballistics identification can be has used to enhance 2D characterization methodology used in the past decade. Moreover, preliminary results indicate that 3D methodologies can significantly outperform 2D based systems. Based on these observations, the objective of the present study present study are two: The first objective is to continue to improve on the state of the art in 3D-based comparison methodologies, and to make such improvements available to the law enforcement community. The second objective involves the development and use of a 3D based ballistic analysis system for the purposes of providing answers to the following questions:

a) Given 3D information from a bullet's surface, what quantitative criteria should be used to establish the individuality of a gun?
b) Given 3D information from a bullet's surface, what quantitative criteria should be used to establish that a suspect gun fired a given evidence bullet?
c) Once such criteria are developed, can the probability of a bullet/gun match being erroneous be estimated?

We begin this report with a brief description of the methods employed for the study. Notation and terminology were developed to facilitate the description of the comparison process through which computerized ballistic analysis systems determine the degree of similarity between two bullets. As will be discussed, the end results of this comparison process are "similarity measure values." Similarity measure values amount to a quantification of the degree of similarity between two bullets, and are at the core of the answers to questions a), b) and c). Therefore, much of our presentation revolves around the statistical characteristics of these values when obtained by comparing bullets fired by the same and different guns. The definition of these values and the methods used to evaluate individuality and classification are discussed in Section 4.2. The question of barrel wear is considered in Section 4.4. The issues of gun individuality and bullet-to-gun classification (or matching) are discussed in Sections 4.5 and 4.6. The damaged bullet results are included in Section 4.7. Finally, we include our conclusions in Section 5.

4.2 Project Design, Data and Methods

The project design, data and methods utilized were aimed at answering questions a) through c) posed in the previous section.

4.2.1 Project Design and Data

This project was designed in four well defined phases. The order of these phases was refined over the duration of the project, but their objectives remained the same. We begin by describing the main phases of this project, and the data generated and used in each of them. We then discuss the methods followed to complete these phases.

Phase I: Preliminary Tasks:
The first phase of the study consisted of three main tasks. These tasks were: a) The manufacture of the Acquisition Station, together with the development of the algorithms necessary for the acquisition and processing of damaged bullets (which had not been developed at the beginning of this study), b) The selection of guns models/makes to be used as part of the study, and c) The selection of the ammunition to be used as part of the study. From the inception of the study it was decided that nine barrel models/makes and two different brands of ammunition would be used.

Phase II: Barrel Wear Study:
The second phase of the study was the Barrel Wear Study. The purpose of this portion of the study was to assess the possible effect of barrel wear on the features transferred between a barrel and the bullets fired by it. It was decided that this portion of the study would be performed first so as to guarantee that any potential barrel wear issues would be understood and appropriately addressed for the remaining phases of the study. To this effect, one barrel of each of the barrel models/makes to be used in the study was dedicated for the evaluation of possible barrel wear effects.

Phase III: Pristine Bullets Study:
The third phase of the study was by far the most demanding phase (in terms of number of sample bullets and analysis) and it is referred to as the "Pristine Bullet Study." The purpose of this portion of the study is to address questions a) through c) discussed earlier for the case of pristine bullets. This phase of the study was further subdivided into two main components:

- Individuality Study: The purpose of the individuality study was to address question a) discussed in the previous section. Namely, to develop a methodology to validate whether the features transferred between a given barrel and the bullets fired by it were sufficiently **individual** to consider the barrel in question identifiable.

- Classification Study: The purpose of the classification study was to address questions b) and c) discussed in the previous section. Namely, to develop methodologies to answer the bullet-to-gun classification question: was this bullet fired by this barrel? (question b). An additional goal of this portion of the study was the means to estimate the probability of error associated with the different approaches (question c).

Phase IV: Damaged Bullets Study:
The fourth phase of the study is referred to as the "Damaged Bullet Study." The purpose of this portion of the study is to validate the degree to which the conclusions reached for pristine bullets apply to damaged bullets.

Bullet samples were created for phases II through IV of the study as follows:

Phase II: Barrel Wear Study:
For each of the nine barrel makes/models selected for the study, eleven barrel samples were obtained (either purchased or donated by the manufacturer). Out of the eleven barrel samples, one barrel was dedicated for the Barrel Wear Study. For each of these barrels, a total of 220 bullets were fired, out of which 80 bullets were retrieved using a water tank. These 80 bullets correspond to the bullets 01 through 50, bullets 101 - 110, bullets 201 - 210, and bullets 211 – 220 fired by each of these barrels. These order in which these bullets were fired was preserved so that possible wear effects could be assessed. All bullets fired for the Barrel Wear Study were of Winchester manufacture. The total number of bullets test fired (and retrieved) as part of this portion of the project was 720 bullets.

Phases III and IV: Pristine Bullet Study and Damaged Bullet Study:
For each barrel brand listed in selected for this study, ten barrels (except in the case of Taurus, where only five barrels were available) were used for these two portions of the study. Twenty four bullets were test fired by each of these 85 barrels. Of the 24 bullets, 12 were of Winchester manufacture and 12 were of Remington manufacture. For each of these sets of 12 bullets, 10 were retrieved in pristine condition and 2 bullets were fired in such a manner as to "damage" them. The total number of bullets test fired (and retrieved) as part of this portion of the project was 2,040 bullets.

Details of the test firing process for the bullets used in these studies can be found in Test Firing Protocol No. 1 (corresponding to the bullets collected for the Barrel Wear Study) and Test Protocol No. 2 (corresponding to the bullets collected for the Pristine Bullet Study and the Damaged Bullet Study) included in Appendix A of Progress Report No. 8.

4.2.2 Definitions and Notation

4.2.2.1 Definitions

We begin our presentation by reviewing the manner in which the current automated system operates. In comparing a pair of bullets, it is necessary to take in consideration all possible *relative orientations* between them (this applies to both automated systems and firearms examiners). Figure 7 shows the cross section of two bullets to be compared. In both of these cross sections, the Land Engraved Areas (LEAs) have been labeled. From Figure 7 one can observe that two bullets can be compared in a number of relative orientations. For example, one such orientation is consistent with comparing LEA 1 of bullet 1 with LEA 1 of bullet 2, LEA 2 of bullet 1 with LEA 2 of bullet 2, up to LEA 5 of bullet 1 with LEA 5 of bullet 2. This is in fact the orientation shown in Figure 7. However, if we "rotate" bullet 2 counter-clockwise by one LEA, the resulting relative orientation would be consistent with comparing LEA 1 of bullet 1 with LEA 2 of bullet 2, LEA 2 of bullet 1 with LEA 3 of bullet 2, up to LEA 5 of bullet 1 with LEA 1 of bullet 2. In other words, **because the pair of bullets under consideration has five rifling grooves, they can be compared in five possible relative orientations**.

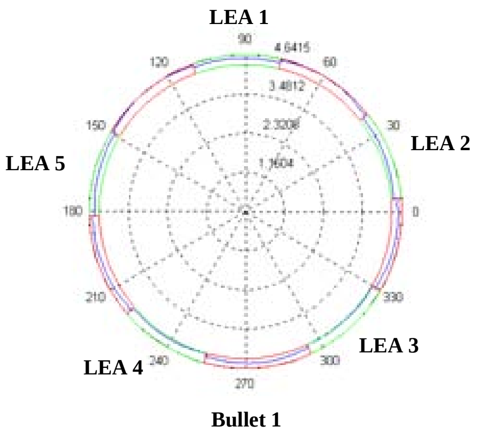


Figure 7: Relative Orientation between a Pair of Bullets

Given a pair of bullets to be compared, the current automated system evaluates each possible relative orientation between these bullets by comparing the appropriate pairs of LEAs. For each of these LEA-to-LEA comparisons, a *LEA-to-LEA similarity measure* is computed. These LEA-to-LEA similarity measures are weight-averaged to compute an *orientation similarity measure* for each possible orientation. In this manner, if a pair of bullets to be compared has *n* rifling impressions (and therefore *n* possible relative orientations,) *n* similarity measure values corresponding to each of the possible relative orientations are obtained. These orientation similarity measures are ranked to identify the best (highest) and second-best (second highest) orientation similarity measure for the pair of bullets under comparison. The relative orientation rendering the best similarity measure is assumed to be the relative orientation at which the two bullets under comparison are aligned. The *overall similarity measure* for the bullet pair under comparison is given by this value.

4.2.2.2 Notation and Terminology

To ease our discussion of the 3D ballistics matching procedures and methodologies developed for this project, it is helpful to first introduce the notation for bullet comparison and the terminology used for gun individuality and classification studies. We first introduce notation for bullet and sets of bullets. We denote a generic i^{th} bullet as

$$b_i \tag{1}$$

and a group of bullets fired by the same barrel as

$$b(gunModel, gunBarrel, \{orderedBullets\}) \tag{2}$$

where *gunModel* corresponds to one of the nine models under consideration, *gunBarrel* corresponds to the barrel number within those of the model specified by *gunModel*, and *{orderedBullets}* corresponds to the ordered set of bullets fired by the *gunBarrel* of the *gunModel* under consideration. As an example, *b(Ruger,1,{1,2,3,4})* refers to the first four bullets fired by Ruger barrel #1.

In the following discussion, it will often be convenient to use a shorthand notation for sets or groups of bullets. To this effect, instead of listing all *orderedBullets* as shown in Eqn (2), we denote

$$G_I(gunModel, gunBarrel) = b(gunModel, gunBarrel, I) \quad (3)$$

where I denotes an indexed set of bullets. It will be convenient to define four such indexed sets corresponding to groups of bullets which we will often refer to:

$$I_{brand}^{condition} \quad (4)$$

where the entry condition indicates pristine (p) or damaged (d) condition, and grand indicates either Remington (R) or Winchester (W) ammunition. For example, the set of Remington bullets in pristine condition will be denoted as I_R^p, while those in damaged condition will be labeled I_R^d. Using this notation, $G_{I_R^p}(Beretta,1)$ denotes all pristine Remington bullets fired by barrel 1 of model Beretta, while $G_{I_W^d}(Beretta,1)$ denotes all damaged Winchester bullets fired by barrel 1 of model Beretta.

We now introduce notation associated with similarity measures and sets of similarity measures. We denote the i^{th} orientation similarity measure obtained from the comparison of b_1 and b_2 as:

$$s_i(b_1, b_2) \quad (5)$$

where we denote the best orientation similarity measure is given by $s_1(b_1, b_2)$, the second best orientation similarity measure is given by $s_2(b_1, b_2)$, etc. Further, as discussed above, $s_i(b_1, b_2)$ is a weighted function of a vector of LEA-to-LEA similarity measure vector to be denoted:

$$\bar{s}_i(b_1, b_2) \quad (6)$$

The "bar" on top of the similarity measure will be used to indicate whether we are talking about a vector of LEA-to-LEA similarity measure or an orientation similarity measure (without bar). Without loss of generality, we will not repeat the notation for both LEA-to-LEA similarity measures and orientation similarity measures, since they are the same except for the "bar."

We extend the notation to similarity measures to that of sets of similarity measures, so that the following notation:

$$s_i(b(gunModel1, gunBarrel1, \{orderedBullets1\}, b(gunModel2, gunBarrel2, \{orderedBullets2\})) \quad (7)$$

refers to the set of i^{th} orientation similarity measures resulting from the comparison of the bullets *orderedBullets1* fired by barrel number *gunBarrel1* of model *gunModel1*, against the set of bullets *orderedBullets2* fired by barrel number *gunBarrel2* of model *gunModel2*.

The *best* and *second-best orientation similarity measures* sets include the similarity measure values obtained by comparing bullets fired by the **same gun** in their best and second best relative orientations. Notice that the best and second-best orientation similarity measure sets can be obtained for a **group** of guns of the same model, as long as the similarity measures included in these sets are restricted to those resulting from the comparison of bullets fired by the same gun. We denote the set of best similarity measure values resulting form the comparison of the indexed set of bullets *I* fired by barrels *{1,…,n}* of model *gunModel* as:

$$s_1(gunModel,\{1,...,n\},I) = \bigcup_{\forall i \in \{1,...,n\}} s_1(G_I(gunModel,i),G_I(gunModel,i)) \qquad (8)$$

The set of second-best similarity measure values resulting form the comparison of the indexed set of bullets *I* fired by barrels *{1,…,n}* of model *gunModel* is denoted:

$$s_2(gunModel,\{1,...,n\},I) = \bigcup_{\forall i \in \{1,...,n\}} s_2(G_I(gunModel,i),G_I(gunModel,i)) \qquad (9)$$

Similarly, we denote the set of matching similarity measure values resulting form the comparison of the indexed set of bullets *I* fired by barrels *{1,…,n}* of model *gunModel* by $s_m(gunModel,\{1,...,n\},I)$:

$$s_m(gunModel,\{1,...,n\},I) = \bigcup_{\forall i \in \{1,...,n\}} s_1(G_I(gunModel,i),G_I(gunModel,i)) \qquad (10)$$

while the set of non-matching similarity measure values associated with the comparison of the indexed set of bullets *I* fired by with *gunBarrels {1,…,n}* of model *gunModel* is denoted by $s_{\overline{m}}(gunModel,\{1,...,n\},I)$:

$$s_{\overline{m}}(gunModel,\{1,...,n\},I) = \bigcup_{\substack{\forall i,j \in \{1,...,n\} \\ i \neq j}} s_1(G_I(gunModel,i),G_I(gunModel,j)) \qquad (11)$$

As an example, the set of matching similarity measure values corresponding to Ruger barrels 1,…,5 obtained by comparing bullets of Remington manufacture will be denoted as $s_m(Ruger,\{1,...,5\},I_R^p)$. Having defined sets of matching and non-matching similarity measure values, we will refer to the *matching distribution,* a *best orientation distribution,* a *second best orientation distribution* and a *non-matching distribution* as the distribution of the respective sets of values.

Observations: Notice best orientation similarity measure sets (such as $s_1(Ruger,\{1,...,5\},I_R^p)$) and matching similarity measure sets (such as $s_m(Ruger,\{1,...,5\},I_R^p)$) are the same. This is simply for convenience, since in some contexts it makes more sense to talk in terms of matching and non-matching, while in others it is advantageous to talk in terms of best and second-best orientation similarity measures. Also, notice that these distributions depend on the ammunition used to compute the similarity values. Further, these distributions are always obtained by comparing bullets in pristine condition.

4.2.3 Methods

In this section we discuss the core methodologies used throughout the study in the evaluation of the different properties of interest (individuality and classification).

4.2.3.1 Gun individuality

Question a) of Section 4.1 can be further refined in the following manner:

Are the features transferred between a given barrel and the bullets fired by it sufficient in number and individuality to allow for the identification of all bullets fired by said barrel?

1) Given a suspect weapon with k rifling impressions, fire m control bullets (m to be determined according to desired level of significance).

2) After acquiring all control bullets, compute all lea-to-lea similarity measures.

3) Create two sets of similarity measures:
   a. Control bullet's best similarity measures (labeled r). This set will have $\left(\frac{m!}{(m-2)!\times 2!}\right)$ elements.
   b. Control bullet's second-best similarity measures (labeled w). This set will have $\left(\frac{m!}{(m-2)!\times 2!}\right)$ elements.

4) Perform statistical test to evaluate the following hypothesis:

   H_0: The probability distributions from which the samples arose are not different from one another,

   H_1: The samples arose from different probability distributions.

5) As a result of the statistical test, we obtain an estimate of the probability of error associated with rejecting the H_0 hypothesis (p-value). If the obtained p-value is lower than a pre-established significance level, the gun will be considered identifiable. If the obtained p-value exceeds the pre-established significance level, the gun will be considered non-identifiable.

Table 1: Procedure to test gun individuality

If presented with this question, a firearms examiner would fire a number of control bullets, and by inspection determine if the striations found on their surface reproduce consistently from control bullet to control bullet. To do so, the firearms examiner must first identify the matching orientation between every pair of control bullets, and then subjectively evaluate the degree of similarity of the matching impressions as compared to non-matching impressions. Table 1 describes an analogous procedure based on the best and second-best orientation distributions.

At the core of the procedure outlined in Table 1 the following reasoning: given a group of control bullets fired by the suspect gun, if the sets of best orientation and the second-best matching scores are statistically undistinguishable, it is not possible to identify the matching relative orientation between pairs of bullets from this group. If the matching orientation between bullets from the control group cannot be determined, matching them will not be possible. If the control bullets fired by the suspect gun cannot be matched, matching the evidence bullet to the control bullets it will be highly unlikely, making the barrel non-identifiable.

For any given gun model *gunModel*, barrel *gunBarrel* and ammunition, let us identify the set of control bullets by the index I_c, We compute the best and second-best similarity value datasets associated with the given control bullets:

$$r = s_1(gunModel, gunBarrel, I_c) \tag{12}$$

$$w = s_2(gunModel, gunBarrel, I_c) \tag{13}$$

The procedure described in Table 1 can also be viewed from a different perspective. We postulate that the second best similarity measure distribution is a good approximation of the non-matching distribution for guns of a given model. From this perspective, we are not only evaluating our ability to identify the best relative orientation between two bullets fired by the suspect gun, but we are also evaluating our ability to differentiate between bullets fired by the suspect gun and other guns of the same manufacture.

Notice that the procedure described in Table 1 requires the specification of a metric (such as a *p*-value) of similarity between distributions (see Step 4). There are a number of methodologies which can be applied to measure the degree of similarity between two distributions. At this point we cannot assume that any of these distributions are normal, therefore, the statistical tests to be used in Step 4) of Table 1 may not rely on the normality-of-data assumption (as will be seen in Section 4.6.2, the best orientation similarity distribution is not normal). We considered two possible approaches:

Rank Sum Test Approach: An effective approach to evaluate the statistical difference between to sets of data for which normality-of-data does not hold is the Rank Sum test. The Rank Sum test allows us to quantify the degree of statistical similarity/difference between sets of data The *p*-value obtained through this test provides an estimate of the probability of obtaining the computed set of best orientation similarity measures (labeled *r*) if the phenomenon that generated these coefficients had the same statistical distribution as that which generated the second-best orientation similarity measures (labeled *w*). We use the *p*-value as a metric of distance between distributions, where the distance is inversely proportional to the computed *p*-value.

Empirical Pe Approach: This approach is probably one of the most classic statistical methods used in conventional Hypothesis Testing problems. Having the empirically generated sets of best and second-best similarity measures for a barrel under consideration, it is possible to compute an optimal boundary or threshold such that if a given orientation yields an orientation similarity measure above the threshold, it is assumed to be the best relative orientation, while any orientation yielding a similarity measure below the threshold is classified as a being in a non-best orientation. The boundary or threshold value is selected to minimize the average probability of error (both false positive and false negative). Figure 8 shows a graphical representation of this approach, where two distributions are shown, a matching distribution corresponding to the best orientation similarity measures, and a non-matching distribution corresponding to the second-best orientation similarity measures. Having identified the optimal threshold, it is possible to estimate the probability of false positive best orientation identification, as well as false negative best orientation identification. We use the probability of error as a metric of distance between distributions, where the distance is inversely proportional to the empirically computed probability of error.

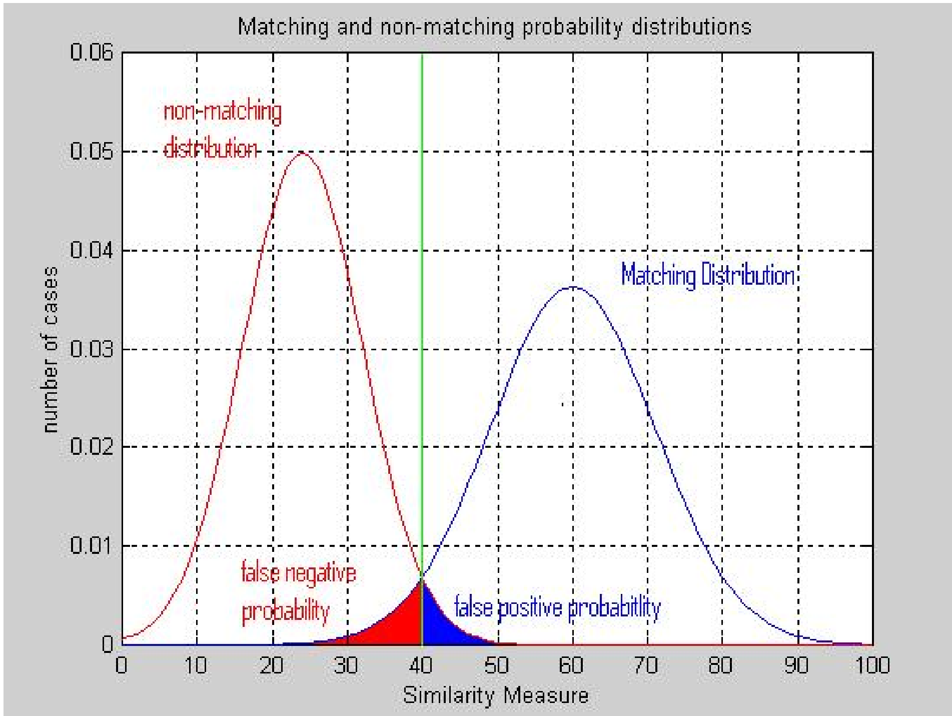


Figure 8: Empirical Estimation of Probability of Orientation Error

The main difference between these two approaches is that the empirical approach makes use of data from a relative large number of test fires, while the Rank Sum test allows us to consider a smaller amount of data, similar to that which firearms examiners currently use in this kind of determination. We begin by describing the manner in which we performed these tests.

4.2.3.2 Bullet-to-gun classification

Once the individuality of the suspect gun has been established to the satisfaction of the firearms examiner, the question of bullet-to-gun classification is equivalent to asking whether the degree of similarity between the evidence bullet and the control bullets – in the presumed matching orientation – merits the conclusion that both evidence and control bullets were fired by the same gun. Questions b) and c) of Section 4.1 can be further refined in the following manner:

Assuming the barrel under consideration satisfies the conditions posed by question a), does the degree of similarity found between the evidence bullet and the control bullets merit the conclusion that these bullets were fired by the same gun? What is the probability of error associated with the decision (either matching or non-matching)?

In order to make such determination, a firearms examiner would compare the evidence bullet against the control bullets, and attempt to identify matching orientations between them. Assuming that such orientations are identified, the firearms examiner would subjectively assess whether the degree of similarity between the evidence bullet and the control bullets merits the conclusion that all bullets were fired by the same gun. In making such assessment, the firearms examiner should not only consider the degree of similarity found between evidence and control bullets, but he/she should contrast this degree of similarity with that achievable by chance among different guns of the same model. To do so, a firearms examiner must accumulate considerable experience about virtually every gun he/she is called to identify.

1) Given a suspect weapon and an evidence bullet, fire *m* control bullets.

2) After acquiring all control bullets, compute best and second-best similarity measures obtained from the comparison of all control bullets among themselves.

3) Create three sets of similarity measures:
   a. Control bullet's best similarity measures (labeled *r*). This set will have $\left(\frac{m!}{(m-2)!\times 2!}\right)$ elements.
   b. Control bullet's second-best similarity measures (labeled *w*). This set will have $\left(\frac{m!}{(m-2)!\times 2!}\right)$ elements.
   c. Control bullets vs. evidence bullet best similarity measure (labeled *e*). This set will have *m* elements (assuming a single evidence bullet).

4) Perform two statistical tests to evaluate the following hypothesis:

   a. Evaluate similarity between the distribution of *r* and the distribution of set *e* through some metric (such as *p*-value).
   b. Evaluate similarity between the distribution of *w* and the distribution of set *e* through some metric (such as *p*-value).

5) Accept the hypothesis associated with the distribution *r* (match) or *w* (non-match) which best resembles that of set *e*. In other words, if set *e* is more similar to set *r*, classify the evidence bullet as matching the control bullets and conversely.

Table 2: Procedure to test bullet-to-gun classification

Ideally, an automated system should emulate this procedure by using both matching scores and non-matching scores. However, due to the fact that compiling a database of non-matching scores is a significant undertaking – because it would require obtaining and comparing control bullets from a large number of guns of the same model as the suspect gun – a procedure that relies on the similarity of the distributions of non-matching scores and the second-best matching scores is presented. Table 2 describes this procedure.

For any given gun model *gunModel* and barrel *gunBarrel*, let us identify the set of control bullets by the index I_c, (so that the control bullets are denoted $b(gunModel, gunBarrel, I_c)$) and the evidence bullet as b_e where $b_e \notin b(gunModel, gunBarrel, I_c)$. Let us compute the best and second-best similarity value datasets associated with the given control bullets:

$$r = s_1(gunModel, gunBarrel, I_c) \tag{14}$$

$$w = s_2(gunModel, gunBarrel, I_c) \quad (15)$$

And the “questioned” similarity values dataset resulting from the comparison of the evidence bullet and the control bullets:

$$e = \bigcup_{i \in I_c} s_1(b_e, G_{I_c}(gunModel, gunBarrel)) \quad (16)$$

At the core of the procedure outlined in Table 2 is the following reasoning: In order to conclude that the evidence bullet and the control bullets were fired by the same gun, the distribution of similarity measures obtained by comparing the evidence bullet against the control bullets – in their best orientation - should be significantly more similar to the distribution of the best orientation similarity measures obtained by comparing the control bullets among themselves, than to the distribution of non-matching scores resulting from the comparison of bullets fired by different barrels of the same make and model. As already mentioned, because of the difficulties associated with obtaining a representative set of non-matching scores, we approximate this distribution with the distribution of the second-best matching scores obtained by inter-comparing the control bullets.

Due to the fact that the distribution of matching scores is not normal (as will be shown in Section 4.6.2), the metric required in Step 4) should be obtained by using a statistical test which does not rely on the normality assumption. We considered three possible approaches:

Hard Threshold or Empirical Approach:
Given the distributions of the sets *r* and *w* as defined in Table 1, it is possible to compute the optimal threshold which minimizes (in an empirical sense) the probability of error associated with a classification decision for these two distributions. Let us denote this threshold by T_{opt}. Having obtained the optimal threshold, the mean of the set of similarity measures *e* is computed. The evidence bullet will be classified as a match if the mean of the similarity measure set *e* is greater than the threshold (or in other words, closer to the best orientation similarity measure distribution *r*); otherwise it will be classified as a non-match.

Closest Mean:
The closest mean criterion is based on the distance between the mean values of the different distributions under consideration. In other words, if $|\bar{r} - \bar{e}| < |\bar{w} - \bar{e}|$ (where $|\bullet|$ denotes absolute value of $\bullet$, and $\bar{\bullet}$ denotes mean of $\bullet$) the evidence is classified as matching the control bullets, otherwise it is classified as non-matching.

Normalized Closest Mean:
The normalized closest mean criterion is similar to the closest mean criterion, except that the “distances” are normalized by the appropriate standard deviations. In other words, if $|\bar{r} - \bar{e}| / \sigma(r) < |\bar{w} - \bar{e}| / \sigma(w)$, (where $\sigma(\bullet)$ denotes standard deviation of $\bullet$), the evidence is classified to be a match with the control bullets, otherwise it is classified as non-match with the control bullets.

Besides the classification approach, the probability of identification error will depend on variables such as barrel quality, ammunition manufacture, number of control bullets etc. In the following section we discuss the effect of all these variables in the resulting probability of error.

4.3 Preliminary Tasks

The main preliminary activities undertaken as part of this study were the manufacture of the 3D Acquisition Station, the selection of guns to be used as part of the study, and the selection of the ammunition to be used for the test firing. In this section we briefly discuss these tasks.

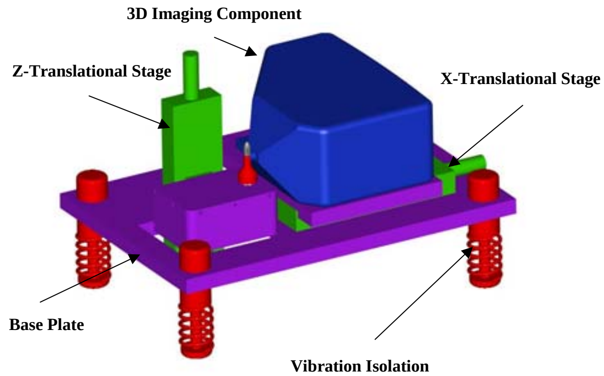


Figure 9: Preliminary Design of Acquisition Station (above), and Manufactured Acquisition Station (below)

4.3.1 Manufacture of Acquisition Station

Based on our experience and evaluation of SCICLOPS™, and based on test results of alternative 3D acquisition hardware, an upgraded 3D ballistic data acquisition system was designed and manufactured. Figure 9 shows both the preliminary design of the acquisition station and a photograph of the actual manufactured station. This design includes the 3D imaging component, X and Z translational stages, and a rotational stage found under the bullet holding mechanism.

This acquisition station is fully computer controlled, and allows for an automated and a manual mode of data acquisition. In either the automatic or the manual mode, the acquisition process starts with the user placing the bullet to be analyzed on the rotational stage. In the automatic acquisition mode, the operator simply selects the acquisition position on the bullet surface (along the bullet's rotational axis) and allows the system to perform the data acquisition process. In the manual mode (most commonly used for damaged bullets), the user adjusts the laser focal spot of the laser both along the bullet's rotational axis and with respect to the individual land impression to be acquired by controlling the X and Z translational stages. Then the system acquires the surface depth data of the individual land impression automatically. The user then proceeds with any other land impression of interest, and repeats the procedure.

4.3.2 Selection of Barrel Brands

In consultation with collaborating firearms examiners, a set of 9 gun models have been selected for this study. These gun models were selected based on the following criteria: a) frequency of association with crime scenes, b) availability of barrels for purchase, c) availability of reliable information regarding their manufacture, and d) degree to which the overall group of selected guns spans the spectrum of commonly used manufacturing techniques. The selection process began with a study of those gun models most commonly associated with crimes, and the manufacturers of those guns. An excellent source of information regarding weapons involved in the perpetration of crime is the Bureau of Alcohol Tobacco and Firearms' (BATF) National Tracing Center (NTC). The National Tracing Center maintains statistics of every gun that is "traced" as part of a crime investigation. At IAI's request, NTC provided the statistics of the twenty-five guns most often associated with crimes over the last twelve years. The list of candidate guns was narrowed based on the availability of guns for purchase, and the availability of information

	Manufacturer	Number of Barrels	Manufacturing Technique	Number of Impressions	Width of Impressions [mm]	Notes
1	Taurus	6	Gang Broach	6	1.3	Consecutive
2	Bryco	11	Gang Broach	6	1.3	Consecutive
3	Beretta	11	Gang Broach	6	2.0	Consecutive
4	HiPoint	11	Button Rifling	6 / 9	1.3/1.6	Consecutive
5	Glock	11	Hammer Forged	6	--	Standard
6	S&W	11	ECM	5	2.5	Standard
7	SIG	12	Hammer Forged	6	1.7	Consecutive
8	Browning	15	Hammer Forged	6	1.8	Sequencial
9	Ruger	11	Gang Broach	6	2.0	Consecutive

Table 3: Selected gun/barrel manufacturers, and manufacturing techniques

regarding the manufacturing techniques involved in the rifling of their barrels. The nine brands selected were a compromise between all these factors.

These gun models were chosen to represent the three most common barrel manufacturing techniques (Gang Broach, Button Rifling, and Hammer Forging). To the extent possible, barrels of different levels of manufacturing quality (poor, average, above average) were chosen for each of the manufacturing techniques under consideration, so that the effect of the quality of manufacture on the individuality of the barrel can also be studied. Table 3 shows the gun manufacturers selected for the study, their corresponding barrel manufacturing technique, the number of barrels obtained, the number of rifling impressions in these barrels, and an approximate land impression width. These parameters are relevant because the amount of data available for comparison is proportional to the overall land impressed area on each bullet. Notice that in the case of the HiPoint barrels two different land widths and number of impressions are recorded. The first number corresponds to the barrel used for the Barrel Wear Study, and the second corresponds to the barrels used in the Pristine and Damaged Bullet Study. Table 3 also includes information regarding the origin of these barrels. This information is coded in the "Notes" column as follows:

a) Consecutive: These barrels were consecutively manufactured as stated by the manufacturer (in principle, this implies that the different manufacturing processes were performed consecutively between barrels).
b) Sequential: these barrels were sequentially manufactured as stated by the manufacturer. As consecutively manufactured barrels require a significant amount of effort, manufacturers are not always willing to provide such level of assurance. The usual interpretation of sequentially manufactured barrels is that these barrels were created in close proximity to each other, and using the same tools, but not necessarily in sequence.
c) Standard: these barrels were purchased through a supplier.

4.3.3 Selection of Ammunition Brands

A well known fact in the firearms community is that the ammunition characteristics have a significant effect on the "quality" of the features found on its surface. The purpose of the ammunition study was to select two different types of ammunition which would be representative of those commonly found in crime scenes. As in the case of the barrels, it was desired to select ammunition that would also provide an indication of the spectrum of the possible levels of performance.

Upon consultation with firearms examiners, it became clear that there is no consensus on the type of ammunitions best suited for identification. Because of this situation, and thanks to the initiative of Firearms Examiner/Detective Mark Ensor and Firearms Examiner Michael Thomas, from Baltimore County Police, we decided to conduct our own study on this subject. A list of the ammunition brands and models used in this evaluation is included in Table 4. With the exception of the Norinco and the Federal (American Eagle) ammunition, all bullets fired in this evaluation were of 115 Gr. weight (at the recommendation of firearms examiners). Ten samples of each type of ammunition were fired. A total of one hundred bullets were fired and retrieved in pristine condition. In addition, two Remington UMC bullets were fired to produce damage similar to that shown in Figure 10 (consistent with the degree of damage to be used as part of this study). These one hundred and two bullets were acquired and compared against each other. The procedure to test gun individuality described in Section 4.2.3.1 was followed with these bullets, and the results of the evaluation using different number of control bullets (or land impressions) for the different brands of ammunitions are summarized in Table 5. Based on these results, and after consultation with firearms examiners, we selected to use Winchester and Remington UMC for this purpose. Both types of ammunitions are very commonly used in crime, and they range from fairly good performance (Winchester) to intermediate performance (Remington UMC) as shown in Table 5.

	Manufacturer	Weight	Model
1	Magtech	115 Gr.	FMC (9A)
2	CCI Blazer	115 Gr.	TMJ (3509)
3	Winchester	115 Gr.	FMJ (Q4172)
4	Winchester	115 Gr.	Winclean (WC91)
5	Remington UMC	115 Gr.	Metal Case (L9MM3)
6	Lellier & Bellot	115 Gr.	Czech
7	American Ammunition	115 Gr.	CCC
8	PMC	115 Gr.	FMJ (9A)
9	Federal American Eagle	124 Gr.	Metal Case (AE9DP)
10	Norinco	124 Gr.	Chinese

Table 4: Ammunition models tested in ammunition selection study

Figure 10: Example of desirable damaged bullet

Whenever an evidence bullet is found in a crime scene and test fires are performed to attempt an identification with a suspect gun, firearms examiners traditionally try to use either the same ammunition brand or a similar ammunition as that of the evidence bullet. An important conclusion of this portion of the study was to validate practice. As shown in Report No. 9, this practice minimizes the overall probability of identification error.

4.4 Barrel Wear Study

The question of barrel wear was originally postulated in terms of whether the first few bullets fired by a given barrel could be reliably matched to bullets fired many firings later. The rationale for asking such question lies in the fact that the very first bullets fired through a barrel may have a relatively significant impact in the rifled surface of the barrel. It has been argued that such bullets may have a “smoothing” effect on the rifling, or that some of the bullet material may fill some crevices on the rifled surface. Such changes in the barrel rifled surfaces might translate into changes in the features transferred by the barrel to bullets subsequently fired, prompting the question as to whether the features found on the very first bullets fired by a barrel are the same as those found on bullets fired later. Therefore, the main focus of Section 4.4.1 will be the effect of barrel wear on the first few bullets fired through a barrel. However, we also tackle a question not originally contemplated as part of the current study. An extension of the barrel wear question is whether the “distance” (in terms of bullets fired) between fires may have an effect on the degree of similarity of the feature transferred between the barrel and the bullets. In other words, will the

100th bullet fired by a barrel be more similar to the 101st than to the 200th bullet fired? This question will be the focus of Section 4.4.2.

		Rank Sum Test p-value			
Number of Test Bullets		2	3	4	5
List Length		6	18	36	60
Ammunition Type	Magtech	2.2E-03	3.0E-07	3.2E-13	N/A
	Winchester	6.5E-02	6.7E-05	1.4E-09	2.2E-16
	PMC	6.5E-02	7.7E-05	1.9E-09	2.2E-16
	FAE	6.5E-02	1.7E-04	5.8E-09	2.2E-15
	Remington UMC	4.5E-01	2.3E-02	8.5E-05	2.3E-08
	L&B	5.9E-01	4.6E-02	5.4E-04	4.9E-07
	Norico	8.2E-01	1.9E-01	1.1E-02	7.1E-05
	CCI	9.4E-01	1.6E-01	8.4E-03	6.2E-05
		Upper bound on p-value for 95% of the trials			

Table 5: Upper bound on *p*-values for 95% of the trials

Let us review the way in which the sample bullets were collected and organized for the purpose of this part of the study. As mentioned in our previous report, the Barrel Wear Study is conducted with a single barrel for each of the nine barrel models under consideration. The bullets fired through each of these barrels were separated in groups of ten according to the order in which they were fired. For example, we consider the first ten test fired bullets as group "1" ($I_1 = \{1,...,10\}$), the next ten bullets are labeled group 2 ($I_2 = \{11,...,20\}$) etc. In this manner, for each gun model $\bullet$, we consider eight different groups:

Group 1: test fired bullets 001 through 010, ($G_{I_1}(\bullet) \ = \ b(\bullet,1,\{001,...,010\})$),

Group 2: test fired bullets 011 through 020, ($G_{I_2}(\bullet) \ = \ b(\bullet,1,\{011,...,020\})$),

Group 3: test fired bullets 021 through 030, ($G_{I_3}(\bullet) \ = \ b(\bullet,1,\{021,...,030\})$),

Group 4: test fired bullets 031 through 040, ($G_{I_4}(\bullet) \ = \ b(\bullet,1,\{031,...,040\})$),

Group 5: test fired bullets 041 through 050, ($G_{I_5}(\bullet) \ = \ b(\bullet,1,\{041,...,050\})$),

Group 6: test fired bullets 101 through 110, ($G_{I_6}(\bullet) \ = \ b(\bullet,1,\{101,...,110\})$),

Group 7: test fired bullets 201 through 210, ($G_{I_7}(\bullet) \ = \ b(\bullet,1,\{201,...,210\})$),

Group 8: test fired bullets 211 through 220, ($G_{I_8}(\bullet) \ = \ b(\bullet,1,\{211,...,220\})$).

It is important to note that the bullets from Group 8 were fired **after** the barrel under evaluation was cleaned. At no other time was the barrel cleaned (except for before the test firing process began). This group was added to our original set of test firings to isolate the effect of barrel wear from the effect of residue buildup in the barrel after we realized that residue buildup could affect the results of our analysis. For this reason, Group 8 is not available for all barrels.

The ability of the present study to answer the questions of interest for the Barrel Wear Study is predicated on two necessary conditions:

a) The barrel transfers **repeatable** features to the bullets fired through it.

b) The **instrumentation and algorithms** used in the current system have the necessary sensitivity to detect the unique features transferred to the bullets.

Condition a) depends on the manufacture of the barrel under consideration (and/or possibly the manufacture of the test fired bullets). If no repeatable features are transferred between barrel and bullets, it is not possible to evaluate whether the barrel is changing by looking at those bullets fired by it. Condition b), on the other hand, depends on the instrumentation used to acquire the data, and the algorithms responsible for their processing and comparison. If the instrumentation and/or algorithms are not sensitive enough to detect the features transferred between the barrel and the bullets, it is not possible to detect changes in these features either. Conditions a) and b) are independent, and have very different implications. If condition a) does not hold, one has to conclude that the barrel under consideration cannot be identified through the bullets fired by it. On the other hand, failure of condition b) to hold implies that the limitation lies in the instrumentation and/or algorithms, and identification should be possible.

Let us assume that the different lands found on any given barrel always have unique microscopic features which are the unintentional result of the manufacturing process. This is a well founded assumption, since no two lands on a given barrel can be manufactured in exactly the same manner. In fact, in all modern manufacturing techniques, different parts of the manufacturing tool will be responsible for creating the different lands found in a given barrel. Assuming, then, that condition a) holds, the bullets fired by such barrel also have unique microscopic features on each of their different land impressions. Assuming that condition b) holds, using the instrumentation and algorithms of the current system it is possible to identify corresponding land impressions (land impressions created by the same land on the barrel are referred to as corresponding land impressions). This in turn implies that using the instrumentation and algorithms of the current system it is possible to identify a unique *best orientation* between pairs of bullets fired by the same barrel. Following this argument, **we conclude that if the current system is capable of identifying the best orientation between pairs of bullets fired by the barrels under consideration, then conditions a) and b) must be satisfied**.

The ability to identify the best orientation similarity is equivalent to being able to detect the statistical difference between the best and second-best similarity measures. Let us consider as an example the results obtained from the comparison of bullets fired through the Ruger barrel. Figure 11 shows the histograms associated with the best ($s_1(G_{I_i}(Ruger), G_{I_i}(Ruger))$) and second-best ($s_2(G_{I_i}(Ruger), G_{I_i}(Ruger))$) orientation similarity measures resulting from the comparison of bullets within each or the groups discussed earlier in this section. The first (top) plot in Figure 11 shows the histograms corresponds to Group 1 ($I_1 = \{1,...,10\}$), the next plot shows the histograms corresponding to Group 2 ($I_2 = \{11,...,20\}$), etc. Looking at these histograms, it is clear that the distribution associated with the best orientation similarity measures is very distinct from that of the second-best orientation similarity measures. Based on our previous discussions, this is strong evidence that in the case of the Ruger barrel conditions a) and b) do hold.

The individuality test described in Table 1 is ideally suited for testing the ability to validate the statistical difference between the best and second best similarity measure distributions. Therefore, in order to validate conditions a) and b), we followed the same approaches discussed in Section 4.2.3.1. The first test was based on using a Rank Sum test. **Conditions a) and b) are assumed to hold for a given barrel model if the average *p*-value taken over 100 iterations is less than 1% for *m* = 3 (i.e. using three control bullets).** The second test was based on an empirical estimation of the probability of a false positive or false negative identification of the best relative orientation for each of the gun models under consideration. **Conditions a) and b) are assumed to hold for a given barrel model if the overall empirically computed probability of error is less (or equal) than 20%.**

Both these tests were applied to each of the barrel models under consideration, and were the basis for our assessing whether conditions a) and b) hold. Table 6 summarizes the results of both the Empirical and Rank Sum based evaluations for all barrel models. These results have been ranked in terms of the barrel models most likely to satisfy conditions a) and b), from Ruger as the most likely to satisfy these conditions, and Bryco as the least likely. The first and second column of this table show the gun brand under consideration, and the label used to identify it through this study. Columns four through eleven (labeled Set 1 through Set 8) show the probability of error computed based on the Empirical approach (middle row), and the average *p*-value obtained by the Rank Sum approach (bottom row). All these values are in percentages. In the case of the Rank Sum test, the values shown correspond to the average *p*-value obtained after repeating the sequence described in Table 1 one hundred times, while using data corresponding to three control bullets each time. The last four columns in Table 6 show the same information as the previous eight columns, except that averaged according to groups of interest: column twelve simply repeats the information regarding Group 1, column thirteen shows the averaged values for Groups 1 through 5, column fourteen shows the results for Group 8, and column fifteen shows the averaged results over all groups. From the perspective of both these statistical evaluations it appears clear that conditions a) and b) are satisfied for the first three or maybe four barrel models (Ruger, Beretta, Smith, Browning).

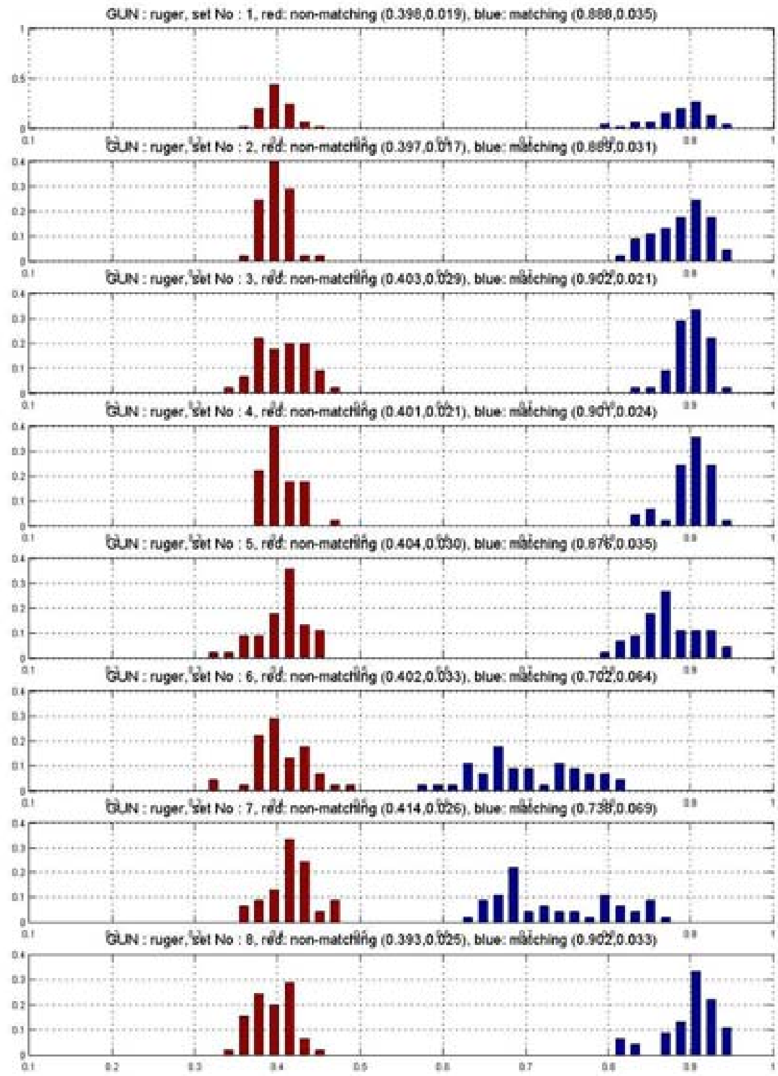


Figure 11: Best and Second-best Orientation Similarity Measure Histograms, Group-by-Group Bullet Comparisons, Ruger

											First	First 5	Set 8	ALL
Ruger	**E**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		pValue	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Beretta	**I**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		pValue	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Smith	**H**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	4.44	12.22	18.89	18.89	13.33	26.67	28.89	20.00	4.44	13.56	20.00	17.92
		pValue	0.00	0.01	0.02	0.04	0.01	2.34	4.90	0.58	0.00	0.02	0.58	0.99
Browning	**G**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	18.89	8.89	17.78	12.22	27.78	27.78	32.22	22.22	18.89	17.11	22.22	20.97
		pValue	0.20	0.00	0.15	0.00	2.55	3.00	8.19	0.19	0.20	0.58	0.19	1.78
Taurus	**A**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	24.44	26.67	25.56	24.44	26.67	32.22	26.67	26.67	24.44	25.56	26.67	26.67
		pValue	2.13	0.50	0.64	1.70	2.15	11.29	2.95	3.50	2.13	1.42	3.50	3.11
Glock	**C**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	30.00	31.11	31.11	33.33	25.56	34.72	28.89		30.00	30.22		30.67
		pValue	7.83	7.35	6.29	7.15	0.99	13.22	3.50		7.83	5.92		6.62
SIG	**F**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	37.78	28.89	34.44	35.56	33.33	28.89	31.11	28.89	37.78	34.00	28.89	32.36
		pValue	11.79	2.30	10.91	8.22	3.85	2.44	7.97	5.40	11.79	7.42	5.40	6.61
HiPoint	**B**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	37.78	25.56	22.22	28.89	26.67	15.56	21.11		37.78	28.22		25.40
		pValue	34.19	1.71	0.19	7.42	7.44	0.03	0.18		34.19	10.19		7.31
Bryco	**D**		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8				
		empirical	37.78	27.78	27.78	26.67	32.22	34.44	34.72		37.78	30.44		31.63
		pValue	23.62	8.11	4.35	5.87	9.36	12.53	11.46		23.62	10.26		10.76

Table 6: Results of Sensitivity Requirement Evaluation

Similar to Figure 11, each of the distributions associated with each of the barrel models under consideration have been plotted (see Appendix A of report No. 10). Visual inspection of these histograms confirms the results discussed in the previous paragraph; i.e. that only for Ruger, Beretta, Smith and Browning barrels can the distribution of the best orientation similarity measures be distinguished from that of the second-best orientation. This indicates that for the rest of the barrel models either condition a) or b) is not satisfied. Which condition does not hold can only be answered by performing similar tests with a better system, and/or by recruiting the assistance of firearms examiners.

It is worth noting that the barrel models for which the distribution of the best orientation similarity measures can be easily distinguished from the second-best orientation similarity measures coincide with the barrel models which have the widest LEAs. As seen in Table 3, the LEA widths are 2.0 mm for Ruger, 2.0 mm for Beretta, 2.5 mm for Smith & Wesson, and 1.8 mm for Browning. This implies that a greater amount of data was available for the system to make LEA-to-LEA comparisons, which most likely improved the effective performance of the system.

Notice also the fact that the distribution of $s_1(G_{I_i}(Ruger), G_{I_i}(Ruger))$ for *i =6* and *i = 7* has a lower mean than all other such distributions associated with the first Ruger barrel. The reason for this appears to be that as the barrel was continuously fired, some fouling of the barrel took place sometime after test fire 50. We reach this conclusion because once the barrel was cleaned (right before $G_{I_8}(Ruger)$ was fired) the distribution of $s_1(G_{I_8}(Ruger), G_{I_8}(Ruger))$ shows approximately the same mean as that of $s_1(G_{I_i}(Ruger), G_{I_i}(Ruger))$ for *i < 6*. For this reason, bullets belonging from groups 6 and 7 may be excluded from certain portions of the study.

An interesting phenomena associated with groups *i =6* and *i = 7* is that the mean of the best orientation similarity measure resulting from the comparison of these bullets was higher than that obtained when comparing, say, group *i =7* and *i = 8*. This indicates that the bullets fired when the barrel was not properly cleaned were more similar among themselves than when compared against bullets fired when the barrel was cleaned. The implications of this result is that bullets fired when the barrel is in a less-than-optimal condition (like in this case, when there was an accumulation of fouling) may be more similar among themselves than when compared with bullets fired when the barrel is clean. This result is of interest, because it validates a common practice of firearms examiners. Whenever possible, firearms examiners will test fire a suspect gun **before and after** cleaning the barrel The phenomena observed with bullets corresponding to *i =6* and *i =7* confirm the soundness of this practice.

4.4.1 Short Term Barrel Wear Analysis

The question to be addressed in this section is the following: **Do the features transferred between a barrel and the bullets fired by it change for the first ten bullets?** This is a question of considerable interest. Based on the discussion of the previous section, we focus our attention in those barrel models for which conditions a) and b) seem most likely to be satisfied for the first 50 test fired bullets. These barrel models are Ruger, Beretta, Smith & Wesson and Browning.

Let us begin by considering the change of a single LEA as a result of barrel wear. Figure 12 shows a conceptual representation of the possible effect of barrel wear on LEAs created by the same barrel land on sequentially fired bullets (Figure 12 only shows a cross section of the LEA). The two columns shown in Figure 12 correspond to the two cases under consideration: Case I, where there is no significant change in the LEA features, and Case II, where there is significant initial change in the LEA features. Each row of the plots shown in Figure 12 shows the evolution of the cross section of the same LEA starting from the first fired bullet at the bottom of the plots (labeled 1); up to the tenth fired bullet at the top of the plots (labeled 10). Notice that these plots are in no way meant to be representative of actual LEAs, but simply a conceptual tool. Figure 12 will be a useful tool for the following discussion.

Let us begin with Case I. In this scenario, there is no significant change between the features transferred to the first few bullets and those fired later. If there is no such change, the features found on the first bullet fired by the barrel model under consideration would not be significantly different to those found on all subsequent bullets fired by the same barrel (as depicted in Figure 12, Case I). If so, the set of best orientation similarity measures resulting from the comparison of the bullet $b(\bullet,1,1)$ against subsequently fired bullets $b(\bullet,1,\{1+k_1,...,1+k_2\})$ would be statistically similar to the set of best orientation similarity measures obtained by comparing bullet $b(\bullet,1,i)$ against bullets $b(\bullet,1,\{i+k_1,...,i+k_2\})$ for any *i* and any $k_1 < k_2$ (they would not be exactly the same because of a variety of random factors such as powder residue, barrel temperature, barrel fouling, etc. which always affect the features transferred between the barrel and the bullet). Therefore, if Case I is true, we conclude that:

a) **The sets of best orientation similarity measures** $s_1(b(\bullet,1,\{i\}), b(\bullet,1,\{i+k_1,...,i+k_2\}))$ **are statistically similar for all** i **and all** $k_1 < k_2$.

Let us now consider the alternative scenario, Case II (depicted in the right column of Figure 12.) In the representation of Case II we have attempted to capture the notion that if the features transferred between the barrel and the bullet change at all during the first few fires, then said change is only significant for a limited number of bullets. In this case, the cross section of the LEA changes significantly for the first five or so bullets, but any significant change stops after bullet six. Let us denote by n the bullet number after which the change of the LEA cross section stops. For the case depicted in Figure 12, $n = 6$.

In the second scenario, the argument developed for Case I holds for those bullets fired after test fire n. In other words, the sets of best orientation similarity measures $s_1(b(\bullet,1,\{i\}), b(\bullet,1,\{i+k_1,...,i+k_2\}))$ are statistically similar for all $i > n$, (where $n = 6$), and any $k_1 < k_2$. What would we expect to happen for $i \le n$? In this case $b(\bullet,1,i)$ will have been fired while the features transferred between barrel and bullets are still undergoing significant change. Selecting $k_2 > k_1 > n$, would guarantee that $b(\bullet,1,\{i+k_1,...,i+k_2\})$ will have been fired while said change is no longer significant. Using the conceptual example shown in Figure 12, and estimating $n = 6$, the LEAs associated with $b(\bullet,1,i)$, for $i \le 6$ are still undergoing significant change, while taking $6 < k_1$ the LEAs associated with $b(\bullet,1,\{i+k_1,...,i+k_2\})$ are not. Therefore, we would expect that the statistical characteristics of the set of best orientation similarity measures $s_1(b(\bullet,1,\{i\}), b(\bullet,1,\{i+k_1,...,i+k_2\}))$ for $i \le n$, $n < k_1 < k_2$ will be different to those obtained for $n < i$. In fact, the average of the best orientation similarity

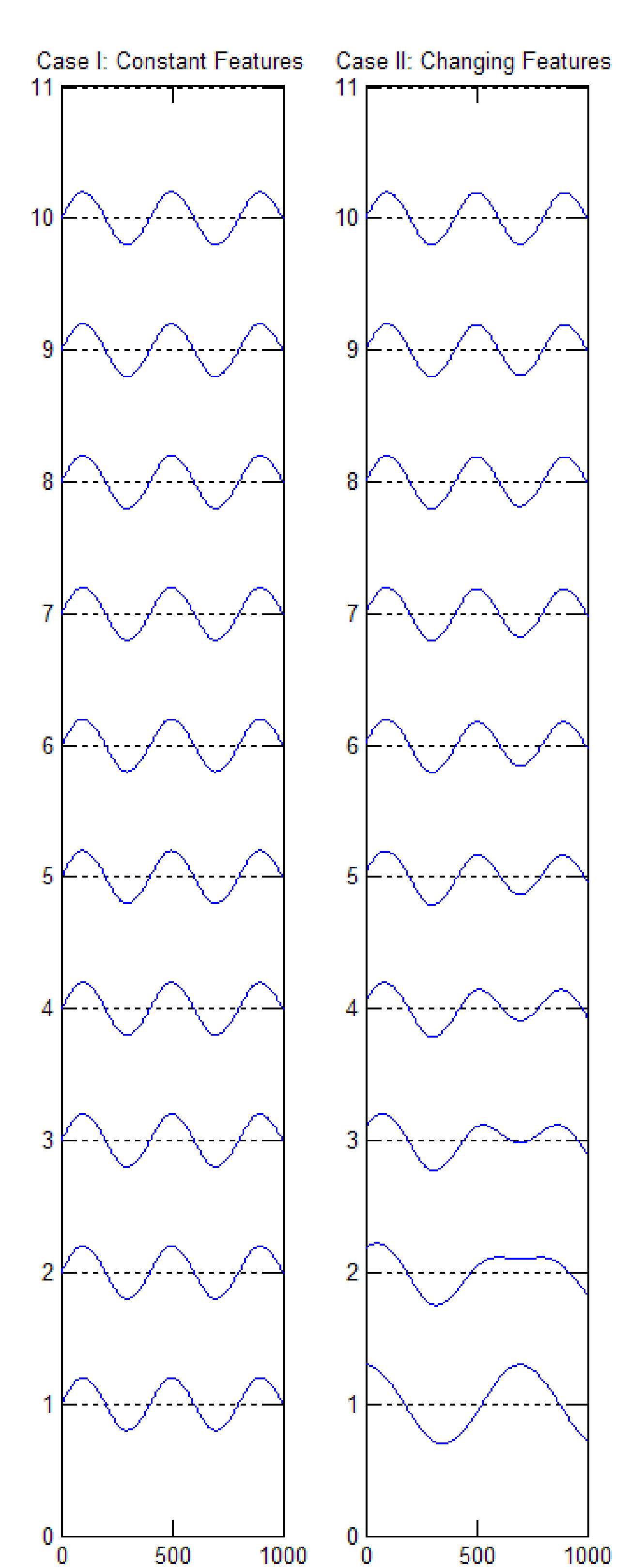


Figure 12: Conceptual Effect of Barrel Wear on Single LEA

measures of $s_1(b(\bullet,1,\{i\}),b(\bullet,1,\{i+k_1,...,i+k_2\}))$ will be lower for $i \le n$ than for $i > n$. In summary, if Case II is true, we conclude that:

a) **The sets of best orientation similarity measures $s_1(b(\bullet,1,\{i\}),b(\bullet,1,\{i+k_1,...,i+k_2\}))$ are statistically similar for any *n* < *i* and any *k₁* < *k₂*.**
b) **The sets of best orientation similarity measures $s_1(b(\bullet,1,\{i\}),b(\bullet,1,\{i+k_1,...,i+k_2\}))$ are statistically different for $i \le n$ and $n < k_1 < k_2$**
c) **The mean of $s_1(b(\bullet,1,\{i\}),b(\bullet,1,\{i+k_1,...,i+k_2\}))$ for $i \le n$ and $n < k_1 < k_2$ is less than the mean of $s_1(b(\bullet,1,\{i\}),b(\bullet,1,\{i+k_1,...,i+k_2\}))$ for $n < i$ and any $k_1 < k_2$.**

Based on the previous arguments, we can propose an empirical methodology to verify which of the two scenarios is most likely for each of the barrels under consideration. For each of these barrels the best orientation similarity measures $s_1(b(\bullet,1,\{i\}),b(\bullet,1,\{i+k_1,...,i+k_2\}))$, their mean and standard deviation were computed for $i = \{1,\ldots,10\}$, $k_1 = 10$, $k_2 = 40$. The median and standard deviation of the best orientation similarity measures resulting from the comparison of all first 50 bullets against each other $s_1(b(\bullet,1,\{1,...,50\}),b(\bullet,1,\{1,...,50\}))$ was also computed as a reference of the overall characteristics of the set. We will make use of these results to assess which of the two possible scenarios (Case I or Case II) is more likely for each barrel under consideration.

For ease of presentation, we introduce the shorthand notation (to be used in this section only):

$$s_{k_1,k_2}(\bullet,i) = s_1(b(\bullet,1,\{i\}),b(\bullet,1,\{i+k_1,...,i+k_2\}))) \tag{17}$$

and

$$s_1^{l-m}(\bullet) = s_1(b(\bullet,1,\{l,...,m\}),b(\bullet,1,\{l,...,m\})) \tag{18}$$

Also, let us define the square of the standardized distance between the sets of best orientation similarity measures $s_{k_1,k_2}(\bullet,i)$ and $s_1^{1-50}(\bullet)$

$$dist^2(s_{k_1,k_2}(\bullet,i),s_1^{1-50}(\bullet)) = \left(\frac{mean(s_{k_1,k_2}(\bullet,i)) - median(s_1^{1-50}(\bullet))}{std(s_1^{1-50}(\bullet))}\right)^2 \tag{19}$$

Figure 13 shows the plots associated with the results computed for Ruger, Beretta, Smith & Wesson and Browning barrels. Let us consider as an example the results obtained for the Ruger barrel (see Figure 13, left column). The top plot shows the average and standard deviation of $s_{k_1,k_2}(Ruger,i)$ for each value of i in an error-bar format. Also shown is this plot is the median (red line) as well as the region bounded by +/- one standard deviation of $s_1^{1-50}(Ruger)$ (green lines). The bottom plot shows the square of the standardized distance between the sets $s_{k_1,k_2}(Ruger,i)$ and $s_1^{1-50}(Ruger)$ for each **i**. This is a measure, in terms of the distribution of $s_{k_1,k_2}(Ruger,i)$ of the distance between the mean of $s_{k_1,k_2}(Ruger,i)$ and the

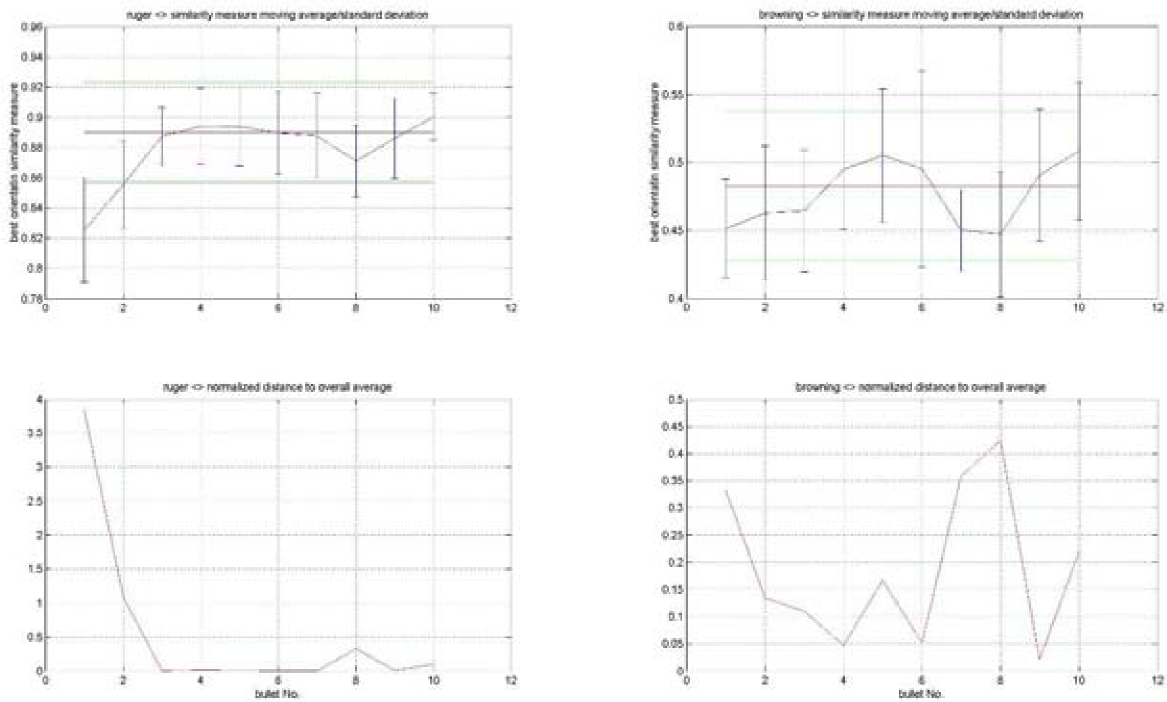


Figure 13: Short Term Barrel Wear Analysis, Ruger (left) vs. Browning (right)

median of $s_1^{1-50}(Ruger)$. This plot helps visualize the difference between $s_{k_1,k_2}(Ruger, i)$ and $s_1^{1-50}(Ruger)$.

Based on the plots shown in Figure 13 and our previous discussion, it seems apparent that the features transferred by the barrel to the first and possibly second bullets fired by the Ruger barrel are to some degree different to the features transferred to the remaining bullets. We reach this conclusion by comparing the results of the empirical data against our conclusions associated with Case I and Case II. The results of the empirical evaluation seem to correspond to the results expected in Case II for $n \approx 2$. Notice that the mean of $s_{k_1,k_2}(Ruger, i)$ is lower for $i \leq 2$ than for *2 < i,* as expected in Case II. Also, notice that $s_{k_1,k_2}(Ruger, i)$ appears to converge or remain approximately constant for *i > 2,* also in agreement with Case II.

In contrast to the results of Ruger, the results of Browning (shown in the right column of Figure 13) suggest no evidence of significant change in the features of the bullets fired by this barrel. Again we reach this conclusion by comparing the results of the empirical data against our conclusions associated with Case I and Case II. In the case of the Browning barrel, $s_{k_1,k_2}(Browning, i)$ appears to remain approximately constant for all *i*, as expected for Case I.

Based on a similar analysis for Beretta and Smith & Wesson barrels, we observe that there is evidence of change in the features found on the surface of the first bullets for Beretta and although very mild, Smith & Wesson. Table 7 summarizes the results of the evaluation. For each of the barrel models under consideration, where the second column shows the computed maximum standardized distance (as defined in equation 8), and the third column shows the maximum relative difference between $s_{k_1,k_2}(\bullet, i)$ and $s_1^{1-50}(\bullet)$. Finally, the forth column shows the value of *n* associated with each gun model if we define as "significant" a standardized distance greater than one ($dist(s_{k_1,k_2}(\bullet, i), s_1^{1-50}(\bullet)) > 1$).

	Max Standardized Distance	Max Relative Decrement	*n*
Ruger	3.84	-7.30%	2
Beretta	3.72	-20.40%	1
Smith	0.40	-8.70%	0
Browning*	0.42	-7.40%	0

Table 7: Summary of Short Term Barrel Wear Analysis Results.

It is important to note that even in the cases where evidence of varying features was observed, the degradation of the similarity measures associated with these changes does not seem to be extreme. As seen in Table 7, in most cases the degradation noticed in the barrels under consideration was less than 10%, and in the most extreme case the degradation was 20.4%. It is doubtful that this effect could be severe enough to prevent identification of the first few bullets and those fired later, although it would be interesting to verify this with the assistance of firearms examiners.

The results obtained for the Browning barrel do not indicate any effect due to barrel wear. The values shown in Table 7 do not even correspond to the first bullet fired, but to the one which displayed the greater maximal standardized distance (bullet 8.) For all other barrels, the values shown in this table applies to *i = 1*; i.e. the maximum standardized distance and relative decrement take place for the first bullet fired.

4.4.2 Long Term Barrel Wear Analysis

Having analyzed the effect of barrel wear on the first bullets fired by a barrel, we are prompted to ask whether the changes taking place in the barrel for the first few bullets stop after a relatively low number of fires, or do they continue indefinitely. In other words, **do the features transferred between a barrel and the bullets fired by it change over a long time period?** It should be noted that the current study was not really designed to test this hypothesis, but we believe that the available data does provide some insight into this issue.

The approach used in this evaluation was rather straight forward. For each barrel model under consideration ($\bullet$) we make use of five of the eight groups of bullets defined earlier. We exclude the first group $G_{I_1}(\bullet)$ to avoid any artifacts due to the variations found on the first few bullets. We also exclude Groups $G_{I_6}(\bullet)$ and $G_{I_7}(\bullet)$ to minimize the potential effect of barrel fouling in the features transferred to the bullets. For each of the groups left, we compute the average best orientation similarity difference between bullet groups defined as follows (labeled $\Delta(i, j)$):

$$\Delta(\bullet, i, j) = mean(s_1(G_{I_i}(\bullet), G_{I_j}(\bullet))) - mean(s_1(G_{I_i}(\bullet), G_{I_i}(\bullet))) \tag{20}$$

as well as the relative average best orientation similarity difference between bullet groups defined as follows (labeled $\Delta_r(\bullet, i, j)$):

$$\Delta_r(\bullet, i, j) = \left(\frac{mean(s_1(G_{I_i}(\bullet), G_{I_j}(\bullet))) - mean(s_1(G_{I_i}(\bullet), G_{I_i}(\bullet)))}{mean(s_1(G_{I_i}(\bullet), G_{I_i}(\bullet)))} \right) \tag{21}$$

Relative Difference	Distance, [n = number of sample group comparisons]						
	10 [n = 6]	20 [n = 4]	30 [n = 2]	170 [n = 2]	180 [n = 2]	190 [n = 2]	200 [n = 2]
Ruger	-0.40%	-0.70%	-0.88%	-1.95%	-0.28%	-1.96%	-1.74%
Beretta	-0.03%	-0.84%	-1.80%	-6.78%	-6.58%	-8.69%	-5.85%
Smith	-4.10%	-8.10%	-10.43%	-11.28%	-11.94%	-10.63%	-11.63%
Browning	-1.06%	-4.22%	-5.58%	-6.45%	-8.11%	-8.22%	-11.91%

Table 8: Summary of Long Term Barrel Wear Analysis Results

The values $\Delta(\bullet, i, j)$ and $\Delta_r(\bullet, i, j)$ respectively quantify in absolute and relative difference - in terms of similarity measures - between two different groups of bullets $G_{I_i}(\bullet)$, $G_{I_j}(\bullet)$ (evaluated by $mean(s_1(G_{I_i}(\bullet), G_{I_j}(\bullet))))$ as opposed to comparing bullets $G_{I_i}(\bullet)$ among themselves (evaluated by $mean(s_1(G_{I_i}(\bullet), G_{I_i}(\bullet))))$. In other words, these values provide a measure of how different is $G_{I_j}(\bullet)$ to group $G_{I_i}(\bullet)$ in relationship to the degree of similarity found within group $G_{I_i}(\bullet)$.

Since the only difference between groups of bullets $G_{I_i}(\bullet)$ and $G_{I_j}(\bullet)$ is the average number of bullets fired between them, $\Delta(\bullet, i, j)$ and $\Delta_r(\bullet, i, j)$ depend only on the average number of bullets fired between these groups. The average number of bullets fires between bullet groups $G_{I_i}(\bullet)$ and $G_{I_j}(\bullet)$ can be readily computed, and will be referred to as the "distance" between these groups. We denote the distance between said bullet groups by $dist(G_{I_i}(\bullet), G_{I_j}(\bullet))$. Shown in Table 8 are the summarized results for all four barrels of interest.

Considering the values in Table 8, there appears to be strong evidence to the fact that the number of fires between bullets does have an effect on their degree of similarity. In all cases of interest (Ruger, Beretta, Smith & Wesson and Browning) we notice relative degradation in the average best orientation similarity measure which increases as the distance between groups of bullets increases. This effect appears to extend to bullet separations of approximately 30 or 40 test fires. However, after said number of fires, the difference between the bullet groups does not seem to increase further. This is of significant importance, because it implies that the features transferred between the barrel and the bullets do not continue to change indefinitely. Finally, it is also important to note that the degree of deterioration even for bullets 200 fires away from each other is minor (on average, never exceeding 12%, see Table 8). It is doubtful that such deterioration would prevent the correct identification of a pair of bullets fired by barrels of the models under consideration.

4.5 Pristine Bullets: Individuality Study

The specifics of the procedure to test gun individuality were described in Section 4.2.3.1. The key to evaluate the individuality of a gun is through a statistical hypothesis test on the distributions of the best and second-best similarity measures obtained from the comparison of bullets fired by the barrel under consideration. The two hypotheses are:

H0: best and second-best distributions are indistinguishable,

H1: best and second-best distributions are different.

Ruger	E	Barrel 1	Barrel 2	Barrel 3	Barrel 4	Barrel 5	Barrel 6	Barrel 7	Barrel 8	Barrel 9	Barrel 10	Barrel 11	Mean	Stdev
Win	empirical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.22	0.00	0.00	0.20	0.67
	pValue	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rem	empirical		3.33	1.11	0.00	0.00	0.00	10.00	1.11	6.67	0.00	0.00	2.22	3.47
	pValue		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Beretta	I	Barrel 1	Barrel 2	Barrel 3	Barrel 4	Barrel 5	Barrel 6	Barrel 7	Barrel 8	Barrel 9	Barrel 10	Barrel 11	Mean	Stdev
Win	empirical	0.00	0.00	1.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.22	0.30	0.72
	pValue	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rem	empirical		0.00	3.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.33	0.67	1.41
	pValue		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Smith	H	Barrel 1	Barrel 2	Barrel 3	Barrel 4	Barrel 5	Barrel 6	Barrel 7	Barrel 8	Barrel 9	Barrel 10	Barrel 11	Mean	Stdev
Win	empirical	18.89	0.00	0.00	0.00	0.00	0.00	0.00	2.22	2.22	0.00	0.00	2.12	5.63
	pValue	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rem	empirical		16.67	27.78	0.00	10.00	22.22	16.67	3.33	30.00	8.89	15.56	15.11	9.83
	pValue		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Taurus	A	Barrel 1	Barrel 2	Barrel 3	Barrel 4	Barrel 5	Barrel 6	Barrel 7	Barrel 8	Barrel 9	Barrel 10	Barrel 11	Mean	Stdev
Win	empirical	25.56	10.00	7.78	10.00	10.00	23.33						14.44	7.83
	pValue	0.00	0.00	0.00	0.00	0.00	0.00						0.00	0.00
Rem	empirical		20.00	8.89	25.56	23.33	15.56						18.67	6.64
	pValue		0.00	0.00	0.00	0.00	0.00						0.00	0.00

Brownin	G	Barrel 1	Barrel 2	Barrel 3	Barrel 4	Barrel 5	Barrel 6	Barrel 7	Barrel 8	Barrel 9	Barrel 10	Barrel 11	Mean	Stdev
Win	empirical	12.22	18.89	10.00	31.11	3.33	31.11	4.44	14.44	1.11	24.44	12.22	14.85	10.51
	pValue	0.00	0.00	0.00	0.28	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.03	0.09
Rem	empirical		25.56	27.78	26.39	25.56	31.11	27.78	36.67	12.22	13.33	30.00	25.64	7.54
	pValue		0.00	0.01	0.01	0.00	0.02	0.00	0.37	0.00	0.00	0.00	0.04	0.12

HiPoint	B	Barrel 1	Barrel 2	Barrel 3	Barrel 4	Barrel 5	Barrel 6	Barrel 7	Barrel 8	Barrel 9	Barrel 10	Barrel 11	Mean	Stdev
Win	empirical		25.56	32.22	33.33	33.33	22.22	24.44	28.89	18.89	30.00	24.44	27.33	4.97
	pValue		0.00	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Rem	empirical		31.11	31.11	32.22	31.11	26.67	31.11	25.56	23.33	23.33	23.33	27.89	3.79
	pValue		0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01

SIG	F	Barrel 1	Barrel 2	Barrel 3	Barrel 4	Barrel 5	Barrel 6	Barrel 7	Barrel 8	Barrel 9	Barrel 10	Barrel 11	Mean	Stdev
Win	empirical	35.56	23.33	35.56	34.44	32.22	33.33	36.67	37.78	31.11	34.44	37.78	33.84	4.08
	pValue	0.07	0.00	0.54	0.47	0.02	0.06	1.13	0.79	0.01	0.09	0.21	0.31	0.38
Rem	empirical		22.22	34.44	27.78	30.00	24.44	32.22	27.78	30.00	31.11	25.56	28.56	3.74
	pValue		0.00	0.04	0.00	0.01	0.00	0.06	0.00	0.01	0.04	0.00	0.02	0.02

Bryco	D	Barrel 1	Barrel 2	Barrel 3	Barrel 4	Barrel 5	Barrel 6	Barrel 7	Barrel 8	Barrel 9	Barrel 10	Barrel 11	Mean	Stdev
Win	empirical	26.67	27.78	33.33	31.11	28.89	31.11	32.22	27.78	31.11	34.44	30.00	30.40	2.45
	pValue	0.00	0.00	0.11	0.03	0.00	0.10	0.08	0.00	0.03	0.84	0.00	0.11	0.25
Rem	empirical		27.78	33.33	30	31.11	31.11	31.11	31.11	36.67	32.22	28.89	31.33	2.45
	pValue		0.01	0.16	0.04	0.08	0.04	0.01	0.05	0.11	0.07	0	0.06	0.05

Table 9 Individuality study for all barrels and ammunitions

The comparison of these distributions has been performed in the same manner as in the Barrel Wear Study, where two alternative approaches to quantify the degree of similarity were used. The first approach was based on the use of a Rank Sum test (p-value) approach. Based on this criterion, a barrel is considered identifiable if the p-value is 1% or less. The second approach was based on the estimation of the empirical probability of identification error (either a false positive or false negative error). Based on this criterion, a barrel is considered identifiable if the empirically estimated probability of error is 20% or less. These tests were performed for both types of ammunition (Winchester and Remington).

A summary of the results of these tests is shown in Table 9. Notice that the data associated with barrel 1 of each gun model only includes data associated with Winchester ammunition. The reason for this is that for these barrels only this kind of ammunition was fired (since these barrels were used in the Barrel Wear Study).

From Table 9 we can observe that the results associated with each of the two tests (Rank Sum test and Empirical) are consistent. The overall evaluation of each gun model is made through the mean of the empirically estimated probability of error and *p*-values computed by the Rank Sum Test. These values are tabulated for both types of ammunition in the last columns of Table 9. The order in which the gun models are listed in Table 9 corresponds roughly to the level of individuality displayed by each of the barrel models, from Ruger as the barrel model displaying the greatest individuality, to Bryco as the barrel model displaying the worse individuality. This ranking result also shows consistency with that found in the Barrel Wear Study.

The results in Table 9 shows that the individuality criteria are definitely satisfied for the first three gun models (Ruger, Berettta, and Smith), not-so-clearly satisfied for the next two models (Taurus and Browning), and definitely not satisfied for the last three models (SIG, HiPoint, Bryco, and most likely Glock). This phenomenon may be due to the fact that the current data acquisition system and comparison algorithms lack the capability to validate the individuality of certain gun models (SIG, HiPoint, and Bryco, not to mention Glock). The reader may find it interesting to inspect the distributions of best and second-best similarity measures of each gun model. These distributions were included in Appendix A in Progress Report No. 11. From the visual inspection of those distributions, it is not difficult to reach the conclusion that the best and second-best distributions for Ruger, Bereta, and Smith barrels are clearly distinguishable for both types of ammunition. In the case of Browning and Taurus, it is possible to distinguish the best and second-best matching distributions for Winchester ammunition, but not so for Remington. Finally, the distinction of these distributions is virtually impossible for the remaining models; SIG, HiPoint and Bryco.

Based on the data shown in Table 9, we can reach a number of conclusions. The ability to obtain an individual and repeatable transfer of features between the barrel and bullets depends significantly on:

a) The manufacture/quality of the barrel. This does not necessarily mean the manufacturing process (gang breach, hammer forged, button rifling, etc.), although it seems that barrels manufactured by gang broach generally display good transfer of features, while barrels manufactured by hammer forging display poor transfer of features.

b) The ammunition manufacture. From Table 9 one can observe that the individuality of the different barrels is almost uniformly better recognized by Winchester as opposed to Remington ammunition.

c) Even within a given barrel model, not all barrels are “created equal.” As an example, Smith and Wesson barrel No. 1 did not show good individuality characteristics for Winchester ammunition, but all other Smith and Wesson barrels displayed very good individuality characteristics with the same ammunition. This is a very important observation, as it implies that general population-wide conclusions may not be applicable to each independent member of the population.

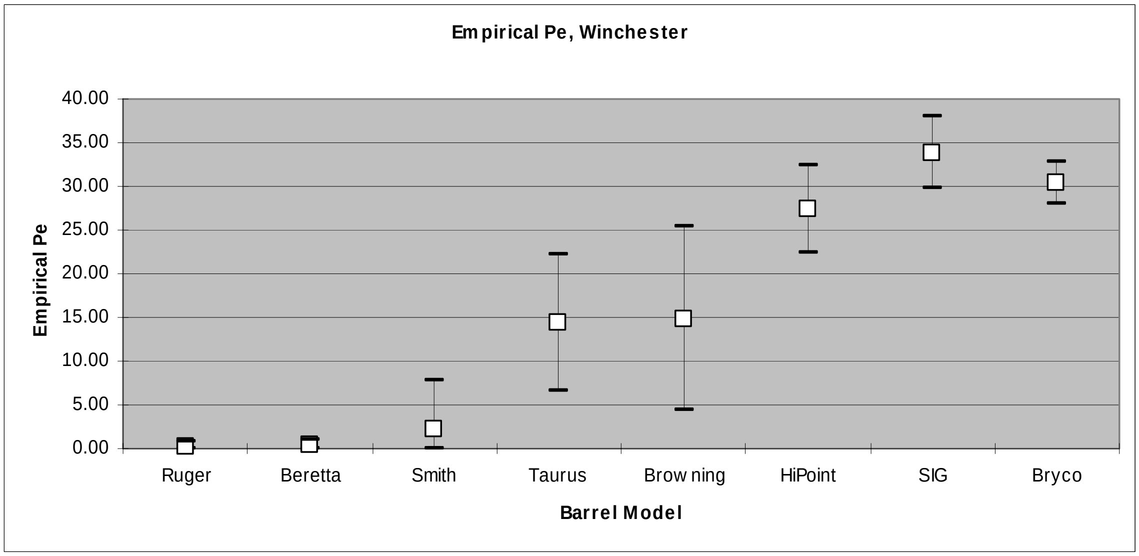


Figure 14:Variation of Estimated Pe Between Different Barrels, Winchester Ammo

In the following sub-sections we discuss the further analyze points a) through c).

4.5.1 Effects of Barrel Manufacture

Figure 14 and Figure 15 display graphically the differences in terms of the empirical estimated probability of error estimated using the best and second-best similarity measures for the different barrel models under consideration. In each of these figures, the center of each bar indicates the average value of the empirically estimated probability of error based on best and second-best similarity measure distributions, while the upper and lower bounds indicate range corresponding to +/- one standard deviation. Figure 14 corresponds to Winchester ammunition, while Figure 15 corresponds to Remington ammunition. From these tables it is clear that there are considerable differences in the degree of individuality attained by each of the barrel models under consideration.

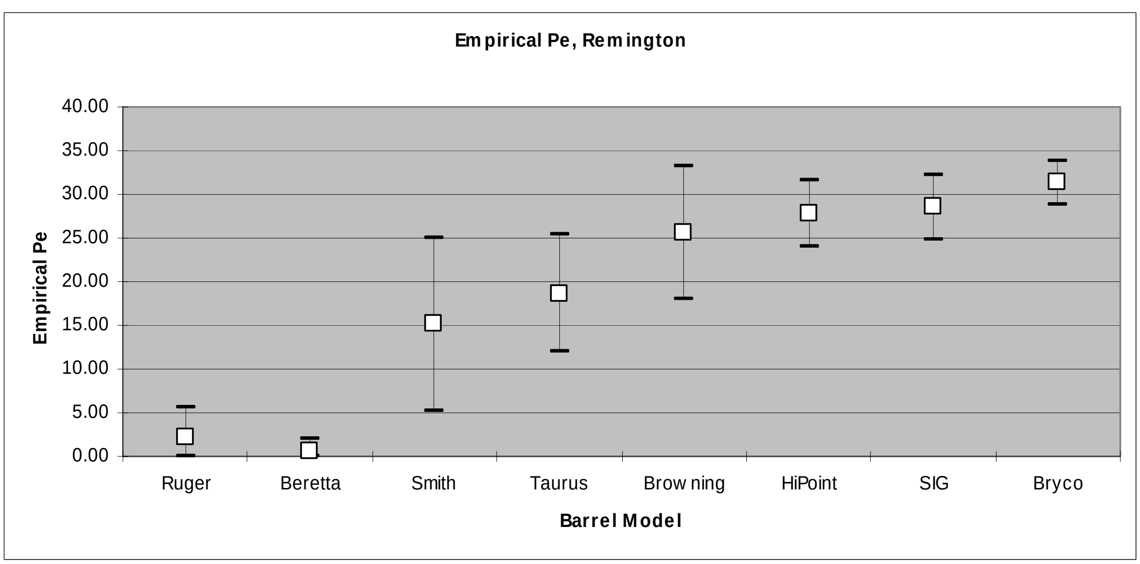


Figure 15: Variation of Estimated Pe Between Different Barrels, Remington Ammo

An alternative and very illustrative way to display the differences in individuality between barrel models is to plot the different Receiver Operator Characteristic (ROC) curves for the different barrel models under consideration. An ROC curve is a parametric curve of the possible False Negative Probability vs. the False Positive Probability. Therefore, it allows us to consider the overall tradeoff between false positive and false negative identifications, as opposed to looking only at a probability of error as in Figure 14 and Figure 15.

The top of Figure 16 shows the ROC curves obtained using the best and second-best similarity measures resulting from the comparison of Winchester bullets for all barrel models, while the bottom of the same figure shows the same results for Remington ammunition. The degree of individuality of a barrel can be quantified by the area under the ROC curve. These plots show that the individuality of each barrel model can vary significantly (as already seen in Figure 14 and Figure 15). Moreover, each barrel model seems to belong to a particular group for each ammunition type. Let us consider the results obtained for Winchester ammunition. For this ammunition, Ruger, Beretta and Smith barrels seem to be grouped at the bottom of the plot; Taurus and Browning barrels are at the middle of the plot, while Hipoint, SIG and Bryco barrels are at the top, displaying very poor performance (in terms of estimated probability of false positive and false negative identification). It is interesting to notice that a similar grouping can be seen when Remington ammunition is used, but there are slight changes in the position of some barrels. For example, Smith barrels have moved from the "best" to the "middle" group, while Browning has moved from the "middle" to the "worse" group. Clearly, the performance of each barrel model depends on the ammunition.

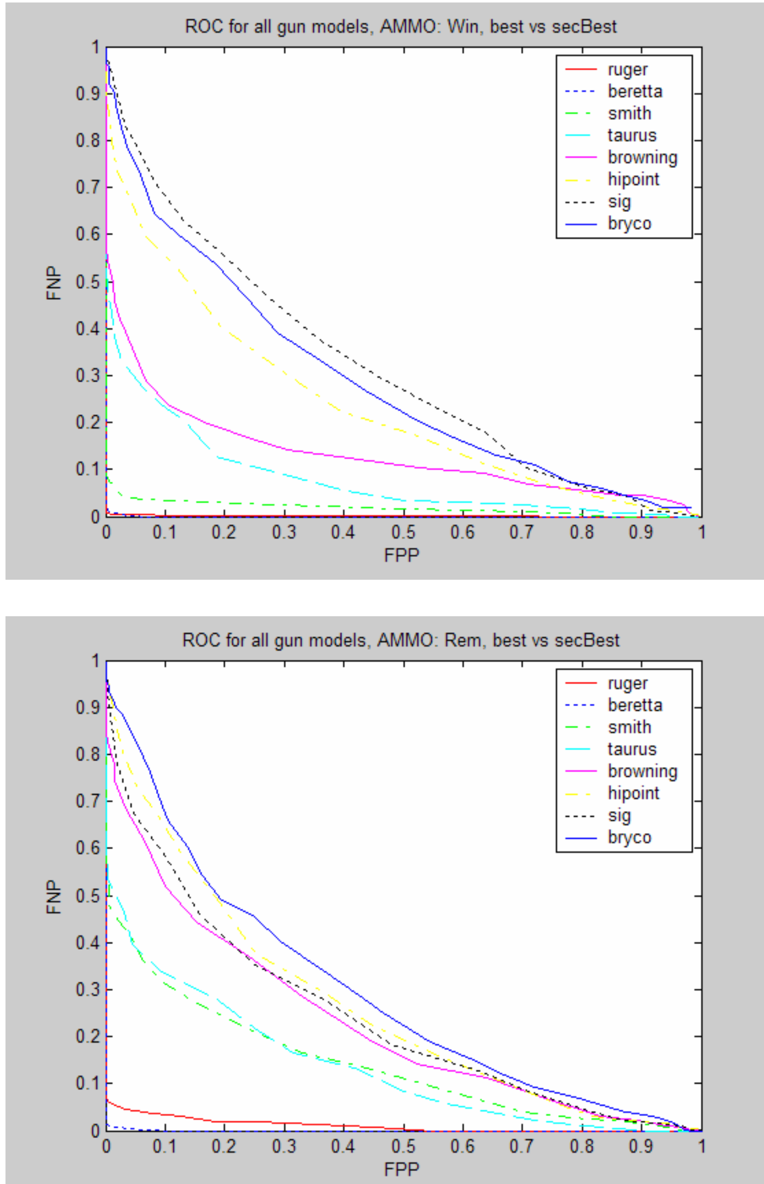


Figure 16: Receiver Operator Characteristic (ROC) curves for each barrel model, Winchester (top) and Remington ammunition (bottom)

The barrel manufacture seems to be the dominant factor in the individuality of the bullets fired by it. For example, bullets fired by Ruger or Beretta barrels were easily identifiable by the system, while the ability of the system to identify bullets fired by HiPoint or SIG Sauer barrels was limited. A systematic understanding of the specific factors which play a role in this phenomenon is beyond the scope of this project. However, based on observations of the characteristic of the features found on the surface of bullets fired by different brands of barrels we can postulate some of the factors which seem to play the most significant roles. The dimensional tolerances of the barrel, the degree to which the interior surface of the barrel is polished during manufacture, and the general quality of the machining all seem to contribute to the overall quality of the barrel. Let us consider the effect of these manufacture characteristics.

- Dimensional Tolerance: A common practice among manufacturers of low cost barrels is to oversize the bore of the barrel. The purpose of this practice is to decrease the pressure to which the barrel is subjected when a bullet is fired. As the bore of the barrel is increased, the seal between the bullet and the barrel is less effective, allowing the high pressure gases generated by the burning of the powder in the cartridge case to flow in between the bullet and the barrel, and decreasing the pressure generated by these gases. The barrel pressure may be a concern to the manufacturer because the quality of the material used in the manufacture of the barrel (and other parts of the gun mechanism) may be relatively inferior, and of doubtful ability to withstand the full pressure generated by a good seal between the bullet and the barrel. By decreasing the internal barrel pressure, the manufacturer decreases the risk of these barrels failing (and possibly exploding) as they are fired. The effect of an oversized bore, however, is to afford a less consistent contact between the barrel interior surface and the bullet fired through it. An undesirable effect of this fact is that the features transferred between the barrel and the bullets do not repeat with the same consistency as in the case of well dimensioned barrels. As a consequence, the identification of bullets fired by low quality barrels can be very challenging.

- General Dimensional Quality: A phenomena that can also be present in barrels of poor manufacturing quality is poor bore consistency. Consider the case where the bore of the barrel is not constant throughout its length. This may happen, for example, in the case of button rifled barrels. The rifling of button rifled barrels is created by a “button” traveling along the bore of a barrel and imprinting on the interior of the barrel the rifling grooves (the rifling grooves are present in the button, and transferred to the barrel as it travels through it). As the button travels through the interior of the barrel, it pushes the barrel walls outwards, and imprints the rifling grooves on them. However, if the material of the barrel is not homogeneous throughout its length, some sections of the barrel may be pushed further than others, resulting in a barrel that has slight bore variations. A barrel with a varying bore will result in the bullets fired by it possibly “skipping” in its interior (as the bullets may loose contact with the barrel momentarily). Such phenomena would result in bullets for which part of the features transferred will vary from bullet to bullet. As a consequence, the identification of bullets fired by such barrels (which are in general low quality barrels) can be very challenging.

- Finishing of Barrel Interior Surface: Manufacturers of good quality barrels often add steps to the barrel manufacture procedure to polish the interior surface of the barrel to a good quality finish. The purpose of this effort is to decrease the friction between the barrel and the bullet. However, as a highly polished barrel interior has less surface imperfections, the number and depth of features transferred between the barrel and the bullet is also decreased. As the depth of the features transferred between the barrel and the bullet decreases, it is more challenging for any type of instrumentation to detect such features. Similarly, as the number of features transferred between the barrel and the bullet decreases, the identification task becomes more challenging. As a consequence, the identification of bullets fired by very high quality barrels can be very challenging.

The three examples given above are by no means an exhaustive list of the ways in which the manufacture of the barrel may influence the degree of difficulty of the identification process. These examples are meant to show that challenging identification situations may arise both in the case of poorly manufactured barrels and very high quality barrels. In the case of high quality barrels these challenges can often be overcome by the use of better, more sensitive instrumentation. In the case of poor quality barrels improving the sensitivity of instrumentation will not be helpful in general. The bullets fired by such barrels pose a different challenge because in general only portions of the features will transfer between two bullets fired by the same gun. In other words, bullets fired by such barrels require smarter algorithms than those so far developed. Such algorithms should be sophisticated enough to seek "portions" of agreement, and not entire agreement.

Of the three parameters discussed above, the only one which can be analyzed within the scope of this study is the surface finishing of the barrel interior. Figure 17 shows a visual comparison of land impression cross section data acquired from bullets fired by barrels of the three levels of quality. On the top of Figure 17, one can see that a very good quality barrel, e.g. SIG, creates a very smooth land impression on the bullets fired by it (the amplitude of the features is minimal, on the sub-micron range). This is most likely due to the high quality of the finishing processes during manufacturing. At the middle of Figure 17, one can see that the land impressions found on a bullet fired by a good quality barrel, e.g. Ruger, show a significantly more roughness (the features in this land impression are on the order of a few microns), and very high repeatability. On the bottom of Figure 17, we see an example of a poor quality barrel, e.g. HiPoint. This type of barrel creates an intermediately rough surface on the land impression of bullets fired by it (on the order of a few microns, or even tens of microns). However, the impressions created by poor quality barrels are hardly repeatable from bullet to bullet.

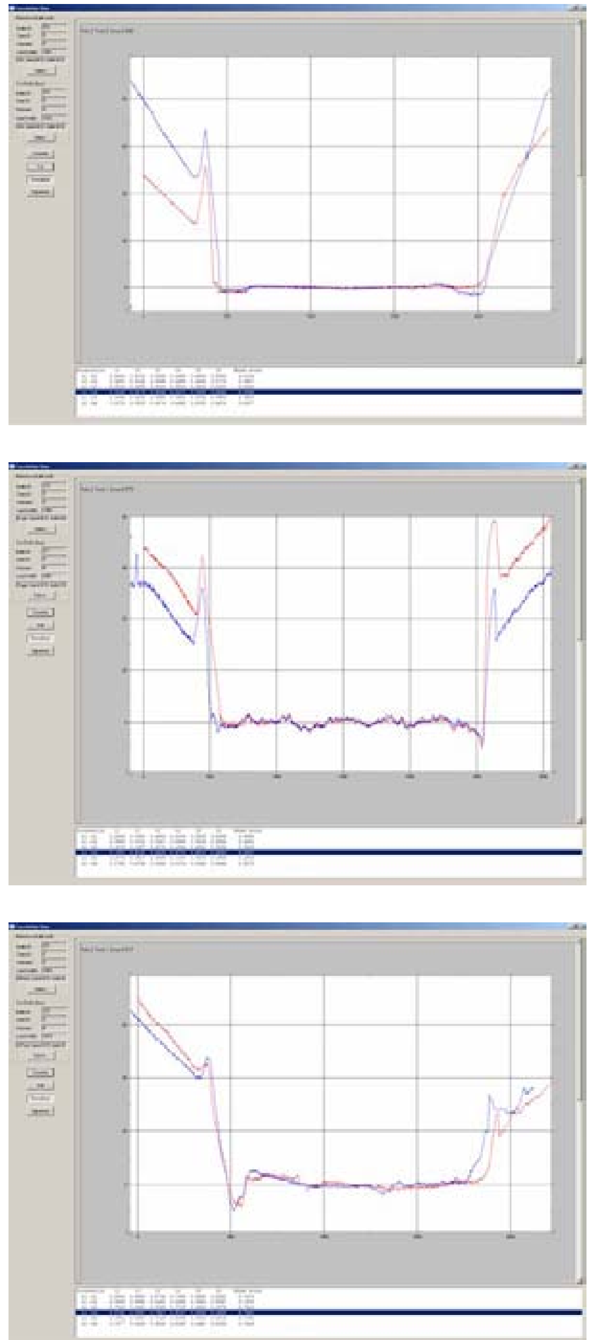


Figure 17 Land impression comparison of very good quality (top, Sig Sauer), good quality (middle, Beretta) and poor quality (bottom, HiPoint)

A quantitative evaluation of the finishing of the barrel interior surface can be performed by computing the median RMS (Root Mean Square) surface roughness (in microns) of the LEAs associated with each barrel brand and ammunition type under consideration. Figure 18 shows a scatter plot of the empirical probability of error as a function of the median RMS surface roughness for each barrel and ammunition under consideration. A least square fit to a power curve has also been included for each of the two types of ammunitions under consideration. As seen in this plot, the estimated Pe generally decreases as the roughness of the LEAs increases (generally, because there are significant exceptions.) In fact, we see a very sharp increase in the estimated Pe as the surface roughness

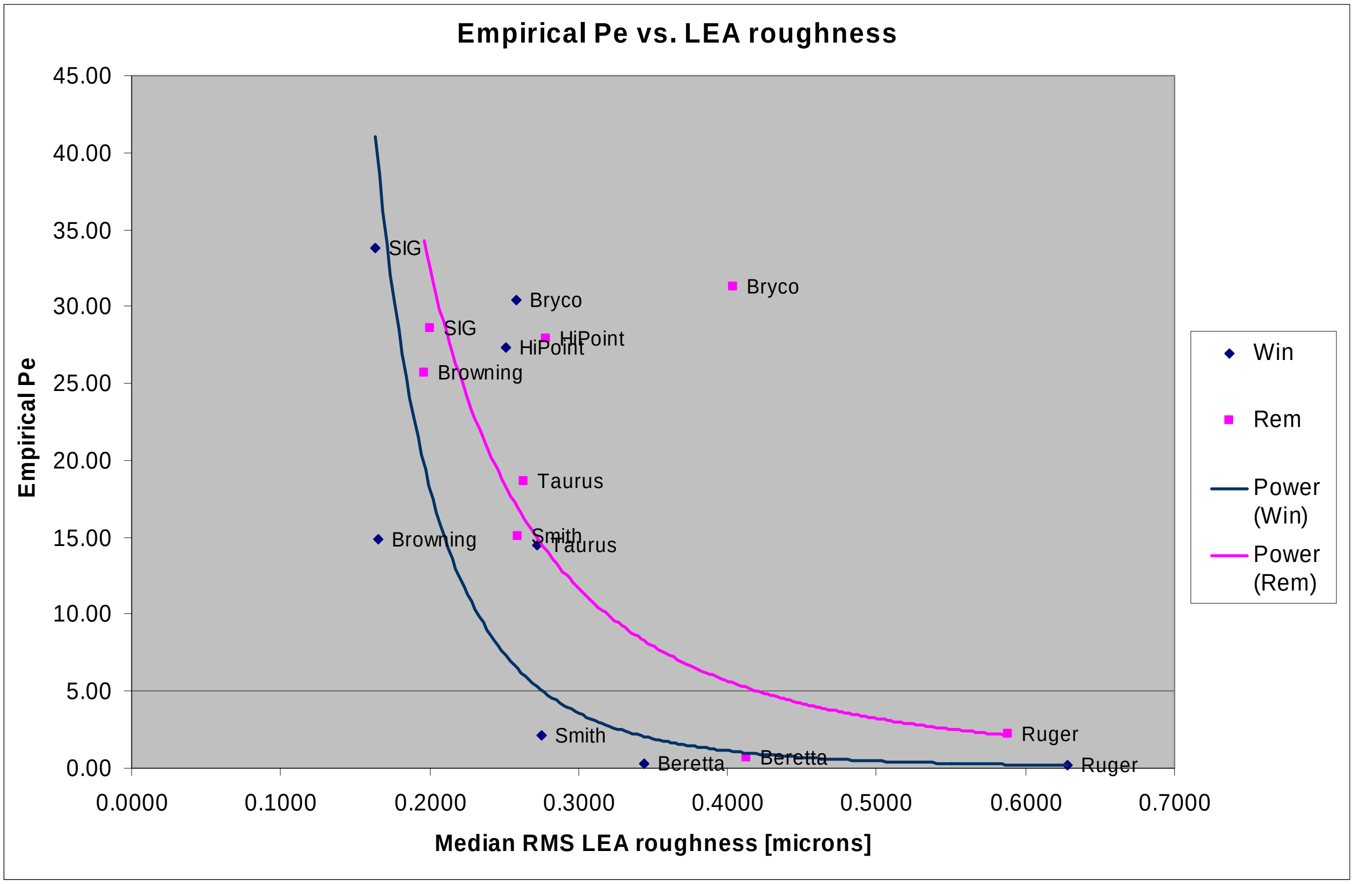


Figure 18: Relationship between Empirical Probability of Error and LEA roughness

goes below 0.3 microns. This indicates that as the size of the features transferred from the barrels to the bullets become smaller, the estimated probability of error becomes significantly larger. We suspect that this phenomenon is due to limitations of the instrumentation currently used to measure the surface of the bullets. Figure 18 seems to indicate that the current instrumentation becomes ineffective below 0.3 microns. Although the specification of the sensor used in this project claims a depth resolution of 0.025 micron, it would seem that the actual resolution of the acquisition platform is on the order of 0.3 – 0.4 microns. This loss of resolution may be due to vibration induced by the motion components during the acquisition process. The consequence of the reduced resolution capability of the acquisition platform is to prevent us from obtaining useful data for many of the barrels/bullets under consideration. It is worth noting that the required acquisition platform resolution was not known at the beginning of this project, since no such measurements over such a varied spectrum of barrels had been made before this project.

An interesting data point in Figure 18 is that corresponding to Bryco barrels using Remington ammunition. This data point is the most noticeable exception to our previous observations. The median roughness for the LEAs associated with Bryco-Remington was approximately 0.4 microns. Nevertheless, the probability of error was over 30%, significantly higher than that achieved for bullets with lower roughness values (such as Winchester bullets fired by Beretta barrels, which have a median roughness of .34 microns, and an empirical Pe of 0.3%.) For this particular set of data, it is reasonable to assume that the features on the surface of the bullet were within the measurement capabilities of the sensor; and yet the probability of error is significantly high. One has to conclude that for this barrel/ammunition combination the features transferred between the barrel and the bullets were not repeatable **for the algorithms used in the current implementation of the system**. We cannot conclude, however, that a

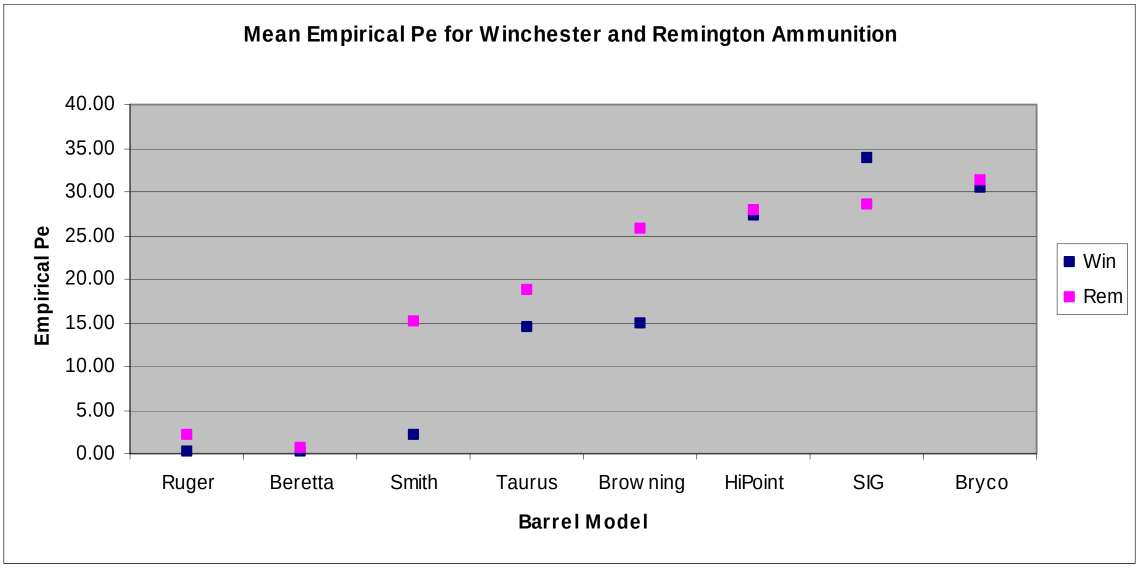


Figure 19 : Effect of Ammunition Manufacture

human examiner or a better set of algorithms would not have been able to perform better. **It is important to keep in mind that the performance achieved by the present implementation of the system provides an upper bound of the achievable performance.** Besides resolution limitations, the amount of data used by the current algorithms is so limited (only five cross sections of the bullet are used) as to preclude identification of good data from unreliable data by the algorithms used in the comparison process. Such limitations could be addressed by a system with better depth resolution and the capability to acquire a greater number of bullet cross sections for analysis.

4.5.2 Effects of Ammunition

The properties of the ammunition submitted for identification can also have a significant effect on the degree to which a bullet can be identified as having been fired by a given gun. Both mechanical characteristics and dimensional tolerances have significant influence in the manner in which features are transferred between the barrel and the bullet. Let us consider the effect of these characteristics:

- Mechanical Characteristics: The mechanical properties of the material used to manufacture the external surface of a bullet (most bullets have an exterior surface coating usually referred to as a “jacket”) will play an important role in the manner in which the features found in the interior of the barrel transfer to the bullets fired by it. As an example, a jacket manufactured with a harder material will accept such features in a different manner than a softer jacket.

- Dimensional Tolerances: As discussed in the case of the barrel bore, the relationship between the bullet diameter and the barrel bore will dictate the amount of pressure between the barrel and the bullet. The pressure between the barrel interior and the bullet will have a significant effect in the manner in which features are transferred between the barrel and the bullet.

Based on the results already presented, it is apparent that there is a significant relationship between the ammunition used and the resulting individuality results. For each of the two ammunition types used in this evaluation, Figure 19 shows the mean of the empirical probability of error for all the barrels of each of the barrel models under consideration. As seen in Figure 19, the empirical probability of error is consistently lower for Winchester ammunition than for Remington ammunition. The only exception to this rule is SIG. However, given that the ability of the instrumentation used in this evaluation to detect the features found on bullets fired by SIG barrels is questionable, the consequence of this exception is questionable too.

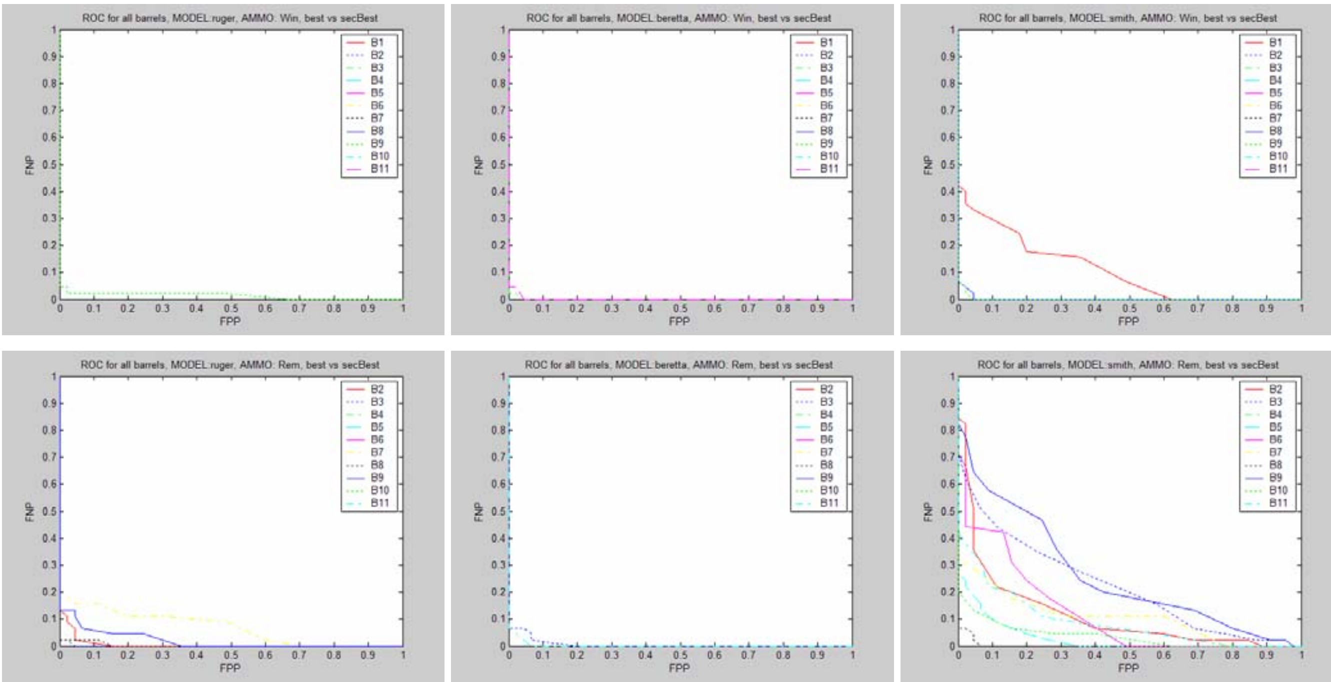
The significance of the results shown in Figure 19 is that the degree of individuality associated with a barrel depends on the ammunition used to evaluate it. This is an important observation because it implies that the likelihood of a correct identification depends not only on the barrel model of the suspect barrel/gun, but also on the evidence ammunition. This conclusion should not come as a surprise. Among the many parameters associated with the manner in which features are transferred between a barrel and a bullet is the relationship between the barrel bore and the bullet diameter. Another such factor is the hardness of the bullet jacket. Different ammunition manufacturers use different tolerances and materials in their bullets. Therefore, it is to be expected that the bullet manufacture plays a significant role in the quality of the features transferred between barrel and bullet. Although beyond the scope of the present effort, it would be of significant interest to develop a better understanding of the parameters which play a role in the quality of the features transferred between barrel and bullet.

4.5.3 Are all Barrels Created Equal?

So far, we have discussed the variation in terms of individuality between different barrel models (or more accurately, different brands of barrels). A related question of interest is the following: Are all barrels of the same model/brand similarly individual? Inspection of the results compiled in Table 9 indicate that even for barrels of the same manufacture there is some variation in the average best and second-best similarity measures. Take as an example Smith and Wesson barrel 1, with best similarity measure average of 0.43, against barrel 4 of the same model, with best similarity measure average of 0.77. Notice also that the average second-best similarity measure is not that different for these barrels; 0.36 for barrel 1, and 0.38 for barrel 4. Clearly, although of the same brand, these barrels show considerable difference.

The variation of individuality between different barrels of the same make and model can be graphically displayed by using ROC curves. Figure 20 shows the ROC curves for each of the barrels under consideration for Ruger, Beretta and Smith and Wesson for both Winchester (top) and Remington (bottom) ammunition. In the case of these barrel models it is clear that Ruger and Beretta display extremely good individuality characteristics (one can barely see the different ROC curves) and very consistent results (although it is questionable whether any difference could be gleamed from these plots since both barrel models perform extremely well). On the other hand, in the case of Smith & Wesson, the performance depends on the ammunition, where very good performance can be seen for Winchester ammunition (except for a single barrel out of 11), but mediocre and inconsistent performance is observed for Remington ammunition (performance varies widely between barrels).

Figure 21 shows the ROC curves for both Browning and Taurus obtained with Winchester and Remington ammunition. Clearly, the variability in performance is much more noticeable for the different barrels of these models than for Ruger, Beretta or even Smith & Wesson. In the case of the last group of barrels; SIG, Bryco and HiPoint we can see that the performance is very poor and inconsistent between barrel and barrel (see Figure 22).

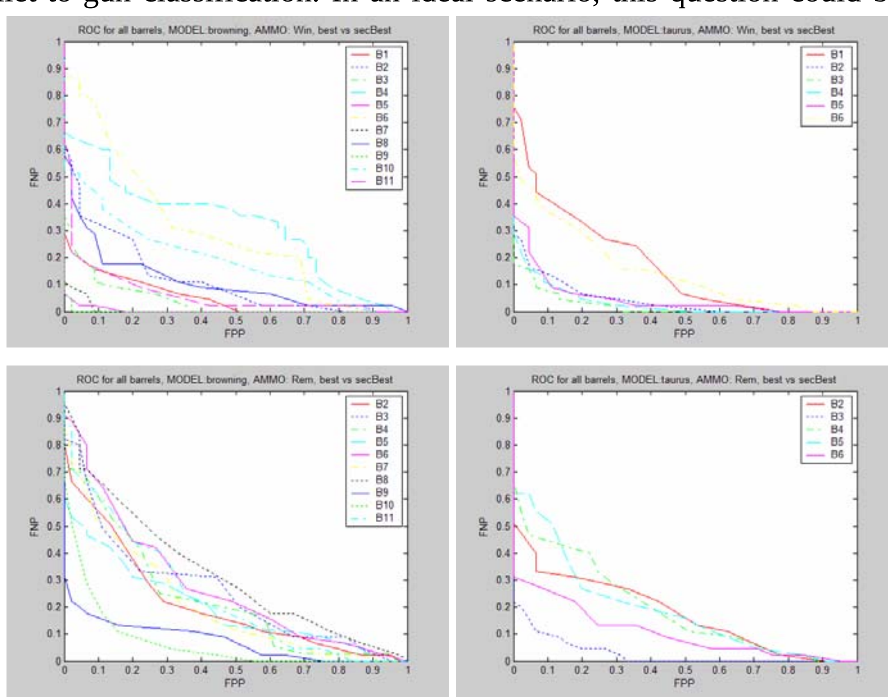


a) b) c)

Figure 20: ROC curves for all barrels; a) Ruger, b) Beretta, c) Smith & Wesson for Winchester (top) and Remington (bottom) ammunition

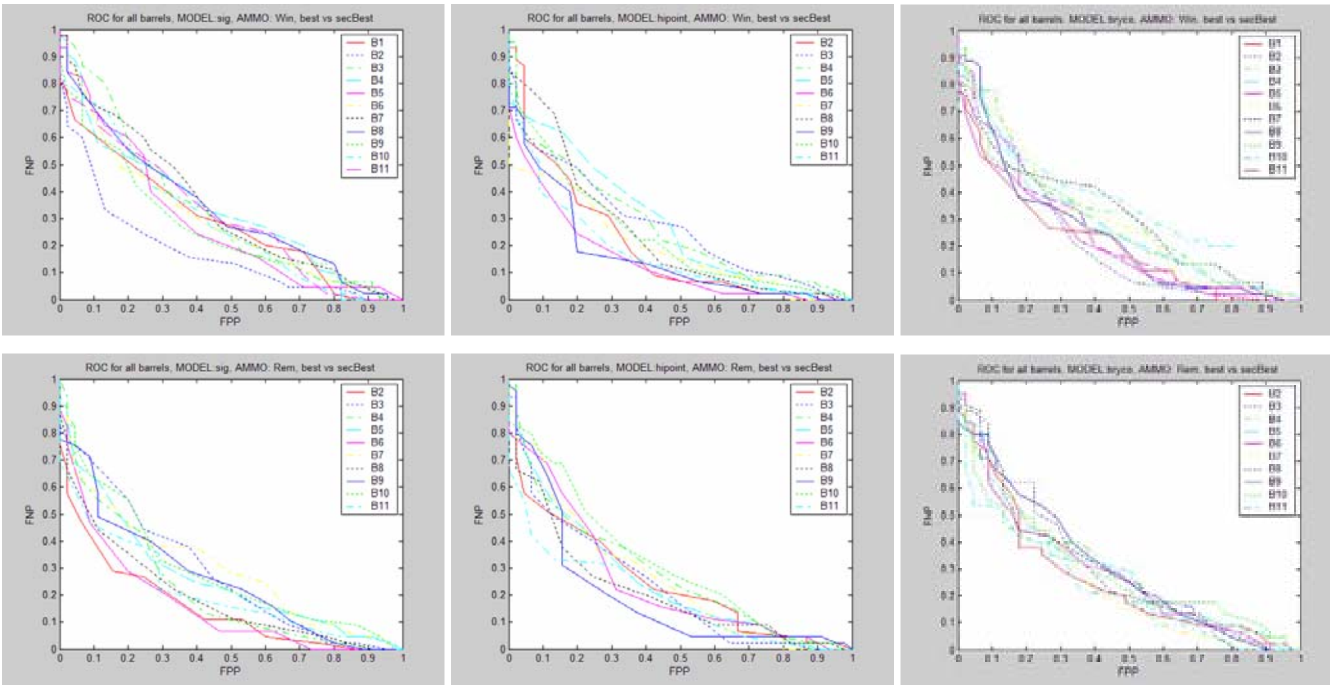
4.6 Pristine Bullets: Classification Study

Let us consider the question of bullet-to-gun classification. In an ideal scenario, this question could be addressed in the following manner: Imagine that a database which contains both the matching and non-matching distributions associated with guns of the make and model of the suspect gun is available. Such database would be the result of performing a large number of comparisons between bullets fired by different guns of the same make and model as that of the suspect gun, and compiling the matching and non-matching similarity measures. When faced with a classification decision, a comparison between control bullets and the evidence bullet could be made to obtain a sample of evidence-to-control similarity measures (at this point we do not know if they are matching or non-matching). The classification



a) b)

Figure 21: ROC curves for all barrels; a) Browning, b) Taurus for Winchester (top) and Remington (bottom) ammunition



a) b) c)

Figure 22: ROC curves for all barrels; a) SIG, b) HiPoint, c) Bryco for Winchester (top) and Remington (bottom) ammunition

decision could then be made by determined whether the sample distribution of evidence-to-control similarity measures most resembles the matching or the non-matching distributions corresponding to the guns of the same make and model as that of the suspect gun obtained from our ideal database. Furthermore, equipped with all this data, it would be possible to also estimate the probability of a false positive or a false negative determination.

In a realistic scenario, our conceptual database of matching and non-matching similarity distributions is not available, and the creation of such a database would be a monumental effort. As a practical alternative, perhaps it would be possible to estimate these distributions using the control bullets? Let us consider the best and second-best orientation similarity measure distributions obtained by comparing the control bullets among themselves. We postulate that these distributions can be used as approximations/estimates of the matching and non-matching distributions for the gun under consideration. The rational for such approximation is the following: As already discussed, the best orientation similarity distribution resulting from the comparison of the control bullets is in fact the same as the matching distribution for a set of guns consisting of the gun under consideration (the suspect gun). Therefore, the matching distribution obtained by comparing the control bullets provides us with a sample of the matching distribution. In terms of the non-matching distribution, we propose that it can be approximated by the second best distribution obtained from the comparison of the control bullets. This is based on the fact that the similarity values obtained from comparing land impressions created by two different barrel lands has a characteristic distribution regardless of whether these lands belong to the same barrel or not.

In the following sections we evaluate the extent to which the proposed approach can be implemented, and the results obtained using this approach. In Section 4.6.1, we analyze the statistical behavior or the non-matching distribution and the second best distribution for the barrels and ammunition under consideration, and we seek to answer the questions whether these distributions can be parameterized (in particular, are they normal?) and whether they are, as suggested, similar.

4.6.1 Second-best distribution vs. non-matching distribution

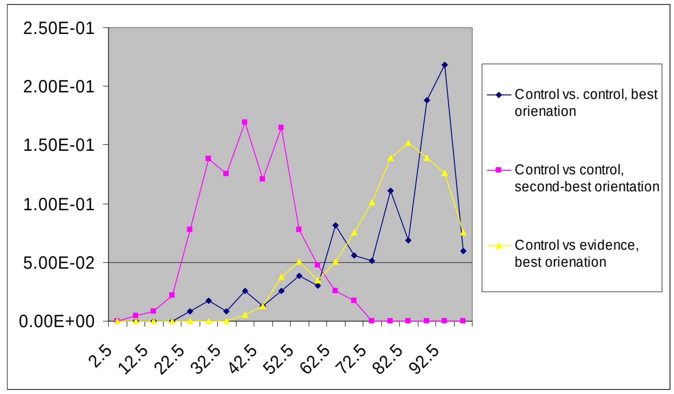


Figure 23: Example of best and second-best orientation similarity measure distributions of among control bullets; and distribution of best orientation similarity measure between control bullets vs. evidence bullet.

In Section 4.2.3, we discussed the basic conceptual approach to the bullet-to-gun classification problem. Let us begin this section with an actual example. Figure 23 shows a typical example of the distribution of the best (blue) and second-best (pink) orientation similarity measures corresponding to control bullets fired by a Ruger gun. Figure 23 also shows the distribution of the similarity measures obtained by comparing an evidence bullet against the control bullets in their best orientation (yellow). As can be seen in this example, the distribution obtained by comparing the evidence bullet against the control bullets is more "similar" to the best orientation similarity measures distribution, suggesting that the evidence bullet was fired by the same gun which fired the control bullets (which was the case). Had the distribution obtained by comparing the evidence bullet against the control bullets been more similar to the second-best orientation distribution, we would have concluded that the evidence bullet was not fired by the same gun as the control bullets. This is the principle applied in our classification algorithms

4.6.1.1 Non-matching Distribution

Visual observation of the histograms of the second best and the non-matching distributions suggests that both distributions are similar and appear to be normal (see Report No. 11, Appendix A). To verify their normality, we applied the Lilliefors test to the data. Lilliefors test is a refinement of the one-sample Kolmogorove-Smirnov (KS) test, which has long been used to evaluate the normality of a data sample. The null hypothesis and its alternate hypothesis to be tested are:

H0: the sample data originates from a normal distribution

H1: the sample data does not originate from a normal distribution

As we are interested in differences for all sample values, we perform a two-tailed test. Furthermore, the significant level α, is fixed at α=0.05, or at the 5% level, throughout the hypothesis test in the following sections.

The Lilliefors test is based on the maximum difference between the empirical cumulative distribution function (CDF) from the sample, and the normal CDF using the sample's mean and standard deviation. Reliance on a single value (the maximum difference) may make this test sensitive to small clusters of data within the sample. To better understand the effects of the sample size on the hypothesis test outcomes, we randomly selected samples with different sizes from the non-matching dataset and applied the Lilliefors test to test their normality. As an example, the hypothesis results from the Ruger model and Winchester ammunition are shown in Table 10, where the experiment was repeated 10 times for each

Sample Size	Test Statistics	Critical Value	Outcome
50	0.0799	0.1253	Can not reject H0
100	0.0649	0.0886	Can not reject H0
150	0.0522	0.0723	Can not reject H0
200	0.0499	0.0626	Can not reject H0
250	0.0487	0.0560	Can not reject H0
300	0.0370	0.0512	Can not reject H0
400	0.0314	0.0443	Can not reject H0
495	0.0305	0.0396	Can not reject H0
1000	0.0235	0.0280	Can not reject H0
2000	0.0260	0.0198	Reject H0
3000	0.0187	0.0162	Reject H0
4000	0.0208	0.0140	Reject H0
5500	0.0232	0.0119	Reject H0

Table 10: Lilliefors test of normality for Ruger and Winchester with different sample sizes (significant level 5%)

Sample Size, and the Test Statistic is the average of the 10 repetitions of the test. The results indicate that with a sample size below 1000, we can not reject the normality hypothesis for the non-matching data. However, we can reject the hypothesis as the sample size increases above 2000. The hypothesis test for other gun models and ammunitions yields similar results. Therefore, the non-matching distribution of all gun models and ammunitions is fairly normal as the sample size is about 1000 or less. This normal approximation is particularly appropriate since the sample size available for the second-best and the matching distribution is significantly less than that available from the non-matching distribution in the ballistic classification process.

Once the normality of non-matching distributions for all gun models and ammunitions is verified, the next question comes up naturally: Are all those non-matching distributions from the same underlying normal distribution? To answer this question, the two-sample Kolmogorov-Smirnov test was used to compare the non-matching distributions of different gun models. The null hypothesis and the alternative hypothesis for this test are:

H0: sample 1 and sample 2 data are drawn from the same distribution

H1: sample 1 and sample 2 data are drawn from different distributions

A two-tailed test at the 5% significance level was performed. As an example, the results of non-matching distribution comparison between the Ruger and the rest of the gun models are listed in Table 11. Similar results have been obtained from the comparison between the non-matching distributions of other gun models and ammunitions. These results indicate that the underlying distributions for the non-matching datasets are different for different guns. This is not entirely surprising. We would anticipate that the mean of the non-matching distribution will vary according to the roughness of the LEAs of the brand under consideration.

	p-Value	Test Statistics	Outcome
Ruger vs. Beretta	2.267E-126	0.2290	Reject H0
Ruger vs. Smith	0	0.6071	Reject H0
Ruger vs. Browning	2.51E-33	0.1171	Reject H0
Ruger vs. Sig	0	0.4793	Reject H0
Ruger vs. Hipoint	0	0.2421	Reject H0
Ruger vs. Bryco	2.92E-7	0.2267	Reject H0

Table 11: Kolmogorove-Smironov test results of non-matching distributions of Ruger and other gun models with the Winchester ammunition

Figure 24 summarizes the comparison of means of non-matching distributions for all gun models and ammunitions under investigation. As seen in Figure 24, the mean of the non-matching distributions vary for the different barrel models. These results suggest that the non-matching distribution needs to be characterized for each a specific gun model. The mean of the non-matching for different ammunitions (Winchester vs. Remington) agree with each other very well in most cases, except in the case of SIG and Bryco barrels. As discussed earlier, the ability of the present system to acquire and process data from these barrel brands is limited, so that it is not clear whether this phenomenon is due to the bullets or the system itself. For all other barrel brands, the mean of the non-matching distribution is relatively independent of the ammunition type.

4.6.1.2 Second Best Distribution

The same approach used to study the normality of the non-matching distribution has been applied to the second best distribution. Again, the data from Ruger with Winchester has been used as the example, and the Lilliefors test results are plotted in Figure 25. Keep in mind that the null hypothesis (the samples are drawn from a normal distribution) can not be rejected as long as the test statistics is less than the critical value. The results shown in Figure 25 indicate that the second best distribution is mostly normal, except

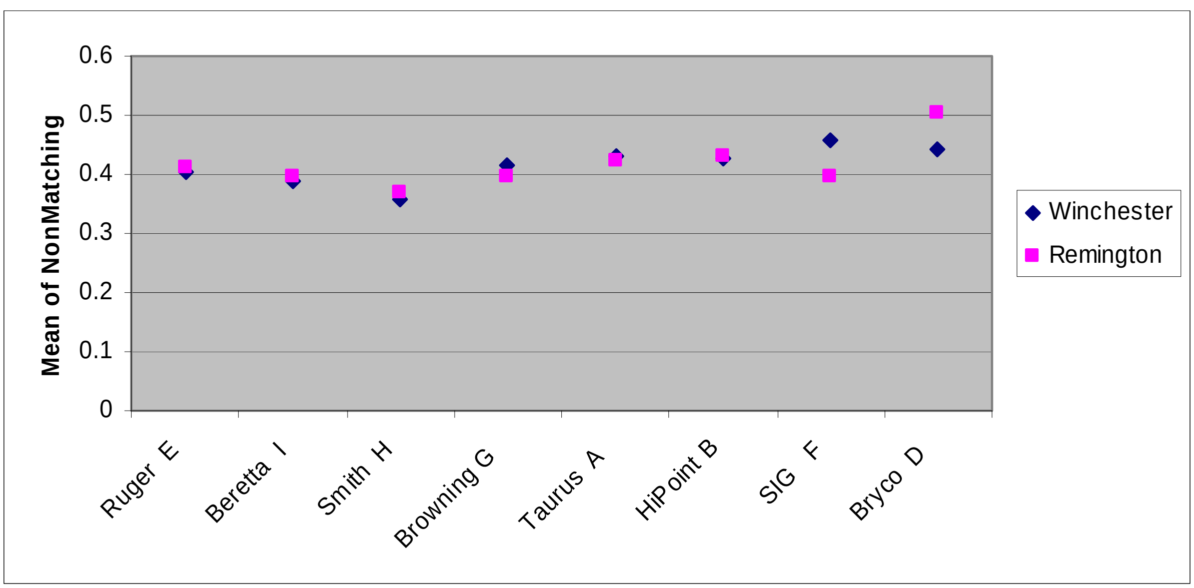


Figure 24: Comparison of non-matching distributions mean for different guns and ammunition

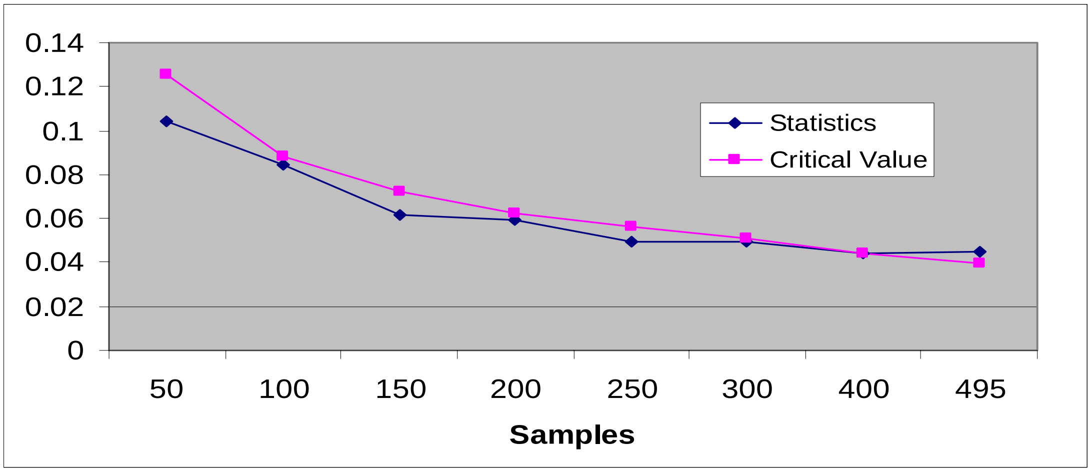


Figure 25: Lilliefors normality test results for second-best distribution for Ruger with Winchester

that last data point with all 495 samples. The same Lilliefors normality test has been applied to the second best distribution of the remaining gun models and ammunitions. Even in the case of all 495 samples are used in the Lilliefors test, 10 out of 15 combinations of gun models and ammunitions confirm the normality of their second best distributions.

Since the second best distribution appears to be normal, parametric analysis tools can be used to compare the second best distributions among different gun models and ammunitions. Figure 26 shows the comparison of means of the second best distributions of all guns and ammunitions. Similar to the appearance in Figure 24, results shown in Figure 24 confirm the findings from non-matching data that the second best distribution depends mostly on the barrel brand, and appears to be independent of the ammunition type. Once again, the exceptions to this rule are SIG and Bryco, a fact which may need further investigation.

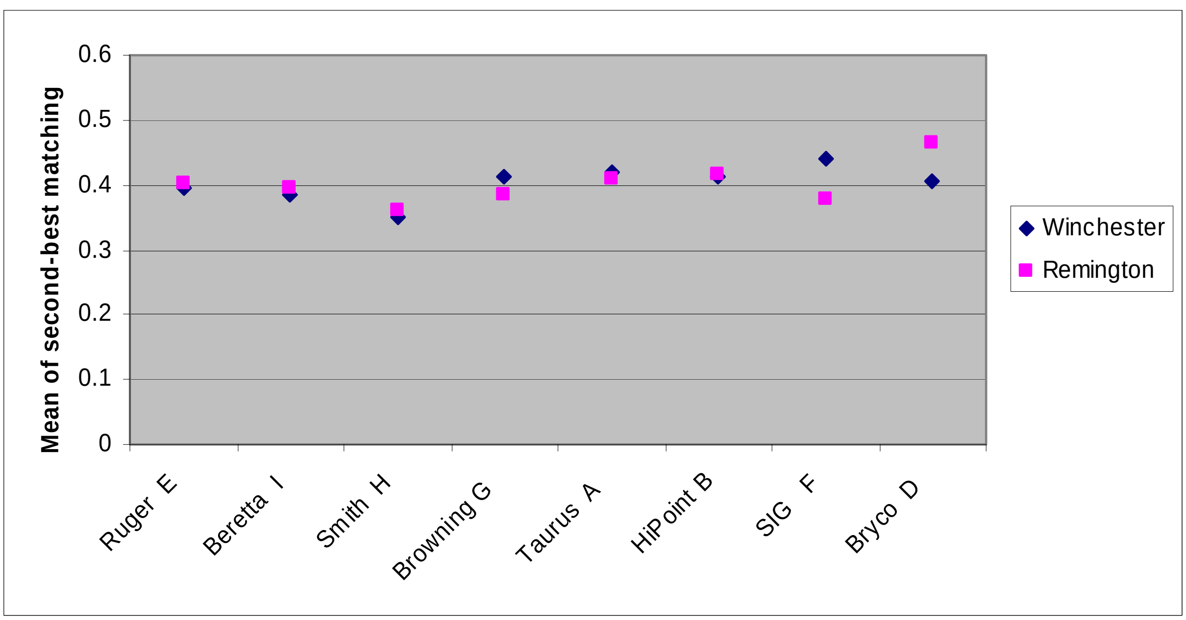


Figure 26: Comparison of second-best distributions mean among different guns and ammunitions

		Lillietest of Normality with Transformation (Alpha=0.05)								
		dataOut = dataIn			dataOut = -log(1-dataIn)			dataOut = -log(dataIn)		
		Match	Non-Match	Sec. Match	Match	Non-match	Sec. match	Match	Non-match	Sec. match
Ruger E	Win	Not normal	Not normal	Not normal	**Normal**	Not normal	Not normal	Not normal	**Normal**	**Normal**
	Rem	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal	**Normal**	**Normal**
Beretta I	Win	Not normal	Not normal	**Normal**	Not normal	Not normal	**Normal**	Not normal	**Normal**	**Normal**
	Rem	Not normal	Not normal	**Normal**	Not normal	Not normal	**Normal**	Not normal	**Normal**	**Normal**
Smith H	Win	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal	**Normal**	**Normal**
	Rem	Not normal	Not normal	**Normal**	Not normal	Not normal	**Normal**	Not normal	**Normal**	**Normal**
Browning G	Win	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal
	Rem	Not normal	**Normal**	**Normal**	Not normal	Not normal	**Normal**	Not normal	Not normal	**Normal**
Taurus A	Win	**Normal**	Not normal	**Normal**	**Normal**	Not normal	**Normal**	**Normal**	**Normal**	**Normal**
	Rem	Not normal	**Normal**	Not normal	Not normal	**Normal**	Not normal	**Normal**	Not normal	Not normal
SIG F	Win	Not normal	Not normal	**Normal**	Not normal	Not normal	**Normal**	**Normal**	Not normal	**Normal**
	Rem	**Normal**	Not normal	**Normal**	**Normal**	Not normal	**Normal**	**Normal**	Not normal	Not normal
HiPoint B	Win	Not normal	Not normal	**Normal**	Not normal	Not normal	**Normal**	Not normal	Not normal	Not normal
	Rem	Not normal	Not normal	**Normal**	Not normal	Not normal	**Normal**	N/A	N/A	N/A
Bryco D	Win	Not normal	Not normal	**Not normal**	Not normal	Not normal	Not normal	N/A	N/A	N/A
	Rem	Not normal	Not normal	Not normal	Not normal	Not normal	Not normal	**Normal**	Not normal	**Normal**

Table 12: Lilliefors normality test results of data sets before and after transformation

4.6.1.3 Relationship between Non-matching and Second-best Distributions

As part of our uniqueness analysis, we have postulated the premise that the second-best similarity measure obtained by comparing bullets fired by the same barrel has a statistical distribution which closely approximates the non-matching distribution. In this section we seek to verify if this assertion is valid. Preliminary observation of these distributions indicates that our assertion seems to be true, at the very least, for those barrels which display satisfactory individuality characteristics, i.e. Ruger, Beretta, and Smith & Wesson.

Previous normality test results for the non-matching and the second best distributions suggest that both distributions be approximately normal. In order to fully utilize the rich analysis features associated with normal distributions, we attempted to identify a transform to convert the data from approximate normal to truly normal for all sample data. The effort was first made with a logarithm function, which was chosen based on the visual observations of their appearance in the distributions. Besides applying the log transformation to the non-matching and the second best similarity measure data, we also applied it to the matching distribution data with a hope to obtain a normal matching distribution as well. The Lilliefors test was employed as the tool for the evaluation of a distribution's normality. The summarized results of the normality test for the matching, non-matching, and second best matching before and after the logarithmic data transformation are listed in Table 12.

The results in Table 12 show that the transformation –log(1-Data) has virtually no effect on the normality of most distributions. However, the transformation –log(Data) shows a significant effect. We notice that after the –log(Data) transformation, both the non-matching and the second-best matching distributions become normal for those guns which display satisfactory individuality (Ruger, Beretta, Smity & Wesson.) Furthermore, both the non-matching and second best distributions obtained of Taurus with Remington have been proven normal after the transformation.

With the normal distributions of transformed non-matching and second best matching distributions in hand, the hypothesis test has been applied to check if both distributions are statistically the same. For two normal distributions comparison, there are many ways to carry out the test. The first approach we chose is the two-sample Kolmogorove-Smirnov test. The null hypothesis for this test is that both sample sets are

		Hypothesis Test of Non-match and Sec. match Distributions after Transformation (Alpha=0.05, dataOut = -log(dataIn))														
		dataOut = -log(dataIn)		Kolmogorov-Smirnov test			t-test						ranksum test			
		Non-match	Sec. match	H	P	KSStats	H	Significance	CI(1)	CI(2)	Tstats	df	H	P	zval	ranksum
Ruger E	Win	**Normal**	**Normal**	1	8.04E-11	0.1614	1	0	-0.0353	-0.0219	-8.3332	5993	1	0	7.8359	1.77E+06
	Rem	**Normal**	**Normal**	1	2.04E-05	0.1178	1	1.09E-06	-0.0262	-0.0112	-4.8805	4948	1	7.80E-06	4.4705	1.24E+06
Beretta I	Win	**Normal**	**Normal**	0	0.164	0.0522	1	0.0256	-0.0137	-0.0009	-2.2327	5993	1	0.0358	2.0988	1.56E+06
	Rem	**Normal**	**Normal**	1	0.0295	0.0713	1	0.0118	-0.0166	-0.0021	-2.5194	4948	1	0.0377	2.0779	1174038
Smith H	Win	**Normal**	**Normal**	1	1.47E-05	0.1134	1	1.88E-04	-0.021	-0.0065	-3.7375	5993	1	1.97E-04	3.7225	1.62E+06
	Rem	**Normal**	**Normal**	1	1.08E-06	0.132	1	2.04E-08	-0.0318	-0.0153	-5.6181	4948	1	2.30E-08	5.5879	1275492
Browning G	Win	Not normal	Not normal													
	Rem	Not normal	**Normal**													
Taurus A	Win	**Normal**	**Normal**	0	0.064	0.086	1	0.0026	-0.0264	-0.0056	-3.0188	1768	1	0.0109	2.5469	258776
	Rem	Not normal	Not normal													
SIG F	Win	Not normal	**Normal**													
	Rem	Not normal	Not normal													
HiPoint B	Win	Not normal	Not normal													
	Rem	N/A	N/A													
Bryco D	Win	N/A	N/A													
	Rem	Not normal	**Normal**													

Table 13: Hypothesis test results of transformed non-matching and the second best data from selected guns and ammunitions.

drawn from the same continuous distribution. The alternative hypothesis is that they are drawn from different continuous distributions. We assign the test result H is 1 if one can reject the hypothesis that the distributions are the same and 0 if one cannot reject that hypothesis.

The second way to compare two normal distributions is the t-test, which determines whether two samples from a normal distribution could have the same mean when the standard deviations are unknown but assumed equal. Again, we assign the test result H a value of 1 if one can reject the null hypothesis that the means are equal at the 0.05 significance level and 0 otherwise.

The third method we chose to compare the non-matching and second best matching distributions is the Wilcoxon ranksum test, which performs a two-sided rank sum test of the hypothesis that two independent samples come from distributions with equal medians. The *p*-value returned from the test is the probability of observing the given result. Small values of *p* cast doubt on the validity of the null hypothesis. The null hypothesis is that two sets of data are assumed to come from continuous distributions that are identical except possibly for a location shift, but are otherwise arbitrary. If the result H is 1, then the null hypothesis, i.e., medians are equal, can be rejected at the 5% level (default).

The test results of comparing the non-matching and the second best distributions of selected guns and ammunitions with three hypothesis tests described above are summarized in Table 13. Clearly, the hypotheses that the non-matching and the second-best similarity measure are drawn from the same distribution, or to have same mean, or to have same median value can be rejected.

To better understand the internal relationship between the non-matching distribution and the second-best distribution, we compared the mean, median, and standard deviation of both distributions, as well as their differences. The results listed in Table 14 show the mean and median of the non-matching distributions are slightly higher than those of the second-best matching distributions. This systematical offset owes to the following fact: Given a barrel with *n* lands (or a bullet with *n*), the non-matching dataset correspond to a set of **best** similarity measures out of ***n*** possible orientations, while the second-best dataset corresponds to a set of second-best similarity values out of *n* possible orientations, or in other words, the **best** similarity measures out of ***n-1*** orientations. Therefore, under the assumption that the distribution of similarity measures obtain from the comparison of non-matching LEAs is the same whether the pair of LEAs belongs to the same barrel or not, one should expect that the highest out of ***n*** possible samples of a given distribution should be higher than the highest of ***n-1*** possible samples. For this reason, the non-matching distribution has a higher average than the second-best distribution. By the same argument, the

		Non-match and Sec. match Distribution Comparison											
		Non-match			Sec. match			Difference (Non-match - Sec.matc			Different percentage (%)		
		Mean	Median	Std	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std
Ruger E	Win	0.40	0.40	0.11	0.40	0.40	0.10	0.01	0.01	0.01	1.93	2.08	8.54
	Rem	0.41	0.41	0.11	0.40	0.40	0.11	0.01	0.01	0.00	1.94	1.78	-0.04
Beretta I	Win	0.39	0.39	0.10	0.38	0.39	0.10	0.00	0.00	0.00	0.68	0.36	1.85
	Rem	0.40	0.40	0.11	0.39	0.40	0.11	0.00	0.00	0.00	0.82	0.42	1.56
Smith H	Win	0.36	0.36	0.09	0.35	0.35	0.09	0.01	0.01	0.00	2.03	2.06	0.89
	Rem	0.37	0.37	0.10	0.36	0.36	0.10	0.01	0.01	0.00	2.28	2.74	2.48
Browning G	Win	0.41	0.42	0.12	0.41	0.41	0.12	0.00	0.00	0.00	0.54	0.65	-0.98
	Rem	0.40	0.40	0.12	0.39	0.39	0.12	0.01	0.01	0.00	2.64	2.64	-0.39
Taurus A	Win	0.43	0.44	0.14	0.42	0.43	0.13	0.01	0.01	0.01	2.69	1.55	5.91
	Rem	0.42	0.43	0.14	0.41	0.42	0.14	0.01	0.02	0.00	3.27	3.61	0.62
HiPoint B	Win	0.43	0.43	0.15	0.41	0.41	0.15	0.01	0.02	0.00	3.38	4.46	-0.05
	Rem	0.43	0.43	0.14	0.41	0.42	0.15	0.02	0.01	0.00	3.96	2.97	-3.04
SIG F	Win	0.46	0.46	0.13	0.44	0.44	0.12	0.02	0.02	0.00	3.66	3.98	3.65
	Rem	0.40	0.40	0.12	0.38	0.38	0.12	0.02	0.02	0.01	5.14	5.13	4.87
Bryco D	Win	0.44	0.45	0.15	0.40	0.40	0.16	0.04	0.04	0.00	8.49	9.93	-1.52
	Rem	0.51	0.51	0.17	0.47	0.46	0.17	0.04	0.04	0.00	7.96	8.85	1.76

Table 14: Statistical comparison of non-matching and second best similarity measure

standard deviation of the second-best distribution should be slightly higher than that of the non-matching distribution. In terms of the mean and median of these distributions, this phenomenon can be clearly seen in Table 14. As far as the variance, this phenomena can be seen for the good barrels; those barrels for which the current system is able to make reliable differentiation between the best and second-best distributions.

4.6.2 Matching distribution

As mentioned earlier, the matching distribution appears to be skewed (and therefore not normal). This postulate has been verified by means of the statistical normality test, or the Lilliefors test as described above. The results in Table 12 show that most matching distributions are not normal, especially for those guns which display good uniqueness properties (Ruger, Beretta, Smith, and Browning). By comparing the histograms of the matching distributions as shown in Appendix A of Report No. 11, the following conclusions can be easily drawn:

- Most matching distributions are not normal
- The matching distribution varies between gun to gun
- The matching distribution varies from barrel to barrel even for the same gun model
- The matching distribution is dependant on the ammunition

Furthermore, the barrel quality has been found to play an important role for the matching distribution. On one hand, the limitation of the current sensor hardware prevents the acquisition of the features transferred to the bullets from high quality barrels (such as Sig and Browning). On the other hand, the limitation of the current matching algorithms hinders the extraction of features on the bullets fired from low quality barrels (such as Taurus and Hipoint). As a result, the barrels with “good” middle of the road quality possess the best identification capability with a matching distribution clearly distinguishable from the non-matching distribution. All these conclusions provide guidance for the future ballistic identification that the matching distribution can only be made from a specific barrel, such as the suspect gun found from the crime scene.

4.6.3 Empirical Results

In order to assess the probability of error associated with a given comparison methodology, barrel model, ammunition manufacture and number of control bullets, we simulated combinations of randomly selected “control bullets,” (index I_c in Eqn 14) and “evidence” bullets (index b_e in Eqn. 16), and applied the

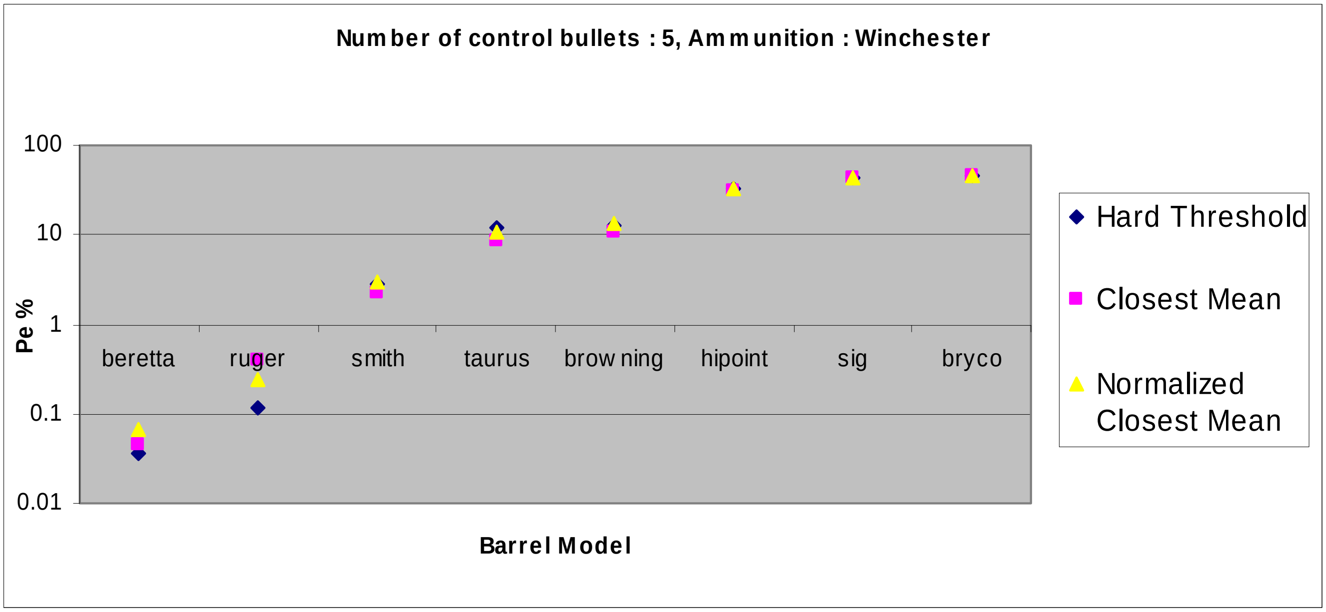


Figure 27: Probability of error as a function of barrel model

different comparison methodologies discussed in Section 4.2.3.2 to the different gun models and ammunitions available. Having a-priori knowledge regarding the correct classification between the selected control and evidence bullets allows us to determine whether the true answer to the classification question is a match or a non-match. In this manner, it is possible to estimate the probability of false positive, false negative and overall error for any combination of classification methodology, barrel, ammunition and number of control bullets. In this section we discuss the results obtained following this methodology.

4.6.3.1 Effects of Barrel Model/Quality

Of all the variables influencing the probability of individuality and classification error, the barrel make/model appears to be the most significant. Figure 27 shows the probability of error computed for all barrel models under consideration and all three comparison methodologies discussed in Section 4.2.3.2 (while assuming 5 control bullets of Winchester brand). From the results shown in Figure 27, it can be seen that the "good" quality barrels (Beretta, Ruger, Smith & Wesson) achieve significantly lower error rates in the classification process (notice that the vertical scale is logarithmic.) Both "very good" quality barrels (SIG, and Browning) and "poor" quality guns (Taurus, HiPoint and Bryco) produce much larger probability of error[1]. As mentioned earlier, we believe that this phenomenon is due to the limitations of the current hardware and software.

4.6.3.2 Effects of Ammunition Manufacture

Since land impressions are created as a result of the interaction between a bullet and a barrel, it is reasonable to expect that the bullet manufacture properties (e.g., dimensions, hardness, elastic properties, etc) will affect the transfer of features from the barrel to the bullet. Figure 28 shows the results of using

[1] The qualification of barrel makes/models as "very good," "good" and "poor" is not meant to be an evaluation of the quality of the barrel or the weapon itself, but only of the repeatability of the features transferred between the barrels of the make/model under consideration and the bullets fired by it.

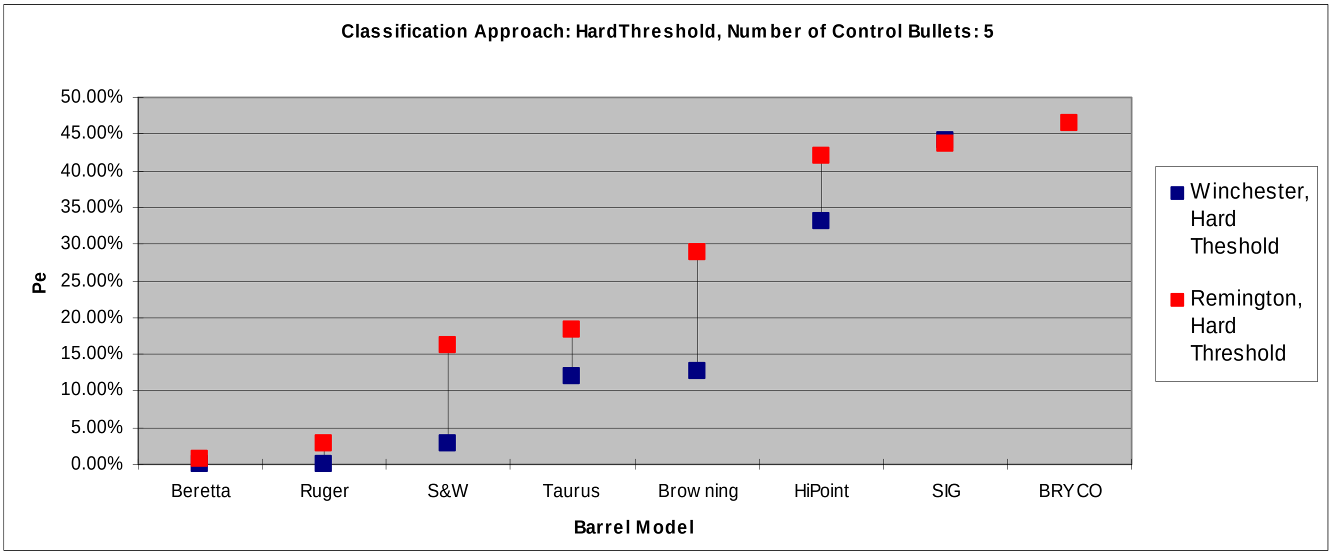


Figure 28: Effect of ammunition manufacture on overall Probability of error

the same classification approach (hard threshold) and number of control bullets (5 control bullets) for all barrel models with bullets of Winchester and Remington manufacture.

As seen in Figure 28, the probability of error for Winchester ammunition is lower than that from Remington for all barrels other than SIG (a similar phenomenon was observed in the individuality evaluation, see Figure 19). However, even in the case of SIG barrels, the probability of error is so high with either type of ammunition (approx 44%), that it is difficult to reach any kind of conclusion from this phenomenon. More important, the difference of probability of error between two ammunitions is very significant even for good quality guns such Ruger, Beretta, and Smith. The ammunition manufacture appears to be the most significant variable affecting the probability of identification second only to the barrel manufacture.

4.6.3.3 Effects of Number of Control Bullets

Common sense dictates that the greater the number of control bullets used in identification, the lower the probability of error should be. A question of interest is then: How significant is the effect of increasing the number of control bullets used in classification? We evaluated the variation in the probability of error for 3, 4, 5 and 6 control bullets. This range of control bullets is consistent with that which a firearms examiner would normally fire during identification.

The results of this analysis in the case of Hard Threshold classification approach are tabulated for both Winchester and Remington ammunition in Table 15. In a similar manner, Figure 29 shows a graphical representation of the same analysis using the normalized closest mean classification approach. As can be seen in both cases, for most barrel models under consideration, the probability of error does decreases as a result of the increase of the number of control bullets. This trend is applicable to both ammunition brands under consideration.

This trend is less clear for those models for which the probability of error is relatively large to begin with (on the order of 30%), which are within groups of very high quality and poor quality barrels, such as the SIG, and HiPoint. This is probably a result from the poor classification ability of the current system for those gun models and ammunitions, which negates the effects of the increased number of control bullets.

Probability of Error		Winchester, Hard Theshold							
		Beretta	Ruger	S&W	Taurus	Browning	HiPoint	SIG	BRYCO
No. Control Bullet	3	0.12%	0.48%	3.46%	14.90%	15.40%	34.02%	44.60%	47.67%
	4	0.05%	0.23%	2.79%	12.12%	15.06%	34.27%	44.65%	46.54%
	5	0.04%	0.11%	2.87%	11.88%	12.73%	33.19%	44.22%	46.41%

Probability of Error		Remington, Hard Threshold							
		Beretta	Ruger	S&W	Taurus	Browning	HiPoint	SIG	BRYCO
No. Control Bullet	3	1.08%	4.78%	17.89%	25.05%	32.18%	43.66%	43.32%	46.77%
	4	0.52%	3.51%	16.73%	20.49%	29.49%	38.81%	43.17%	45.36%
	5	0.74%	2.74%	16.28%	18.40%	28.91%	42.08%	43.64%	46.44%

Table 15: Comparison of probability of error with respective to the number of control bullets for each gun model using different ammunition.

One could also say that if bad data is provided to the classifier, the amount of data is irrelevant (trash in, trash out). By the same token, if the data provided to the classifier is good, the effect of increasing the number of control bullets is significant (see Figure 29, where the scale is logarithmic). Notice also that very low probabilities of error can be achieved as the number of control bullets is increased (as low as 0.04% for Beretta with 5 control bullets, and Winchester ammunition). This is a very encouraging result.

The probability of classification error is inversely proportional to the number of control bullets used in the classification process. Does this mean that it is possible to achieve any arbitrarily low probability of error (by increasing the number of control bullets)? Do all barrels have the same behavior in terms of probability of classification error decay as a function of the number of control bullets? In order to analyze the different barrel models side-by-side, it is convenient to normalize the improvements with respect to a baseline number of control bullets. Having done that, we can study the **relative** change in the probability of classification error as the number of control bullets is increased. Figure 30 shows a graphical representation of the decay of the probability of classification error as a function of the number of control bullets, where the baseline number of control bullets is three. As seen in this plot, the addition of control bullets decreases the probability of error for all barrel models, **but not to the same extent**. Consider for

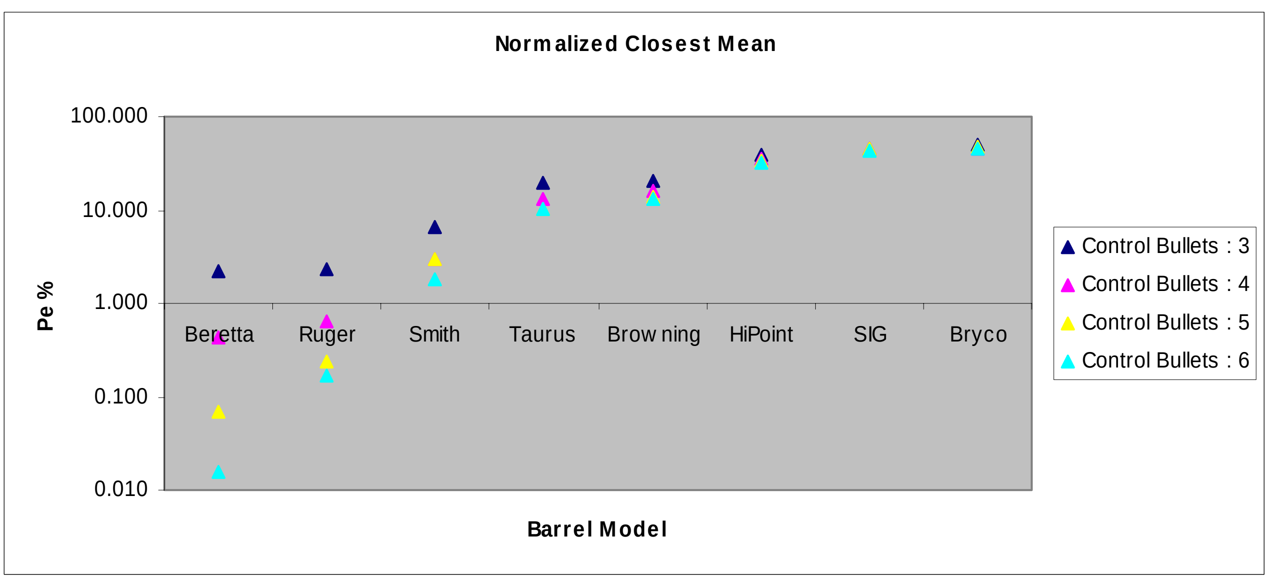


Figure 29: Effect of Number of Control Bullets

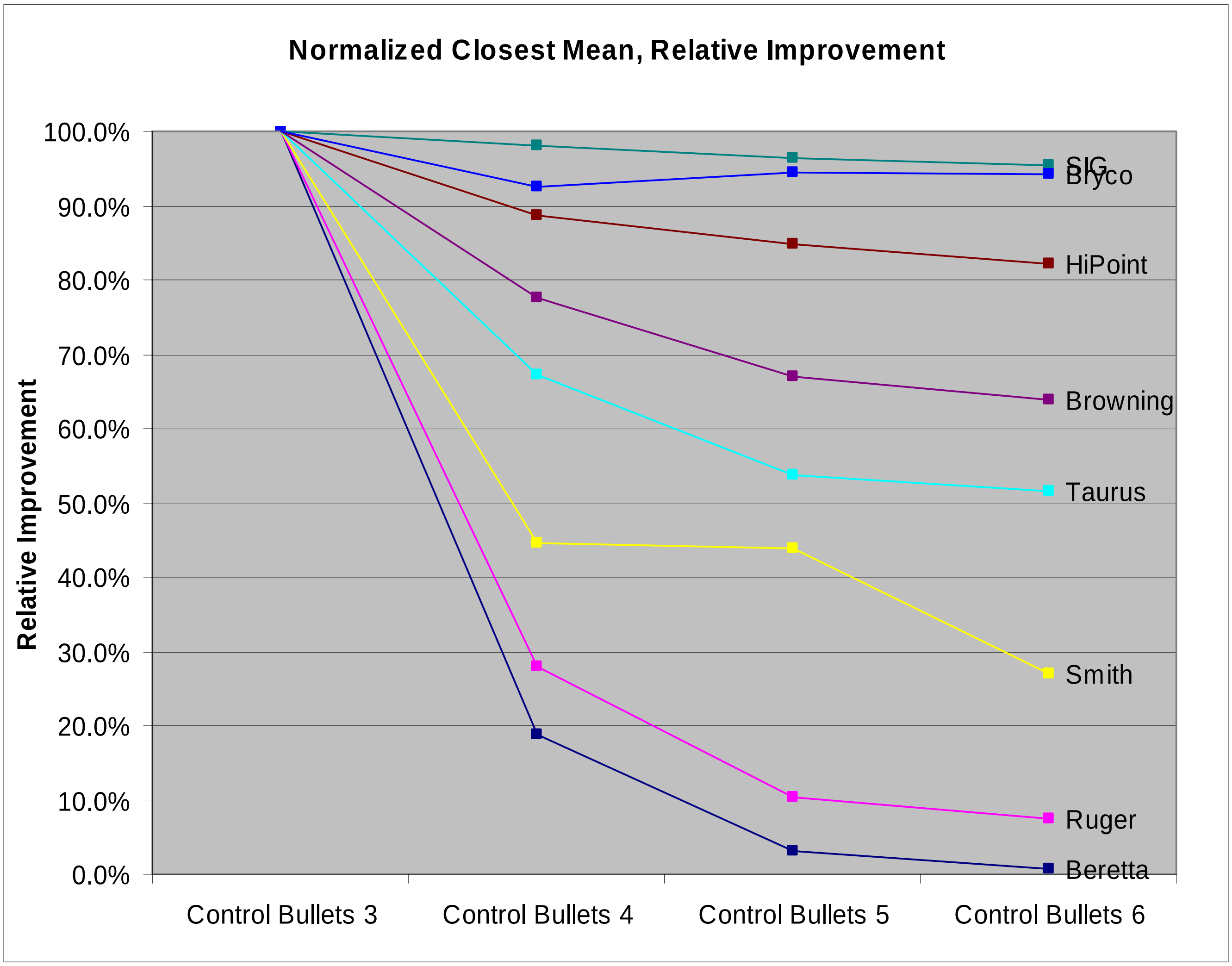


Figure 30: Relative Improvement of Probability of Error as a Function of Number of Control Bullets; Normalized Mean, Winchester Ammunition.

example the effect of adding a forth control bullet. While the probability of classification error for Browning guns decreases to approximately 78% when the number of control bullets increases from three to four, the probability of error for Beretta guns decreases to about 19%. Moreover, the degree of improvement depends on the quality of the barrel. In the case of good quality barrels, the relative improvement is very significant, while in the case of very good or poor quality barrels, the decrease is minimal. Notice too that the decay seems to be exponential.

Figure 30 also suggests that it is not be possible to decrease the probability of error to any arbitrarily small number by increasing the number of control bullets, since the plots in Figure 30 seem to have an asymptotic behavior. Furthermore, Figure 30 suggests that it should be possible to model the expected probability of error as a function of the number of control bullets for any barrel model by obtaining a relatively minimal amount of empirical data. Although a preliminary model of this behavior was explored, time limitations prevented us from completing such investigation.

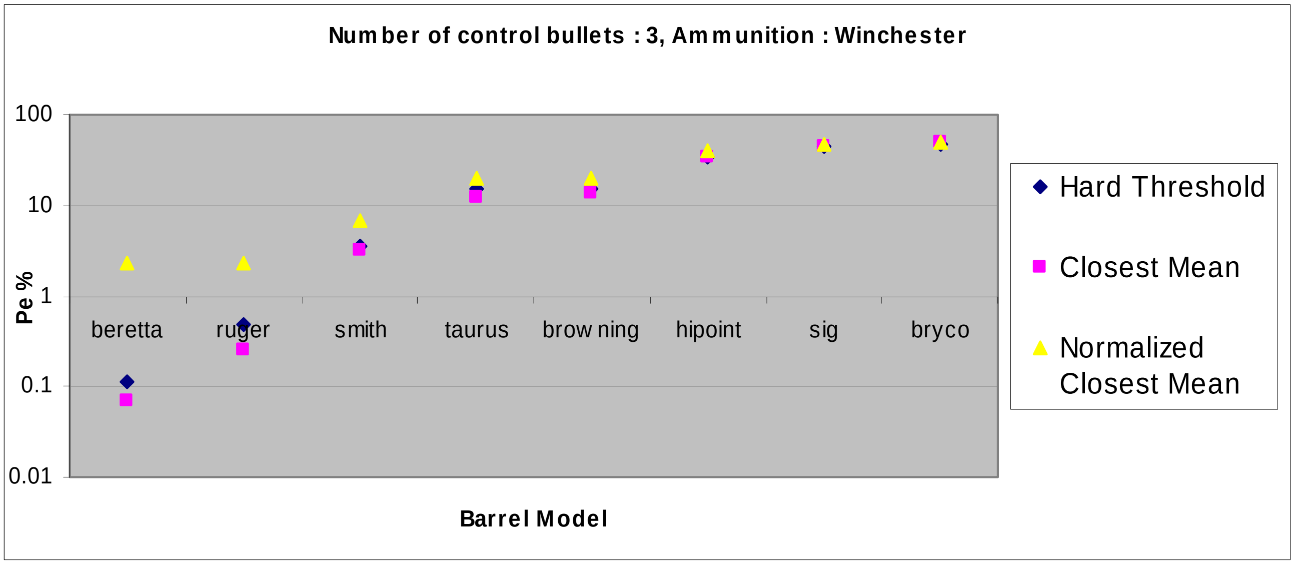


Figure 31: Effect of Classification Approach for 3 Control Bullets

4.6.3.4 Effects of Classification Approach

It is not surprising that the metric used to measure the degree of similarity between two distributions (and used in Step 4 of Table 2) plays a role in the bullet classification. Nevertheless, as seen in Figure 27 the effect of the classification approach (among the three approaches discussed in Section 4.2.3.2) is secondary when compared to that of the barrel model and ammunition manufacture.

Figure 31 shows the variation in the probability of error for all barrels under consideration, and for all three classification approaches when 3 bullets are used for control purposes and the ammunition under consideration is Winchester. As seen in Figure 31, the classification approach has a significant effect for barrel models of "good" quality (Beretta, Ruger, Smith & Wesson). The effect of the classification approach on the remaining barrels is almost negligible. However, the significance of this phenomenon is questionable, since the bullets fired by these barrels present such challenges to the acquisition hardware and processing algorithms, that no classification approach would have been successful with them. In particular, notice that the probability of error for HiPoint and SIG barrels is almost 50%, meaning that the classification outcome is almost random.

As the number of control bullets is increased, we have found that the different classification approaches tend to perform more or less equivalently. This phenomenon can be seen in Figure 32, where the same type of plot shown in Figure 31 has been repeated, but using in this case 5 control bullets. Notice how the difference in performance between the classification approaches is in this case minor even for those barrels considered as "good." The reason for this result can most likely be explained by the fact as the number of control bullets is increased, the best and second-best similarity measures obtained by comparing the control bullets better approximates the associated matching and non-matching distributions. Therefore, the classification results are more accurate, almost regardless of the classification approach used.

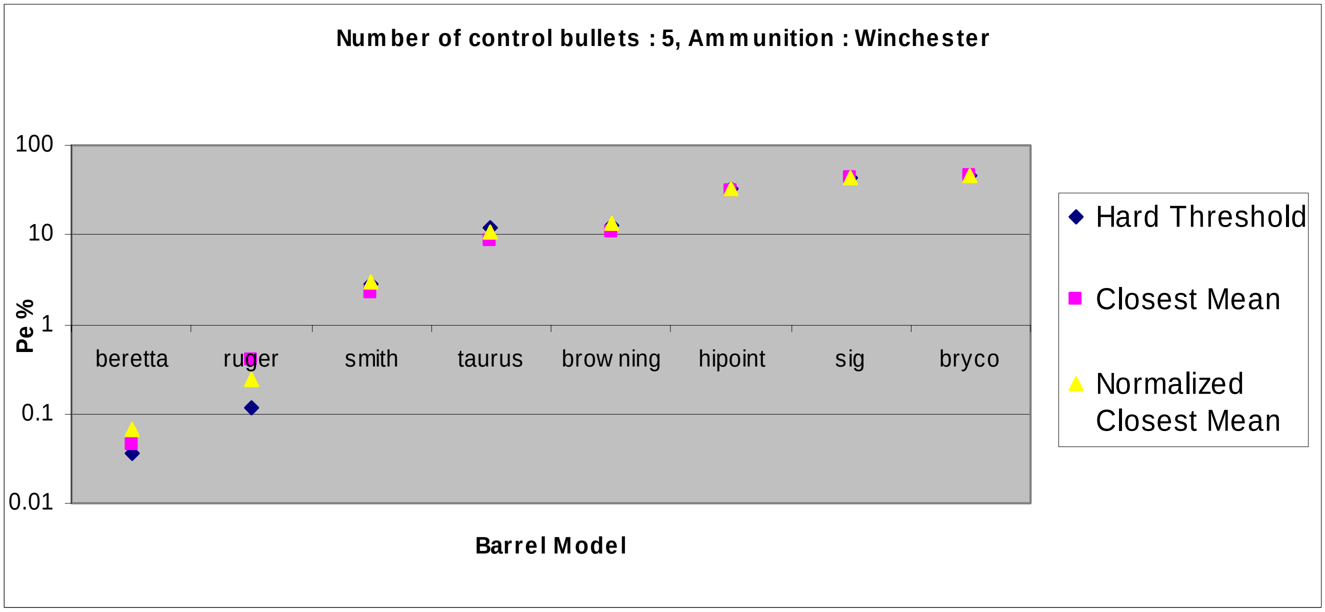


Figure 32: Effect of Classification Approach for 5 Control Bullets

4.6.3.5 Effects of Allowing "Inconclusive" Determinations

For most gun models under consideration, the probabilities obtained of identification errors shown in are less than satisfactory. As already discussed, part of the problem lies in the fact that the current 3D sensor and software are not capable to handle the very good and poorly manufactured barrels. However, there is another factor which contributes to the high number of errors. So far, the only two options available to the classifier have been to declare a "match" or a "non-match" between an evidence bullets and the set of control bullets. This does not reflect all the options available to firearms examiners, who can also declare an "inconclusive" identification. Inconclusive identifications may be the result of a variety of factors, such as the condition of the evidence bullet, the condition of the suspect gun, or more relevant to our case, the lack of barrel individuality. The question is then: how would the performance of the proposed classifier change if given the option to declare an "inconclusive" identification. The approach to be followed is to take advantage of the definition of barrel individuality to decide if the barrel under consideration is sufficiently individual for identification. If not, the classifier will simply declare an "inconclusive" identification. During the final periods of this project, we begun to explore this possibility and our results are preliminary.

Appendix C in Progress Report No. 12 includes the results of a large number of simulations performed to evaluate the effect of allowing for inconclusive classifications. In evaluating this option, a variety of approaches were tested ranging form a rank-sum based approach to a normalized closest mean approach. The normalized closest mean approach seemed to perform in the most consistent way, and it is defined as follows:

Normalized Closest Mean:

The normalized closest mean criterion measured the number of standard deviations between the mean of the matching and non-matching distributions. If the means of these distributions were not sufficiently "far away," the barrel under consideration was declared not sufficiently unique, and the comparison "inconclusive." More precisely, given a positive number γ, consider the following two conditions:

$$\left|\bar{r}-\bar{w}\right|/\sigma(r)>\gamma \tag{22}$$

$$\left|\bar{w}-\bar{r}\right|/\sigma(w)>\gamma \tag{23}$$

If either of these two conditions was not satisfied, the barrel was considered insufficiently unique, and the classification was inconclusive. Otherwise, the classification proceeded.

Table 16 shows the variation of the probability of error as a function of gamma for those barrel which have been found to be "good" (where the term "good" indicates quality of transfer of features between barrel and bullet). This table considers the case where a normalized closest mean classification approach is taken, 5 control bullets are used, and the ammunition is of Winchester manufacture. Overall, this is considered to be a relatively "benign" case, where good ammunition and good barrels are used. The rows labeled "Pe" correspond to the probability of error in the classification process, while the rows labeled "Pinc" correspond to the percentage of all classification attempts where the comparison was deemed "inconclusive." The probability of error data is also plotted in Figure 33. As can be seen in Figure 33, increasing the value of gamma as defined in Equation (22) and (23) does, in general, decrease the probability of error. However, this effect is not always monotonic, as in the case of Taurus, where the probability of error increases between Gamma = 2 and Gamma = 3.

Although we have invested a considerable amount of effort trying to integrate the possibility of an "inconclusive" classification, Table 16 shows the "inefficiency" of the approach so far undertaken. As expected, as Gamma increases, the percentage of comparison cases which is classified as inconclusive increases (since higher Gamma implies a stricter requirement for the best and second-best distributions of the control bullets). If we consider the case of Taurus as an example, in order to decrease the probability of error from 10.43% to 4.17% it was necessary to classify 98.23% of the comparisons as inconclusive. This percentage of inconclusive classifications seems excessive. One would expect that if the initial probability of error is 10.43% (with no inconclusive classifications) then something on the order of 10.43% inconclusive classifications would lead to approximately 0% probability of classification error. In other words, it would seem like the presented approach is too aggressive, and does not offer an optimal way to identify those classifications which should be labeled inconclusive.

The reason for the apparent deficiency of the presented approach may be the following: The proposed approach uses the best and second-best similarity measures resulting from the comparison of the control bullets as samples of the matching and non-matching similarity measures of the barrel model under consideration. It is possible that the number of samples provided by the control bullets (in the case of 5 control bullets, only 10 sample values are available) is too small to estimate the mean and standard deviation (see Equations (22) and (23)) of these distributions. This limitation

		Beretta	Ruger	Smith	Taurus
Gamma 0	Pe	0.068	0.555	2.909	10.430
	Pinc	0.000	0.000	0.000	0.000
Gamma 1	Pe	0.068	0.555	2.809	10.475
	Pinc	0.000	0.000	1.260	5.958
Gamma 2	Pe	0.068	0.422	1.813	6.830
	Pinc	1.240	1.351	15.174	59.625
Gamma 3	Pe	0.061	0.111	2.723	9.566
	Pinc	6.946	4.430	27.864	90.889
Gamma 4	Pe	0.000	0.113	1.197	4.167
	Pinc	27.603	6.612	47.831	98.236

normalized Closest Mean 5 control bullets

Table 16: Effect of Allowing for Inconclusive Classifications

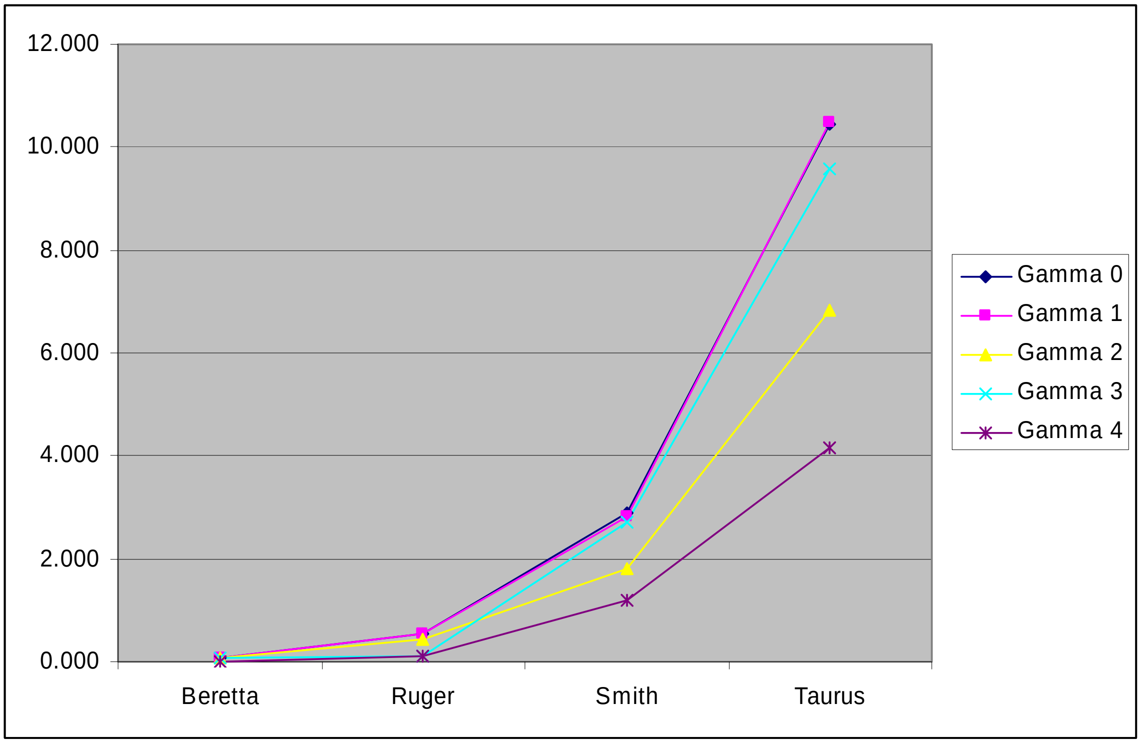


Figure 33: Effect of Allowing for "Inconclusive" Classifications

will become more evident as the matching and non-matching distributions are closer to each other, as is the case of Taurus. Alternative approaches to integrate an inconclusive classification have been considered as part of this project, but we have not had the chance to evaluate their effectiveness due to time limitations. As an example, one possible approach would be to modify the normalized closest mean criterion so that in order to classify a match between the evidence and the control bullets, the normalized closest mean requirement becomes $\rho|\bar{r}-\bar{e}|/\sigma(r) < |\bar{w}-\bar{e}|/\sigma(w)$ (for $\rho > 1$), while in order to classify a non-match, the normalized closest mean requirement becomes $|\bar{r}-\bar{e}|/\sigma(r) > \rho|\bar{w}-\bar{e}|/\sigma(w)$. If neither of these conditions is satisfied, then the comparison is classified as an "inconclusive." This is a generalization of the original approach, where $\rho = 1$ and no inconclusive classifications are allowed.

4.7 Damaged Bullet Study

The identification of even moderately damaged bullets presents a greater challenge than the identification of pristine bullets. While this is the case for both human firearms examiners and automated systems, it is particularly true in the case of an automated system since damaged bullets will not in general conform to a specific shape. Figure 34 shows an example of what a moderately damaged bullet may look like as compared to a pristine bullet. While a pristine bullet has a round cross-section which closely approximates a circle, a damaged bullet will in general have unpredictable cross sections which do not fit a simple mathematical description. This complicates the processing of such data, since no assumptions can be made regarding their general shape. Damaged bullets also pose challenges at the acquisition stage, again due to their unpredictable shape.

As part of this project, software and hardware to address the challenges associated with damaged bullets were developed. From the perspective of hardware, different bullet holders were manufactured for both pristine and damaged bullets (see Figure 35.) However, the main hardware problem presented by a damaged bullet is created by the fact that as it is rotated in range of the depth sensor (the laser depth sensor can be seen in Figure 35) it becomes increasingly difficult to maintain the bullet surface within the sensor range. From the software perspective, the fact that damaged bullets are not cylindrical (nor is their shape known a-priori) necessitated the development of manual LEA-by-LEA acquisition software and processing routines.

Figure 34: Comparison of Pristine and Damaged Bullet

Although challenging, the challenges mentioned above can be overcome. A more significantly difference between pristine and damaged bullets is that while pristine bullets will in most cases have all their LEAs intact and available for acquisition, damaged bullets will normally have only a reduced number of LEAs suitable for acquisition (since the striations on those LEAs which make impact with a target will be completely obliterated), and even those LEAs may be deformed. From a statistical perspective, this implies that the amount of data available for each comparison may vary significantly from comparison to comparison. Based on the bullets which were generated for this study, the number of LEAs which could normally be acquired was 3 (compared to 5 or 6 in most pristine bullets.) For this reason, in studying the statistical properties of the distribution of matching and non-matching pairs of bullets it makes sense to consider LEA-to-LEA comparisons as opposed to bullet-to-bullet comparisons.

4.7.1 Comparison of Relevant Distributions for Pristine and Damaged Bullets

Figure 36 shows the distributions of best, second-best and non-matching LEA-to-LEA similarity measure values for Ruger barrels both for Winchester (left) and Remington (right) ammunition for both pristine bullet comparisons (blue) and damaged bullet comparisons (red). It is important to note that while the pristine bullets distributions were obtained by comparing pristine bullets among themselves, the damaged bullet distributions are the result of comparing damaged bullets against pristine bullets. We consider the comparison of damaged bullets against pristine bullets (as opposed to damaged bullets against themselves) because these comparisons are more representative of those made by firearms examiners. In most

Figure 35: Acquisition of Pristine and Damaged Bullets

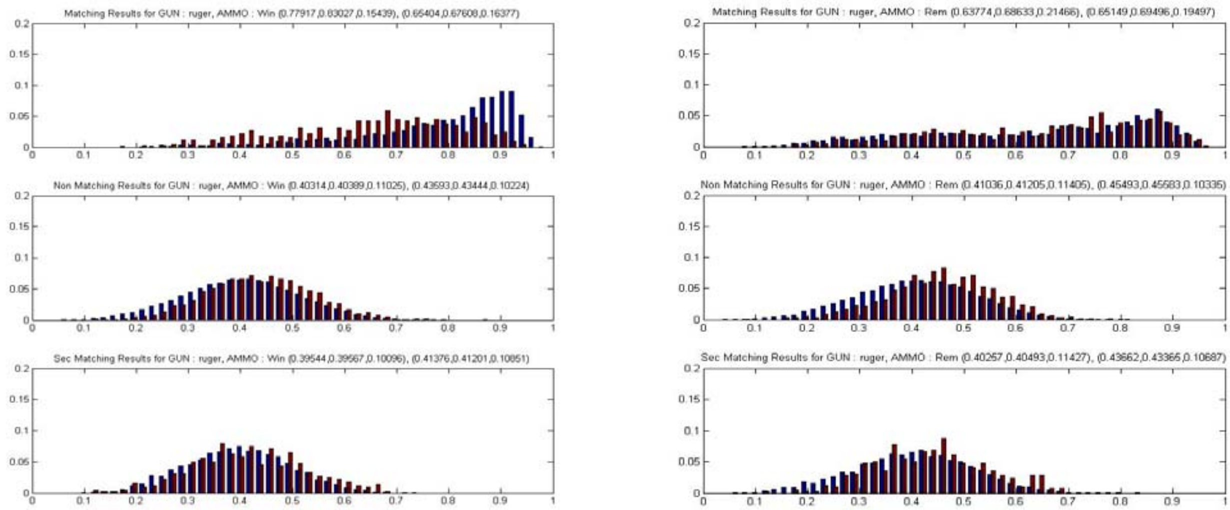


Figure 36: Distributions of best, second-best and non-matching LEA-to-LEA similarity measure values obtained for Ruger barrels for Winchester (left) and Remington (right) ammunition

situations, an evidence bullet found in a crime scene will show some degree of damage. However, the control bullets fired by the examiner for identification purposes will be retrieved in pristine condition. Therefore, at the time of performing a comparison the examiner will compare the damaged evidence bullet against the pristine control bullets.

The main statistical parameters (mean, median and standard deviation) associated with the distributions shown in Figure 36 are summarized in Table 17. As seen by these results (both tabulated and graphical), the quantitative nature of these distributions has not been significantly affected by the damage of the bullets. The most notable change takes place for the mean and median of the matching distribution for Winchester ammunition, where the pristine bullets achieve a median of .83, while the damaged bullets achieve a median of .68. In the case of the non-matching and second best LEA-to-LEA distributions minor shifts can also be noticed. However, these shifts do not seem to affect the satisfactory degree of individuality of these barrels. In the case of Remington ammunition, it is almost surprising to see that the matching distribution is almost the same for pristine and damaged bullets. The same type of analysis was performed with bullets fired by Beretta and Smith and Wesson (these are the guns which display the best degree of individuality.) The results in the case of Beretta barrels are shown in Figure 38, and for Smith and Wesson in Figure 37.

As seen in these distributions, most of the highest scores achieved using pristine bullets become much less likely to be achieved for damaged bullets (this is particularly noticeable in the case of Ruger barrels). As mentioned before, the challenge associated with the acquisition and analysis of damaged bullets is

			best			**non-matching**			**second-best**		
			mean	median	std	mean	median	std	mean	median	std
Ruger	Win	pristine	0.779	0.830	0.154	0.403	0.404	0.110	0.395	0.396	0.100
		damaged	0.654	0.676	0.164	0.436	0.434	0.102	0.414	0.412	0.110
	Rem	pristine	0.638	0.686	0.215	0.410	0.412	0.114	0.403	0.405	0.114
		damaged	0.655	0.695	0.195	0.455	0.456	0.103	0.437	0.434	0.107

Table 17: Main statistical parameters of similarity value distributions for pristine and damaged bullets fired by Ruger bullets

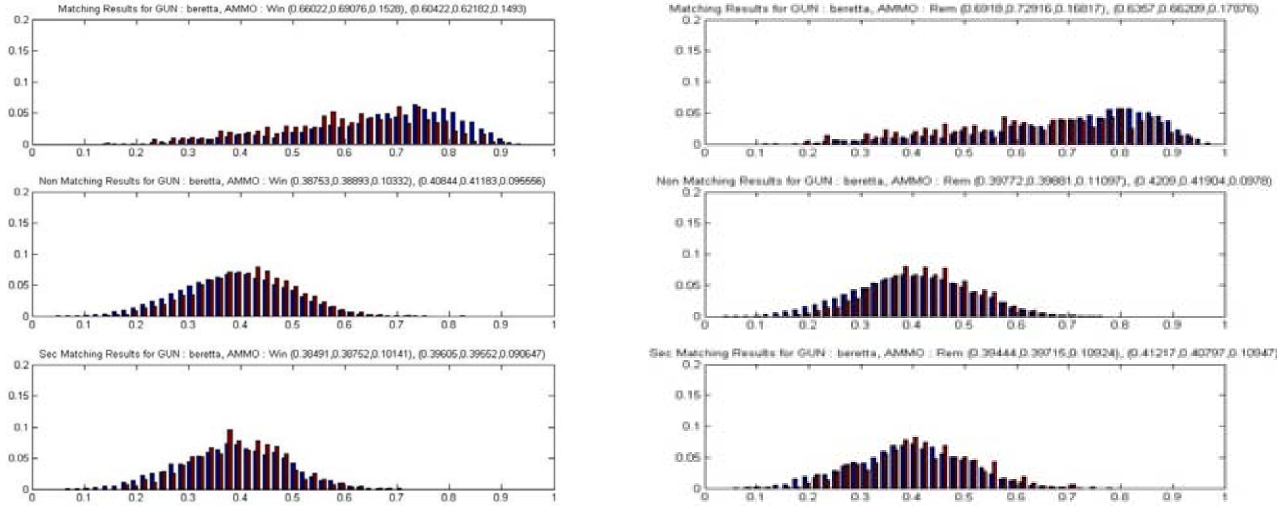


Figure 38: Distributions of best, second-best and non-matching LEA-to-LEA similarity measure values obtained for Beretta barrels for Winchester (left) and Remington (right) ammunition

significant. Therefore, this result is not surprising. Moreover, the highest similarity measure values achieved for pristine bullets occur for some extraordinarily repeatable LEAs. Since damaged bullets will not have all LEAs available for acquisition (in most cases, less than half are available), it becomes less likely for some of the highest LEA-to-LEA similarity values be achieved.

Overall, the distributions achieved with damaged bullets are less "attractive" than those achieved with pristine bullets (i.e. their properties towards accurate classification are less favorable). This is particularly true in the case of Smith and Wesson using Winchester ammunition. The matching distribution achieved by damaged Smith and Wesson bullets of Winchester manufacture is noticeable worse than that achieved with pristine bullets. However, this phenomenon did not repeat for Remington ammunition. At this point we have not been able to explain this difference.

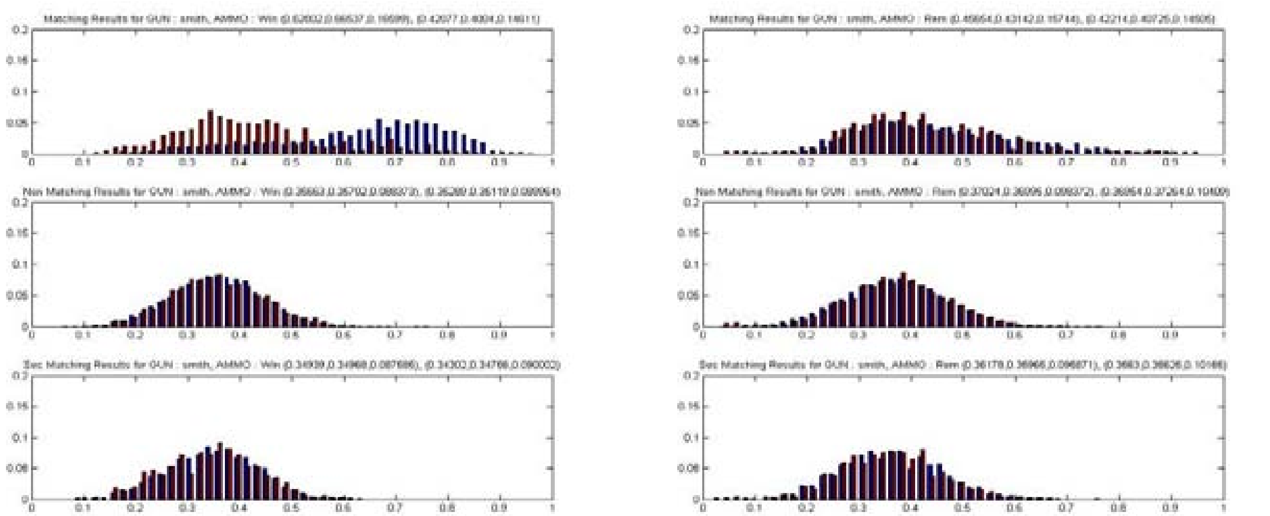


Figure 37: Distributions of best, second-best and non-matching LEA-to-LEA similarity measure values obtained for Smith barrels for Winchester (left) and Remington (right) ammunition

		Winchester				Remington			
		Orientation		LEA to LEA		Orientation		LEA to LEA	
		PFP	PFN	PFP	PFN	PFP	PFN	PFP	PFN
Ruger	pristine	0.38	0.25	0.00	0.20	7.12	2.40	0.59	4.55
	damaged	3.06	18.33	4.57	13.06	25.69	6.11	29.85	5.56
Beretta	Pristine	1.04	0.10	0.17	0.40	1.01	1.05	0.13	3.10
	Damaged	1.67	20.59	4.87	9.41	3.73	18.82	4.64	14.41
Smith	pristine	0.54	1.50	0.04	2.10	24.67	8.00	19.49	9.90
	damaged	16.67	70.00	2.31	63.61	26.53	42.75	39.64	34.00

Table 18: Effect of bullet damage on probability of classification error

4.7.2 Comparison of Classification Results for Pristine and Damaged Bullets

Having compared the distributions of LEA-to-LEA similarity measures obtained for pristine and damaged bullets, we turn to the question of greatest interest. To what extent does this difference affect the ability to perform an accurate match/non-match classification? In order to explore this issue, we performed a comparison of the probability of false positive and false negative identification obtained using pristine and damaged bullets. This comparison was performed in following two very similar approaches. The first approach was the same followed in Section 4.6. The second approach introduces a minor but significant variation of this approach. In Section 4.6 we made us of the best and second-best **orientation similarity measures** to perform a classification. In this section, due to the fact that damaged bullets often have a small number of LEAs suitable for comparison, we turn to the **best and second best LEA-to-LEA similarity measures**.

With each of these two versions of the procedure described in Table 2, we performed a bullet-to-gun classification experiment assuming four control bullets, one evidence bullet and using the normalized closest mean approach described in Section 4.2.3. Furthermore, in order to evaluate the possible effect of bullet damage, we repeated this experiment under two scenarios: one scenario where the evidence bullet was in pristine condition and a second scenario where the evidence bullet was damaged. The results of this experiment for Ruger, Beretta and Smith and Wesson are tabulated in Table 18 (where the numbers are in percentage of error) for both Winchester and Remington ammunition.

Table 18 provides a significant amount of information. Some of this information is the result of considering the probability of false positive and false negative classification errors separately. In the case of pristine bullets, Table 18 suggests that the use of LEA-to-LEA similarity measures (as opposed to orientation similarity measures) generally decreases the probability of false positive identification errors, while it increases the probability of false negative identifications errors. This observation indicates that it may well be possible to improve upon the results obtained in Section 4.6 for pristine bullets.

Of greater interest for this section is the fact that the presence of damage in the evidence bullet significantly increases the probability of both false positive identifications and false negative identifications with respect to the results obtained with pristine evidence bullets. The increase in the number of false negative identifications is not surprising, since it is to be expected that the features found on damaged bullets will not have the same quality as those found on pristine bullets, resulting in a decrease of similarity measure values, and therefore a false negative classification. The increase of false positive identifications, however, is surprising. This result seems to be due to the fact that the number of

LEAs available for comparison in the case of damaged bullets is lower than that available in the case of pristine bullets. Due to this fact, the standard deviation of the mean of LEA-to-LEA similarity measures obtained when comparing the evidence bullet against the control bullets is higher in the case of damaged bullets than in the case of pristine bullets. Therefore, the probability of obtaining a relatively high LEA-to-LEA similarity measure average is greater in the case of damaged evidence bullets than in the case of pristine evidence bullets, causing a larger number of false positive identifications. **This observation is important because it would appear that the problem lies in the classification approach, not on the bullet data**. More work remains to be done to improve the classification approach.

5. Conclusions

This study was structured into three main components: The first component dealt with the effect of barrel wear. As a result of this portion of the study, it was demonstrated that barrel wear does not pose a significant challenge to firearms identification. The second component of the study dealt with the development of methodologies to address two main issues: a) the evaluation of the degree of individuality of barrels by looking at the bullets fired through them, and b) the estimation of the probability of error in bullet-to-barrel classifications. Both these two components of the study were analyzed using bullets in pristine condition. As a result of this portion of the study it was demonstrated that it is feasible to apply a classification approach to identify bullets fired by the same barrel as opposed to bullets fired by different barrels of the same manufacture and make. The third component of the study focused on the degree to which the conclusions of the previous sections can be applied to damaged bullets. As a result of this portion of the study, it was demonstrated that it is possible to link a damaged bullet to a barrel with a high degree of certainty. However, the achievable overall classification accuracy is inferior to that attained with pristine bullets.

The key technical questions to be addressed by this study were the following:

a) Given 3D information from a bullet's surface, what quantitative criteria should be used to establish the individuality of a gun?
b) Given 3D information from a bullet's surface, what quantitative criteria should be used to establish that a suspect gun fired a given evidence bullet?
c) Once such criteria are developed, can the probability of a bullet/gun match being erroneous be estimated?

As part of this study, quantitative criteria to establish the individuality of a gun (question a), and to perform a classification (question b) were developed. These criteria were implemented and empirically tested with an unprecedented number of sample bullets. Estimates of the probability of error associated with bullet/gun classification (question c) were obtained, and the feasibility of identifying a bullet as having been fired by a given gun was validated based on well established statistical practices. A number of fundamental questions associated with bullet identification were considered and answered. Some of these questions were associated with the effect of a number of variables such as: the effect of barrel manufacture, condition and wear; the effect of bullet manufacture and condition (damaged bullets); the effect of the number of control bullets used in identification; the effect of various methods of classification, etc. Other questions were related to the soundness of common practices of firearms examiners. Some of these practices are related to the choice of ammunition used in an identification (to the extent possible, we have showed that it should be the same as that of the evidence), the test firing of a suspect barrel before and after it is cleaned, and other factors such as recommendations regarding the number of control bullets to use in an identification.

Overall, this study provides a solid validation of the foundations of ballistic identification. However, due to limitations of the current system, these conclusions could not be verified for all barrel models under consideration. The limitations of the current system can be identified as hardware and algorithmic limitations. In terms of hardware, it has become evident that the depth resolution of the system used for this project was not sufficient for the features found on the bullets fired by barrels for which the surface finish of the rifling is above average. The second significant limitation of the hardware used in this project was the amount of data which could be obtained in a reasonable acquisition time. Figure 39 demonstrates the significance of this limitation. The top plot of this figure shows a comparison of two

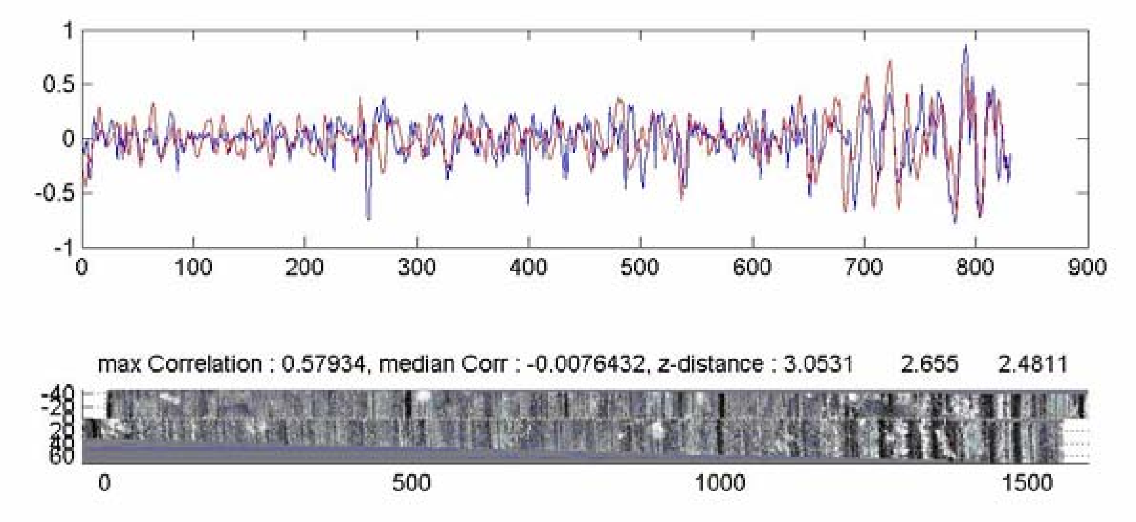


Figure 39: Preliminary Results Obtained with Confocal Microscope

cross sections of LEAs based on the data attainable with the current system. This data is composed of five cross sections of the LEAs under consideration. Notice that based on this data, it is impossible to judge which portions, if any, contain useful data. The bottom of Figure 39 shows data of the same LEAs imaged on the top, but obtained with a confocal microscope. This data amounts to a sequence of patches providing a total of 512 cross sections of the LEAs of interest (the image shown is a pseudo-realistic rendering of the LEA surface data as seen from above the LEA). Moreover, the depth resolution of this instrument is an order of magnitude better than that of the current system. Notice that the data obtained with the confocal microscope is sufficiently dense to allow for an identification of useful portions of data (those portions which show clear striations). The availability of data of this quality and quantity would allow not only to detect smaller features than those currently detected, but it would also allow the development of significantly more sophisticated algorithms for ballistic identification, addressing the limitations of the current algorithms. Therefore, it should be clear that there is significant room for improvement terms of the achievable classification performance.

The limitations of the current system to classify bullets fired by certain brands of barrels should not be extrapolated to suggest that firearms examiners (or an improved system) would not be able to attain lower probabilities of error. In fact, even with the data available, there is ample evidence that the results presented in this report can be improved upon. This issue may be the subject of future research.

Another important consequence of this project was to bring topographical methods of bullet identification to the law enforcement community. In May 2000 IAI and FTI reached an agreement to develop a commercial system which would integrate both 2D and 3D surface data. A prototype of such system was unveiled to the firearms examiner's community at the main exhibit floor of AFTE 2003 as part of the FTI booth. By the end of 2004, FTI began the commercialization of BulletTRAX^{TM}-3D, a 3D based ballistic analysis system developed as a result of the collaboration between IAI and FTI. BulletTRAX^{TM}-3D has received high praise from the firearms examiner's community, and brings the benefits in performance of topographical analysis of firearms evidence to the law enforcement community.