

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



**Grado en Ingeniería de Tecnologías y Servicios de
Telecomunicación**

TRABAJO FIN DE GRADO

**Investigación reproducible: uso de la plataforma BEAT para la
evaluación tecnológica de algoritmos de reconocimiento
biométrico**

**Roberto Daza García
Tutor: Aythami Morales Moreno
Ponente: Julián Fierrez Aguilar**

Mayo 2016

**Investigación reproducible: uso de la plataforma BEAT para la
evaluación tecnológica de algoritmos de reconocimiento
biométrico**

**AUTOR: Roberto Daza García
TUTOR: Aythami Morales Moreno**

**Biometric Recognition Group - ATVS
Dpto. de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Mayo 2016**

Resumen

La reproducibilidad es un tema de gran preocupación dentro de la comunidad científica. Gran cantidad de los artículos científicos que se publican en la actualidad carecen del suficiente detalle, código o datos para garantizar la reproducibilidad de los mismos por otros investigadores. En este trabajo se presenta una herramienta de evaluación tecnológica desarrollada a partir de la plataforma *Biometric Evaluation and Testing* (BEAT) orientada a fomentar la experimentación reproducible dentro del área del reconocimiento biométrico. Para ello, se ha propuesto una competición internacional (*Keystroke Biometric Ongoing Competition - KBOC*) centrada en la evaluación de sistemas de reconocimiento de usuarios a través de dinámica de tecleo.

Los principales objetivos de esta competición son impulsar la investigación en dinámica de tecleo (atraer a nuevos investigadores), impulsar la plataforma BEAT y ser una base común donde comparar los sistemas biométricos.

La competición incluye una de la mayores base de datos en dinámica de tecleo con más de 300 usuarios y 4 sesiones diferentes (escriben nombre y apellido).

Hay dos maneras de participar en KBOC (misma base de datos para las dos): Competición *Ongoing* (permanecerá activa por tiempo indefinido), desarrollada en la plataforma BEAT y competición *Offline* que sirve de referencia para la competición *Ongoing*.

En este TFG se introduce la plataforma (los recursos creados para la competición, etc.), se detallan las características de la competición y se presentan los resultados *ongoing* (experimentos facilitados por la competición) y los resultados y sistemas de los participantes en la competición *offline*.

Palabras clave

Experimentación Reproducible, Reconocimiento Biométrico, Biometría, Dinámica de Tecleo, Plataforma BEAT, Competición, KBOC16, Competición Ongoing, Competición Offline, Base de Datos.

Abstract

The reproducibility of the research is a worrisome topic among the scientific community.

In this work, a tool developed under BEAT platform for technological evaluation is presented, which fosters the reproducible research inside the biometric field. In order to do that, an international competition (Keystroke Biometric Ongoing Competition) focused on the assessment of the recognition systems has been proposed by means of keystrokes dynamic.

The ongoing evaluation tool has been developed using the BEAT platform and includes keystroke sequences (fixed-text) from 300 users acquired in 4 different sessions under realistic conditions.

In this work, we introduce the platform, we detail the features of the competition and we present the results. In addition, the results of a parallel offline competition based on the same data and evaluation protocol are presented. The results reported have achieved EERs as low as 5.32%, which represent a challenging baseline for keystroke recognition technologies to be evaluated on the new publicly available KBOC platform.

Keywords

Reproducible Research, Biometric Recognition, Keystroke, BEAT Platform, Competition, KBOC16, Competition Ongoing, Competition Offline, Data Base.

Agradecimientos

En primer lugar, darle las gracias a mi tutor Aythami por introducirme y enseñarme el apasionante mundo de la biometría. Como dijo Einstein "El verdadero arte de un profesor, es despertar la alegría por el trabajo y el conocimiento", sin duda lo has conseguido. Gracias por toda la ayuda prestada y por el buen equipo de trabajo que hemos formado.

Cuando no tienes nada, nada tienes que perder. Yo tengo mucho que perder con una familia tan estupenda. Gracias a mi padre por decirme cuando me equivoco y ayudarme a ser mejor persona. A mi madre, porque sin ti no hubiera sido posible, desde pequeño me has ayudado para llegar hasta aquí. A mi hermano, por ser el mejor amigo que se puede tener y el compañero que nunca te decepciona. Y a mi hermana, que siempre creyó en mí, incluso cuando yo mismo dudaba.

Gracias a ti Carmen, por hacerme un poco mejor cada día. Eres una parte fundamental en este proyecto, porque tú eres capaz de transformar el carbón en diamante.

A mi amigo Paco, por las muchas horas de trabajo que hemos pasado juntos, por los muchos Red Bull bebidos y por supuesto, por las risas y fiestas que hemos compartido.

A mi amigo Ayman, porque la distancia que nos separa no podrá hacer olvidar tantos recuerdos, risas, y como no, el famoso teorema de Sergey.

INDICE DE CONTENIDOS

1 Introducción	7
1.1 Importancia de la Experimentación Reproducible	7
1.2 Reconocimiento Biométrico	12
1.3 Dinámica de tecleo.....	15
1.4 Dinámica de Tecleo: Evaluaciones tecnológicas.....	18
1.5 Motivación y Objetivos	20
1.6 Contribuciones del TFG	22
1.7 Organización de la memoria	22
2 KBOC. Competición en dinámica de tecleo.....	23
2.1 Introducción	23
2.2 Conjunto de datos de referencia.....	24
2.3 Competición ongoing y competición offline	25
2.4 Introducción a la plataforma BEAT.....	26
2.4.1 Características de la plataforma BEAT	27
2.5 Estructura de la competición ongoing en la plataforma BEAT.	29
2.5.1 Algoritmos de clasificación (<i>baseline</i>).....	36
2.6 Promover la participación	37
2.7 Ayuda a la mejora de la plataforma BEAT	39
3 Resultados.....	41
3.1 Resultados obtenidos con los sistemas baseline.....	41
3.2 Descripción de los sistemas presentados por los participantes	41
3.2.1 Resultados de la competición <i>offline</i>	43
4 Conclusión	45
Referencias	47
Glosario	49
Anexos	- 2 -

INDICE DE FIGURAS

FIG. 1-1. NÚMERO DE ARTÍCULOS RETIRADOS POR CAUSAS ESPECIFICADAS POR AÑO DE RETIRADA. FIGURA MODIFICADA DE [5].	9
FIG. 1-2. DISTRIBUCIÓN DE CITAS (2004-2005) UTILIZANDO DIAGRAMA DE CAJA. FIGURA MODIFICADA DE [11].	10
FIG. 1-3. DISPONIBILIDAD EN LÍNEA DE CÓDIGO (A) Y DATOS (B) EN FUNCIÓN DEL NÚMERO DE CITAS. EL TAMAÑO DE LOS PUNTOS INDICA LA MULTIPLICIDAD. FIGURA MODIFICADA DE [2].	11
FIG. 1-4. DISTINTAS MEDIDAS EXPUESTAS EN EL LIBRO <i>IDENTIFICATION ANTHROMÉTRIQUE</i> [13].	12
FIG. 1-5. SISTEMAS SEGÚN SU FISIOLÓGÍA (A) Y SEGÚN SU COMPORTAMIENTO (B)	13
FIG. 1-6. RASGOS BIOMÉTRICOS FRENTE A MEDIDAS DE CALIDAD. FIGURA MODIFICADA DE IBG.	14
FIG. 1-7. MERCADO BIOMÉTRICO (2007-2015) E INGRESOS DE LA INDUSTRIA BIOMÉTRICA EN ESTADOS UNIDOS	15
FIG. 1-8. PARÁMETROS MÁS UTILIZADOS EN LA DINÁMICA DE TECLEO. FIGURA MODIFICADA DE [22]	16
FIG. 1-9. ESQUEMA PARA OBTENER EL RENDIMIENTO DE UN SISTEMA BIOMÉTRICO	17
FIG. 2-1. ESQUEMA DE LA COMPETICIÓN USADO PARA INCENTIVAR LA MOTIVACIÓN. LA FIGURA ORIGINAL ESTÁ ESCRITA EN INGLÉS.	23
FIG. 2-2. SE MUESTRAN LOS CONCEPTOS GENERALES QUE DIERON LUGAR A LA PLATAFORMA Y LOS OBJETIVOS DE BEAT.	26
FIG. 2-3. MUESTRA QUE BEAT UTILIZA PYTHON (CÓDIGO LIBRE) PARA HACER MÁS ACCESIBLE LOS ALGORITMOS.	27
FIG. 2-4. ESQUEMA DE UN DIAGRAMA DE BLOQUES UTILIZADO EN LA PLATAFORMA BEAT	28
FIG. 2-5. INTERFAZ DE LA PLATAFORMA BEAT	29
FIG. 2-6. <i>TOOLCHAIN</i> CREADO PARA LA COMPETICIÓN KBOC EN BEAT	30
FIG. 2-7. <i>TOOLCHAIN</i> DE LA COMPETICIÓN INDICÁNDOSE EL ORDEN DE EJECUCIÓN DE CADA BLOQUE A TRAVÉS DE LOS NÚMEROS Y LOS COLORES QUE INDICAN LA SINCRONIZACIÓN DE LOS BLOQUES.	31
FIG. 2-8. EJEMPLO DE FORMATO DE DATO CREADO PARA LA COMPETICIÓN KBOC	32
FIG. 2-9. IMAGEN EXPLICATIVA DE LA SINCRONIZACIÓN Y ENTREGA DE DATOS EN LOS BLOQUES DE LA COMPETICIÓN	33
FIG. 2-10. EXPERIMENTO FACILITADO POR LA COMPETICIÓN, DONDE SE MUESTRAN LOS RESULTADOS PARA LA DISTANCIA MANHATTAN ESCALADA MODIFICADA	34

FIG. 2-11. TABLA DE CLASIFICACIÓN A 26 DE MAYO DEL 2016	35
FIG. 2-12. ESQUEMA DE LA COMPETICIÓN EN LA PLATAFORMA BEAT	35
FIG. 2-13. IMÁGENES DEL TUTORIAL AVANZADO INTERACTIVO CREADO PARA LA PLATAFORMA BEAT.....	37
FIG. 2-14. IMAGEN DEL VIDEO TUTORIAL CREADO PARA LA COMPETICIÓN.....	38
FIG. 2-15. INTERFAZ ANTIGUA DE LA PLATAFORMA BEAT	39
FIG. 3-1. IZQUIERDA: CURVA DET CON TODAS LAS SESIONES Y SISTEMAS PRESENTADOS; DERECHA: MEJOR SISTEMA ENVIADO POR CADA ENTIDAD, DIBUJADO PARA LA SESIÓN 2 (LÍNEA DISCONTINUA) Y PARA LA SESIÓN 4 (LÍNEA CONTINUA)	44

INDICE DE TABLAS

TABLA 1-1: ESTUDIO DE LAS BASES DE DATOS DISPONIBLES PÚBLICAMENTE EN DINÁMICA DE TECLEO.....	19
TABLA 2-1: RESUMEN DE LAS PRINCIPALES CARACTERÍSTICAS DE LA BASE DE DATOS UTILIZADA EN LA COMPETICIÓN	24
TABLA 2-2: ENTIDADES COLABORADORAS CON LA PLATAFORMA BEAT.....	27
TABLA 3-1: RESULTADOS OBTENIDOS PARA LAS TRES DISTANCIAS EN BEAT.	41
TABLA 3-2: RESUMEN DE LAS CARACTERÍSTICAS DE LOS MEJORES SISTEMAS PRESENTADOS EN LA COMPETICIÓN.....	44
TABLA 3-3: RESULTADOS FINALES PARA LA COMPETICIÓN. OBTENIDOS CON LAS MUESTRAS DE ENTRENAMIENTO (SESIÓN 1) Y LAS TRES SESIONES (TODAS LAS MUESTRAS TIPO TEST).	44
TABLA 3-4: RESULTADOS OBTENIDOS CON ERR_G . USANDO LAS MUESTRAS DE ENTRENAMIENTO (SESIÓN 1) Y LAS MUESTRAS TEST DE LA SESIÓN UNO O SESIÓN 4.	44
TABLA 3-5: EER_G OBTENIDO PARA TODOS LOS SISTEMAS PRESENTADOS. OBTENIDO CON LAS MUESTRAS DE ENTRENAMIENTO (SESIÓN 1) Y LAS TRES SESIONES (TODAS LAS MUESTRAS TIPO TEST).....	44

1 Introducción

En el presente proyecto Fin de Grado se ha desarrollado una herramienta de evaluación de tecnologías de reconocimiento biométrico basadas en la dinámica de ratón. La herramienta será desarrollada a partir de la plataforma BEAT, la cual garantiza la completa reproducibilidad de todos los experimentos evaluados con la herramienta.

1.1 Importancia de la Experimentación Reproducible

Para darnos cuenta del alcance que supondría facilitar un método sencillo donde se pudieran reproducir los experimentos publicados en artículos científicos, tenemos que mencionar el famoso último teorema de Fermat [1] que marcó un hito en la historia de las matemáticas. Sería Pierre de Fermat en 1637 quien enunciara dicho teorema:

$$x^n + y^n = z^n \quad (1)$$

donde si n es un número entero mayor que 2, entonces no existe números enteros positivos (x,y,z) , tales que cumpla la formula (1).

Pero lo que realmente subyace de este teorema es que nos encontramos con un caso de experimento no reproducible, ya que Fermat escribió en el margen de un ejemplar de la Arithmetica de Diofanto:

"Cubum autem in duos cubos, aut quadrato quadratum in duos quadrato quadratos, et generaliter nullam in in²nitum ultra quadratum potestatem in duos ejusdem nominis fas est dividere; cujus rei demonstrationem mirabile sane detexi. Hanc marginis exiguitas non caperet."-Pierre Fermat [1]

Donde hace mención que había conseguido una demostración realmente admirable a su teorema, pero el margen del libro era demasiado pequeño para ponerla. Esto supuso que hasta 1995 nadie pudiera demostrarlo. Sería Andrew Wiles quien 358 años después lo demostraría.

Hicieron falta 358 años para demostrar un teorema que supuestamente ya había sido demostrado por Fermat. Como este ejemplo hay muchos más, algunos muy conocidos y otros no tanto. Todo esto supone un tremendo atraso para la comunidad científica, que en el caso del último teorema de Fermat supuso 358 años de retraso. ¿Acaso esa falta de espacio en el libro no es semejante a la que sufren los investigadores en la actualidad para exponer la manera de reproducir sus experimentos en sus publicaciones científicas?, ¿No es cierto que si Fermat hubiera tenido la tecnología que existe actualmente o un sistema sencillo para exponer la manera de reproducir su experimento, la comunidad matemática se hubiera horado 358 años en demostrar lo que era obvio para su creador?

La realidad es que en algunas áreas son pocas las publicaciones que facilitan la reproducibilidad de los resultados presentados [2]. Por ejemplo, en el ámbito de la Biometría es habitual en las publicaciones científicas la ausencia del algoritmo que genera dichos resultados. Pero no sólo es este el problema principal, sino que también se suelen omitir las bases de datos sin las cuales es imposible reproducir el trabajo de investigación. Hay diversos motivos por los cuales se prescinde de su inclusión, entre los cuales podemos destacar la falta de espacio a la que se está condicionado por las publicaciones y la falta de

unanimidad a la hora de programar, siendo una limitación al desconocer este tipo de lenguaje, añadiendo que puede que sean programas de pago (Matlab, etc.). También puede deberse a que el autor no esté interesado en facilitar su código por diversas razones, como la privacidad de sus códigos o la falta de auto-disciplina. En general, se podría decir que hay pocos incentivos que empujen a los investigadores a presentar trabajos que faciliten la reproducibilidad a otros investigadores.

Sin duda es importante la reproducibilidad de un trabajo de investigación, pero **¿cuál es exactamente el significado?** Una definición formal podría ser: "capacidad de un experimento para ser reproducido con exactitud por otros investigadores ajenos a éste". Consecuentemente, para conseguir una gran reproducibilidad (objetivo primordial en el método científico) es esencial que el científico publique sus descubrimientos y los describa con precisión, para que otro investigador los pueda reproducir con los mismos resultados.

Ya desde la antigüedad era conocido el gran valor de la reproducibilidad en diferentes ámbitos (matemáticas, física, etc.), cada uno con sus peculiaridades. Por ejemplo, en el ámbito de las matemáticas hay constancia que alrededor de 500 años antes de la actualidad, investigadores mandaban cartas a compañeros en los cuales contenían retos para que reprodujeran sus resultados, muchos con la intención de comprobar si realmente su trabajo con los datos proporcionados era sencillo de reproducir, para posteriormente realizar una publicación de calidad, donde su experimento fuera fácilmente reproducible.

Sin embargo, en ciencias como física, química o biología, existe una tradición de dar descripciones detalladas del montaje experimental y como reproducirlos.

Ya en 1637 Descartes, a través de su obra "Discurso del método" [3] en la sexta parte hacía alusión al valor de la ciencia como una obra colectiva, ya que el progreso científico necesita la comunicación de las experiencias de otras personas. Es más, establece el marco para el método científico donde la reproducibilidad es un componente fundamental (solo se ciñe a los hechos).

Si nos centramos en la actualidad hay otro gran problema que afecta a la comunidad científica: la falsificación de resultados o simplemente errores cometidos en su proceso que se dan por correctos, los cuales son publicados en revistas científicas. ¿Cómo se pueden publicar estos resultados? La realidad es que muy pocas revistas dan la posibilidad de poder reproducir los experimentos expuestos y en muchos casos tampoco exigen las bases de datos con las cuales han sido realizados. Esto conlleva tener que creer ciegamente en los resultados expuestos en el artículo sin poder verificarlos, pudiendo estos contener errores o ser manipulados. En los últimos años, los avances informáticos también han afectado a este proceso, ya que programas de edición gráfica como Adobe Photoshop han hecho que sea más sencillo manipular resultados científicos en los artículos, sumado al afán de incrementar el número de publicaciones científicas por el enorme reconocimiento que supone en la actualidad. Como ya dijo Phil Clapham "Publicar o morir" [4] resume en una breve frase lo antes mencionado, llegando a ser tentador para los investigadores publicar artículos manipulados.

Es por este motivo que cada vez más estudios corroboran un mayor número de artículos fraudulentos que se asocian a errores cometidos. Es más, muchos de estos artículos no son identificados por las revistas ni por compañeros investigadores por la ausencia de experimentos reproducibles y los datos con los que se han generado.

Sirva de ejemplo dentro del campo de la biomedicina, un estudio publicado en la revista *Proceedings of the National Academy of Sciences* [5], donde se reafirma esta hipótesis de un incremento del número de artículos fraudulentos publicados. En este estudio se dan los resultados del análisis de 2047 artículos que fueron retirados de las revistas hasta Mayo de 2012, siendo el primer artículo publicado en 1973 y retirado en 1977. Los datos de esta investigación son alarmantes. Solo el 21,3% fueron atribuidos a

error, mientras que el 67,4% se asociaba a fraude. Es más, según se desprende del referido estudio, el aumento en los últimos años de artículos fraudulentos es alarmante.

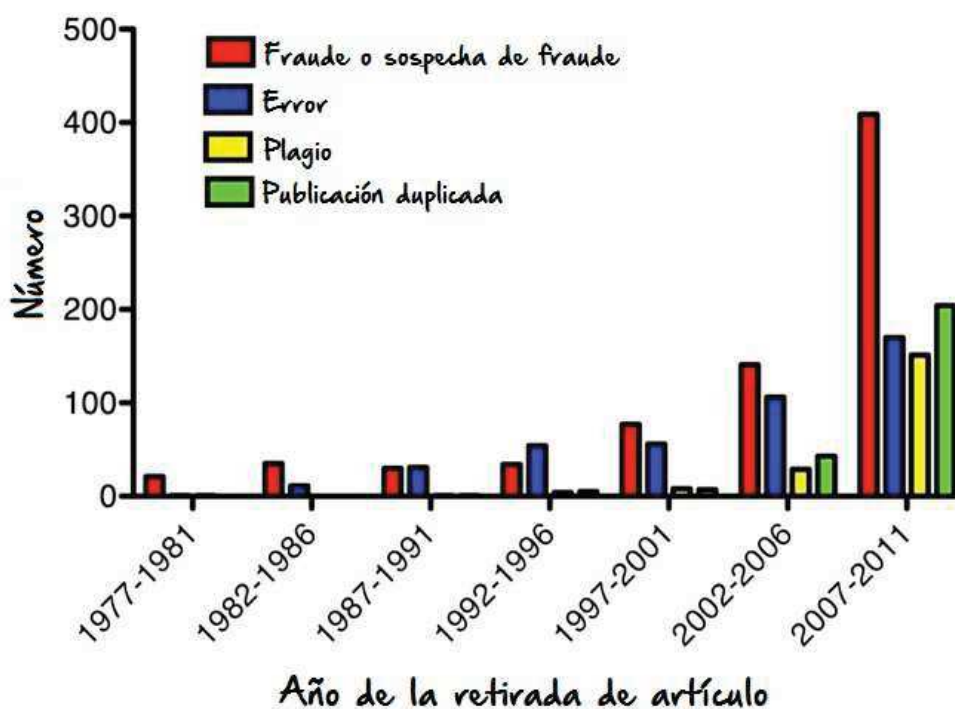


Fig. 1-1. Número de artículos retirados por causas especificadas por año de retirada. Figura modificada de [5].

La Fig. 1-1, aunque pertenece a un estudio en el ámbito de la biomedicina, se puede aplicar a otras muchas materias científicas puesto que diversos estudios en otras materias revelan resultados semejantes. Como podemos ver en la Fig. 1-1, el fraude desde 1977 se ha multiplicado aproximadamente por 10. Son datos realmente alarmantes y que no pueden justificarse en base a la mayor cantidad de artículos publicados, dado que el aumento de estos no es proporcional al número de fraudes.

Ante esta situación, cabe destacar la alta responsabilidad de los que autorizan la publicación en las revistas científicas de artículos sin constatar previamente la veracidad de lo publicado. Lo cierto es que la mayoría permiten que ni siquiera se expongan las bases de datos utilizadas, aunque revistas de gran importancia como *Nature* normalmente solicitan realizar los experimentos con bases de datos reconocidas. Además, si los experimentos presentan dudas razonables, son previamente reproducidos por otros científicos antes de su publicación. También se someten a otras fuertes cribas para ser publicados, pero a pesar de todas estas medidas ha habido grandes escándalos por publicar resultados manipulados, como el caso de la memoria del agua o la fusión fría.

Incluso premios Nobel han incurrido en falsificación de resultados, como Linda Buck que publicó un artículo en la revista *Nature* que no se podía reproducir y fue retirado [6]. Pero no hay que irse tan lejos, en España un artículo publicado en *Science* por un grupo de investigación del CSIC de Granada fue retirado por contener datos manipulados [4], [7].

Lo que parece obvio es que la comprobación de resultados por investigadores independientes ayuda a detectar el fraude. Un ejemplo práctico es el de Hwang Woo-Suk. [8].

Por este motivo estamos convencidos que la experimentación reproducible es una solución al fraude o muy útil a la hora de simplemente detectar errores de otros

investigadores. Es más, si nos centramos en el ámbito de las telecomunicaciones y más concretamente en el de la Biometría, ramas muy modernas donde normalmente los experimentos suelen estar representados por códigos que en un simple clic pueden reproducirse. Por tanto, en estos ámbitos donde encontramos tantas facilidades, es esencial impulsar experimentación reproducible.

Está claro que la comunidad científica solo tiene que ganar con la experimentación reproducible, **¿pero tiene algún beneficio para el autor compartir sus datos y hacer reproducible su experimento?**

La realidad es que suministrar los datos usados y hacer experimentación reproducible reduce la barrera de entrada en las publicaciones, dado que los lectores podrán obtener los resultados mucho más fácil si el código y los datos han sido suministrados. De este modo la publicación tendrá un mayor impacto social. Esto también es aplicable para el propio autor, puesto que un trabajo reproducible conlleva tener todo muy organizado y documentado, facilitando que en un futuro se retome el trabajo de una manera práctica, sencilla y fiable.

¿Existen estudios que demuestren lo manifestado anteriormente? Para empezar, diversas publicaciones constatan que los artículos que están libremente disponibles online son mucho más citados, lo que verifica que dar facilidades al lector se ve siempre recompensado [9], [10].

Parece obvio que si damos facilidades al suministrar los datos y experimentos reproducibles, obtendremos una ganancia en el número de citaciones. La realidad así lo corrobora, pues diversos estudios lo demuestran como "*Sharing Detailed Research Data Is Associated with Increased Citation Rate*" [11], que examina 85 publicaciones de ensayos de microarrays de cáncer, a los efectos de poder constatar la relación existente entre las veces en que han sido citados y el hecho de estar a disposición del lector en internet los datos primarios detallados (bases de datos). Los resultados de esta publicación son realmente alentadores para la investigación reproducible.

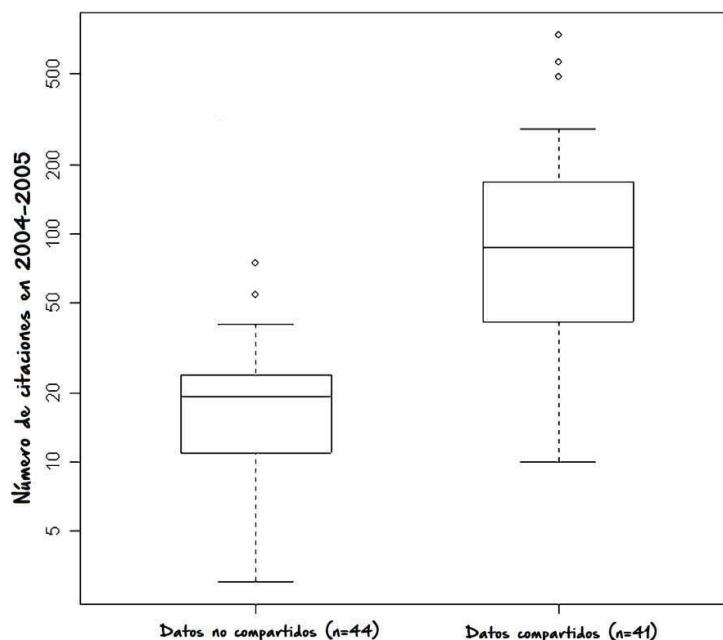


Fig. 1-2. Distribución de citas (2004-2005) utilizando diagrama de caja. Figura modificada de [11].

Los datos muestran que las publicaciones que comparten las bases de datos son alrededor de un 70% más citadas que las publicaciones que no lo hacen. Esto se mantenía también para publicaciones con perfil bajo. Todas las publicaciones de gran impacto

compartían los datos en el estudio. Si nos fijamos en la Fig. 1-2 podemos apreciar como son mucho más citadas las que comparten datos. Además, apreciamos como en las publicaciones que no comparten datos su máximo en citaciones (sin contar los valores atípicos), se encuentra en el 25% de citaciones en relación con las publicaciones que si comparten sus datos. Un dato revelador de la importancia de mostrar las bases de datos, es que el 85% de las citaciones se encontraban en los artículos que las compartían.

Un estudio más cercano a las telecomunicaciones es el realizado en la *Ecole Polytechnique Fédérale de Lausanne* [2]. En este primer estudio se vio tanto las descargas que se realizaban de los artículos con experimentos reproducibles, como de los códigos suministrados para reproducirse. Además, comprobaron el número de citaciones que recibían los artículos teniendo presente si publicaban código, datos, ambas cosas o ninguna de ellas.

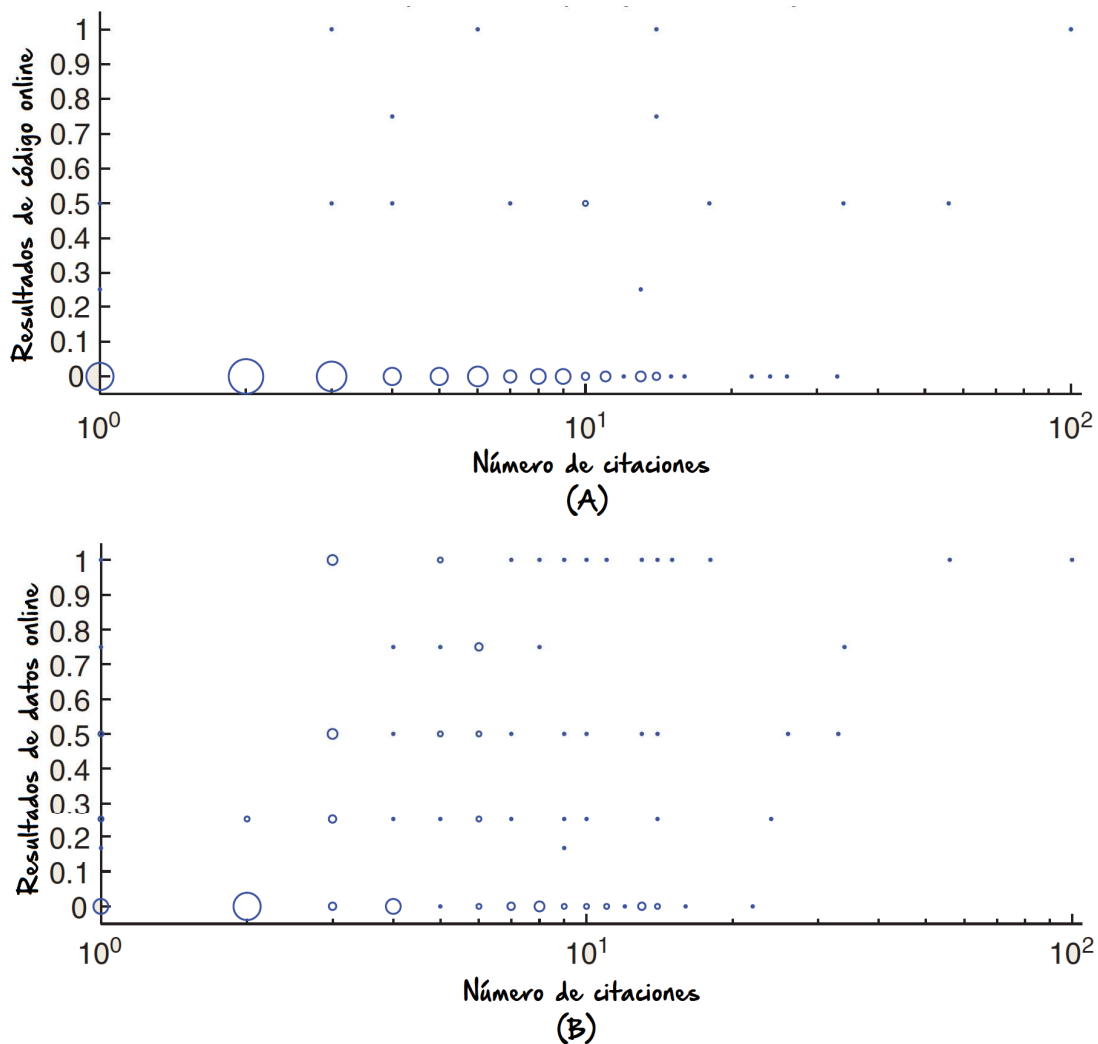


Fig. 1-3. Disponibilidad en línea de código (a) y datos (b) en función del número de citas. El tamaño de los puntos indica la multiplicidad. Figura modificada de [2].

Los resultados expuestos en la Fig. 1-3 son contundentes, cuanto más citados son los artículos más se encuentran disponibles online el código y los datos. En cambio, cuantos menos datos se encuentren disponibles (la mayoría de los casos), menos citados son estos

artículos. Existen algunos caso que contradice lo antes expuesto, pero esto es normal si tenemos presente que también depende de la calidad de la publicación.

Todos los datos anteriormente mencionados nos muestran la gran importancia de la experimentación reproducible, por ello es fundamental para incentivar el uso de protocolos reproducibles facilitar a los autores mecanismos que solucionen los problemas antes mencionados. Y es éste precisamente el punto de partida para nuestro trabajo. En este caso nos centraremos en un sub-área muy concreta de la investigación de sistemas de reconocimiento biométrico: el reconocimiento a partir de dinámica de tecleo.

1.2 Reconocimiento Biométrico

El término biometría procede del griego "bios" vida y "metron" medida. Estas palabras en si definen la función de la biometría, que como su nombre indica es el reconocimiento único de individuos a través de un conjunto de medidas.

Existen evidencias de que en el siglo XIV usaban en la cultura China (según Joao de Barros) la característica corporal de la palma de la mano para imprimirla en papel como método para distinguir entre personas. Si retrocedemos hasta varios siglos antes de Cristo, existe constancia en la antigua Persia de la utilización de huellas dactilares para verificar la identidad [12].

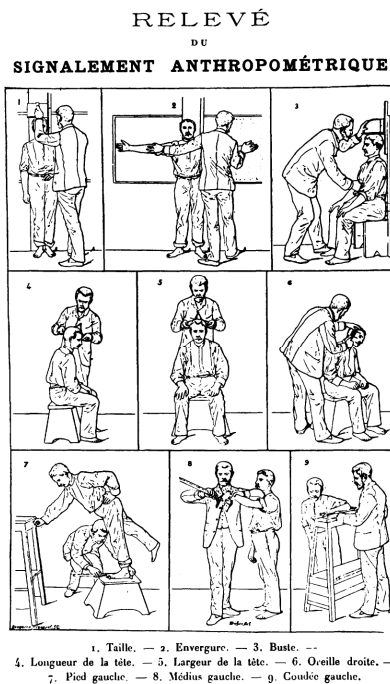


Fig. 1-4. Distintas medidas expuestas en el libro *Identification anthropométrique* [13].

En Occidente no sería hasta finales del siglo XIX cuando Alphonse Bertillon [13], jefe del departamento fotográfico de la policía de París, desarrolló un sistema de medidas corporales (sistema antropométrico). Se basaba en medir de forma precisa longitudes y anchuras de la cabeza y el cuerpo como se puede ver en Fig. 1-4, también registraba marcas individuales como tatuajes. Este sistema presentaba diversos problemas, como la utilización de métodos distintos de medidas y cambios de medida, lo que originó que fuera sustituido por la huella dactilar [14]. Podemos considerar la antropometría como el inicio de la biometría actual.

La biometría actual, tal y como la conocemos hoy en día, no tiene más de 30 años. En los últimos años, sobre todo a partir del atentado contra las torres gemelas del 11-S y del aumento tecnológico, se ha originado un crecimiento drástico en este sector. Se debe al interés por la sociedad de utilizar patrones biométricos para identificar personas [15] y al hecho de un incremento de la capacidad de cómputo e interacción de los usuarios con los ordenadores, que permiten al campo de la biometría ser realmente una opción dentro de los sistemas de identificación.

Aunque la biometría en un inicio surge para el cumplimiento de la ley (identificar criminales, extranjeros ilegales, etc.), en los últimos tiempos ha trascendido este campo para ser también muy utilizada en aplicaciones civiles (identificación del propietario de un teléfono móvil, etc.).

La gran ventaja de la biometría comparada con otros sistemas de identificación es sin duda las facilidades que ofrece, pues podemos identificar a una persona por "sus características fisiológicas y de comportamiento" en lugar de por la "posesión que el usuario tiene" (contraseña, tarjeta de identificación, etc.).

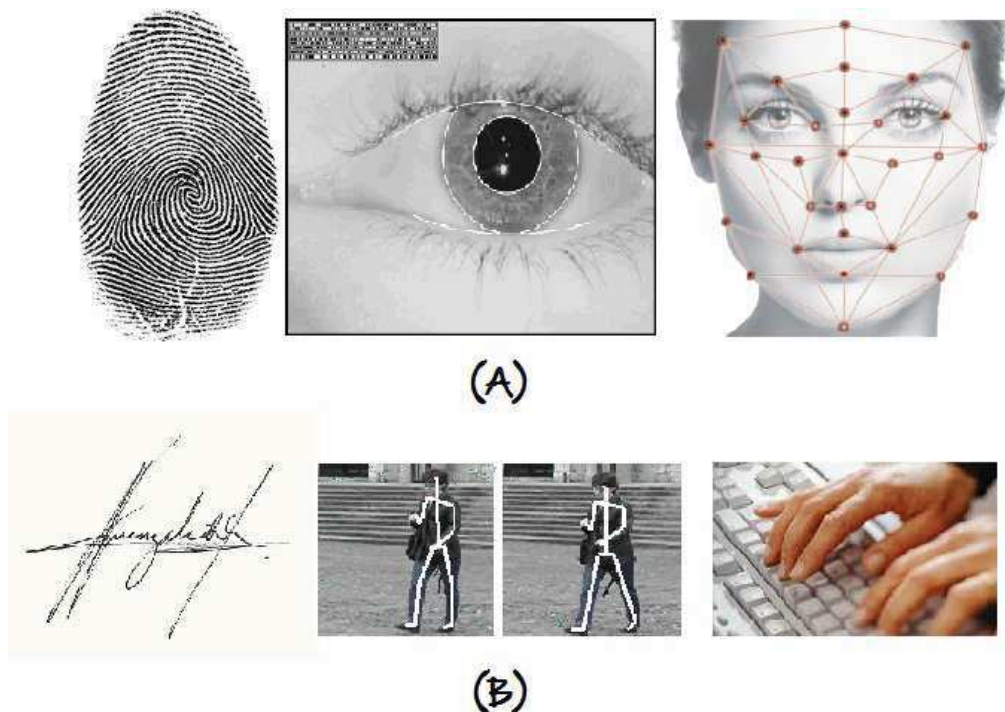


Fig. 1-5. Sistemas según su fisiología (A) y según su comportamiento (B)

El reconocimiento biométrico en si es el reconocimiento automático de los individuos en función de su fisiología (p. ej. huella digital, cara, iris) representados en la Fig. 1-5 (A) y/o características de su comportamiento [16] (p. ej. firma, el movimiento al caminar, pulsaciones de teclas) representados en la Fig. 1-5 (B). Por tanto, reconocimiento biométrico es un área de investigación muy amplia que incluye a investigadores de las comunidades de reconocimiento de patrones y aprendizaje automático.

Cada sistema biométrico es diferente, teniendo ventajas y desventajas comparado con otros sistemas. Un estudio publicado por *International Biometric Group* (IBG) muestra los rasgos biométricos más representativos respecto a unas medidas de calidad.

Estas medidas de calidad son:

- **Aceptación por el Usuario (AU):** Describe como el usuario percibe la prueba como una intromisión.
- **Eficacia del Sistema (ES):** Capacidad de identificar a una persona.
- **Equipamiento y Coste (EC):** Necesidad de equipamiento (sensores, lectores, etc.) y el coste que suponen.
- **Esfuerzo para el Usuario (EU):** Esfuerzo que realiza el usuario durante la toma de medidas.

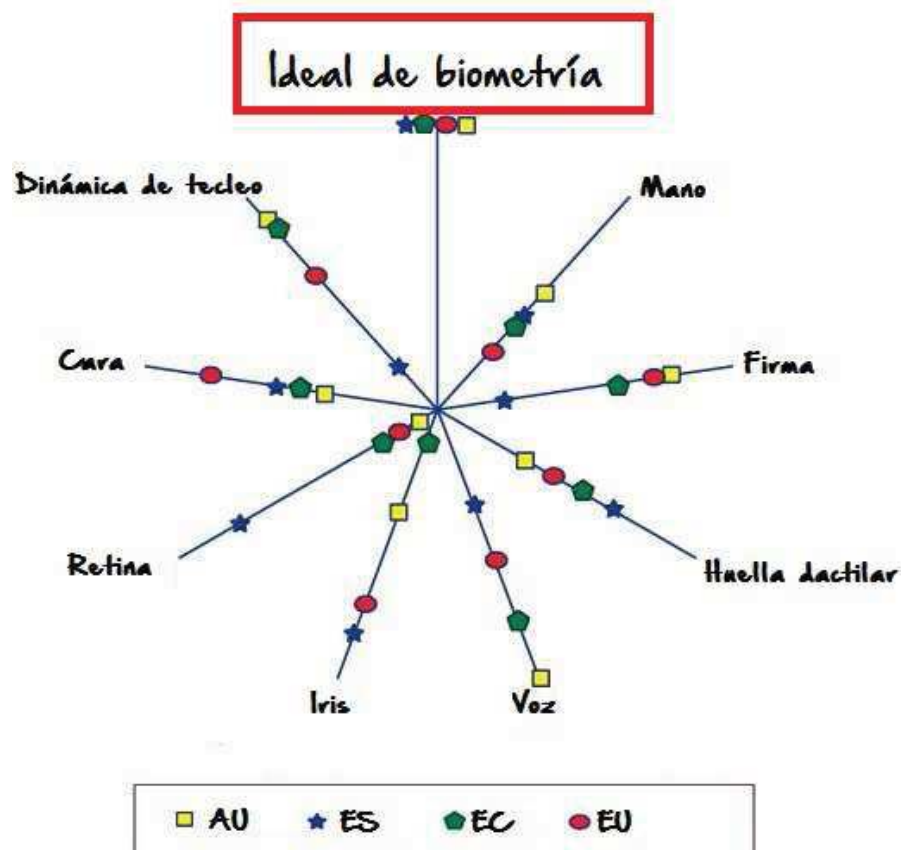


Fig. 1-6. Rasgos biométricos frente a medidas de calidad. Figura modificada de IBG.

Medidas como la eficacia del sistema o coste son comprobadas empíricamente, por tanto obtenidas a partir de la experimentación [17]. En cambio medidas como la aceptación de los usuarios son subjetivas y dependen de aspectos culturales o sociológicos.

Como podemos ver en la Fig. 1-6 ningún rasgo biométrico es superior al resto, sino que cada uno tiene unos beneficios determinados y unas desventajas. Dependiendo en el entorno en que nos encontremos (acceso y servicios domóticos, seguridad comercial, etc.) puede ser más útil uno que otro.

En el caso de la dinámica de tecleo vemos que los valores de AU y EC son excelentes. Es lógico puesto que el coste es mínimo ya que un teclado es un dispositivo que se encuentra hoy en día en casi todos los hogares (ausencia de equipamiento costoso, se basa en software), además no es nada intrusiva (el tecleo puede ser capturado sin constancia del

usuario). También se obtiene un valor más alto de la media en EU, sin embargo el valor ES es bastante bajo, por lo tanto si conseguimos mejorar la eficacia en dinámica de tecleo tendremos un sistema de identificación óptimo.

En esta última década el mercado biométrico no para de crecer y se espera que en los próximos cinco años crezca un 118% según ABI Research, debido a diferentes causas como las tensiones entre Oriente Medio y Europa, electrónica de consumo, etc.

Los estudios publicados por IBG [18], [19] muestran los datos referentes a los ingresos del mercado biométrico en 2009-2017 en Estados Unidos y la implantación de las tecnologías biométricas en los años 2007-2015. Estos datos han sido representados en las siguientes gráficas.

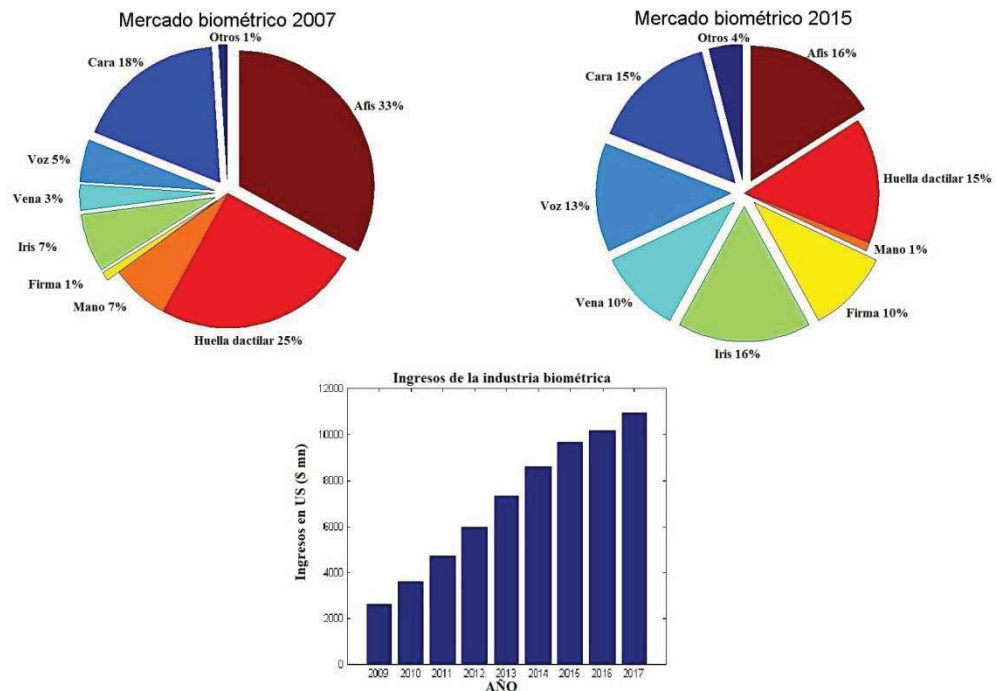


Fig. 1-7. Mercado biométrico (2007-2015) e ingresos de la industria biométrica en Estados Unidos

Como se puede ver en la Fig. 1-7, los sistemas AFIS y los basados en huella dactilar copan el 58% del mercado en 2007. Son sistemas comercializados desde hace más de 15 años con una presencia consolidada en el mercado.

Sin embargo, si nos fijamos en 2015, AFIS y huella dactilar copan el 31%. Este descenso es natural ya que nuevas propuestas han surgido y otras menos desarrolladas en 2007 empiezan a consolidarse. En la sección de otras tecnologías vemos como se ha pasado de un 1% a un 4%, debido en parte a la dinámica de tecleo y otras propuestas.

1.3 Dinámica de tecleo

La dinámica de tecleo es un rasgo biométrico de comportamiento que consiste en analizar la forma o el ritmo en que un usuario escribe en un teclado.

Un usuario puede realizar diferentes versiones de un patrón de tecleo debido a diversos motivos como: teclado empleado, posicionamiento ante el teclado, etc. Pero estas medidas ofrecen la suficiente información para ser discriminatorias, dado que las personas son predecibles en su comportamiento al desempeñar tareas rutinarias o repetitivas.

La dinámica de tecleo puede ser cambiante a lo largo del tiempo (biometría dinámica), lo que puede ser una ventaja. Recientes estudios han demostrado que la biometría estática (huella dactilar, etc.) no es lo suficientemente segura (fotografías tomadas con teléfonos inteligentes o dedos de plástico pueden copiar las huellas dactilares [20]).

Diversos parámetros pueden servir para identificar a una persona a través de dinámica de tecleo, latencia entre pulsaciones sucesivas, presión aplicada a las teclas [21], sonido al teclear, etc.

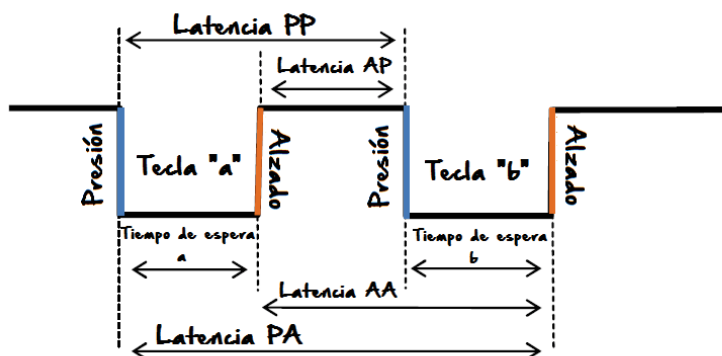


Fig. 1-8. Parámetros más utilizados en la dinámica de tecleo. Figura modificada de [22]

No obstante, el parámetro fundamental es el tiempo. En la Fig. 1-8 podemos ver los parámetros de tiempo más utilizados (Tiempo de espera, Latencia Alzado-Presión, Latencia Presión-Presión, Latencia Alzado-Alzado, Latencia Presión-Alzado), que obtienen resultados muy competitivos por sí mismos como podemos ver en [22].

Como hemos mencionado en la sección anterior, la dinámica de tecleo tiene muchas ventajas. Presenta un coste bajo, es poco intrusiva y además tiene un alto grado de reemplazabilidad (en caso de robo solo hay que generar otra contraseña aplicando la dinámica de tecleo).

La dinámica de tecleo puede ser tanto estática como dinámica. En este trabajo abordaremos la dinámica de tecleo estática. Este método solo se centra en un momento específico, por ejemplo al escribir una contraseña de seguridad. Existen diversas formas y entre ellas cabe la posibilidad de permitir que los usuarios puedan repetir contraseña. No obstante, en este trabajo no se permitirá su repetición (una contraseña por cada usuario).

Este método puede ser muy llamativo comercialmente, prueba de ello es que numerosos sectores se han visto atraídos por esta tecnología (bancos, identificación en correos electrónicos, etc.).

Una de las empresas actuales que tiene sistemas implementados con dinámica de tecleo es *BehavioSec*¹, que expone en un informe la importancia del uso de biometría de comportamiento [23]. El estudio que muestra es contundente: un 64% de personas usan la misma contraseña en todas partes. Además, un 37% reconoce a ver visto las cuentas personales de otras personas, por sesiones abiertas o conocer las contraseñas. Diversos estudios sugieren que las contraseñas sean complejas (longitud elevada, caracteres extraños, números, etc.) y sean cambiadas con regularidad. La realidad es que un número muy reducido de personas cumple estas sugerencias [22], [24].

¹ <https://www.behaviosec.com/>

Debido a estos motivos la dinámica de tecleo es una excelente opción como expone *BehavioSec*. Supone un suplemento de seguridad añadido a la contraseña [24] y en el caso de utilizar la dinámica de tecleo como único método, nos proporciona la ventaja de no tener que memorizar contraseñas (pudiendo ser el correo electrónico o tu propio nombre, datos que no se olvidan).

Para obtener el rendimiento de un sistema con dinámica de tecleo o cualquier sistema biométrico, primero hay que obtener una base de datos con el rasgo biométrico a estudiar de diferentes usuarios (en el caso de dinámica de pulsaciones, tiempos de la Fig. 1-8, teclas pulsadas, etc.).

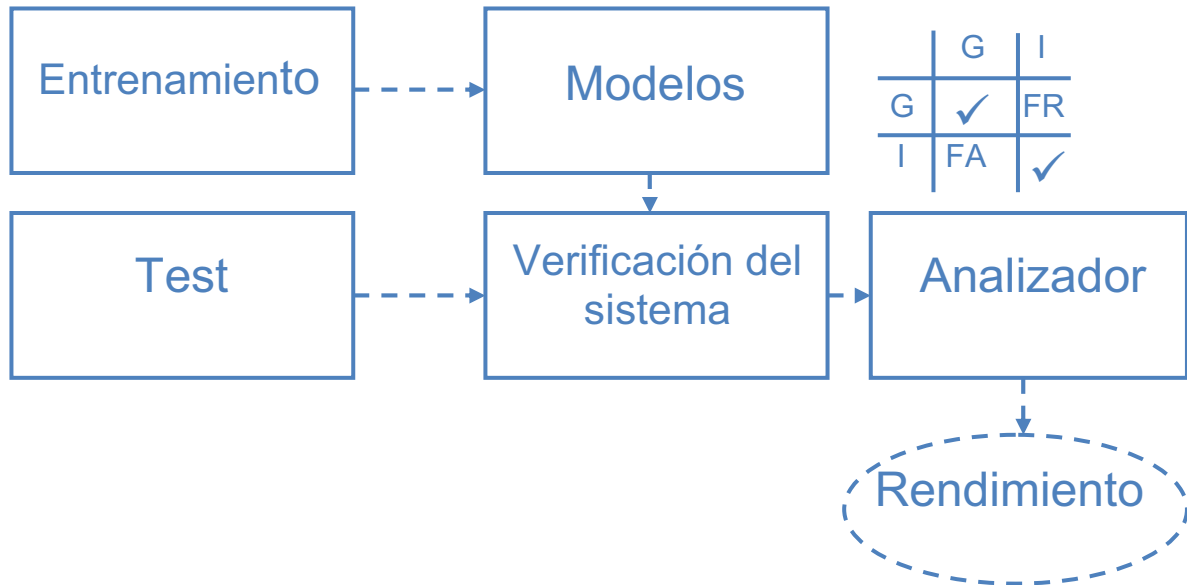


Fig. 1-9. Esquema para obtener el rendimiento de un sistema biométrico

En la Fig. 1-9 se ve un esquema para la obtención del rendimiento de un sistema. Consiste en:

- **Entrenamiento:** Son muestras que son tomadas a cada usuario introduciendo un parámetro de ellos mismos (por ejemplo, su nombre, su contraseña, etc.). Cada usuario repite este proceso un número de veces adecuado (cuanto más muestras más información se tiene del usuario, sin embargo a partir de un número de muestras no se aporta información relevante).
- **Test:** Son muestras tomadas tanto de usuarios que introducen sus propios datos (muestras genuinas), como de personas que suplantan la identidad de otro usuario introduciendo parámetros que no les pertenecen (muestras impostoras). Por ejemplo, por cada usuario puede haber 12 muestras, siendo 8 genuinas y 4 impostoras.
- **Modelos:** Son los algoritmos aplicados a las muestras de entrenamiento para obtener un sistema eficiente que sea capaz de identificar a una persona.

- **Verificación del sistema:** Una vez obtenida la suficiente información de las muestras de entrenamiento adquiriéndose el patrón de cada usuario, se verifica la eficiencia del sistema a través de algoritmos matemáticos, los cuales obtienen una distancia para cada muestra de la base de datos Test (por ejemplo, distancia Manhattan [22]).
- **Analizador:** El analizador tiene las diversas posibilidades con cada muestra, las cuales vienen representadas en la Fig. 9 en la parte derecha. Si es una muestra genuina y la reconoce como una impostora será un Falso Rechazo (FR). En cambio, si es una impostora y la identifica como una genuina será una Falsa Aceptación (FA). Otra posibilidad es la ausencia de error genuina como genuina e impostora como impostora. Se lleva a cabo estableciendo un valor umbral (por ejemplo, cuando una distancia es menor que el valor umbral se reconocerá como genuina, en cambio cuando una distancia es mayor se reconocerá como impostora)
- **Rendimiento:** Se utilizan parámetros como Tasa de Falsa Aceptación (FAR), Tasa de Falso Rechazo (FRR), Tasa de Igual Error (ERR) y Curva Característica Operativa del Receptor (ROC).

Los inicios de la dinámica de tecleo se remontan al código morse en los telégrafos, donde se percataron que cada persona tenía una manera diferente de teclear los mensajes (ritmos distintos). Este descubrimiento sería de gran ayuda en la Segunda Guerra Mundial permitiendo identificar a los emisores del mensaje. La técnica utilizada era llamada "El Puño del Emisor", una forma única de teclear "puntos" de un mensaje y "rayas", creando un ritmo que podía ayudar a distinguir aliados de enemigos.

Hay referencias de trabajos de dinámica de tecleo desde la década de 1970 [25], [26], pero no sería hasta 1980 cuando Gaines et al. [27] publicaran un artículo que fue fundamental para la dinámica de tecleo.

En las últimas décadas han sido muchas las aplicaciones de reconocimiento biométrico basadas en la dinámica de pulsación de teclas que se han investigado. Esta área es interesante para la comunidad científica por diversas razones, entre las que cabe destacar los desafíos asociados a la modelización y la búsqueda de secuencias dinámicas con alta variabilidad intra-clase, o el comportamiento humano, fuertemente dependiente del usuario y que varía significativamente entre los sujetos. Además, la naturaleza simple de los datos (secuencias de tiempo), hace que la biometría de pulsaciones de teclas sea un buen campo para introducir nuevos investigadores (sin experiencia previa en aplicaciones biométricas) en esta área desafiante que atrae tanto a académicos y profesionales.

Por este motivo el trabajo presentado se centrara en el reconocimiento a partir de dinámica de tecleo.

1.4 Dinámica de Tecleo: Evaluaciones tecnológicas

La comunidad científica que trabaja en reconocimiento basado en dinámica de teclas es heterogénea e incluye investigadores de diferentes disciplinas [28]. Muy pocas son las propuestas que nos ofrecen la oportunidad de comparar los sistemas en las mismas condiciones. Este problema es importante, ya que el rendimiento de los sistemas de dinámica de teclas es tremendamente dependiente de la aplicación (texto libre o fijo) y bases de datos (diferentes usuarios muestran comportamientos muy diferentes), además del

problema añadido por la inexistencia de un enfoque genérico debido al gran número de algoritmos.

La literatura ofrece bases de datos públicas que posibilitan comparar los sistemas en las mismas condiciones, siendo dos principalmente las más populares en dinámica de teclas: CMU [29] y GREYC [30]. Aunque estas bases de datos son de utilidad para la comunidad científica, nos encontramos con dos problemas importantes. Por un lado el pequeño número de usuarios disponibles en cada base de datos (51 en CMU y 133 para GREYC) y por otro el escenario de reconocimiento, en el cual todos los usuarios comparten la misma contraseña. Pensamos que este escenario limita mucho a estas propuestas, pues los usuarios resultan insuficientes, más cuando sabemos que en pulsaciones de tecleo se ve altamente afectado el resultado por este número reducido de participantes, dentro de un escenario que no es el más realista (normalmente no todas las personas comparten contraseña).

Tabla 1-1: Estudio de las bases de datos disponibles públicamente en dinámica de tecleo.

Base de datos	Usuarios	Muestras	Sesiones	Propiedades
CMU [29]	51	400	8	Igual contraseña para todos los usuarios: “.tie5Roanl”
MIMOS [31]	100	10	*	Igual contraseña para todos los usuarios: “try4-mbs”
Clarkson [32]	39	20	1	Igual contraseña para todos los usuarios:3 contraseñas “yesnomaybe”, “bahaNe312!” y “ballzonecart”
GREYC [30]	133	5-107	*	Igual contraseña para todos los usuarios: “greyc laboratory”
BeiHang [33]	117	4-16	1	Diferentes contraseñas por usuario
ATVS [22]	63	120	2	Diferentes contraseñas por usuario

*Depende de cada usuario.

En la Tabla 1-1 podemos ver las bases de datos públicas más populares en dinámica de tecleo. Como mencionamos anteriormente, la mayoría (CMU, MIMOS, GREYC) utilizan la misma contraseña para todos los usuarios. Realmente la única ventaja que ofrece compartir la contraseña es la posibilidad de utilizar las muestras de los usuarios también como impostoras. Además, estas muestras utilizadas como impostoras ofrecen una situación realista, pues los usuarios manejan la contraseña con soltura (un impostor real anteriormente abría practicado con la contraseña).

Aunque utilizar todas las muestras como impostoras tampoco es una situación conveniente puesto que puede afectar negativamente a los resultados de los sistemas como expresa el artículo [29].

Por otra parte, hay dos bases de datos (BeiHang, ATVS) que no tienen esta restricción (cada usuario tiene una contraseña diferente). En el caso de BeiHang la adquisición de las muestras se hizo sin ningún tipo de supervisión y el número de muestras por usuario (4-16) hace difícil proponer un sistema eficiente y sencillo.

El artículo publicado con la base de datos ATVS es una propuesta muy reciente y nos servirá de punto de partida para nuestro trabajo, no obstante su número reducido de usuarios (63), supone una limitación. Aunque la base de datos ATVS consta de 7560 muestras en las cuales cada usuario introduce nombre, apellido, email, nacionalidad y DNI, obteniéndose resultados realmente esperanzadores para la dinámica de tecleo, no son estos los motivos de nuestro gran interés. Lo realmente llamativo en este artículo es el uso de una plataforma orientada a la experimentación reproducible en el ámbito de la biometría. Esta será sin duda nuestra motivación en el presente trabajo.

En este trabajo se presenta una plataforma de evaluación de tecnologías de reconocimiento biométrico basadas en dinámica de tecleo desarrollada sobre la plataforma BEAT: Biometrics Evaluation and Testing (BEAT²) [34], fruto del proyecto europeo homónimo cuyo principal objetivo es promover la investigación reproducible. Como aplicación, se propone una competición que formará parte de las competiciones oficiales del congreso internacional The IEEE Eighth International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2016).

Por otro lado, solo tenemos constancia de un precedente de competición de reconocimiento de dinámica de tecleo que tuvo lugar en la conferencia internacional de Biometría BTAS 2015: “One-handed Keystroke Biometric Identification Competition” [35]. En esta competición se contaba con un conjunto de datos tomados de 64 usuarios que se ven obligados a usar solo una mano para escribir muestras de texto largo mecanografiadas libremente. Como podemos ver, la competición evalúa el rendimiento de los algoritmos en condiciones de más dificultad (una sola mano), cuando lo normal es utilizar las dos.

Entendemos del todo necesario que los usuarios usen dos manos, pues no hay que olvidar que se busca poder evaluar sistemas reales para poder implementarlos en la realidad.

En otras sub-áreas de la biometría existen competiciones *ongoing* con un gran éxito, es el caso *FVC-onGoing Competition* [36] con más de 900 participantes y más de 4000 algoritmos evaluados desde el año 2009 para huella digital, cara y huella de la palma de la mano. Se han obtenido resultados realmente competitivos, por ejemplo en huella digital se consiguieron valores de ERR 0,7%.

Por estas razones, *FVC-onGoing Competition* es un punto de referencia en el presente trabajo.

1.5 Motivación y Objetivos

La experimentación reproducible es imprescindible en el mundo científico (sección 1.1). Para la comunidad científica supone un ahorro de tiempo (en el último teorema de Fermat 358 años) y una mayor calidad en los trabajos (mayor organización y claridad, ausencia de errores y manipulaciones en los resultados). Además del beneficio para el propio autor, pues los artículos que hacen públicas las bases de datos utilizadas y los algoritmos que generan los resultados son mucho más citados.

Teniendo presente estas circunstancias, **parece una necesidad incentivar la experimentación reproducible en el ámbito de la biometría** (sección 1.2). El área de la biometría está en continuo crecimiento (en los próximos cinco años se espera que su mercado se duplique), siendo numerosos los artículos que exponen nuevos sistemas y resultados. Por este motivo, creemos necesario una plataforma que promueva la experimentación reproducible en el área de la biometría, sobre todo cuando los experimentos se basen en algoritmos (con un simple clic se pueden obtener los resultados).

Parece idónea para esta tarea la plataforma BEAT que nace con el objetivo de promover la investigación reproducible en el ámbito de la biometría.

Para realizar esta tarea es necesario elegir un rasgo biométrico. La dinámica de tecleo parece ser el indicado (sección 1.3), pues tiene una proyección enorme de futuro a pesar de

² <https://www.beat-eu.org/platform/>

no haber tenido una gran aceptación en sus inicios (si se mejora su eficiencia será un rasgo muy competitivo). La naturaleza simple de los datos (secuencias de tiempo), permiten introducir a nuevos investigadores, sin experiencia previa en biométrica.

Actualmente son escasas las bases de datos públicas en dinámica de tecleo (sección 1.4), adoleciendo de un planteamiento poco realista (misma contraseña para todos los usuarios) y un número reducido de participantes (los sistemas se ven altamente afectados por el número limitado de usuarios). Además, conjuntamente solo encontramos una competición en dinámica de tecleo que adolece de no ser demasiado realista, dado que los datos se obtenían mediante el uso de una sola mano.

Debido a todos los motivos anteriormente expuestos, los objetivos principales del presente trabajo son:

- Realizar una competición KBOC que forme parte de las competiciones oficiales del congreso internacional BTAS 2016.
- Realizar una competición en un marco completamente reproducible utilizando la plataforma BEAT.
- Introducir a nuevos investigadores en aplicaciones biométricas.
- Incentivar la dinámica de tecleo a través de la competición, proponiendo un entorno realista donde comparar los sistemas propuestos por los investigadores.
- Hacer pública una nueva base de datos en dinámica de tecleo con 300 usuarios que escriben su nombre y apellido repetidas veces para un total de 8400 secuencias de tecleo de las cuales 3600 son ataques de impostores.

Como objetivos secundarios, destacan:

- Ofrecer diversas formas de participar en la competición: *participación offline* y *participación ongoing*.
- Dar a conocer la competición KBOC mediante la creación de una página web y su posterior difusión.
- Presentar la plataforma BEAT a investigadores procedentes del área del reconocimiento biométrico.
- Estimular la participación ongoing (mediante BEAT) ofreciendo algoritmos de sistemas de dinámica de tecleo realizados explícitamente para el concurso en la plataforma BEAT y dar a conocer los resultados obtenidos.
- Mantener por tiempo indefinido la competición ongoing para poder comparar los sistemas con diferentes investigadores, creándose una tabla de posiciones para conocer los sistemas más eficientes.
- Análisis y comprensión de los resultados obtenidos en la competición.

1.6 Contribuciones del TFG

Las contribuciones de este Trabajo Fin de Grado se resumen en:

- Desarrollo de una herramienta de evaluación de tecnologías de reconocimiento de dinámica de tecleo basada en la plataforma BEAT.
- Desarrollo de una competición internacional que sirva para validar dicha herramienta y pueda ser utilizada como un nuevo recurso por la comunidad científica.
- Desarrollo de material suplementario que fomente la participación: página web, tutorial, video tutorial, algoritmos de desarrollo.
- Análisis de los resultados obtenidos por los participantes en la competición.
- Validación de la plataforma BEAT a través de la competición.
- Envío de dos artículos a congresos que resumen las principales aportaciones del TFG:

Aythami Morales, Julian Fierrez, Marta Gomez-Barrero, Javier Ortega-Garcia, Roberto Daza, Jhon V. Monaco, Jugurta Montalvão, Jânio Canuto, Anjith George, "KBOC: Keystroke Biometrics OnGoing Competition", *IEEE Eighth International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2016)*, **Under Review**.

Roberto Daza, Aythami Morales, Julian Fierrez, Marta Gomez-Barrero y Javier Ortega-Garcia, "KBOC: Plataforma de Evaluación de Tecnologías de Reconocimiento Biométrico Basadas en Dinámica de Tecleo", *XXXI Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2016)*, **Under Review**.

1.7 Organización de la memoria

La memoria del presente trabajo se compone de los siguientes capítulos:

- **Capítulo 1: Introducción:** En este capítulo se explica la importancia de la experimentación reproducible, introduciéndonos en el ámbito de la biometría y del sub-área de la dinámica de tecleo. La motivación para el desarrollo de este proyecto y los objetivos a cumplir durante la ejecución del mismo, forman parte de este capítulo, así como la estructura de esta memoria.
- **Capítulo 2: KBOC. Competición en dinámica de tecleo:** Se explica todo lo referente a la competición KBOC (base de datos, manera de promover la participación, material ofrecido, diferentes maneras de participar (offline, ongoing). También se presenta la plataforma BEAT a la comunidad científica, se exponen los objetivos fundamentales de la plataforma y se expresa la ayuda dada para mejorar la plataforma. Además se muestra como se ha realizado la competición *ongoing* en la plataforma BEAT (*Toolchain, Data forms, Algorithms, etc.*)
- **Capítulo 3: Resultados:** Se exponen los resultados obtenidos por la competición en los algoritmos dados de ejemplo. También se muestran los sistemas de los participantes con sus correspondientes resultados y finalmente se analizan.
- **Capítulo 4: Conclusión:** En este capítulo se presentan las conclusiones extraídas del proyecto realizado.

2 KBOC. Competición en dinámica de tecleo

En este capítulo se presenta la competición KBOC. Se introduce a la plataforma BEAT y se detallan las herramientas facilitadas por la competición en la plataforma BEAT.

2.1 Introducción

La competición presentada en este trabajo, *Keystroke Biometrics OnGoing Competition*, forma parte del congreso internacional *The IEEE Eighth International Conference on Biometrics: Theory, Applications, and Systems*.

Se centra en el área de la biometría, concretamente en el sub-área de la dinámica de tecleo.

La competición KBOC trata de ser un punto de referencia en el sub-área de la biometría, intentando complementar las experiencias anteriores (sección 1.4, GREYC, OKBIC, etc.) con:

1. Una nueva base de datos, la cual (que tengamos constancia) contiene el mayor número de usuarios (300) en dinámica de tecleo hasta la fecha. Estos usuarios escriben su nombre y apellido repetidas veces.
2. Al contrario de anteriores competiciones, KBOC es una competición *ongoing* que permite a los investigadores seguir compitiendo con sus sistemas de dinámica de tecleo de manera online.
3. Una competición en un marco completamente reproducible utilizando la plataforma BEAT.

La competición KBOC emula un escenario realista a diferencia de otras competiciones, además de permitir a investigadores y profesionales de otros campos iniciarse en dinámica de tecleo (no es necesario ser un experto). Siendo una competición innovadora, pues la mayoría de competiciones están activas por un corto periodo de tiempo, KBOC *ongoing* estará de forma permanente (aun habiendo finalizado BTAS16) y los resultados se actualizarán automáticamente (BEAT).

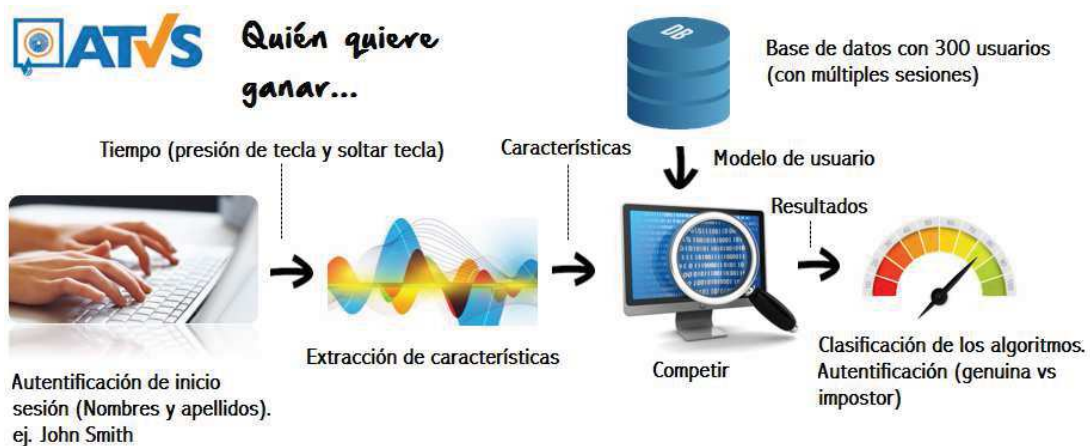


Fig. 2-1. Esquema de la competición usado para incentivar la motivación. La figura original está escrita en inglés.

2.2 Conjunto de datos de referencia

La base de datos utilizada en la competición KBOC es parte del conjunto de datos de BiosecurID [37] y consiste en muestras de tecleo tomadas a 300 usuarios en 4 sesiones diferentes, en un espacio temporal de 4 meses. En cada sesión los usuarios repiten 4 veces su nombre y apellido (separado por la tecla espacio), usando tanto mayúsculas como minúsculas, no permitiéndoseles errores (pulsar tecla retroceso). En el caso de que el usuario se equivoque tendrá que iniciar el proceso de nuevo. Tampoco se les permiten errores a los impostores (3 muestras impostoras por sesión son tomadas para cada usuario).

La dinámica de tecleo de cada secuencia viene definida por las marcas de tiempo referentes a dos eventos: pulsar y soltar teclas. Se suministra las marcas de tiempo en milisegundos así como el identificador de tecla pulsada.

Tabla 2-1: Resumen de las principales características de la base de datos utilizada en la competición

Características	
Número de usuarios (conjunto de prueba)	300
Número de sesiones	4
Muestras de entrenamiento por usuario	4
Muestras de Test por usuario	20
Muestras Genuinas por usuario*	entre 8-12
Muestras Impostoras por usuario*	entre 8-12
Total de secuencias genuinas de la base de datos	4800
Total de secuencias genuinas utilizadas en KBOC	3028
Total de secuencias impostoras de la base de datos	3600
Total de secuencias impostoras utilizadas en KBOC	2972
Separación media entre las sesiones	1 mes
Número medio de teclas pulsadas	25,55

*Los participantes solo disponen de las etiquetas de 4 muestras genuinas de la primera sesión de cada usuario. El resto de muestras se utilizan para evaluar el rendimiento del sistema y los participantes desconocen si pertenecen a usuarios genuinos o impostores (el número de muestras genuinas e impostoras de test suman 20 para todos los usuarios pero su distribución varía según usuario).

Las principales estadísticas del conjunto de datos del concurso se resumen en la Tabla 2-1. El protocolo de experimentación se resume en los siguientes pasos:

- La dinámica de tecleo de cada usuario se modela a partir de 4 muestras genuinas (muestras de entrenamiento) de cada usuario (pertenecientes a la primera sesión).
- Se obtiene una puntuación por cada una de las 20 muestras de test (muestras genuinas e impostoras de las sesiones 2 a 4).
- El rendimiento final se da en forma de curvas de falsa aceptación, curvas de falso rechazo, curvas DET y tasas de igual error obtenidas a partir de las puntuaciones genuinas e impostoras de todos los usuarios.

2.3 Competición *ongoing* y competición *offline*

Hay dos formas de participación en la competición, se describen a continuación:

- **Participación *offline*:** Debido a que la base de datos y protocolos de experimentación son los mismos para ambos tipos de participaciones, la participación *offline* sirve de referencia para los sistemas evaluados en la plataforma *ongoing*.

Las muestras de test sin etiquetar se facilitaron a los participantes (los participantes no conocían si las muestras eran genuinas o impostoras, pero si a que usuario correspondía la muestra). También se suministraron las muestras de entrenamiento (4 muestras genuinas por cada usuario). Se proporciono un conjunto pequeño de las muestras de test (10 usuarios con muestras etiquetadas) y un algoritmo de referencia (código matlab, completamente comentado). Los 10 usuarios facilitados sirven a los participantes para probar sus sistemas antes de mandarlos a la competición.

Los participantes no obtuvieron el rendimiento de sus sistemas hasta el final de la competición. Era posible enviar 15 sistemas diferentes los cuales se ejecutaron en las máquinas de los participantes, tras lo cual se envió un fichero de resultados con un formato predefinido que fue procesado por la organización de la competición.

- **Participación *ongoing*:** Era la opción recomendada, se realizo a través de la plataforma BEAT y los participantes obtuvieron el rendimiento de sus sistemas al instante y por lo tanto no se precisaba del envío de ficheros de puntuaciones. Simplemente tenían que hacer público el experimento (poner público o privado el algoritmo era elección del participante). No había restricción frente al número de sistemas creados por los participantes.

La base de datos utilizada es la misma que en la participación *offline* (en la base de datos test no conocen si las muestras son genuinas o impostoras). En la competición *ongoing* no era necesario facilitar los 10 usuarios con las muestras etiquetadas, ya que los usuarios obtenían el rendimiento de sus sistemas al instante desde la plataforma BEAT.

Para evitar malas conductas o sobreajustes de los sistemas, en un inicio solo se facilitó los 100 primeros usuarios, ya que se comprobó que eran representativos del conjunto completo (300 usuarios). En la última semana de la competición se habilitó la base de datos con los 300 usuarios.

La competición *ongoing* no tiene fecha de finalización programada (permanecerá activa por tiempo indefinido).

En las siguientes secciones se introduce la plataforma BEAT, se detalla cómo se ha llevado a cabo la competición *ongoing* en la plataforma, y se expone el material facilitado a los participantes.

2.4 Introducción a la plataforma BEAT

La plataforma BEAT ha sido pieza fundamental en el desarrollo de este TFG. Dada su reciente aparición (Junio de 2015) creemos oportuno introducirla en este capítulo para entender mejor así las contribuciones del TFG. La plataforma BEAT surge por un proyecto UE con la idea de promover la investigación reproducible en Biometría, pero podemos afirmar que trasciende este campo ya que engloba cualquier proceso de señales.

BEAT proporciona todas las herramientas necesarias para evaluar los sistemas biométricos, tanto para entidades académicas como comerciales. Supone un gran avance, facilitando la comunicación entre ambas entidades y ofreciendo una base común donde poder comparar los sistemas biométricos.

Inexistencia de un método sencillo para hacer reproducibles los experimentos biométricos. Evaluar los sistemas biométricos de una forma común (criterio unificado).

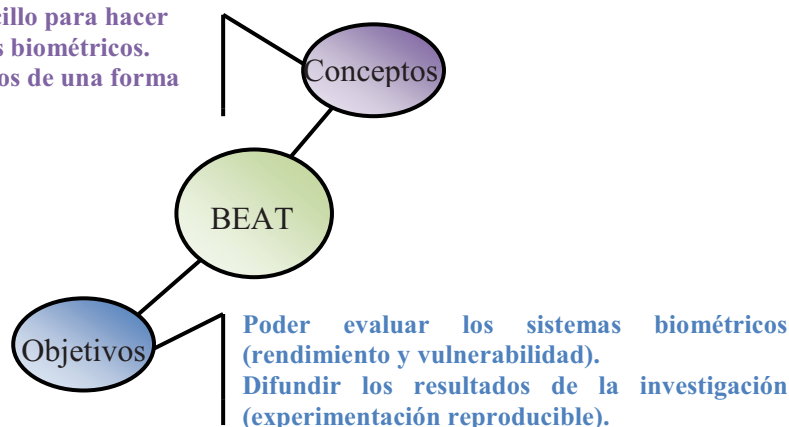


Fig. 2-2. Se muestran los conceptos generales que dieron lugar a la plataforma y los objetivos de BEAT.

Para conseguir todos los objetivos anteriormente mencionados, el proyecto BEAT se ayuda de colaboraciones de entidades especializadas en biometría. Hay un primer grupo de entidades (IDIAP, UAM, UNIS, MORPHO) con una amplia experiencia en el desarrollo de sistemas biométricos. Un segundo grupo de entidades (CEA, TÜViT) con experiencia en el estudio sobre vulnerabilidad en sistemas biométricos. Y por último, un tercer grupo (EPLF, TÜBITAK) con extensa experiencia en criptología. La Tabla 2-2 muestra la importancia en el área de la biometría de las entidades participantes en la plataforma BEAT, pudiéndose beneficiar de su experiencia y conocimientos.

Destacar que el proyecto BEAT se compone de agencias gubernamentales de toda Europa (Alemania, España, Francia, Inglaterra y Australia), trabajando juntas para conseguir los ambiciosos objetivos que propone.

BEAT ofrece multitud de posibilidades, desde compartir nuestros resultados manteniendo el código privado hasta compartir nuestro algoritmo, etc. Además nuestra experiencia nos indica que nosotros mismos no somos capaces de reproducir un experimento antiguo a causa de que hemos hecho nuevas versiones olvidando las anteriores. BEAT da una solución a este problema guardando un historial de nuestros algoritmos, pudiéndose llegar a cualquier versión anterior.

La plataforma BEAT ofrece una propuesta innovadora en el ámbito de la biometría, con una serie de características que la hacen idónea para facilitar la experimentación reproducible (solucionando los problemas mencionados en la sección 1.1). Dichas características serán descritas en la siguiente sección.

Tabla 2-2: Entidades colaboradoras con la plataforma BEAT

Entidades colaboradoras con BEAT	Características
UAM ³	Varias tecnologías implementadas debido a su labor investigadora (huella dactilar, dinámica de tecleo, etc.)
IDIAP ⁴	Desarrollo de tecnologías de reconocimiento de la cara y basadas en el habla (BioLogin ⁵).
UNIS ⁶	Sistemas de biometría centrados en los labios, sistema de seguimiento con reconocimiento facial, etc.
EPFL ⁷	El laboratorio LASEC en EPFL tiene una larga experiencia en sistemas cifrado, autenticación (varias patentes).
TÜBITAK ⁸	Varios grupos trabajan en biometría en TÜBITAK, gran reconocimiento en habla, gramática y semántica.
CEA ⁹	Proporciona la experiencia en sistemas de huella dactilar (metodología de evaluación, técnicas de suplantación, etc.)
MORPHO ¹⁰	Primer dispositivo multimodal huella dactilar y vena (utiliza Hitachi). Sistemas de reconocimiento por iris, etc.
TÜViT ¹¹	Desde 2008 tiene presencia en el ámbito evaluación de la vulnerabilidad en sistemas biométricos (suplantación de identidad).

2.4.1 Características de la plataforma BEAT

BEAT es una plataforma innovadora, proporcionándonos las herramientas necesarias para implementar sistemas de Biometría, incluyendo:

- Acceso libre para la comunidad científica (simple registro añadiendo correo electrónico, nombre y usuario).
- No es necesario la instalación de ningún software, todo se realiza vía web.
- Utiliza uno de los lenguajes más populares y de mayor proyección de futuro (Python), el cual es libre.



Fig. 2-3. Muestra que BEAT utiliza Python (código libre) para hacer más accesible los algoritmos.

³ <http://atvs.ii.uam.es/listpublications.do>

⁴ <https://www.idiap.ch/scientific-research/research-groups/biometric-person-recognition>

⁵ <http://www.idiap.ch/biologin/>

⁶ <http://www.surrey.ac.uk/cvssp/research/biometrics/>

⁷ <http://lasec.epfl.ch/>

⁸ <http://www.tubitak.gov.tr/>

⁹ <http://www.leti.fr/en>

¹⁰ <http://www.morpho.com/>

¹¹ <https://www.tuvit.de/en/downloads-1898.htm>

- Incorpora la mayoría de bibliotecas más populares en reconocimiento de patrones actualizadas a su última versión (scipy, numpy, etc.). Además se ha creado una biblioteca con las funciones más usadas en los trabajos relacionados con el reconocimiento de patrones (bob¹²).
- La plataforma trabaja con diagramas de bloques que permiten una sencilla programación por módulos y una eficiente reutilización de bloques en diferentes aplicaciones.

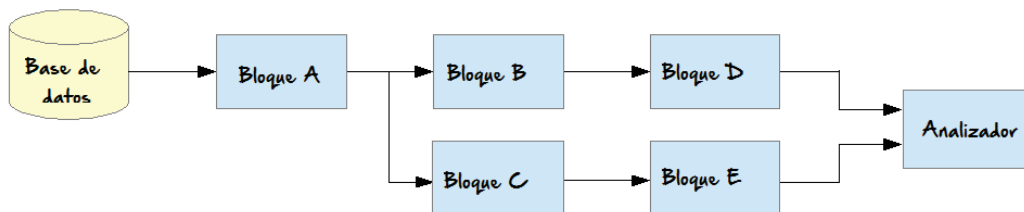


Fig. 2-4. Esquema de un diagrama de bloques utilizado en la plataforma BEAT

- La computación es de muy alto rendimiento, debido en parte a la utilización de diagramas de bloques, lo cual permite a la plataforma dividir el trabajo en diferentes bloques, dando más importancia a los que requieren un mayor cómputo. Para hacer esta tarea BEAT se organiza en lo que llaman ranuras. Cada usuario tienen asignado un número de recursos para evitar saturar la plataforma BEAT. Este número de recurso puede ser cambiado por el administrador de la plataforma. Otro rasgo fundamental de la plataforma BEAT es que contiene una cache de 10 Tb, lo que hace ejecutar los experimentos mucho más rápido. Una vez ejecutado un experimento, si posteriormente se modifica alguno de los bloques, los demás sin modificar se ejecutarán al instante al tenerlos guardados en la cache.
- Posibilidad de hacer público o privado los trabajos. Esta es una de las principales ventajas de BEAT. Debido a que la plataforma trabaja con diagramas de bloques nos permite muchas alternativas en este sentido. Por ejemplo, compartir nuestros resultados sin tener que compartir nuestro algoritmo o simplemente compartir la base de datos utilizada, etc. Cada campo en BEAT (experimento, algoritmo, bases de datos, etc., se comentarán en la sección 2.5) tiene las siguientes alternativas en la plataforma:
 - Público: Puede hacerse público para todos los usuarios o simplemente para algunos usuarios seleccionados.
 - Privado: No podrán tener acceso los demás usuarios (todos los campos una vez que son guardados serán privados por configuración para una mayor seguridad, pudiéndose cambiar a las otras alternativas)
 - Ejecutable*: Permite que utilicen este campo (ejemplo, algoritmo), pero no pueden ver su contenido. Pueden utilizarlo todos los usuarios o simplemente para algunos seleccionados.

¹² <https://pypi.python.org/pypi/bob/1.0.4>

* La opción ejecutable no aparece en todos los campos puesto que no siempre tiene sentido. Por ejemplo, en el campo experimentos donde se recogen los resultados del trabajo.

- Posibilidad de crear equipos en BEAT. Permite compartir los resultados fácilmente con el equipo y mostrarse en la plataforma como una entidad.
- Un mecanismo de certificación de los resultados.

Estas son algunas de las características que hacen a BEAT una gran opción en el mundo de la biometría. En la siguiente sección se analizará como está estructurada la competición *ongoing* en la plataforma BEAT.

2.5 Estructura de la competición ongoing en la plataforma BEAT.

Es importante mencionar que hemos realizado el primer tutorial avanzado de la plataforma BEAT (recientemente BEAT ha hecho su propio tutorial). Este tutorial está orientado para comprender y manejar todos los campos de BEAT a la vez que se enseña como participar en la competición KBOC. También se muestran los ejemplos creados para dicha competición. Este tutorial tiene la particularidad de ser interactivo y completamente en inglés, lo que sin duda favorecerá su uso a nivel internacional.

Para una mayor comprensión de BEAT y percibir todas las herramientas utilizadas en la competición *ongoing*, es aconsejable ver el tutorial que se encuentra en el siguiente enlace¹³ (diferentes formatos).

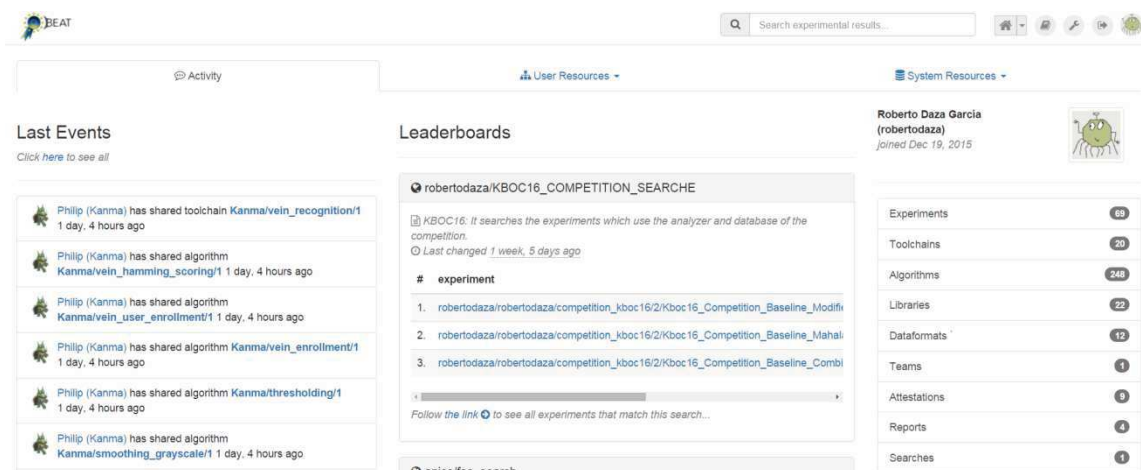


Fig. 2-5. Interfaz de la plataforma BEAT

En la Fig. 2-5 podemos ver la interfaz de la plataforma BEAT, donde observamos a la izquierda los últimos eventos (usuarios que comparten código, etc.), en el centro la tabla de clasificación creada exclusivamente para la competición y en la parte derecha las diferentes secciones a las que se puede acceder dentro de la plataforma BEAT (*Experiments, Toolchains, Algorithms, Libraries, etc.*).

¹³ <https://www.dropbox.com/s/lstdedyzgzs62oid/Tutorial%20BEAT.zip?dl=0>

La plataforma BEAT se estructura en los siguientes campos principalmente: *Algorithms, Toolchains, Dataformats, Experiments, Attestations*. Esto se debe a que utiliza programación orientada a bloques, permitiendo dividir fácilmente los algoritmos en tareas.

Por tanto en este TFG se utilizan los campos anteriormente mencionados para llevar a cabo la competición *ongoing* en la plataforma BEAT. A continuación se detallan todas las herramientas utilizadas en la plataforma para realizar la competición *ongoing*.

- **Toolchains:** Es el diagrama de los bloques donde se indican las entradas y salidas de cada bloque. Hay tres diferentes bloques:
 - **Bloques de bases de datos:** Son bloques que solo contienen salidas y albergan la base de datos del experimento (ejemplo, muestras de dinámica de tecleo). La competición desarrolla dos bloques de bases de datos (*templates* y *probes*)
 - **Bloques de procesamiento:** Son bloques con entradas y salidas, destinados a realizar las operaciones de nuestro sistema biométrico. La competición diseña el *Toolchain* con un solo bloque de procesamiento (*Participant_Block*) para ofrecer un sistema más sencillo a los participantes.
 - **Bloques de análisis:** Solo tienen salidas, son los bloques encargados de obtener el rendimiento de nuestro sistema (ERR, FAR, ROC, etc.). Se diseña un analizador para la competición (*analyzer_private*).

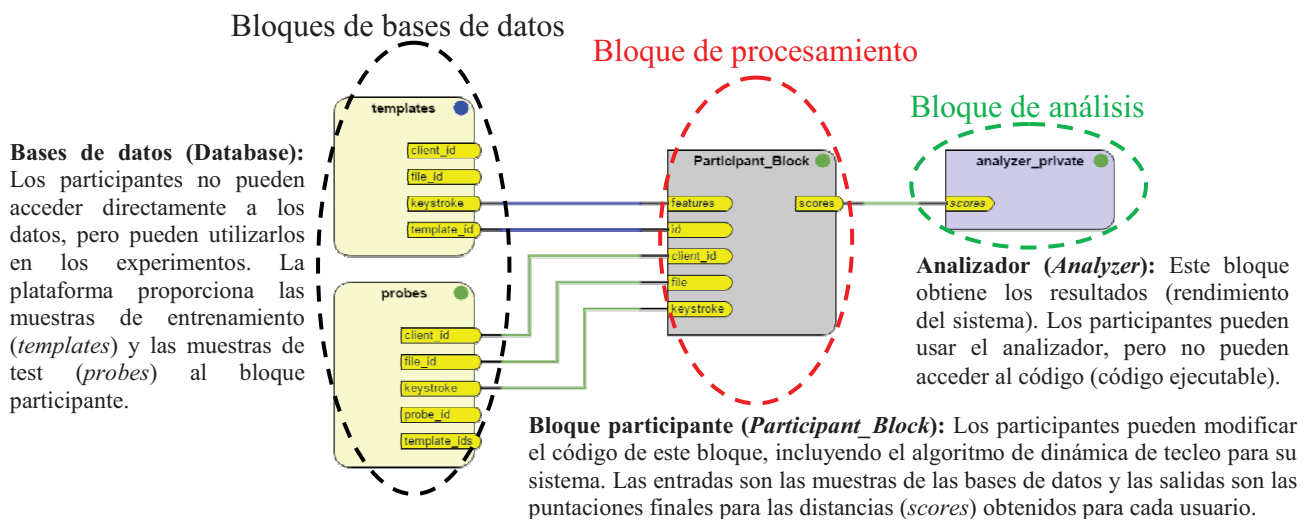


Fig. 2-6. *Toolchain* creado para la competición KBOC en BEAT

La Fig. 2-6 indica el *Toolchain* creado para la competición donde se puede ver los tres tipos de bloques y como se conectan las entradas y salidas. Se puede acceder al *Toolchain* de la competición desde el siguiente enlace¹⁴.

Hay que mencionar una peculiaridad de BEAT, cada bloque se ejecuta siguiendo una prioridad indicada por colores que elige el usuario. Para entenderlo lo explicaremos a través del *Toolchain* de la competición.

¹⁴ https://www.beat-eu.org/platform/toolchains/robertodaza/competition_kboc16/2/

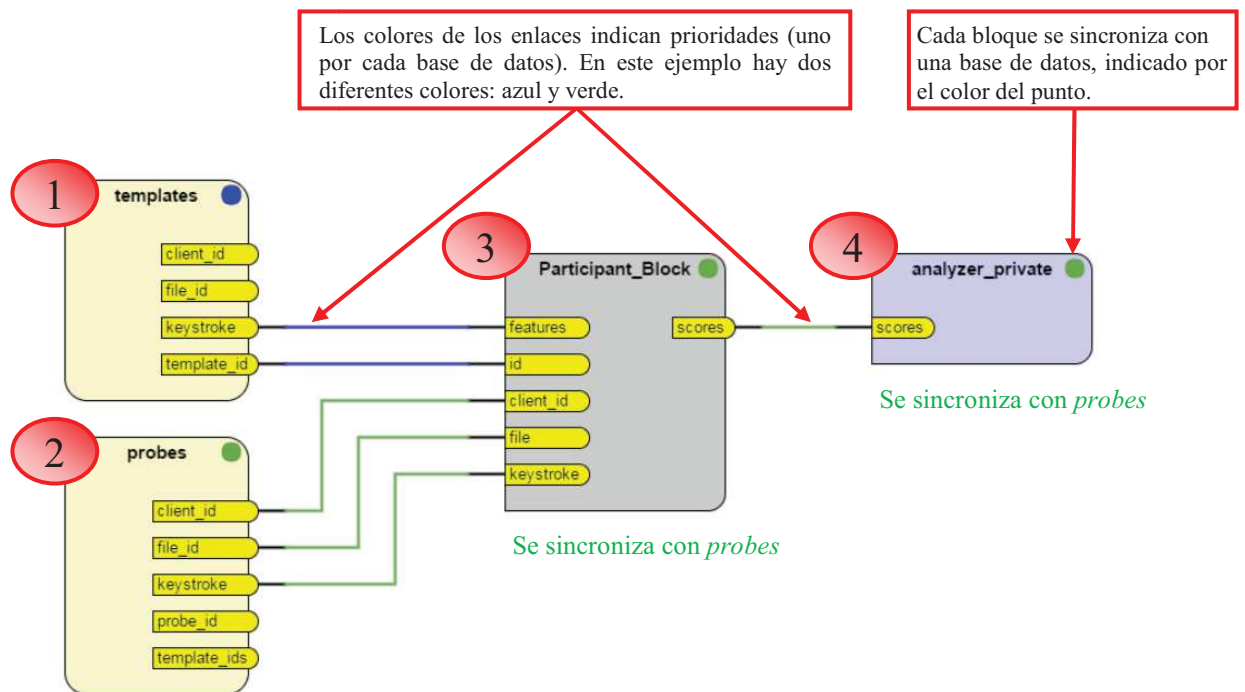


Fig. 2-7. Toolchain de la competición indicándose el orden de ejecución de cada bloque a través de los números y los colores que indican la sincronización de los bloques.

Los bloques de bases de datos definen un canal de sincronización único (en el caso de la Fig. 2-7 hay dos: azul y verde). Por ejemplo, el bloque *Participant_Block* esta sincronizado con el bloque *probes*, por tanto cada vez que el bloque *probes* se ejecuta se obtiene una salida en el bloque *Participant_Block*. Más adelante se hace una explicación más avanzada de como se sincronizan los bloques en la competición.

➤ **Bloques de bases de datos de la competición:** Hay dos bloques:

- Bloque *templates*: Corresponde a las muestras de entrenamiento.
 - Salidas: *keystroke*, *template_id*
- Bloque *probes*: Corresponde a las muestras de test.
 - Salidas: *Client_id*, *file_id*, *keystroke*

El **bloque *templates*** se encarga de suministrar los datos de la base de entrenamiento por las siguientes salidas:

La salida *keystroke* manda las características (vector de tiempos y vector con los indicadores de la tecla pulsada por cada muestra) de cada usuario a la entrada *features*.

La salida *template_id* indica el número de usuario que la base de datos manda (de 1 a 300).

El **bloque *probes*** se encarga de suministrar los datos de la base de test por las siguientes salidas:

La salida *keystroke*: contiene exactamente la misma información que en *templates*, pero para las muestras de test.

La salida *client_id*: indica el número de usuario que manda la base de datos y la salida *file_id*, la cual contiene el número de muestra. Pongamos un ejemplo para entenderlo mejor: si tenemos el usuario 3 con la muestra 1, *client_id* mandará el número 3 y *file_id* mandará el número 49 (debido a que va sumando las 24 muestras de los demás usuarios -

cuenta también las de entrenamiento-, es sencillo a través del *file_id* saber que es la muestra 1).

➤ **Dataformats:** Son los tipos de formatos de datos de entradas y salidas de los bloques. Es muy importante que una entrada y salida conectadas tengan el mismo tipo de formato de datos.

Los usuarios deben declarar el tipo de dato utilizado. BEAT dispone de los siguientes tipos de datos:

- Enteros con signo: int8, int16, int32, int64
- Enteros sin signo: uint8, uint16, uint32, uint64
- Números en coma flotantes: float32, float64
- Números complejos: cpmplex64, complex128
- Valor booleano: bool
- Una cadena de caracteres: string

```
1 {
2   "client_identity": "string", ←-----Es el número de usuario con formato string
3   "#description": "Score for verification", ←---- Descripción del formato de datos (es obligatorio)
4   "scores": [
5     0,
6     {
7       "score": "float64",
8       "file_id": "string"
9     }
10  ]
11 }
```

Se define un diccionario que contiene dos variables.
"score" es la distancia final para el usuario.
file_id indica el número de muestra del usuario con formato string.

Fig. 2-8. Ejemplo de formato de dato creado para la competición KBOC

En la competición se crea un formato de datos para cada entrada y salida. La Fig. 2-8 representa un formato de dato creado para la competición KBOC para la salida scores o la entrada scores (salidas y entradas conectadas contienen mismo formato), donde podemos ver como se definen (utilizando el formato de objetos en código Python) y en el que encontramos tres valores y la descripción del formato de datos. Una información más detallada aparece en el tutorial.

➤ **Algorithms:** Este campo es el indicado para el código (Python) a utilizar en los bloques.

BEAT trata a los algoritmos como objetos definidos en código Python, donde se define un *process ()* que procesa las entradas y produce las salidas.

Pero su mayor punto de diferenciación de un compilador normal que utilice Python estriba en la manera que sincroniza el flujo de datos de los bloques, siendo su mayor dificultad. La plataforma se encarga de extraer de las bases de datos las muestras (videos, voz, etc.) y mandarlas a los demás bloques según la prioridad, que viene indicada por el color en el *Toolchain*. Por lo tanto, en BEAT es muy importante la sincronización entre entradas.

Debido a estas razones BEAT no requiere de ningún tipo de bucle. Nos referimos a lo siguiente: en un compilador que utilice Python haríamos un *for* para pedir las muestras, en cambio en BEAT esto no es posible, las muestras son suministradas por la plataforma.

Pongamos el ejemplo de la competición para el bloque *Participant_Block*. Representa un caso complejo puesto que las entradas no están sincronizadas entre sí:

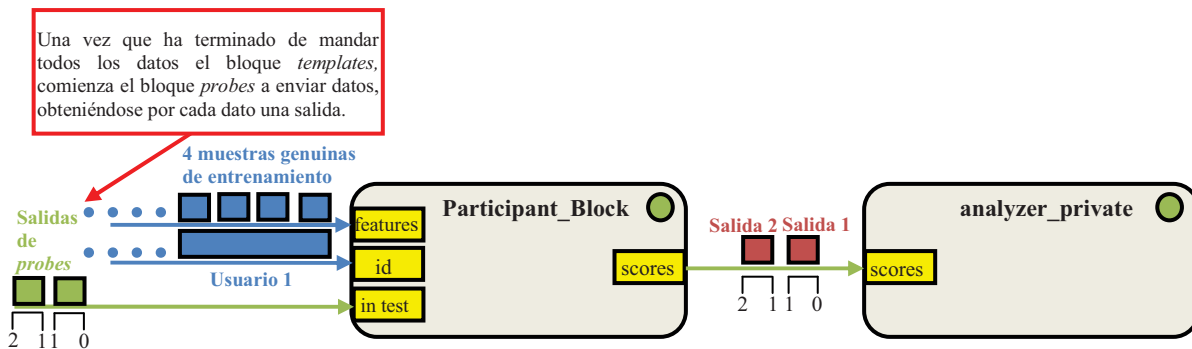


Fig. 2-9. Imagen explicativa de la sincronización y entrega de datos en los bloques de la competición

En la Fig. 2-9 se muestra un esquema de cómo funciona la sincronización en los bloques creados para la competición. Lo primero que hay que tener claro es el orden en que se ejecutan los bloques (representado en la Fig. 2-7). Primero será el bloque *templates*, segundo el bloque *probes*, después el bloque *Participant_Block* y por último *analyzer_private*. Por tanto, como vemos en la Fig. 2-9 la entrada *features* recibe las muestras de entrenamiento (4 muestras por usuario) y la entrada *id* recibe la etiqueta de usuario (durante las 4 muestras no cambiara para indicar que pertenecen a un usuario en concreto). Un dato importante es que la plataforma no tiene obligación de mandar los usuarios ordenados. Una vez que *templates* ha enviado todas las muestras de usuario, se empieza ejecutar *probes* enviando las muestras tipo test. Por cada muestra se obtiene una salida en el bloque *Participant_Block* que recibe el *analyzer_private*. El *analyzer_private* las procesa y obtiene un resultado representado en el campo experimentos.

Para realizar lo antes descrito es necesario un algoritmo en el bloque *Participant_Block*, así que KBOC proporciona un código de referencia para facilitar la participación. Dicho código puede ser modificado por los participantes añadiendo sus sistemas con simples conocimientos de Python (sin conocer las instrucciones especiales de BEAT).

El algoritmo se encuentra en el siguiente enlace¹⁵, completamente comentado para ayudar a los participantes.

Por tanto, la competición desarrolla dos algoritmos para llevar a cabo la competición:

- *Participant_Block*: Este algoritmo se encarga de simular el sistema de dinámica de tecleo propuesto.
 - o Entradas: *features* (dos parametros: int32 y string), *id* (string), *client_id* (string), *file* (uint64), *keystroke* (dos parametros: int32 y string)
 - o Salidas: *score_file* (explicado en *dataform*: string, float64 y string)
- *Analyzer_private*: Este algoritmo se encarga de ofrecer los resultados en base a parámetros de rendimiento populares.
 - o Entradas: *scores* (explicado en *dataform*: string, float64 y string)

El algoritmo *Participant_Block* consta de 5 entradas que se corresponde con los mismo datos de las salidas de las bases de datos. Este algoritmo primero obtiene las muestras de entrenamiento de cada usuario, creando una matriz por cada usuario con cuatro muestras.

¹⁵ https://www.beat-eu.org/platform/algorithms/robertodaza/KBOC16_PARTICIPANT_BLOCK/2/

Una vez que se tienen todas las muestras de entrenamiento, el bloque *probes* empieza a mandar las muestras test de cada usuario. Con cada muestra test y la matriz de entrenamiento del usuario en cuestión, se aplica un algoritmo de clasificación a través de una librería, obteniéndose una distancia final ($s_i = -d_i$). Hay que mencionar que se facilita una librería con tres algoritmos de clasificación (*baseline*) (ver sección 2.5.1). La salida llamada *score_file* manda al analizador la distancia final de cada usuario, el número de usuario y el número de muestra.

El algoritmo **analyzer_private** es ejecutable debido a que contiene información confidencial, como una matriz para indicar que muestras son genuinas o impostoras. Debido a este motivo la entrada *scores* del analizador contiene el número de usuario y de muestra para comprobar en la matriz si $s_i = -d_i$ corresponde a una muestra impostora o genuina, donde las filas son el número de usuario (1 a 300) y las columnas el número de la muestra (1 a 24).

Se crean dos vectores para distancias (*scores*, s_i), correspondientes a muestras genuinas y muestras impostoras. A los vectores se les aplica diversas funciones para obtener la eficiencia del sistema (ERR, FAR, FRR, ROC).

Además el analizador contiene partes de código que hacen la función de detectar errores o intentos de engaño en los datos enviados por los sistemas de los participantes. Por ejemplo, se calculan las distancias enviadas (6000 para 300 usuarios y 2000 para 100 usuarios, en caso de no ser estos valores se indica como error), se comprueba que las distancias no obtengan valores problemáticos (NaN y $\pm\infty$ producían un ERR igual a cero, por lo tanto se indicaban como error), etc.

- **Experiments:** Une todos los campos anteriores y nos facilita los resultados. Además es la fase de depuración (en caso de error en el código, formatos de datos, etc., se indica en esta sección).

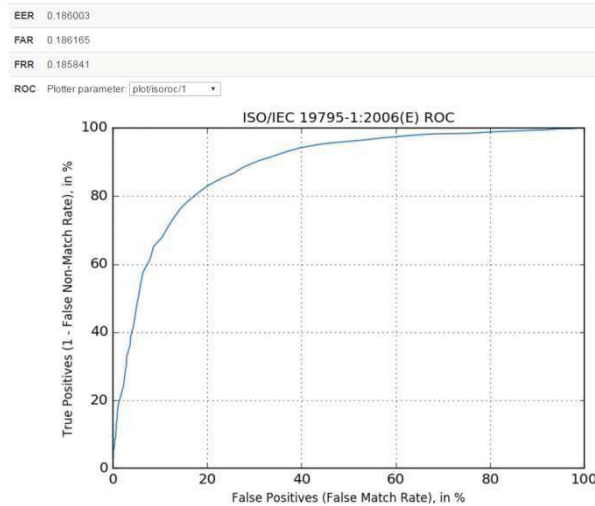


Fig. 2-10. Experimento facilitado por la competición, donde se muestran los resultados para la distancia Manhattan escalada modificada

- **Attestations:** Los certificados incluyen un experimento, haciendo que no pueda ser modificado y dan una confirmación de que los resultados se obtuvieron en la plataforma. Los certificados son muy ventajosos para las publicaciones científica (experimentación reproducible). La competición incluye un attestation realizado con uno de los algoritmos baseline.
- **Tabla de clasificación (Leaderboard):** Debido a que la competición KBOC no tiene fecha de finalización programada (permanecerá activa por tiempo indefinido), el

grupo de trabajo KBOC crea en la plataforma BEAT una tabla de clasificación, la cual se actualiza inmediatamente cada vez que se publica un nuevo sistema. La tabla de clasificación posiciona los sistemas según el ERR (cuanto menor ERR el sistema obtendrá una posición mejor).

Updated	Name	Databases/Protocols	EER	FRR	FAR
April 7, 2016	anjithgeorge/robertodaza/competition_kboc16/2/Kboc16_Compensation_ITKGP_ANJITH_APRIL7_20_D	kboc16/2@D	0.154502	0.154376	0.154629
April 7, 2016	anjithgeorge/robertodaza/competition_kboc16/2/Kboc16_Compensation_ITKGP_ANJITH_APRIL7_20_A	kboc16/2@A	0.165167	0.16518	0.165153
April 7, 2016	aythamimmi/robertodaza/competition_kboc16/2/Kboc16_Compensation_Modified_Scaled_Manhattan_Distance_Complete_Dataset	kboc16/2@A	0.18	0.180046	0.179953
April 7, 2016	anjithgeorge/robertodaza/competition_kboc16/2/Kboc16_Compensation_ITKGP_ANJITH_APRIL7_13	kboc16/2@A	0.188573	0.124876	0.25227

Fig. 2-11. Tabla de clasificación a 26 de Mayo del 2016

Por tanto, los pasos seguidos para realizar la competición en la plataforma BEAT son los siguientes:

- Primero se crea un diagrama de bloques especificando las entradas y salidas de cada bloque (se crea el *Toolchain* de la competición llamado *competition_kboc16*).
- Una vez creado el *Toolchain*, especificamos el formato de las entradas y salidas (en data formats, ej. *competition kboc16* para la entrada score, explicado en la Fig. 2-8) ¡Esto es muy importante!
- Creamos un algoritmo (Python) para cada bloque (*Participant_Block*, *analyzer_private*), especificando en cada entrada y salida el tipo de dato. Trabajamos simultaneamente con ‘*Data Formats*’ y ‘*Algorithms*’.
- Opcionalmente podemos usar *Libraries* para tener los algoritmos más limpios. En el caso de la competición se crea la librería *kboc16_baseline_matchers*.
- Creamos un experimento con los campos anteriores y obtenemos los resultados.

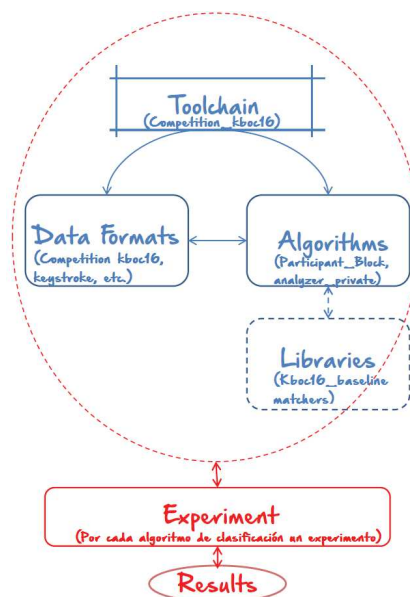


Fig. 2-12. Esquema de la competición en la plataforma BEAT

En la Fig. 2-12 se hace referencia a un esquema donde podemos ver como se utilizan los campos de BEAT para desarrollar la competición. BEAT nos ofrece muchas ventajas

que hacen necesaria utilizarla en la competición, consiguiendo con ello explotar el potencial de la plataforma como hemos podido ver. Se puede acceder a la plataforma de evaluación de tecnologías basadas en dinámica de tecleo a través de este enlace¹⁶, así como a los resultados de todos los sistemas evaluados a través de este otro¹⁷.

2.5.1 Algoritmos de clasificación (*baseline*)

En el código de la plataforma BEAT se suministra una librería¹⁸ con tres algoritmos de *baseline* [22]. La librería dispone de tres distancias populares en los algoritmos de reconocimiento basados en dinámica de tecleo. Se asume $\mathbf{f} = [f_1, f_2, \dots, f_M]$ como el vector de características de una muestra con identidad desconocida (con M características) y $\mathbf{g}^k = [g_1^k, g_2^k, \dots, g_M^k]$ $k \in 1, \dots, T$ el conjunto de entrenamiento de un usuario (en el caso de la competición $T = 4$). Las tres distancias dadas son:

- **Combinación Manhattan-Mahalanobis:**

$$d_1^k = \sum_{i=1}^M |\hat{f}_i - \hat{g}_i^k| \quad (2)$$

donde $\hat{\mathbf{f}} = \mathbf{S}^{-1/2} \mathbf{f}^T$ y $\hat{\mathbf{g}}^k = \mathbf{S}^{-1/2} (\mathbf{g}^k)^T$, \mathbf{S} es la covarianza y $(\cdot)^T$ es la transpuesta. La distancia d_1 es la obtenida para la mínima de las diferentes k .

- **Mahalanobis + vecino más cercano:** basada en la distancia de Mahalanobis y el algoritmo de vecino más cercano:

$$d_2^k = (\mathbf{f} - \mathbf{g}^k) \mathbf{S}^{-1} (\mathbf{f} - \mathbf{g}^k)^T \quad (3)$$

siendo \mathbf{S}^{-1} la inversa de la matriz de covarianzas de los datos de entrenamiento. La distancia final d_2 es obtenida como la mínima para las diferentes k .

- **Manhattan escalada modificada:** una modificación de la anterior definida como:

$$d_3^k = \sum_{i=1}^M |f_i - \bar{g}_i| / \sigma'_i \quad (4)$$

donde \bar{g} es la media de los datos de entrenamiento y $\sigma' = [\sigma'_1, \sigma'_2, \dots, \sigma'_M]$ es una modificación de la desviación estándar:

$$\sigma'_i = \begin{cases} \frac{0.2}{M} \sum_{j=1}^M \sigma_j & \text{si } \sigma_i < \frac{0.2}{M} \sum_{j=1}^M \sigma_j \\ \sigma_i & \text{resto} \end{cases} \quad (5)$$

¹⁶https://www.beat-eu.org/platform/experiments/robertodaza/robertodaza/competition_kboc16/2/Kboc16_Compensation_Baseline_Modified_Scaled_Manhattan_Distance/

¹⁷https://www.beat-eu.org/platform/search/aythamimm/KBOC16_COMPETITION_SEARCH/

¹⁸https://www.beat-eu.org/platform/libraries/robertodaza/kboc16_baseline_matchers/5/

y σ_i es la variación estándar de $\{\mathbf{g}^k\}_{k=1}^T$. Por otra parte $\bar{\mathbf{g}} = \frac{1}{T} \sum_{k=1}^T \mathbf{g}^k$. Esta simple modificación intenta paliar los efectos de las muestras con una escasa variación durante la normalización.

La puntuación final para las 3 distancias, s_i , se obtiene como la inversa de las mismas: $s_i = -d_i$.

2.6 Promover la participación

Una de las mayores dificultades para una competición es incentivar la participación de los investigadores. Como ya mencionamos en la sección 1.4, *FVC-onGoing competition* es una referencia en este sentido (más de 900 participantes).

Es por ello que la competición KBOC incentiva la participación de diversas maneras:

- Se crea una página web¹⁹ exclusivamente para la competición, lo que fomenta su mayor difusión.
- Hay dos formas de participación en la competición:
 - Participación *offline* (sección 2.3)
 - Participación *ongoing* (sección 2.3)

INTRODUCTION
1.2. Explanation data KBOC16 .

- To start in reproducible keystroke biometrics, look the article where you will learn the first steps.
- The example of the article is more difficult than the competition. You have more data (Name, Email, Nationality, ID and Family Name).
- To download the article click on the picture.
- For the competition KBOC16 the database consists:
 - 300 users with 24 samples per user
 - The time elapsed between consecutive key events (press and release) is provided as keystroke dynamics sequence.
 - There are two types of sequences (Genuine, Impostor)

Characteristics	#
Number of users (Testing Set)	300
Number of users (Development Set)	10
Number of sessions	4
Samples per user (including genuine and impostors)	24
Average separation between sessions	1 month

Example given in BEAT "Modified Scaled Manhattan distance" with EER 0.186003

KeyStroke Dynamics Recognition based on Personal Data:
A Comparative Experimental Evaluation Implementing Reproducible Research
Ryffana Morales¹, Mario Falanga², Julian Ferrer¹, Carlo Sansone², Javier Ortega-García¹
¹UVS, EPS, Universidad Autónoma de Madrid, C. Francisco Tomás y Valiente, 11, 28049 Madrid, Spain
²DIETI, University of Naples Federico II, Via Claudio 21, 80132 Naples, Italy
(ryffana.morales, julian.ferrer, javier.ortega)@uam.es, mario.falanga@di.unipi.it, carlo.sansone@unipi.it

Abstract
This work proposes a new benchmark for keystroke dynamics recognition on the basis of fully reproducible research. Instead of traditional authentication approaches based on complex passwords, we propose a novel biometric recognition based on typing patterns from personal data. We present a new dataset made up with the keystroke patterns of 61 users and 760 samples. The proposed approach estimates the similarity to recognize complex passwords considering that we have to recognize them by personal data (something that we are). The results encourage to further explore this new application scenario and the availability of data and source code represent a new valuable resource for the research community.

1. Introduction
Cyber security is a critical concern for governments, companies and end-users. The secure access has become a goal in a global society concerned by internet and Cyber Security is defined as the body of technologies, services and practices designed to protect computers, servers and data from damage, attack or unauthorized access [1]. Password-based authentication is one of the most popular secure access approaches. The security of these systems depends on the strength and the security policies adopted. A typical security policy includes recommendations as (i) to avoid simple passwords; (ii) to change passwords regularly; (iii) to use different passwords for different accounts, systems and applications; and (iv) to store passwords securely (i.e. to memorize your passwords). The number of users who comply with these security policies is low as they decrease the usability of the system. As an example, "PASSWORD" is a password generated by the automatic password generator of the EFFCASS (Credential Management System). The request to reset passwords because they have been forgotten are common. In this context, keystroke dynamics authentication systems have attracted the interest of both researchers and industry [2-5]. Keystroke biometrics are proposed to improve the security of traditionally password-based systems based on passwords or PIN numbers. Biometric recognition is commonly related to "something that users are" instead of "something that users have" such as passwords. In the case of biometric systems, the typical approach based on fixed password authentication considers complex passwords and one biometric dynamic biometric. The password acts as a primary security level and the user access is not allowed until the correct password is entered. The role of the biometric system is a secondary security level which try to detect intruders who are spoofing the identity of the legitimate user. With user and device biometric measurements on the primary and only security level. It is possible a relative recognition by replicating the way the specific type complex passwords (something that we have) by the way they type these users (something that we are). The work requires biometric dynamic authentication based on personal data as opposed to complex passwords. The idea underlying this work is that help us to design complete password, we will never forget our family or group name. This authentication approach try to improve the security of the access as well as its usability. The main advantage of this biometric recognition based on personal data is the learning curve [4] is minimized because personal data is information that can be accessed by users. This usability is improved by estimating the similarity of complex passwords which comply with the security policies but are difficult to memorize. On the other hand, the main disadvantage of this approach use of small amount of information available to recognize per users (short words) and the learning curve of the application is also minimized when the information is based on common names or simple words instead of complex passwords. The main contribution of this work can be summarized as follows: (i) a new dataset with keystroke dynamics of 61 users acquired on the basis of a real application context; (ii) user results in keystroke dynamics recognition on personal data and (iii) a fully reproducible and public available benchmark including the dataset and the source code necessary to reproduce all the results reported in this work.
https://sites.google.com/view/keystroke_biometric

TOOLCHAINS
2.1. Concept.

Click the picture to go to the 'toolchain'

The blocks 'templates' and 'probes' are databases. In 'templates' are the first 4 samples of the users. In 'probe' you have 20 samples, can be impostor or Genuine. 'template_id' is always the user number, e.g., '2'. The databases do not send users ordered, e.g., '4' after '18'. 'Keystroke' are the times for a sample.

This block works as 'scoring' and 'enroll' with an algorithm. We can see two different colors for the blocks, blue and green. This is very important! The colors indicate synchronization and it is chosen in 'toolchains'. The blue runs before the green.

The analyzer is a block with only inputs. You can see the outputs at screen. In competition you can not change and will not be public.

Fig. 2-13. Imágenes del tutorial avanzado interactivo creado para la plataforma BEAT

- Se proporcionan códigos de referencia para ayudar a los participantes en sus primeros pasos en la competición:
 - A los participantes *offline* se les proporciona un código en Matlab, creado exclusivamente para la competición. Este material se puede solicitar desde la página web de la competición.

¹⁹ <https://sites.google.com/site/btas16kbcoc/>

- A los participantes *ongoing* se les suministra un código²⁰ creado en la plataforma BEAT exclusivamente para la competición (completamente comentado para una mayor comprensión para los participantes)

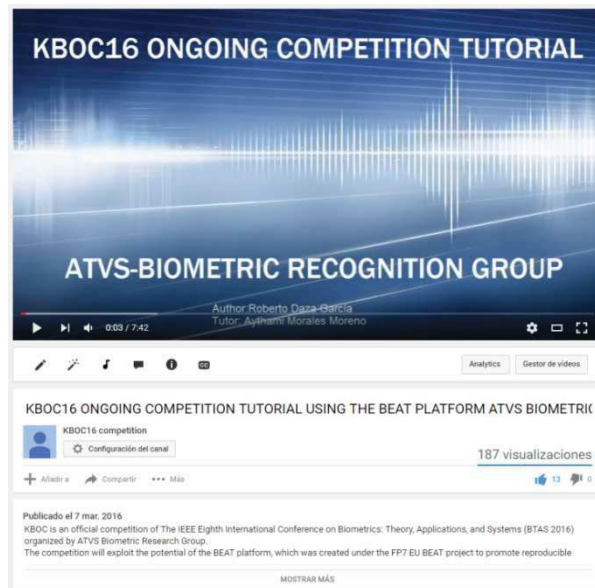


Fig. 2-14. Imagen del video tutorial creado para la competición

Se suministra diverso material de la plataforma BEAT (este material está disponible en la página web de la competición) para un fácil aprendizaje:

- **Un tutorial avanzado de la plataforma BEAT** (sección 2.5): Nuestro grupo de trabajo realiza un tutorial avanzado (interactivo) de la plataforma BEAT, enseñando como participar en KBOC y los ejemplos creados para la competición. La Fig. 2-13 muestra dos páginas del tutorial.
- **Un tutorial simple**²¹: Se realiza un tutorial menos complejo y más orientado a la participación en la competición (solo aparecen los campos fundamentales para participar).
- **Videos tutoriales**²²: Se crean videos exclusivamente para la competición, completamente subtítulos en inglés, mostrando como participar a través de BEAT. La Fig. 2-14 muestra una captura del video tutorial creado para la competición.

- La participación en la competición no implica la publicación del código para los participantes (la organización no tiene acceso al código).
- Las participantes pueden permanecer en el anonimato si lo desean, solo es obligatorio una descripción del sistema presentado.

²⁰ https://www.beat-eu.org/platform/algorithms/robertodaza/KBOC16_PARTICIPANT_BLOCK/2/#viewer

²¹ <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbxidGFzMTZrYm99jGd4OjNIZGFINzMOjNGFjODAyNjI>

²² <https://www.youtube.com/watch?v=lcZE6XvwGxw&feature=youtu.be>

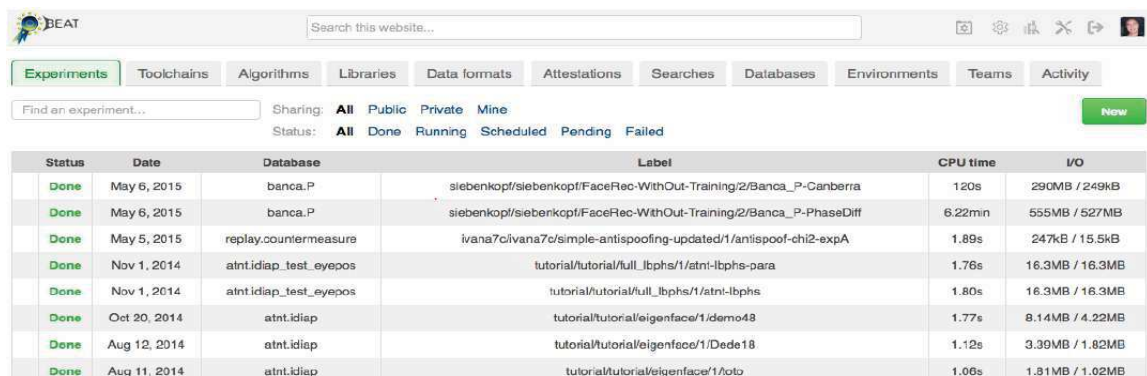
- Se difunde la competición entre los grupos de investigación en biometría más prestigiosos del mundo, además de divulgarse por otras instituciones más alejadas a la biometría (Universidad de Granada, etc.).
- El intervalo de tiempo en el que permanece activa la competición offline comprende desde enero a abril de 2016. La plataforma *ongoing* estará disponible para cualquier investigador por tiempo indefinido.

2.7 Ayuda a la mejora de la plataforma BEAT

La plataforma BEAT es un proyecto muy reciente (finalizado en Febrero de 2014). Cuando empezamos a trabajar en BEAT todavía se encontraba en fase de prueba (beta). Además, como hemos hecho mención en el apartado 2.4, el grupo de investigación Área de Tratamiento de Voz y Señales (ATVS) de la Universidad Autónoma de Madrid colabora con BEAT. Por este motivo hemos ofrecido ayuda en todo lo posible a la plataforma BEAT, una continua comunicación ha existido entre los desarrolladores de BEAT y nuestro grupo de trabajo.

Hay que mencionar que la plataforma BEAT presentaba una serie de inconvenientes en los que hemos colaborado para su correcta solución. Uno de los más destacados, consistía en que los experimentos al hacerlos públicos también hacían público el código, aunque previamente se hubiese marcado que siguiera privado. También existían problemas con la opción ejecutable que convertía en públicos los algoritmos. Estos errores, como puede constatarse, son un grave problema para una plataforma centrada en la experimentación reproducible, aunque han sido corregidos rápidamente una vez que los comunicamos.

La plataforma BEAT, al utilizar diagramas de bloques y sincronizar el flujo de datos (ausencia de bucles) puede resultar compleja por su innovación, más cuando no existe suficiente material de apoyo (tutorial avanzado). Unido a que la plataforma no ofrecía una depuración de los algoritmos (no indica los errores en el código) hacían muy complejo trabajar en la misma. Debido a estas razones se mantuvo un contacto permanente con los desarrolladores de BEAT para obtener más información y crear un tutorial para la competición. Hay que mencionar que BEAT ha subsanado estos inconvenientes ofreciendo una depuración del código y un tutorial avanzado.



Status	Date	Database	Label	CPU time	I/O
Done	May 6, 2015	banca.P	siebenkopt/siebenkopt/FaceRec-WithOut-Training/2/Banca_P-Canberra	120s	290MB / 249KB
Done	May 6, 2015	banca.P	siebenkopt/siebenkopt/FaceRec-WithOut-Training/2/Banca_P-PhaseDiff	6.22min	555MB / 527MB
Done	May 5, 2015	replay.countermeasure	ivana7/civana7c/simple-antispooofing-updated/1/antispooof-chi2-expA	1.89s	247KB / 15.5KB
Done	Nov 1, 2014	atnt.idiap_test_eyepos	tutorial/tutorial/full_lbphs/1/atnt-lbphs-para	1.76s	16.3MB / 16.3MB
Done	Nov 1, 2014	atnt.idiap_test_eyepos	tutorial/tutorial/full_lbphs/1/atnt-lbphs	1.80s	16.3MB / 16.3MB
Done	Oct 20, 2014	atnt.idiap	tutorial/tutorial/eigenface/1/demo48	1.77s	8.14MB / 4.22MB
Done	Aug 12, 2014	atnt.idiap	tutorial/tutorial/eigenface/1/Dede18	1.12s	3.39MB / 1.82MB
Done	Aug 11, 2014	atnt.idiap	tutorial/tutorial/eigenface/1/oto	1.06s	1.81MB / 1.02MB

Fig. 2-15. Interfaz antigua de la plataforma BEAT

También se dieron sugerencias para hacer más intuitiva la interfaz. En la Fig. 2-15 se muestra la interfaz antigua de BEAT, la cual era menos intuitiva.

En la actualidad BEAT ha corregido todos los errores del pasado y ha mejorado su interfaz, siendo una plataforma mucho más asequible y con una infinidad de posibilidades.

3 Resultados

3.1 Resultados obtenidos con los sistemas baseline

Antes de exponer los resultados obtenidos por los participantes, se mostrarán los resultados de los 3 *baseline* (Algoritmos de clasificación) facilitados por la competición.

Tabla 3-1: Resultados obtenidos para las tres distancias en BEAT.

Resultados	d_1	d_2	d_3
ERR	29.60%	27.35%	18.60%

En la Tabla 3-1 se facilitan los resultados para las tres distancias, se omite FAR y FRR debido a su similitud y falta de relevancia.

Hay que mencionar que los resultados son fácilmente mejorables, debido a que algunas de las muestras incluyen errores en teclas de pulsaciones adicionales, pero estas teclas no cambian el texto (principalmente la tecla shift). Ante esta situación se tomó una decisión muy sencilla y fácil de mejorar: tomar como referencia la primera muestra y a las sucesivas igualarlas al tamaño de la primera muestra (recortando o añadiendo ceros). Mencionar que para obtener estos resultados tampoco se hizo uso de la información referente a las teclas pulsadas, la cual puede ser de gran ayuda. Por tanto, para obtener los resultados se aplicó directamente al tiempo de pulsar y soltar las distancias anteriormente descritas.

Como podemos ver en la Tabla 3-1, la distancia d_3 tiene un EER de 0.186, siendo esta distancia la que ofrece un menor ERR²³. Esto era de esperar vistas publicaciones anteriores, donde se podía ver el alto rendimiento [22]. En cambio, con la distancia d_1 obtenemos los peores resultados respecto a este valor.

3.2 Descripción de los sistemas presentados por los participantes

Un total de 12 instituciones de 7 países diferentes (5 de Estados Unidos, 2 de India y 1 de Noruega, Argelia, Los Países Bajos, Brasil y China) se registraron en la competición. Finalmente 4 instituciones presentaron sus sistemas, habiendo un total de 37 diferentes sistemas presentados. En esta sección se describen los mejores sistemas presentados en la competición KBOC.

- **U.S. Army Research Laboratory (ARL):** Un total de 15 sistemas fueron presentados por ARL, 6 de los sistemas utilizaban la distancia Manhattan con un detector de anomalías utilizando el tiempo de latencia entre Presión-Presión (PP) y tiempo de espera (TE, *Hold time*). Normalmente la distancia Manhattan usada en los detectores de anomalías [29] ofrece muy buenos resultados con baja ERR, aunque

²³ <https://www.beat-eu.org/platform/attestations/1495360336/%23results>
(resultados y código de la distancia d_3)

también puede atribuirse a un pre-procesado de los datos, normalización de los datos y normalización de las distancias finales ($s_i, scores$).

Como hemos mencionado en la anterior sección, algunas muestras incluyen errores en teclas de pulsaciones adicionales (la competición calcula que en torno al 10%). ARL plantea un simple algoritmo para solucionar esta situación. Primero selecciona la muestra de entrenamiento con longitud mínima (muestra diana) y en el caso de existir varias que difieran en solo un carácter se elige al azar. Una vez elegida se aplica una modificación Dynamic Time Warping (DTW) [38] para igualar las otras muestras a la longitud M de la muestra diana (tanto de entrenamiento como de test).

Se extrajeron las M-1 latencias PP y M tiempos de espera (*key-hold*). Se aplica una normalización a los tiempos de espera y latencias PP. Se normalizan por la Desviación Típica (DT) y la media de las muestras de entrenamiento.

Una vez extraídas las características se calcula la distancia Manhattan. Las distancias obtenidas para las muestras tipo test se normalizaron por $\pm 2DT$ y la media.

Este proceso para las diez muestras dadas de ejemplo en la competición *offline* obtienen un ERR $6.5 \pm 1.8\%$. Siguiendo los criterios presentados en este trabajo de reproducibilidad, el código es facilitado en el siguiente enlace²⁴.

- **Universidade Federal de Sergipe (UFS):** UFS se autoimpone tres restricciones con el fin de simular un sistema biométrico real:
 - R1: No se conoce el número de muestras tipo test por usuario.
 - R2: La proporción en las muestras tipo test de genuinas y impostoras es desconocida.
 - R3: No se puede usar muestras tipo test futuras para mejorar resultados anteriores.

Como en [39], PP y tiempo de espera se ecualizan con los parámetros $\mu_{PP} = -1.61$, $\sigma_{PP} = 0.64$, $\mu_H = -2.46$ y $\sigma_H = 0.33$ respectivamente, a través de un mapeo no lineal:

$$y = \frac{1}{1 + \exp\left(-\frac{1.7(\log_e(x) - \mu)}{\sigma}\right)} \quad (6)$$

donde x representa un intervalo de tiempo en segundos.

Para cada usuario se tomaron un conjunto de vectores PP y TE. Se calcula la distancia Manhattan para cada vector obteniéndose d_{PP} y d_H . La distancia final será la suma de las anteriores dándole más peso d_{PP} .

$$d = 0.75d_{PP} + 0.25d_H \quad (7)$$

Las muestras tipo test que obtenían una puntuación inferior a 0,14 eran añadidas a las muestras de entrenamiento, influyendo en los resultados futuros para las siguientes muestras. Además también realiza un pre-procesado.

UFS destaca que se pueden conseguir resultados mejores sin poner las restricciones anteriores, pero los sistemas no serían realistas. Por ejemplo, mediante el uso de las

²⁴ <http://vmonaco.com/software/>

10 muestras etiquetadas obtuvieron un ERR $8,0\% \pm 1\%$, sin embargo si no cumplían la R3 obtenían un ERR $6.5\% \pm 1\%$.

- **Indian Institute of Technology Kharagpur (IIT):** IIT presenta un total de 5 sistemas. Utiliza un vector concatenado con las características de TE y latencia AP. No realiza ningún tipo de pre-procesado. Se asume $\mathbf{f} = [f_1, f_2, \dots, f_M]$ como el vector de características de una muestra con identidad desconocida (con M características) y $\mathbf{g}^k = [g_1^k, g_2^k, \dots, g_M^k]$ $k \in 1, \dots, T$ el conjunto de entrenamiento de un usuario.

Obtiene la distancia de la siguiente manera:

$$\mathbf{AD}(i, k) = |g_i^k - f_i|, \quad k = 1, \dots, T \text{ and } i = 1, 2, 3, \dots, M \quad (8)$$

$$\mathbf{md}(i) = \min_{k \in [1, \dots, T]} ad_{ik}, \quad i = 1, 2, 3, \dots, M \quad (9)$$

Donde ad_{ik} es un elemento de la matriz $\mathbf{AD}(i, k)$, la distancia final se obtiene de:

$$d = \text{mean}(\mathbf{md}) + \text{median}(\mathbf{md}) \quad (10)$$

3.2.1 Resultados de la competición *offline*

En esta sección se presentan los resultados de la competición *offline*. Los resultados de la competición ongoing se pueden ver desde BEAT. Se obtienen los siguientes indicadores como resultados:

- **Tasa de error igual global (ERR_G):** Es el ERR utilizando tanto en las muestras genuinas como impostoras, con un único umbral (*score*) para todos los usuarios.
- **Tasa de error igual para cada usuario (ERR_U):** Se calcula un umbral diferente para cada usuario (300 umbrales), obteniendo un ERR diferente para cada usuario. ERR_U es la media de todos los ERR.
- **FMR₁₀₀:** Es el valor que obtiene la tasa de falso rechazo cuando la tasa de falsa aceptación es igual 1%.
- **Gráfica *Detection-Error Tradeoff* (DET):** Permite una fácil comparación de los sistemas para cualquier umbral de decisión. En el eje "x" se representa la tasa de falsa aceptación y en el eje "y" la tasa de falso rechazo.

En la siguiente página se muestran los resultados obtenidos para los participantes con la base de datos de los 300 usuarios. En la tabla 3-2 se representan las características de los sistemas presentados. La tabla 3-3 muestra los resultados obtenidos para los mejores sistemas presentados por las entidades participantes, con todas las sesiones (sesión 1 son las muestras de entrenamiento y sesiones 2, 3 y 4 muestras de test). Expone como la entidad P4 presenta el mejor sistema, además de ofrecer una FMR₁₀₀ (este valor puede ser muy útil para un banco que necesita mucha seguridad) muy superior a los demás sistemas presentados. El ERR_G oscila entre 17,90% y 5,32% y el ERR_U entre 13,66% y 4,72%. Los resultados plasman la importancia de la normalización. También muestran la importancia del procesado, estimándose un máximo de mejoría entorno al 10% en la competición KBOC.

La tabla 3-4 incluye el rendimiento con ERR_G, utilizando las muestras de la sesión 2 o las muestras de la sesión 4. Se aprecia una degradación de los sistemas con el tiempo, aunque el sistema de la entidad P4 es muy estable (sesión 2 un ERR_G de 5,09% y sesión 4

un ERR_G de 5,10%). La tabla 3-5 muestra el rendimiento (ERR_G) de todos los sistemas presentados.

La Fig. 3-1 muestra las curvas DET para todos los sistemas (Fig. 3-1, izquierda) y los mejores sistemas de cada entidad representados utilizando las muestras de la sesión 2 o las muestras de la sesión 4. En las gráficas se aprecia que los sistemas enviados por una misma entidad difieren muy poco, pareciendo más bien pequeños cambios de un sistema inicial.

Tabla 3-2: Resumen de las características de los mejores sistemas presentados en la competición.

Participantes	Pre procesado	Carac.	Carac. norm.	Distancia	Norm. score
P1- <i>Indian Institute of Technology Kharagpur</i>	NO	TE+AP	NO	Combinada	NO
P2 - <i>Federal University of Sergipe</i>	SI	TE+PP	NO	Manhattan	NO
P3 - Participante anónimo	NO	AP	NO	Kendall's tau	NO
P4 - <i>U.S. Army Research Laboratory</i>	SI	TE+AP	SI	Manhattan	SI

Tabla 3-3: Resultados finales para la competición. Obtenidos con las muestras de entrenamiento (sesión 1) y las tres sesiones (todas las muestras tipo test).

ID	ERR_G	ERR_U	FMR_{100}
P1	15,73%	11,95%	51,13%
P2	11,82%	7,96%	54,65%
P3	17,90%	13,66%	64,60%
P4	5,32%	4,72%	28,36%

Tabla 3-4: Resultados obtenidos con ERR_G . Usando las muestras de entrenamiento (sesión 1) y las muestras test de la sesión uno o sesión 4.

ID	Segunda sesión	Cuarta sesión
P1	15,28%	16,13%
P2	11,60%	11,96%
P3	17,01%	18,21%
P4	5,09%	5,10%

Tabla 3-5: EER_G obtenido para todos los sistemas presentados. Obtenido con las muestras de entrenamiento (sesión 1) y las tres sesiones (todas las muestras tipo test).

Sistemas	P1	P2	P3	P4
1	19,33%	12,9%	17,9%	7,82%
2	16,82%	11,85%		6,46%
3	16,52%	12,12%		7,32%
4	16,47%	13,48%		7,35%
5	15,73%	12,25%		8,02%
6		13,92%		5,32%
7		11,82%		7,95%
8		13,03%		8,08%
9		13,03%		5,68%
10		14,66%		5,91%
11				10,35%
12				10,89%
13				11,2%
14				11,23%
15				6,26%

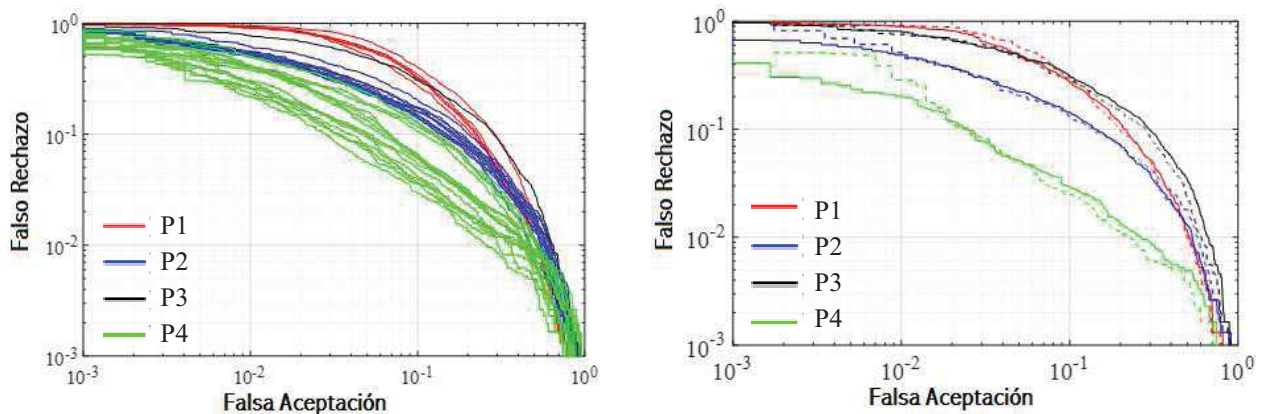


Fig. 3-1. Izquierda: Curva DET con todas las sesiones y sistemas presentados; Derecha: Mejor sistema enviado por cada entidad, dibujado para la sesión 2 (línea discontinua) y para la sesión 4 (línea continua)

4 Conclusión

Desde nuestro punto de vista las publicaciones deben facilitar la reproducibilidad de los trabajos. Esto sin duda aporta numerosas ventajas a la comunidad científica: un ahorro de tiempo para los investigadores, experimentos más claros y organizados, evita la falsificación de resultados, detección de errores en resultados presentados, etc. Pero no solo tiene un beneficio para la comunidad científica, también para el propio autor, pues diversos estudios corroboran que las publicaciones que facilitan la reproducibilidad son más citados (hasta 7 veces más) que los que no. Además, facilitan que el autor pueda retomar su trabajo con sencillez al tenerlo correctamente organizado.

Por tanto, en el área de la biometría que es una ciencia muy reciente, es esencial impulsar la experimentación reproducible para avanzar más rápido (sin cometer los errores del pasado), al mismo tiempo que facilita poder compartir bases de datos que mejoran los sistemas y los hacen más realistas (teniendo una base común donde poder compararlos).

Creemos que la plataforma BEAT es una gran oportunidad para la comunidad científica, permitiendo la experimentación reproducible y obteniendo muchas otras ventajas (ej. certificación de los resultados).

En lo que respecta al ámbito de reconocimiento de usuarios basado en la dinámica de tecleo, es fundamental una competición (KBOC) que impulse la experimentación reproducible, que permita comparar sistemas de pulsaciones de teclas y que suministre una base de datos con suficientes participantes y de forma realista (dos manos).

KBOC es una competición completamente innovadora en el ámbito de la dinámica de tecleo. Primero, ofrece una base común donde poder comparar los sistemas (una de las mayores bases de datos de usuarios), tanto para entidades académicas como comerciales.

Segundo, ofrece dos tipos de participaciones (participación *ongoing* y participación *offline*), cada una con una función específica.

La participación *offline* sirve de referencia para la participación *ongoing*, obteniéndose resultados alentadores (37 sistemas presentados). El mejor sistema presentado consigue un EER_G de 5,32%.

Con la competición *ongoing* conseguimos impulsar la experimentación reproducible mediante la plataforma BEAT. Destacar que los resultados se actualizan inmediatamente en la plataforma. Además, destacar el abundante material facilitado por la competición (tutoriales, video tutoriales, algoritmos, toolchain, etc.), unido a los sistemas presentados en la participación *offline*, los cuales sirven de referencia. Todo esto posibilita que la competición *ongoing* sea una de las mejores propuestas en el área de la biometría, tanto para iniciarse en dinámica de tecleo, como en la plataforma BEAT, a la vez que sirve de base común para comparar los sistemas de dinámica de tecleo.

Debido a que la competición KBOC no tiene fecha de finalización programada (permanecerá activa por tiempo indefinido), pensamos que es una enorme oportunidad para comparar nuestros sistemas con los de los demás investigadores y así poder aprender unos de otros y avanzar de forma conjunta en el área.

Referencias

- [1] H. Darmon, F. Diamond and R. Taylor, "Fermat's last theorem," *Current Developments in Mathematics*, vol. 1, pp. 1-154, 1995.
- [2] P. Vandewalle, J. Kovačević and M. Vetterli, "Reproducible research in signal processing," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 37-47, 2009.
- [3] R. Descartes, *Discurso del método*, Ediciones Colihue SRL, 2004.
- [4] J. Tudela and J. Aznar, "¿Publicar y/o morir? El fraude en la investigación y las publicaciones científicas," *Persona y Bioética*, vol. 17, no. 1, pp. 12-27, 2013.
- [5] F. C. Fang, R. G. Steen and A. Casadevall, "Misconduct accounts for the majority of retracted scientific publications," *Proceedings of the National Academy of Sciences*, vol. 109, no. 42, pp. 17028-17033, 2012.
- [6] Z. Zou, L. F. Horowitz, J. P. Montmayeur, S. Snapper and L. B. Buck, "Genetic tracing reveals a stereotyped sensory map in the olfactory cortex," *Nature*, vol. 452, no. 7183, pp. 120-120, 2008.
- [7] A. Beloqui et al., "Reactome array: forging a link between metabolome and genome," *Science*, vol. 326, no. 5950, pp. 252-257, 2009.
- [8] W. S. Hwang et al., "Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst," *Science*, vol. 303, no. 5664, pp. 1669-1674, 2004.
- [9] S. Lawrence, "Online or invisible," *Nature*, vol. 411, no. 6837, p. 521, 2001.
- [10] S. Hitchcock, "The effect of open access and downloads ('hits') on citation impact: a bibliography of studies," 2004.
- [11] H. A. Piwowar, R. S. Day and D. B. Fridsma, "Sharing detailed research data is associated with increased citation rate," *PloS one*, vol. 2, no. 3, p. e308, 2007.
- [12] A. Morales, "Estrategias para la identificación de personas mediante biometría de la mano sin contacto," 2011.
- [13] A. Bertillon, "Instructions signalétiques: identification anthropométrique," *Imprimerie administrative*, 1893.
- [14] A. Jain, P. Flynn and A. A. Ross, "Handbook of biometrics," *Springer Science & Business Media*, 2007.
- [15] M. Tapiador and J. A. Sigüenza, *Tecnologías biométricas aplicadas a la seguridad*, RAMA, 2005.
- [16] A. Jain, A. Ross and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4-20, 2004.
- [17] P. J. Phillips, A. Martin, C. L. Wilson and M. Przybocki, "An introduction evaluating biometric systems," *Computer*, vol. 33, no. 2, pp. 56-63, 2000.
- [18] International Biometric Group, "Biometrics Market and Industry Report 2007-2012," 2007.
- [19] International Biometric Group, "Biometrics Market and Industry Report 2010-2015," 2010.
- [20] T. Matsumoto, H. Matsumoto, K. Yamada and S. Hoshino, "Impact of artificial gummy fingers on fingerprint systems," in *Optical Security and Counterfeit Deterrence Techniques IV*, San Jose, 2002.

- [21] H. Saevanee and P. Bhattarakosol, "Authenticating user using keystroke dynamics and finger pressure," in *IEEE 6th Conference on Consumer Communications and Networking*, Las Vegas, 2009.
- [22] A. Morales, M. Falanga, J. Fierrez, C. Sansone and J. Ortega-Garcia, "Keystroke dynamics recognition based on personal data: A comparative experimental evaluation implementing reproducible research," in *IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Arlington, 2015.
- [23] BehavioSec, "BehavioSec," 28 Abril 2016. [Online]. Available: <https://www.behaviosec.com/>. [Accessed 10 Mayo 2016].
- [24] R. Moskovitch et al., "Identity theft, computers and behavioral biometrics," in *IEEE International Conference on Intelligence and Security Informatics*, Dallas, 2009.
- [25] G. E. Forsen, M. R. Nelson and R. J. Staron Jr, "Personal attributes authentication techniques," *Pattern Analysis and Recognition Corp Rome Ny*, 1977.
- [26] R. Spillane, "Keyboard apparatus for personal identification," *IBM Technical Disclosure Bulletin*, 1975.
- [27] R. S. Gaines, W. Lisowski, S. J. Press and N. Shapiro, "Authentication by keystroke timing: Some preliminary results," *DTIC Document*, 1980.
- [28] A. Peacock, X. Ke and M. Wilkerson, "Typing patterns: A key to user identification," *IEEE Security & Privacy*, vol. 2, no. 5, pp. 40-47, 2004.
- [29] K. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *IEEE/IFIP International Conference on Dependable Systems and Networks*, Lisbon, 2009.
- [30] R. Giot, M. El-Abed and C. Rosenberger, "GREYC keystroke: A benchmark for keystroke dynamics biometric systems," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, Washington DC, 2009.
- [31] C. C. Loy, W. K. Lai and C. P. Lim, "Keystroke patterns classification using the ARTMAP-FD neural network," in *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kaohsiung, 2007.
- [32] E. Vural, J. Huang, D. Hou and S. Schuckers, "Shared research dataset to support development of keystroke authentication," in *IEEE International Joint Conference on Biometrics*, Clearwater, 2014.
- [33] Y. Li et al., "Study on the BeiHang keystroke dynamics database," in *International Joint Conference on Biometrics*, Washington, 2011.
- [34] S. Marcel, "BEAT—biometrics evaluation and testing," *Biometric Technology Today*, vol. 2013, no. 1, pp. 5-7, 2013.
- [35] J. Monaco et al., "One-handed Keystroke Biometric Identification Competition," in *International Conference on Biometrics (ICB)*, Phuket, 2015.
- [36] R. Cappelli, M. Ferrara, D. Maltoni and F. Turrone, "Fingerprint verification competition at IJCB2011," in *International Joint Conference on Biometrics (IJCB)*, Washington DC, 2011.
- [37] J. Fierrez et al., "BiosecrID: a multimodal biometric database," *Pattern Analysis and Applications*, vol. 13, no. 2, pp. 235-246, 2010.
- [38] M. Müller, *Information retrieval for music and motion*, Berlin: Springer, 2007.
- [39] J. Montalvão, E. O. Freire, M. A. Bezerra Jr and R. Garcia, "Contributions to empirical analysis of keystroke dynamics in passwords," *Pattern Recognition Letters*, vol. 52, pp. 80-86, 2015.

Glosario

AFIS	Automated Fingerprint Identification System
ATVS	Área de Tratamiento de Voz y Señales
BEAT	Biometrics Evaluation and Testing
BTAS	Biometrics: Theory, Applications, and Systems
CMU	Carneige Mellon University
DET	Detection Error Trade-off
DTW	Dynamic Time Warping
EER	Equal Error Rate
EEUU	Estados Unidos
EPFL	École Polytechnique Fédérale de Lausanne
EU	Unión Europea
FA	False Acceptance
FR	False Rejection
IBG	International Biometric Group
KBOC	Keystroke Biometric Ongoing Competition
Latencia AA	Latencia Alzado-Alzado
Latencia AP	Latencia Alzado-Presión
Latencia PA	Latencia Presión-Alzado
Latencia PP	Latencia Presión-Presión
ROC	Receiver Operating Characteristic
TE	Tiempo de Espera
UAM	Universidad Autónoma de Madrid

Anexos

A continuación como anexo al presente trabajo fin de grado, se adjuntan dos artículos enviados a congresos que resumen las principales aportaciones a este TFG:

- Aythami Morales, Julian Fierrez, Marta Gomez-Barrero, Javier Ortega-Garcia, Roberto Daza, Jhon V. Monaco, Jugurta Montalvão, Jânio Canuto, Anjith George, "KBOC: Keystroke Biometrics OnGoing Competition", *IEEE Eighth International Conference on Biometrics: Theory, Applications, and Systems* (BTAS 2016), **Under Review**.
- Roberto Daza, Aythami Morales, Julian Fierrez, Marta Gomez-Barrero y Javier Ortega-Garcia, "KBOC: Plataforma de Evaluación de Tecnologías de Reconocimiento Biométrico Basadas en Dinámica de Tecleo", *XXXI Simposium Nacional de la Unión Científica Internacional de Radio* (URSI 2016), **Under Review**.

KBOC: Keystroke Biometrics OnGoing Competition

Aythami Morales¹, Julian Fierrez¹, Marta Gomez-Barrero¹, Javier Ortega-Garcia¹, Roberto Daza¹,
Jhon V. Monaco², Jugurta Montalvão³, Jânio Canuto³, Anjith George⁴

¹ATVS-Universidad Autonoma de Madrid, C\ Francisco Tomás y Valiente, 11, 28049, Madrid, Spain

²U.S. Army Research Laboratory, Aberdeen Proving Ground, USA

³UFS - Universidade Federal de Sergipe, 49100-000, São Cristóvão, Brazil

⁴Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India

{aythami.morales, julian.fierrez, marta.barrero, javier.ortega}@uam.es, john.v.monaco2.ctr@mail.mil,
jmONTALVAO@ufs.br, anjith2006@gmail.com

Abstract

This paper presents the first Keystroke Biometrics Ongoing evaluation platform and a parallel Competition (KBOC) organized to promote reproducible research and establish a baseline in person authentication using keystroke biometrics. The ongoing evaluation tool has been developed using the BEAT platform and includes keystroke sequences (fixed-text) from 300 users acquired in 4 different sessions under realistic conditions. In addition, the results of a parallel offline competition based on the same data and evaluation protocol are presented. The results reported have achieved EERs as low as 5.32%, which represent a challenging baseline for keystroke recognition technologies to be evaluated on the new publicly available KBOC platform.

1. Introduction

Biometric recognition is a wide research area which includes researchers from pattern recognition and machine learning communities. Biometric technologies are usually divided into physiological (e.g. fingerprint, face, iris) and behavioral (e.g. signature, gait, keystroke) according to the nature of the biometric trait used. Behavioral biometrics have attracted the interest of researchers and industry because of its ease of use, transparency and large number of potential applications.

Keystroke biometric applications have been investigated over the past several decades, attracting both academics and practitioners. These technologies present several challenges associated to modeling and matching dynamic sequences with high intra-class variability (e.g. human behavior is strongly user-dependent and varies significantly between subjects). In addition, the simple nature of the data (time sequences) makes keystroke biometrics a good field to

introduce new researchers (without previous experience on biometric applications) in this challenging area.

From the industry's point of view keystroke technologies offer authentication systems capable of improving the security and trustworthiness of web services (e.g. banking, mail), digital contents (e.g. databases) or new devices (e.g. smartphones, tablets). The keystroke recognition community is heterogeneous and includes researchers from different disciplines [1]. The number of algorithms and approaches is large and it is difficult to establish a baseline. As a behavioral biometric trait, the performance of keystroke biometrics systems is strongly dependent on the application (e.g. fixed or free text) and databases (e.g. different users show very different performances). Public benchmarks have been proposed, offering the opportunity to compare systems under the same conditions. Some of the most popular keystroke benchmarks are CMU [2] and GREYC [3]. Even though these benchmarks represent valuable resources, they suffer from two important limitations: the small number of subjects (51 and 133 for CMU and GREYC respectively) and their application scenario, which assumes that all users share the same password ("tie5Roan" and "greyc laboratory"). In real applications, the most probable scenario is the one in which different users have different passwords.

To the best of our knowledge, there is only one previous keystroke recognition competition that was hosted during the IAPR International Conference on Biometrics 2015: "One-handed Keystroke Biometric Identification Competition" [4]. In this competition, keystroke technologies were evaluated in a free-text scenario involving the response of 63 students to three online exams. The competition analyzed the performance of person authentication algorithms under challenging conditions, in which users were forced to type using only one hand instead of a more natural way, using two hands.

Traditional biometric competitions give a static snapshot of the state-of-the-art in a specific research area. The main problem is how to encourage researchers to invest their resources and time to participate in these competitions (usually operative during a short window of time). Without the participation of the main players, the snapshot will be inaccurate. In contrast, the ongoing competitions provide a dynamic view constantly updated by the community. The FVC-onGoing competition [6] is a successful example with more than 900 participants and more than 4000 algorithms evaluated since 2009 for fingerprint technologies.

The keystroke competition described in the present work tries to complement the previous experiences by: (i) proposing the first keystroke **ongoing competition** which overcomes the limitations of traditional competitions based on a static snapshot of the state-of-the-art; (ii) disclosing a public benchmark involving 4800 keystroke sequences from **300 users**, simulating a realistic scenario in which each user types his own sequence (given name and family name) and 3600 impostor attacks (users who try to spoof the identity of others); (iii) being an online competition carried out over a **fully reproducible** framework based on the BEAT platform¹ [5] and an offline competition as baseline.

The rest of the paper is organized as follows: Section 2 describes the database and evaluation protocols. Section 3 presents the best systems submitted by the participants to the offline competition. Section 4 reports the experiments and results. Finally, Section 6 summarizes the conclusions.

2. Dataset and Protocols

2.1. Dataset and Evaluation Protocol

The dataset proposed for the competition is part of the BiosecurID multimodal database [8] and consists of keystroke sequences from 300 subjects acquired in four different sessions distributed in a four month time span. Thus, three different levels of temporal variability are taken into account: (i) within the same session (the samples are not acquired consecutively), (ii) within weeks (between two consecutive sessions), and (iii) within months (between non-consecutive sessions).

Each session comprises 4 case-sensitive repetitions of the subject's name and surname (2 in the middle of the session and two at the end) typed in a natural and continuous manner. No mistakes are permitted (i.e., pressing the backspace), if the subject gets it wrong, he/she is asked to start the sequence again. The names of three other subjects in the database are also captured as forgeries, again with no mistakes permitted when typing the sequence. The time (in milliseconds) elapsed between key events (press and release) is provided as the keystroke dynamics sequence. Imitations are carried out in a cyclical way, i.e., all the

Table 1. Summary of the main statistics of the database proposed for the competition.

Characteristics	#
Number of users (Testing Set)	300
Number of users (Development Set)	10
Number of sessions	4
Training samples per user	4
Test samples per user	20
Genuine samples per user*	8-12
Impostor samples per user*	8-12
Total genuine comparisons	3028
Total impostor comparisons	2972
Average separation between sessions	1 month
Average length of the key sequence	25.55

*In order to increase the difficulty, the number of genuine and impostor samples per user varies depending on the user. Participants do not know this number.

subjects imitate the previous subjects, and the first imitate the last subjects. The main statistics of the dataset proposed for the competition are summarized in Table 1.

The test samples remained sequestered (i.e., participants did not know whether they are genuine or impostors samples). In addition, a small development set (10 users with labeled samples) and baseline algorithms were provided to the participants.

The experimental protocol was based on the following steps, for each user: i) participants have 4 training samples (genuine samples from the 1st session) as enrollment data; ii) 20 test samples (genuine and impostor samples from 2nd to 4th sessions) are used to evaluate the performance of the systems; iii) each test sample is labeled with its corresponding user model and performance is evaluated according to the verification task (1:1 comparisons).

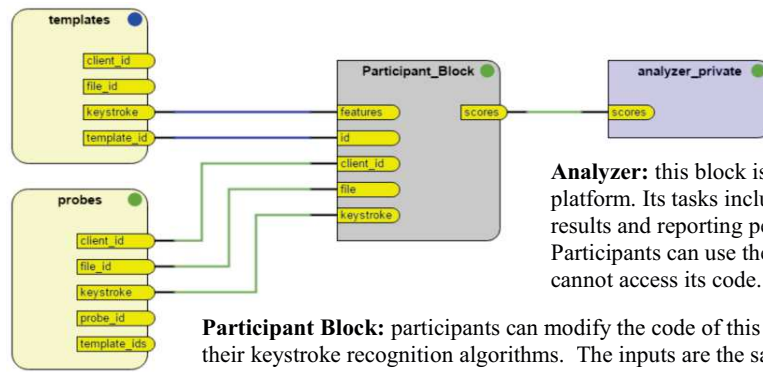
There are two modes of participation: ongoing and offline. Dataset and evaluation protocols of both modes of participation are exactly the same. The performance of the offline evaluation (detailed in Section 4) will be used as baseline for the ongoing competition.

2.2. Ongoing Competition

The competition exploits the potential of the BEAT platform, which was created under the FP7 EU BEAT project to promote reproducible research in biometrics. The BEAT platform is a European computing e-infrastructure for Open Science proposing a solution for open access, scientific information sharing and re-use including data and source code while protecting privacy and confidentiality. The platform is a web-application allowing experimentation and testing in pattern recognition.

¹ <https://www.beat-eu.org/platform/>

Database: participants cannot access directly the data but they can use it in the experiments. The platform automatically provides the training samples (labeled data) and test samples (unlabeled data) to the Participant Block.



Participant Block: participants can modify the code of this block including their keystroke recognition algorithms. The inputs are the samples of the database (training and test samples), and the output are the similarity scores.

Analyzer: this block is the output of the platform. Its tasks include analyzing results and reporting performance. Participants can use the analyzer but cannot access its code.

Figure 1. Toolchain of KBOC developed on BEAT (https://www.beat-eu.org/platform/toolchains/robertodaza/competition_kboc16/2/)

KBOC provides the data and modules necessary to run the evaluation and the BEAT platform ensures that the system is correctly executed, also providing the results. Different algorithms and systems can be easily compared. Figure 1 shows the toolchain of KBOC with the modules involved in the evaluation. The platform also provides an attestation mechanism for your reports (e.g., scientific papers, technical documents or certifications). The competition website provides instructions and examples to facilitate the participation of researchers without previous experience on BEAT. There is no limit regarding the number of systems evaluated, and the results are automatically provided to the participants on the platform (i.e., the performance of the systems is available in real time). The platform will be available beyond the offline competition and it is a new valuable resource for the keystroke recognition community (see the website² for details, tutorial and extra material). The ongoing platform is available online³ as well as results of all systems evaluated⁴.

It should be noted that participating in the ongoing competition and using of the platform does not imply the publication of your code and confidentiality is in any case granted. The organizers have no access to the private code evaluated by the platform but only to the results obtained. Reproducibility is granted by allowing execution permission without code access, thereby preserving confidentiality.

2.3. Offline Competition

In addition to the ongoing evaluation platform, a traditional offline competition was proposed to promote reproducible research serving as baseline for the ongoing evaluation. The training set (4 genuine samples from each of the 300

subjects) and test set (unlabeled data including genuine and impostor samples) were available at KBOC website. The keystroke recognition algorithms were executed at the participant premises according to the competition protocol. The scores (comparisons between user models and genuine/impostors samples) obtained by the participants were sent to the KBOC organization. To avoid overfitting, the number of submissions was limited to 15 different systems that were evaluated after the submission deadline.

3. Description of Participating Systems

There was a total of 12 institutions from 7 different countries registered for the competition (5 from USA, 2 from India and 1 from Norway, Argelia, The Netherlands, Brazil and China). Four from the registered institutions finally submitted their systems for a total number of different systems evaluated equal to 37. This section presents the descriptions of the best systems evaluated during the offline competition.

3.1. U.S. Army Research Laboratory

System 6 submitted by ARL used a Manhattan distance anomaly detector with keystroke duration and press-press (PP) latency features. While Manhattan distance generally yields relatively low EER among distance-based anomaly detectors [2], the low EER can also be attributed to preprocessing, feature normalization, and score normalization.

The raw data of each sample was first converted to a sequence of keystroke events with each event described by the key, press time, and duration. Although the data collection procedure of the test set did not allow mistakes or backspace by the genuine users, a variety of keystrokes

² <https://sites.google.com/site/btas16kboc/>

³ https://www.beat-eu.org/platform/experiments/robertodaza/robertodaza/competition_kboc16/2/Kboc16_Compensation_Baseline_Modified_Scaled_Manhattan_Distance/

⁴

https://www.beat-eu.org/platform/search/aythamimm/KBOC16_COMPETITION_SEARCH/

in both template and unknown samples could be observed for each claimed identity. A simple algorithm was developed to establish a correspondence of features between samples for each claimed identity.

The target keystroke sequence was selected as the minimum length sequence in the template samples. In the case where there were multiple minimum length sequences that differed by a permutation, the target sequence was selected randomly. A modified dynamic time warping (DTW) algorithm then matched the keystrokes of every other sample to the length- M target sequence sorted by press time. The M key-hold durations and $M - 1$ PP latencies were then extracted from both the template and query samples according to the keystrokes in the target sequence. While the PP latency features for the target sequence were strictly positive, the PP latency features for the other samples were negative for permuted keystrokes. The duration and PP latency features of each claimed identity were then normalized to within one standard deviation (SD) of the mean duration and mean PP latency, respectively, of the genuine samples.

Following feature extraction, the Manhattan distance to the mean template feature vector was calculated. The distances from unknown samples to each claimed identity were then normalized to within ± 2 SD of the mean, with distances outside that range clipped to $[0,1]$. This procedure yielded an EER of $6.5 \pm 1.8\%$ on the development set, obtained through a Monte Carlo validation procedure. Following the reproducibility criteria of the competition, the code is available at⁵.

3.2. Universidade Federal de Sergipe

The UFS team self-imposed three restrictions in order to properly simulate an actual biometric system:

- R1: The number of test samples per subject is not known beforehand.
- R2: The proportion of genuine and impostor amongst the test samples are unknown.
- R3: Sequential test samples simulate system interrogation through time, therefore a score obtained at a given time cannot be used to improve previous scores.

As in [9], Press-Press (PP) and Hold-time (H) time intervals were equalized with parameters $\mu_{PP} = -1.61$, $\sigma_{PP} = 0.64$, $\mu_H = -2.46$ and $\sigma_H = 0.33$ respectively, through a non-linear mapping:

$$y = \frac{1}{1 + \exp\left(-\frac{1.7(\log_e(x) - \mu)}{\sigma}\right)} \quad (1)$$

where x stands for a time interval (in seconds).

For each enrolled subject, a set of PP and H vectors were taken as user templates. During interrogation, the unlabeled PP vector was compared to a single minimum profile obtained through the so-called shuffling procedure (Bleha, 1988), for it is slightly better than using mean profile, yielding d_{PP} . Likewise, d_H stands for the minimum distance between templates and the tested H vector. All distances were sums of absolute differences (Manhattan distance) divided by the length of the sample. In case of inconsistent vector lengths, the shorter one was compared to each sub segment of the longer one and the minimum distance is kept (i.e., only time intervals are considered, not character mismatches). Final test score was computed as:

$$d = 0.75d_{PP} + 0.25d_H \quad (2)$$

Moreover, the template set was automatically appended with new samples every time a score lower than 0.14 is found, thus influencing future scores.

We highlight that it is possible to improve performance if no restrictions are imposed (i.e. online template adaptation), but enrollment-interrogation simulation would be less realistic. For instance, by using the training set, through 50 independent runs (i.e. independent choices of 4 training signatures per run), an average EER of $8.0\% \pm 1\%$ (standard deviation) was obtained. By contrast, if R3 is violated and an a posteriori score normalization is done, the EER drops to $6.5\% \pm 1\%$, for the same system.

3.3. Indian Institute of Technology Kharagpur

The raw data provided contained the scan codes and the time gap between key press and key release events. In this approach, the time intervals between consecutive key events are used as the feature vectors. The raw data obtained consisted of a sequence of N keys. Let M be the number of key events (key-press and key-release events) in the sequence. The feature vector was modelled based on the time between two consecutive key events irrespective of press or release events. The formulation is described below.

Let the feature vector of a test sample be, $\mathbf{f} = [f_1, f_2, \dots, f_M]$, where, f_i is the time interval between $(i - 1)^{th}$ and i^{th} key events, where $i = 1, 2, 3, \dots, M$.

Similarly consider the enrolment set $\{\mathbf{g}^k\}_{k=1}^T$, where $\mathbf{g}^k = [g_1^k, g_2^k, \dots, g_M^k]$ $k \in 1, \dots, T$, with $T = 4$ samples and M the number of features for each keystroke sequence. Two distance measures were computed between the feature vector \mathbf{f} of the test sample and the enrolment set $\{\mathbf{g}^k\}_{k=1}^T$. The two distance metrics used find the absolute distance to the nearest neighbor in each feature dimension independently. A combination of mean and median of these distances was used as the final distance metric. The distance measures were computed as:

$$\mathbf{AD}(i, k) = |g_i^k - f_i|, \quad k = 1, \dots, T \text{ and } i = 1, 2, 3, \dots, M \quad (3)$$

⁵ <http://vmonaco.com/software/>

$$\mathbf{md}(i) = \min_{k \in [1, \dots, T]} ad_{ik}, \quad i = 1, 2, 3, \dots, M \quad (4)$$

where ad_{ik} is an element of matrix $\mathbf{AD}(i, k)$ and the final distance was obtained as:

$$d = \text{mean}(\mathbf{md}) + \text{median}(\mathbf{md}) \quad (5)$$

4. Results

This section presents the final results of the offline competition while the ongoing results can be seen at the BEAT platform⁶. As it is an ongoing competition, the results will be automatically updated with any new submission.

Regarding the offline competition, the participants were allowed to submit up to 15 different systems before the deadline. As previously mentioned, the algorithms were compared after the deadline, thus being the performance of all systems reported after the submission period ended, according to the following indicators:

- Global Equal Error Rate (EER_G): unique EER calculated using all genuine and impostor scores and only one threshold for all users.
- User-dependent Equal Error Rate (EER_U): the EER is calculated independently for each of the 300 subjects (300 different thresholds). EER_U is the average individual EER from all subjects. This EER is common in the keystroke dynamics literature [2][3].
- FMR_{100} : the lowest False Non-Match Rate for False Match Rate equal to 1%.
- Detection-Error Tradeoff (DET) curve: a plot of FMR and FNMR that reports system performance at any possible operating point (matching threshold).

It should be highlighted that participants have developed their systems on the basis of a development set with only 10 users, which were then evaluated on 300 sequestered users. Table 2 summarizes the most important characteristics of the best system submitted by each participant, while Table 3 presents the top results achieved across all their submissions (training with first session and testing with remaining three). The results show clear differences between the systems proposed by the participants, whose corresponding EER ranged between 5.32% and 17.90% for the Global EER (EER_G) and 4.72% and 13.66% for the user-dependent EER (EER_U). The large difference between EER_G and EER_U of those systems without score normalization (P1, P2 and P3) suggests the importance of this step, especially when a unique threshold is employed [11][12]. The system based on DTW alignment (with feature and score normalization) proposed by the U.S. Army Research Laboratory is the best competing approach.

Table 4 includes the performance (EER_G) obtained using the genuine samples from the second and fourth session

(maximum time lapse) for testing and first session for training. The results show a marginal degradation of the performance, which suggest the stability of the patterns (note that users typed their given name and family name). Figure 2 shows the DET curves for all submissions (Fig. 2 Left) and best submissions according to the session evaluated (Fig. 2 Right). The curves show how the submissions made by the participants tend to cluster in different performance ranges and the high robustness against the time lapse (2 months between second and fourth session).

5. Conclusions

This paper presented the first keystroke biometrics ongoing evaluation and the results of an associated offline competition used as baseline. The evaluation, developed on the BEAT, platform comprises one of the largest fixed-text keystroke database available. The main characteristics of the competition can be summarized as:

- Small development database (10 subjects) to setup the algorithms.
- Large evaluation database with 300 users and 7200 keystroke sequences acquired during 4 different sessions with time lapse of 4 months
- Multisession database with 4 different sessions. Training using samples from the first session and testing with the remaining 3.
- Realistic scenario with different passwords for each user (given name and family name) with case-insensitive acquisition.
- Performance metrics based on either global EER or user-dependent EER.

The performances achieved by the participants are encouraging with a best EER of 5.32%, which could be used as a challenging baseline in the ongoing platform.

Acknowledgment

A.M. and M. G.-B. are supported by a JdC contract (JCI-2012-12357) and FPU Fellowship from Spanish MINECO and MCD respectively. This work was partially funded by the projects: CogniMetrics (TEC2015-70627-R) from Spanish MINECO and BEAT (FP7-SEC-284989) from EU.

References

- [1] A. Peacock, X. Ke, and M. Wilkerson, "Typing patterns: A key to user identification", *IEEE Security and Privacy*, 2(5):40–47, 2004.
- [2] K. S. Killourhy and R. A. Maxion, "Comparing Anomaly Detectors for Keystroke Dynamics", in *Proc. of the 39th Ann. Int. Conf. on Dependable Systems and Networks*, Estoril, Lisbon, Portugal, IEEE CS Press, pp. 125-134, 2009.

⁶ https://www.beat-eu.org/platform/search/aythamimm/KBOC16_COMPETITION_SEARCH/

Table 2. Summary of the characteristics of the best approaches submitted by the participants.

Participant	Preproc.	Features	Feature norm.	Matcher	Score norm.
P1- Indian Institute of Technology Kharagpur	no	Hold+RP	no	Combined	no
P2 - Federal University of Sergipe	yes	Hold+PP	no	Manhattan	no
P3 - Anonymous participant	no	RP	no	Kendall's tau	no
P4 - U.S. Army Research Laboratory	yes	Hold+PP	yes	Manhattan	yes

Table 3. Final results (best systems) for the KBOC16 offline competition: EER_G (user-independent threshold), EER_U (user-dependent-threshold), FMR_{100} . Training with first session and testing with remaining 3 sessions.

ID	EER_G	EER_U	FMR_{100}
P1	15.73%	11.95%	51.13%
P2	11.82%	7.96%	54.65%
P3	17.90%	13.66%	64.60%
P4	5.32%	4.72%	28.36%

Table 4. Best EER_G for the KBOC16 offline competition according to the session used for testing. Training with first session and testing with second and fourth sessions.

ID	Second Session	Fourth Session
P1	15.28%	16.13%
P2	11.60%	11.96%
P3	17.01%	18.21%
P4	5.09%	5.10%

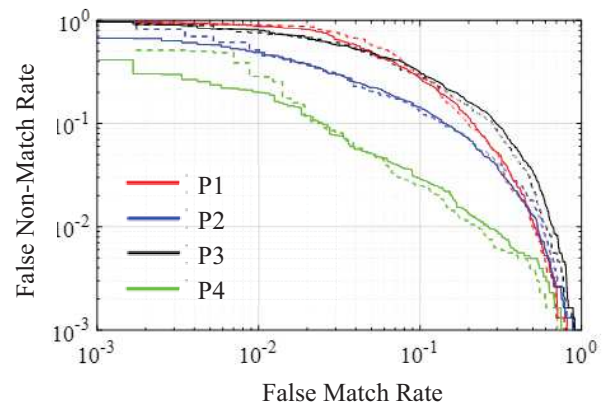
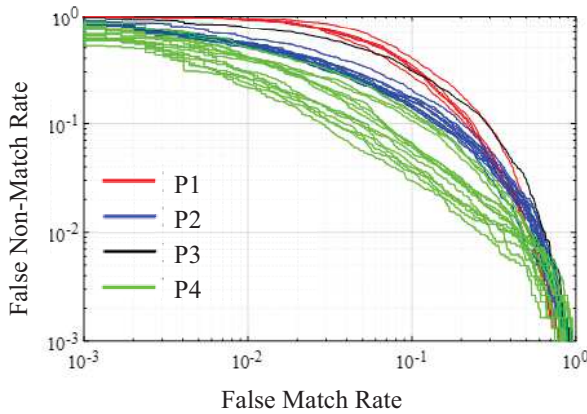


Figure 2. Left: DET curves obtained from all submissions (training with first and testing with remaining 3 sessions); Right: time lapse between enrollment (first session) and testing with second (dotted line) and fourth (continuous line) sessions.

[3] R. Giot, M. El-bed and R. Christophe, "Greyc keystroke: a benchmark for keystroke dynamics biometric systems", in *Proc. of IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems*, Washington DC, pp. 1-6, 2009.

[4] J. V. Monaco, G. Perez, C. C. Tappert, P. Bours, S. Mondal, S. Rajkumar, A. Morales, J. Fierrez and J. Ortega-Garcia, "One-handed Keystroke Biometric Identification Competition", in *Proc. IEEE/IAPR Int. Conf. on Biometrics*, Phuket, Thailand, pp. 58-64, 2015.

[5] S. Marcel, "BEAT biometrics evaluation and testing", *Biometric Technology Today*, pp. 5-7, 2013.

[6] R. Cappelli, M. Ferrara, D. Maltoni and F. Turrone, "Fingerprint Verification Competition at IJCB2011", in *Proc. Of the IEEE/IAPR Int. Joint Conference on Biometrics (IJCB11)*, Washington DC, pp. 1-6, 2011.

[7] A. Morales, M. Falanga, J. Fierrez, C. Sansone and J. Ortega-Garcia, "Keystroke Dynamics Recognition based on Personal Data: A Comparative Experimental Evaluation Implementing Reproducible Research", in *Proc. of the IEEE Seventh Int. Conf. on Biometrics: Theory, Applications and Systems*, Arlington, Virginia, USA, pp. 1-6, 2015.

[8] J. Fierrez, *et al.*, "BiosecurID: A Multimodal Biometric Database", *Pattern Analysis and Applications*, 13(2):235-246, 2010.

[9] J. Montalvão, E. O. Freire, M. A. Bezerra Jr., R. Garcia, "Contributions to empirical analysis of keystroke dynamics in passwords", *Pattern Recognition Letters*, 52(15):80-86, 2015.

[10] S. Bleha, Recognition systems based on keystroke dynamics, Ph.D. dissertation, Univ. Missouri-Columbia, May 1988.

[11] A. Morales, E. Luna, J. Fierrez and J. Ortega-Garcia, "Score Normalization for Keystroke Dynamics Biometrics", *Proc. 49th Annual Int. Carnahan Conf. on Security Technology*, pp. 1-6, Taipei, Taiwan, 2015.

[12] J. Fierrez-Aguilar, J. Ortega-Garcia and J. Gonzalez-Rodriguez, "Target dependent score normalization techniques and their application to signature verification", *IEEE Trans. on Systems, Man & Cybernetics - Part C*, 35(3):418-425, 2005.

KBOC: Plataforma de Evaluación de Tecnologías de Reconocimiento Biométrico Basadas en Dinámica de Tecleo

Roberto Daza, Aythami Morales, Julian Fierrez, Marta Gomez-Barrero y Javier Ortega-Garcia
roberto.daza@estudiante.uam.es, {aythami.morales, julian.fierrez, marta.barrera, javier.ortega}@uam.es
ATVS, EPS, Universidad Autonoma de Madrid, C\ Francisco Tomás y Valiente, 11, 28049 Madrid, Spain

Abstract- La reproducibilidad de la investigación es un tema de gran preocupación dentro de la comunidad científica. En este trabajo se presenta una plataforma de evaluación tecnológica (BEAT) orientada a fomentar la experimentación reproducible dentro del área del reconocimiento biométrico. Para ello, se ha propuesto una competición internacional centrada en la evaluación de sistema de reconocimiento de usuarios a través su de dinámica de tecleo. La competición incluye una base de datos con más de 300 usuarios y 4 sesiones diferentes. En este trabajo se introduce la plataforma, se detallan las características de la competición y se presenta los resultados iniciales.

I. INTRODUCCIÓN

¿Es sencillo reproducir los resultados presentados en un trabajo de investigación? ¿Podemos confirmar realmente la veracidad de estos resultados?

La realidad es que en algunas áreas son pocas las publicaciones que facilitan la reproducibilidad de los resultados presentados [1]. Por ejemplo, en el ámbito de la Biometría es habitual en las publicaciones científicas la ausencia del algoritmo que genera dichos resultados, pero no solo este es el problema principal, sino que también se suelen omitir las bases de datos sin las cuales es imposible reproducir el trabajo de investigación. Hay diversos motivos por los cuales se prescinde de su inclusión, entre los cuales podemos destacar: la falta de espacio a la que se está condicionado por las publicaciones, la falta de unanimidad a la hora de programar, siendo una limitación al desconocer este tipo de lenguaje, añadiendo que puede que sean programas de pago (Matlab, etc). También puede deberse a que el autor no esté interesado en facilitar su código por diversas razones, como la privacidad de sus códigos o la falta de auto-disciplina. En general, se podría decir que hay pocos incentivos que empujen a los investigadores a presentar trabajos que faciliten la reproducibilidad a otros investigadores.

Para incentivar el uso de protocolos reproducibles hay que facilitar a los autores mecanismos que solucionen los problemas antes mencionados, y es éste precisamente el punto de partida para nuestro trabajo. En este caso nos centraremos en un sub-área muy concreta de la investigación de sistemas de reconocimiento biométrico: el reconocimiento a partir de dinámica de tecleo.

El reconocimiento biométrico en si es el reconocimiento automático de los individuos en función de su fisiología (p. ej. huella digital, cara, iris) y/o características de su comportamiento [2] (p. ej. firma, el movimiento al caminar, pulsaciones de teclas). Por lo tanto, reconocimiento biométrico es un área de investigación muy amplia que incluye a investigadores de las comunidades de reconocimiento de patrones y aprendizaje automático.

En las últimas décadas han sido muchas las aplicaciones de reconocimiento biométrico basadas en la dinámica de pulsación de teclas que se han investigado (identificación de los usuarios a partir de patrones derivados de la formas en la que teclean). Esta área es interesante para la comunidad científica por diversas razones como los desafíos asociados a la modelización y la búsqueda de secuencias dinámicas con alta variabilidad intra-clase, como el comportamiento humano, fuertemente dependiente del usuario y que varía significativamente entre los sujetos. Además, la naturaleza simple de los datos (secuencias de tiempo), hace que la biometría de pulsaciones de teclas sea un buen campo para introducir nuevos investigadores (sin experiencia previa en aplicaciones biométricas) en esta área desafiante que atrae tanto a académicos y profesionales.

En este trabajo se presenta una plataforma de evaluación de tecnologías de reconocimiento biométrico basadas en dinámica de tecleo desarrollada sobre la plataforma BEAT: Biometrics Evaluation and Testing (BEAT¹) [3] fruto del proyecto europeo homónimo cuyo principal objetivo es promover la investigación reproducible. Como aplicación, se propone una competición que formará parte de las competiciones oficiales del congreso internacional *The IEEE Eighth International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2016)*.

II. KBOC: KEYSTROKE BIOMETRIC ONGOING COMPETITION

La comunidad en reconocimiento basado en dinámica de teclas es heterogénea e incluye investigadores de diferentes disciplinas [4]. Sin embargo, nos encontramos que el mercado no ha tenido el auge que es de esperar para el potencial que posee. Muy pocas son las propuestas que nos ofrecen la

¹<https://www.beat-eu.org/platform/>

oportunidad de comparar los sistemas en las mismas condiciones. Este problema es importante, ya que el rendimiento de los sistemas de dinámica de teclas es tremendamente dependiente de la aplicación (texto libre o fijo) y bases de datos (diferentes usuarios muestran comportamientos muy diferentes), además del problema añadido por la inexistencia de un enfoque genérico debido al gran número de algoritmos.

La literatura ofrece bases de datos públicas que posibilitan comparar los sistemas en las mismas condiciones, siendo dos principalmente las más populares en dinámica de teclas: CMU[5] y GREYC[6]. Aunque estas bases de datos son de utilidad para la comunidad científica, nos encontramos con dos problemas importantes, por un lado el pequeño número de participantes (51 en CMU y 133 para GREYC) y por otro el escenario de reconocimiento, en el cual todos los usuarios comparten la misma contraseña. Pensamos que este escenario limita mucho a estas propuestas, pues los usuarios resultan insuficientes, más cuando sabemos que en pulsaciones de teclado se ve altamente afectado el resultado por este número reducido de participantes, dentro de un escenario que no es el más realista (normalmente no todas las personas comparten contraseña).

Por otro lado, solo tenemos constancia de un precedente de competición de reconocimiento de dinámica de teclado que tuvo lugar en la conferencia internacional de Biometría BTAS 2015: “One-handed Keystroke Biometric Identification Competition”[7]. En esta competición se contaba con un conjunto de datos tomados de 64 usuarios que se ven obligados a usar solo una mano para escribir muestras de texto largo mecanografiadas libremente. Como podemos ver la competición evalúa el rendimiento de los algoritmos en condiciones de más dificultad (una sola mano), cuando lo normal es utilizar las dos.

La competición KBOC trata de complementar las experiencias anteriores con: (i) una nueva base de datos con 300 usuarios que escriben su nombre y apellido repetidas veces para un total de 8400 secuencias de teclado de las cuales 3600 son ataques de impostores (usuarios que intentan imitar la identidad de los otros usuarios); (ii) al contrario de anteriores competiciones, KBOC es una competición ongoing que permite a los investigadores seguir compitiendo con sus sistemas de dinámica de tecla de manera online; (iii) una competición en un marco completamente reproducible utilizando la plataforma BEAT.

A. Promover la participación.

Una de las mayores dificultades en una competición es incentivar la participación de los investigadores. Existen competiciones ongoing con un gran éxito de participación, como FVC-Ongoing competition[8] con más de 900 participantes y más de 4000 algoritmos evaluados desde el año 2009 para huella digital, cara y huella de la palma de la mano.

Es por ello que la competición KBOC incentiva a la participación de diversas maneras: Hay dos formas de participación en la competición: (i) **participación offline**, los participantes no obtienen el rendimiento de sus sistemas hasta el final de la competición. Es posible enviar 15 sistemas diferentes los cuales se ejecutan en las máquinas de los participantes tras lo cual se debe enviar un fichero de

resultados que será procesado por la organización de la

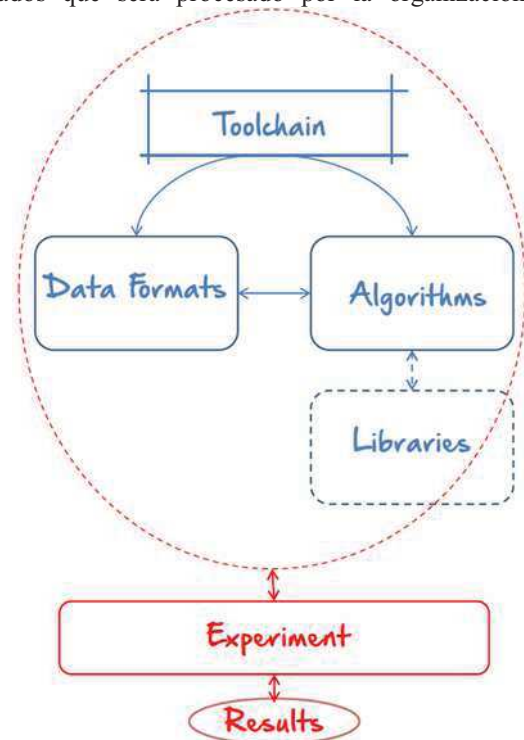


Fig. 1. Esquema de la plataforma BEAT

competición; (ii) **participación ongoing**, es la opción recomendada, se realiza a través de la plataforma BEAT y los participantes obtienen el rendimiento de sus sistemas al instante.

Debido a que la base de datos y protocolos de experimentación son los mismos para ambos tipos de participaciones, la participación offline servirá de referencia para los sistemas evaluados en la plataforma ongoing.

Además, se suministra diverso material de la plataforma BEAT para un fácil aprendizaje (tutoriales, videos, etc.), se crea una página web² para su mayor difusión y se proporcionan códigos de referencia para ayudar a los participantes en sus primeros pasos en la competición (junto con un conjunto de desarrollo de 280 secuencias de teclado similares a las usadas en la competición).

El intervalo de tiempo en el que permanece activa la competición offline comprende desde Enero a Abril de 2016. La plataforma ongoing estará disponible a cualquier investigador por tiempo indefinido.

III. PLATAFORMA BEAT

La plataforma BEAT surge por un proyecto UE con la idea de promover la investigación reproducible en Biometría, pero podemos afirmar que trasciende este campo ya que engloba cualquier proceso de señales. Es de gran interés para el mundo de la investigación BEAT, ya que nos permite infinidad de posibilidades, desde compartir nuestros resultados manteniendo el código privado hasta compartir nuestro algoritmo, etc. Además nuestra experiencia nos indica que nosotros mismos no somos capaces de reproducir un experimento antiguo a causa de que hemos hecho nuevas

²<https://sites.google.com/site/btas16kboc/>

versiones olvidando las anteriores. BEAT da una solución a este problema guardando un historial de nuestros algoritmos, pudiéndose llegar a cualquier versión anterior.

BEAT es una plataforma innovadora, proporcionándonos las herramientas necesarias para implementar sistemas de Biometría, incluyendo: (i) Acceso libre para la comunidad científica; (ii) No es necesario la instalación de ningún software, todo se realiza vía web; (iii) Utiliza uno de los lenguajes más populares y de mayor proyección de futuro (Python), el cual es libre; (iv) Incorpora la mayoría de bibliotecas más sonadas en reconocimiento de patrones; (v) La computación es de muy alto rendimiento; (vi) La plataforma trabaja con diagramas de bloques que permiten más posibilidades; (vii) Posibilidad de hacer público o privado los trabajos.

La plataforma BEAT se estructura en los siguientes campos principalmente: algorithms, toolchains, experiments, attestations. Esto se debe a que utiliza programación orientada a bloques, permitiendo dividir fácilmente los algoritmos en tareas:

- Toolchains: Es el diagrama de los bloques donde se indican las entradas y salidas de cada bloque (bases de datos, analizador, etc.).
- Algorithms: Este campo es el indicado para el código (Python) a utilizar en los bloques.
- Experiments: Une todos los campos anteriores y nos facilita los resultados.
- Attestations: Los certificados incluyen un experimento, haciendo que no pueda ser modificado nunca más y dan una confirmación de que los resultados se obtuvieron en la plataforma.

En la Fig. 1 se hace referencia a un esquema donde podemos ver como se conectan los principales campos de BEAT. Es por todas estas ventajas que se utiliza BEAT en la competición para conseguir explotar el potencial de la plataforma. Se puede acceder a la plataforma de evaluación de tecnologías basadas en dinámicas de tecleo a través de este enlace³ así como a los resultados de todos los sistemas evaluados a través de este otro⁴.

IV. CONJUNTO DE DATOS DE REFERENCIA.

La base de datos utilizada en la competición KBOC es parte del conjunto de datos de BiosecurID[9] y consiste en muestras de tecleo tomadas a 300 usuarios en 4 sesiones diferentes, en un espacio temporal de 4 meses. En cada sesión los usuarios repiten 4 veces su nombre y apellido (separado por la tecla espacio), usando tanto mayúsculas como minúsculas, no permitiéndoseles errores (pulsar tecla retroceso). En el caso de que el usuario se equivoque tendrá que iniciar el proceso de nuevo. Tampoco se les permiten errores a los impostores.

La dinámica de tecleo de cada secuencia viene definido por las marcas de tiempo referentes a dos eventos: pulsar y soltar teclas. Se suministra las marcas de tiempo en milisegundos así como el identificador de tecla pulsada. Las

³https://www.beat-eu.org/platform/experiments/robertodaza/robertodaza/competition_kboc16/2/Kboc16_Compensation_Baseline_Modified_Scaled_Manhattan_Distance/

⁴https://www.beat-eu.org/platform/search/aythamimm/KBOC16_COMPETITION_SEARCH/

Tabla 1. Resumen de las principales características de la base de datos utilizada en la competición

Características	
Número de usuarios (conjunto de prueba)	300
Número de sesiones	4
Muestras Genuinas por usuario*	entre 8-12
Muestras Impostoras por usuario*	entre 8-12
Total de secuencias genuinas	4800
Total de secuencias impostoras	3600
Separación media entre las sesiones	1 mes

*Los participantes solo disponen de las etiquetas de 4 muestras genuinas de la primera sesión de cada usuario. El resto de muestras se utilizan para evaluar el rendimiento del sistema y los participantes desconocen si pertenecen a usuarios genuinos o impostores (el número de muestras genuinas e impostoras de test suman 20 para todos los usuarios pero su distribución varía según usuario).

principales estadísticas del conjunto de datos del concurso se resumen en la Tabla 1.

El protocolo de experimentación se resume en los siguientes pasos: i) la dinámica de tecleo de cada usuario se modelará a partir de 4 muestras genuinas de cada usuario (pertenecientes a la primera sesión); ii) se obtendrá una puntuación por cada una de las 20 muestras de test; iii) en el caso de la competición offline el fichero de puntuaciones se enviará a la organización de la competición siguiendo un formato predefinido. Los rendimientos de todos los sistemas serán notificados una vez acabada la competición. En la competición ongoing, el participante puede evaluar el rendimiento de sus sistemas de forma automática a través de la plataforma y por lo tanto, no se precisa del envío de ficheros de puntuaciones. El rendimiento final se da en forma de curvas de falsa aceptación, curvas de falso rechazo, curvas DET y tasas de igual error obtenidas a partir de las puntuaciones genuinas e impostoras de todos los usuarios.

V. ALGORITMO FACILITADO EN LA COMPETICIÓN

Se proporciona en la competición un código de referencia para facilitar la participación, tanto para los participantes offline (código Matlab) como para los participantes ongoing (plataforma BEAT, código Python). En el código de la plataforma BEAT se suministra una librería con tres algoritmos de *baseline*[10].

La librería dispone de tres distancias populares en los algoritmos de reconocimiento basados en dinámica de tecleo. Se asume $\mathbf{f}=[f_1, f_2, \dots, f_M]$ como el vector de características de una muestra con identidad desconocida (con M características) y $\mathbf{g}^k=[g_1^k, g_2^k, \dots, g_M^k]$ $k \in 1, \dots, T$, el conjunto de entrenamiento de un usuario (en el caso de la competición $T=4$). Las tres distancias dadas son:

- Combinación Manhattan-Mahalanobis:

$$d_1^k = \sum_{i=1}^M |\hat{f}_i - \hat{g}_i^k| \quad (1)$$

donde $\hat{\mathbf{f}} = \mathbf{S}^{-1/2} \mathbf{f}^T$ y $\hat{\mathbf{g}}^k = \mathbf{S}^{-1/2} (\mathbf{g}^k)^T$, Ses la covarianza y $(\cdot)^T$ es la traspuesta. La distancia d_1 es la obtenida para la mínima de las diferentes k .

- Mahalanobis + vecino más cercano: basada en la distancia de Mahalanobis y el algoritmo de vecino más cercano:

$$d_2^k = (\mathbf{f} - \mathbf{g}^k) \mathbf{S}^{-1} (\mathbf{f} - \mathbf{g}^k)^T \quad (2)$$

siendo \mathbf{S}^{-1} la inversa de la matriz de covarianzas de los datos de entrenamiento. La distancia final d_2 es obtenida como la mínima para las diferentes k .

- Manhattan escalada modificada: una modificación de la anterior definida como:

$$d_3^k = \sum_{i=1}^M |f_i - \bar{g}_i| / \sigma'_i \quad (3)$$

donde \bar{g} es la media de los datos de entrenamiento y $\sigma' = [\sigma'_1, \sigma'_2, \dots, \sigma'_M]$ es una modificación de la desviación estándar:

$$\sigma'_i = \begin{cases} \frac{0.2}{M} \sum_{j=1}^M \sigma_j & \text{si } \sigma_i < \frac{0.2}{M} \sum_{j=1}^M \sigma_j \\ \sigma_i & \text{resto} \end{cases} \quad (4)$$

y σ_i es la variación estándar de $\{\mathbf{g}^k\}_{k=1}^T$. Por otra parte $\bar{\mathbf{g}} = \frac{1}{T} \sum_{k=1}^T \mathbf{g}^k$. Esta simple modificación intenta paliar los efectos de las muestras con una escasa variación durante la normalización.

La puntuación final para las 3 distancias, s_i , se obtiene como la inversa de las mismas: $s_i = -d_i$.

VI. RESULTADOS.

En el momento de escribirse esta publicación, la organización de la competición no tiene los resultados finales de los participantes (4 participantes y más de 30 sistemas enviados) debido a que la misma no ha llegado a su fin, es por este motivo que se exponen los resultados obtenidos con los sistemas *baseline* en la Tabla 2, los cuales han sido facilitados de ejemplo a los participantes.

Tabla 2. Resultados obtenidos para las tres distancias en BEAT.

Resultados	d_1	d_2	d_3
EER	29.60%	27.35%	18.60%

Hay que mencionar que los resultados son fácilmente mejorables, debido a que algunas de las muestras incluyen errores en teclas de pulsaciones adicionales (*case-insensitive*, principalmente la tecla shift). Ante esta situación se tomó una decisión muy sencilla y fácil de mejorar: tomar como referencia la primera muestra y a las sucesivas igualarlas al tamaño de la primera muestra (recortando o añadiendo ceros). Mencionar que para obtener estos resultados tampoco se hizo uso de la información referente a las teclas pulsadas, la cual puede ser de gran ayuda. Por lo tanto, para obtener los resultados se aplicó directamente al tiempo de pulsar y soltarlas distancias anteriormente descritas.

Como podemos ver en la Tabla 2, la distancia d_3 tiene un EER de 0.186, siendo esta distancia la que ofrece un menor ERR⁵. Esto era de esperar por publicaciones anteriores, donde

se podía ver el alto rendimiento [10]. En cambio, con la distancia d_1 obtenemos los peores resultados respecto a este valor.

VII. CONCLUSIONES

Desde nuestro punto de vista las publicaciones deben facilitar la reproducibilidad de los trabajos. Diversos estudios corroboran que las publicaciones que facilitan la reproducibilidad son más citadas (hasta 7 veces más) que las que no.

Creemos que la plataforma BEAT es una gran oportunidad para la comunidad científica, permitiendo la experimentación reproducible y obteniendo muchas otras ventajas (ej. certificación de los resultados). En lo que respecta al ámbito de reconocimiento de usuarios basado en la dinámica de tecleo era fundamental una competición (KBOC) que impulsara la experimentación reproducible, que permitiera comparar sistemas de pulsaciones de teclas y que suministra una base de datos con suficientes participantes y de forma realista (dos manos).

Debido a que la competición KBOC no tiene fecha de finalización programada (permanecerá activa por tiempo indefinido), pensamos que es una enorme oportunidad para comparar nuestros sistemas con los de los demás investigadores y así poder aprender unos de otros y avanzar de forma conjunta en el área.

AGRADECIMIENTOS

A.M. disfruta de un contrato JdC del MINECO (JCI-2012-12357). Este trabajo está parcialmente financiado por los proyectos: CogniMetrics (TEC2015-70627-R) del MINECO español y BEAT (FP7-SEC-284989) de la UE.

REFERENCIAS

- [1] P. Vandewalle, J. Kovačević and M. Vetterli, "Reproducible research in signal processing," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 37-47, 2009.
- [2] A. Jain, A. Ross and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4-20, 2004.
- [3] S. Marcel, "BEAT-biometrics evaluation and testing," *Biometric Technology Today*, vol. 2013, no. 1, pp. 5-7, 2013.
- [4] A. Peacock, X. Ke and M. Wilkerson, "Typing patterns: A key to user identification," *IEEE Security & Privacy*, vol. 2, no. 5, pp. 40-47, 2004.
- [5] K. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *IEEE/IFIP International Conference on Dependable Systems and Networks*, Lisbon, 2009.
- [6] R. Giot, M. El-Abed and C. Rosenberger, "GREYC keystroke: A benchmark for keystroke dynamics biometric systems," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, Washington DC, 2009.
- [7] J. Monaco, G. Perez, C. Tappert, P. Bours, S. Mondal, S. Rajkumar, A. Morales, J. Fierrez and J. Ortega-Garcia, "One-handed Keystroke Biometric Identification Competition," in *International Conference on Biometrics (ICB)*, Phuket, 2015.
- [8] R. Cappelli, M. Ferrara, D. Maltoni and F. Turrone, "Fingerprint verification competition at IJCB2011," in *International Joint Conference on Biometrics (IJCB)*, Washington DC, 2011.
- [9] J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos, D. T. Toledano, J. Gonzalez-Rodriguez, J. A. Siguenza, J. Garrido-Salas, E. Anguiano, G. Gonzalez-de-Rivera, R. Ribalda, M. Faundez-Zanuy, J. A. Ortega, V. Cardenoso-Payo, A. Viloria, C. E. Vivaracho, Q. I. Moro, J. J. Igarza, J. Sanchez, I. Hernaiz, C. Omite-Uruñuela, F. Martinez-Contreras and J. J. Gracia-Roche, "BiosecuRID: a multimodal biometric database," *Pattern Analysis and Applications*, vol. 13, no. 2, pp. 235-246, 2009.
- [10] A. Morales, M. Falanga, J. Fierrez, C. Sansone and J. Ortega-Garcia, "Keystroke dynamics recognition based on personal data: A

⁵<https://www.beat-eu.org/platform/attestations/1495360336/#results> (resultados y código de la distancia d_3)

comparative experimental evaluation implementing reproducible research," in *IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Arlington, 2015.