Emerging NUI-based Methods for User Authentication: A New Taxonomy and Survey

Napa Sae-Bae, Jonathan Wu, Nasir Memon, *Fellow, IEEE*, Janusz Konrad *Fellow, IEEE*, and Prakash Ishwar, *Senior member, IEEE*

Abstract—As the convenience and cost benefits of Natural User Interface (NUI) technologies are hastening their wide adoption, computing devices equipped with such interfaces are becoming ubiquitous. Used for a broad range of applications, from accessing email and bank accounts to home automation and interacting with a healthcare provider, such devices require, more than ever before, a secure yet convenient user authentication mechanism. This paper introduces a new taxonomy and presents a survey of "point-of-entry" user-device authentication mechanisms that employ a natural user interaction. The taxonomy allows a grouping of the surveyed techniques based on the sensor type used to capture user input, the actuator a user applies during interaction, and the credential type used for authentication. A set of security and usability evaluation criteria are then proposed based on the Bonneau, Herley, Van Oorschot and Stajano framework. An analysis of a selection of techniques and, more importantly, the broader taxonomy elements they belong to, based on these evaluation criteria, are provided. This analysis and taxonomy provide a framework for the comparison of different authentication alternatives given an application and a targeted threat model. Similarly, the taxonomy and analysis also offer insights into possibly unexplored, yet potentially rewarding, research avenues for NUI-based user authentication that could be explored.

Index Terms—Authentication, Emerging interfaces, Natural User Interaction, Security and usability.

1 Introduction

The landscape of user interfaces has dramatically changed in the last two decades and transformed the way we interact with computing devices. We have become accustomed to interaction involving touch, gesture or speech, and use them daily on smartphones, laptops, smart speakers, wearables, augmented-reality sets, and even some Internet-of-Things (IoT) devices (Fig. 1). The interfaces that leverage these natural, intuitive, everyday user behaviors are known as Natural User Interfaces (NUIs). In addition to touch, speech and gesture, future NUIs are also expected to utilize eye gaze and thought as well.

The emergence of NUIs has enabled novel functionalities. For example, information on the internet can now be accessed *via* smart windows and refrigerators. Multi-finger gestures on a touchscreen or touchpad have simplified scrolling and zooming of electronic documents (two-finger swipe and pinch). Voice recognition has become common in homes and cars, facilitating home automation and hands-free operation of a smartphone or a GPS unit while driving. NUIs are also expected to play an important role in the future in situations where conventional interfaces, such as the keyboard or mouse, may be inconvenient or inappropriate (e.g., in a wet lab).

As the use of such NUI-enabled devices becomes more convenient and prevalent, the risk of sensitive information being exposed increases. NUI-enabled devices can be used to authorize sensitive transactions. For example, a smartwatch [1] can now be used as a digital wallet to enable an e-payment transaction. Smart speakers now facilitate e-commerce transactions, and more devices with such features are being added regularly. A convenient and robust authentication mechanism is needed to protect sensitive information and to prevent such devices from being used by unauthorized individuals.

Most early computing devices had a keyboard as the input interface. Consequently, alphanumeric passwords became the *de facto* authentication standard. The security and usability problems associated with passwords have since been extensively studied [2], [3]. However, new security and usability issues arise if one attempts to mimic keyboard-based password entry on NUI-equipped devices. For example, touch interfaces are not designed for typing-oriented applications. Research has shown that creating and entering a text password on a virtual keyboard on a touchscreen is much more difficult and, therefore, slower [4], [5] due to "fat finger" errors [6], user distraction [7], and limited touch precision [8], [9]. Therefore, it comes as no surprise that most users prefer unlocking mechanisms that are not based on an alphanumeric password on their multi-touch devices [10].

Authentication techniques based on NUIs can leverage the unique capabilities of a device interface, such as the ability to record higher-dimensional data. In traditional interfaces only a sequence of keystrokes, mouse pointer movements and mouse-click events are typically captured. In contrast, a modern touch surface can capture multiple touch locations as well as touch area, touch pressure, and touch shape. A microphone can capture a voice signal at a very high sampling rate. A camera can record images of user interaction, using the whole body or a part thereof, such as the hand. Furthermore, in addition to extracting a user's knowledge ("what you know"), NUIs can also capture biometric information ("what you are") thereby facilitating a variety of multi-factor approaches to authentication. The captured information can be used not only

N. Sae-Bae is with the Department of Computer Science, RMUTSB, Nonthaburi, Thailand, e-mail: (napa.s@rmutsb.ac.th, benapa@gmail.com)
N. Memon is with New York University, NY, USA.
J. Wu, J. Konrad and P. Ishwar are with Boston University, MA, USA.
Manuscript received September 24, 2018; revised December 17, 2018.

Fig. 1: Examples of devices equipped with natural user interfaces

on a local device but potentially over the Internet, thanks to the effort to standardize web authentication protocols under the FIDO alliance [11].

To date, numerous studies have been performed to assess the benefits of natural interactions in authentication. However, many of the proposed authentication schemes were developed for devices whose usage pattern has changed over time and the possible threats and vulnerabilities were not fully understood at the time of original publication. In response to new adversarial threats, some authors have re-evaluated the merits of their original schemes and, consequently, several variants have been published over time. In order to sort out this wealth of methods, a number of surveys have been published to date. These surveys are typically limited to either a specific mode of access, or a specific type of authentication input. For example, Bonneau *et al.* [12] published a comparative study of authentication schemes for web or remote access. The schemes considered were primarily limited to textual or graphical passwords. Biddle *et al.* [13] and Suo *et al.* [14] surveyed a number of graphical password approaches. Similarly, Bhattacharyya *et al.* [15] and Unar *et al.* [16] reviewed a number of popular and trending biometrics used for authentication.

The purpose of this paper is to review point-of-entry user-device authentication on NUI interfaces. This review complements previous work in two ways. First, it proposes a taxonomy of authentication methods along three axes:

- *sensor* type used to capture user input,
- *actuator* type (body trait) a user applies during interaction,
- *credential* type used for authentication.

Second, it provides a survey of fixed "point-of-entry" authentication methods[1] on NUI devices while clustering them into distinct categories in the *sensor-actuator-credential* coordinates. Then, for each category, we briefly describe methodologies typically used, and discuss security (performance, threats, etc.) and usability (convenience, use in public spaces, etc.). We also point out avenues for future research within each category with the goal of identifying and highlighting areas where further studies are needed. Another potentially-useful byproduct of this survey is the compilation of information about public datasets collected for user authentication research, that we have published on-line[2] [18].

This survey does have its limitations. First, not every method surveyed is exhaustively evaluated against every criterion discussed. This is mainly due to the non-availability of such results in the surveyed literature. However, this points to future research opportunities which could "fill-in" missing pieces. Second, this survey leans more on being descriptive than prescriptive. The aim is to systematize the study by grouping different systems using the proposed taxonomy. Thus, no attempt is made to establish a gold-standard benchmark against which all emerging NUI-based authentication methods can be evaluated. Candidates for such a benchmark could be fingerprint and face recognition systems or PIN- and pattern-lock systems. However, they may not be viable for a specific NUI under consideration. For example, face recognition would not be an option when considering an authentication mechanism for a smart speaker. Furthermore, these systems, are widely used, and their strengths and vulnerabilities have been extensively studied [19], [20]. Hence, comparisons against them can be made by a system designer when choosing a scheme. This survey focuses on the emerging NUI-based authentication systems that have not been as well studied yet.

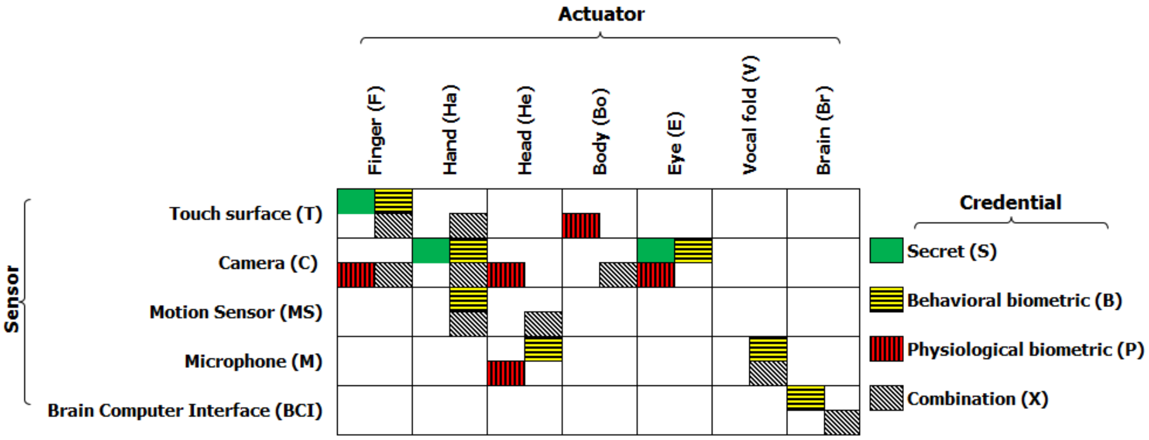
The remainder of this paper is organized as follows. We begin by describing the proposed taxonomy in Section 2. Next, in Section 3, we discuss a common evaluation framework, focusing primarily on how security, usability, and performance pertain to authentication mechanisms. Then, from Section 4 to Section 8 we review authentication methods with each section focusing on one sensor type: touch surface, camera, motion sensor, microphone, and brain computer interface (BCI). In each section, we group the methods by actuator and credential type, and discuss methodology used, security and usability strengths and weaknesses, as well as future research directions. Finally, in Section 9, we summarize our survey in two tables, one focused on functionality and one focused on performance, and discuss viable research directions for the near future.

2 A Taxonomy for NUI-based User Authentication

NUI-based user authentication can be viewed as a process where a user's identity is verified using a *credential* input gleaned from an *actuator* action that is captured and recorded by an NUI *sensor*. Hence, in this survey, NUI-based user authentication systems are classified based on three distinct components: sensor, actuator, and credential, as shown in Fig. 2.

[1] Many of the methods we review are also applicable to continuous authentication, where a user is continuously monitored and authenticated throughout the duration of a session. We refer the interested reader to a recent survey by Patel *et al.* specifically focused on continuous authentication for mobile devices [17].

[2] http://isis.poly.edu/~benapa/authen_dataset_index.html.



Sensor \ Actuator	Finger (F)	Hand (Ha)	Head (He)	Body (Bo)	Eye (E)	Vocal fold (V)	Brain (Br)
Touch surface (T)	🟩 ▤ ▨	▨		🟥			
Camera (C)	🟥 ▨	🟩 ▤ ▨	🟥	▨	🟩 ▤ 🟥		
Motion Sensor (MS)		▤ ▨	▨				
Microphone (M)			▤ 🟥			▤ ▨	
Brain Computer Interface (BCI)							▤ ▨

Credential

🟩 Secret (S)

▤ Behavioral biometric (B)

🟥 Physiological biometric (P)

▨ Combination (X)

Fig. 2: A taxonomy of user authentication approaches – actuator versus sensor versus credential

2.1 Sensor

We consider five types of NUI sensors: touch surface, camera, motion sensor, microphone and brain-computer interface. Below, we briefly describe each sensor type.

- **Touch surface (T):** Most touchscreen surfaces have an overlaid array of sensors, usually capacitive or resistive. They are used to capture a touch gesture, e.g., a fingertip and hand motion on a 2-dimensional surface. The recorded signal is a time series of touch-point attributes, such as $x-y$ coordinates, time stamps, and sometimes (depending on the surface) pressure information. They are commonly used today in smartphones, tablets, smartwatches [21], large interactive displays [22], and more recently in multi-touch coffee table surfaces [23] and augmented-reality headsets [24].
- **Camera (C):** Color cameras are widely available in commodity devices such as smartphones, tablets, smartwatches, laptops, and headsets (e.g., Google Glass). A camera can serve as a sensor to capture either an image or a video representing a variety of input credentials from different actuators. Hand gestures recorded by a camera have been used as an interface to execute various tasks as well as for authentication [25]–[28]. A depth or 3-D sensing camera can capture precise volumetric information from gestural interaction [29]. Depth cameras are widely used as a 3-D control interface. For example, the Kinect depth sensor allows tracking of human body parts, while the Leap Motion sensor allows tracking of one's fingers.

  A camera can also be used as a gaze- or eye-tracking interface; to control a device, a user needs to merely change gaze direction, which is a hands-free operation. Such hands-free interfaces are expected to become more common in consumer electronic devices and smart vehicles [30], [31]. In general, there are two types of hardware for eye trackers: wearables and stationary (fixed) ones [32]. To accurately capture a user's eye movement, the sensor hardware usually includes a near-infrared emitting diode and optical imaging sensors.

  Finally, a camera can be used to capture a user biometric, such as face, iris or fingerprint.
- **Motion sensor (MS):** In addition to cameras, accelerometers and gyroscopes have also been used to capture hand and body gestural information [33]. These two sensor types are frequently embedded into various portable devices ranging from smartphones, tablets, smartwatches, instrumented gloves and augmented-reality headsets to wireless game controllers, such as those used in the Nintendo Wii, Microsoft Xbox and Sony PlayStation. While an accelerometer detects linear acceleration of the device, a gyroscope detects its rotation around a fixed axis. Together, these sensors can be used to determine the orientation, movement, and position of a device in 3-D space [34], [35].
- **Microphone (M):** A microphone captures voice and thus allows hands-free interaction with a device. It has been used as one of the main communication channels between users and devices in several wearable-device categories including smartwatches (e.g., Samsung Galaxy Gear, Sony SmartWatch), augmented-reality headsets with a heads-up display (e.g., Google Glass, Vuzix m100) and many IoT devices. In addition, the interface has been embedded into many portable devices (e.g., mobile phones, GPS units, tablets, laptops) and smart-home units (e.g., Amazon Alexa, thermostats [36]).
- **Brain-computer interface (BCI):** Brain-computer interface is a communication channel between a user and a machine [37] that does not require any physical movement. Currently, electroencephalography (EEG) is used to measure an electrical signal of the brain's activity. The EEG signal is a multi-dimensional waveform captured by a set of electrodes operating at a sampling frequency of several hundred Hertz. This technology forms the basis of low-cost commercial-grade BCI products such as Neurosky's MindWave [38]. Intel is also preparing to offer a BCI product in the near future [39]. In addition, tiny and unobtrusive earbuds can also be used to record EEG data which, in turn, can be sent wirelessly to a smartphone and used for interaction [40].

2.2 Actuator

In a user authentication system, different parts of the human body may be used as an actuator to interact with a NUI sensor to present an authentication credential. We consider seven types of authentication actuators currently in use or under consideration, namely a finger, hand, head, body, eye, vocal fold (voice), and brain (thought). We discuss them below.

- **Finger (F):** A gesture executed by a finger is a non-verbal interaction mechanism that can be realized *via* a touch surface, camera, or motion sensor [41], [42]. On a touch surface, fingertips' positions and, sometimes, pressure information are detected *via* an array of sensors, usually capacitive or resistive. When interacting with a camera, fingertips' positions are extracted from a series of image frames of a hand silhouette (if a 3-D camera is used, depth information for each pixel is also retrieved).
- **Hand (Ha):** A hand gesture is also a non-verbal interaction mechanism that can be captured by a touch surface, camera, or motion sensor. Typically, there are two types of hand gestures: stationary hand pose and hand motion. They are usually captured by a color (RGB) or color-and-depth (RGBD) camera, but can also be registered by a touch surface or a motion sensor [43], [44].
- **Head (He):** Head movement is another non-verbal interaction that can be registered indirectly by a camera or directly by motion sensors [45]. While a motion sensor attached to the head can directly measure head orientation and velocity, they can only be indirectly estimated from the video stream captured by a camera by using appropriate algorithms, e.g., optical flow for velocity and detection of facial landmarks for orientation.
- **Body (Bo):** Body gesture, a movement of the torso, head and body limbs, is yet another type of interaction typically captured indirectly by a color or color-and-depth camera [46], or directly by a motion sensor [47]. Again, since such movements are not directly measurable by a camera, algorithmic solutions are needed to recover motion parameters from the recorded video stream , e.g., by detecting and tracking joints or "interest-points" using random forests as in the Kinect sensor.
- **Eye (E):** In addition to being a source for a physiological biometric, such as an iris pattern, eye's movements and gaze can be tracked using a special camera. This can provide a fast and convenient way for users to interact with the system [48]. The idea of using gaze movement as an authentication factor has been around for several years [49], [50]. The most common information extracted from both low-cost and expensive camera eye-tracking systems is a time series of $x-y$ coordinates of the centers of the left and right pupils. Other information may be extracted as well, including pupil diameters.
- **Vocal fold (V):** Voice produced by vocal folds (cords) is a natural interaction mechanism that we use on a daily basis. Advances in speech recognition technology have made it possible to use voice as a human-computer interface and also as a speaker recognition mechanism [51].
- **Brain (Br):** Neural activity in the brain, that is loosely referred to as thought, can be directly recorded using EEG [52]. Both unintentional "background" thought processes, with their user-specific characteristics, as well as intentional thoughts can be used to authenticate users.

2.3 Credential

Typically, the credentials provided by a user can be broadly classified into one of three types: "something you know", "something you are", or "something you have". The last one is considered inconvenient as it requires users to carry an object (e.g., swipe card, proximity card) and present it to the system to gain access. This is further complicated by the fact that this object can be misplaced or lost. Authentication credentials that do not require a user to carry anything can be classified into one of the following categories: secrets, behavioral biometrics, physiological biometrics, and any combination thereof. Each type is discussed below.

- **Secret (S)**: Secret, or user-specific knowledge, could be used as an authentication credential. One of the most common forms of this credential in use today is a text password entered *via* a physical or virtual keyboard. However, with NUI-enabled devices, secret credentials can also be input by other mechanisms such as a touch pattern, gaze or thought.
- **Behavioral biometrics (B)**: The emergence of NUIs has enabled new types of user interaction. Many studies have indicated that the characteristics of these interactions (e.g., swipes, gestures, signatures, speech) are significantly distinct among individuals and can serve as a behavioral biometric. Therefore, these patterns have been proposed for use as an authentication factor.
- **Physiological biometrics (P)**: There are several authentication approaches that do not rely on behavior but instead use physiological biometric information (e.g., fingerprint, iris, face) captured through the interfaces described in Section 2.1.
- **Combination (X):** Various combinations of a secret, behavioral biometric and physiological biometric can also be used to form a multi-factor authentication system.

3 Evaluation criteria in authentication systems

The ultimate goal of user authentication on a device is to prevent unauthorized access. This requires the authentication mechanism to be secure against relevant threat models. However, no matter how secure the system is, it will not be effective if it is difficult to use and hence not adopted by users. Thus, it is widely accepted that both the security and the usability of an authentication system need to be investigated. However, authentication is not limited to the scenario where a user is sitting in front of a desktop computer. For mobile or wearable devices, the context of use (or operational characteristics) and environmental constraints at the time of authentication are additional important factors to consider when evaluating a mechanism. That is, a particular scheme may be very effective when the user is stationary but may not be as effective when performing a physical activity, such as jogging. Similarly, schemes may or may not be effective in the presence of loud noise, poor lighting conditions, etc.

One recent framework for evaluating authentication schemes was proposed by Bonneau *et al.* [12], with a focus on evaluating web-based authentication. In their framework, a number of security, usability, and deployability metrics were identified. These metrics can be divided into two groups: metrics that correspond to the authentication mechanism and metrics that relate to aspects of device-server communication. In this survey, since we are focussed on point-of-entry user-device authentication, we use the metrics that are directly associated with the authentication mechanism and ignore those associated with device-server communication. Furthermore, in terms of usability evaluation, we align metrics from Bonneau *et al.*'s framework with ISO definitions of usability. We propose four context-of-use characteristics that affect usability, with a focus on user attention and environmental requirements which are not included in their framework. Note that, there are other criteria that need to be considered when selecting an authentication mechanism, e.g., frequency of use and deployment constraints. However, these are application-dependent and are not specific to the mechanism. Therefore, they are not included in the evaluation criteria here.

3.1 Security

The security of an authentication system refers to its ability to thwart unauthorized access. Below, we consider three threats derived from Bonneau *et al.*'s work. To make the correspondences clear, we label the elements of the threat model considered by Bonneau *et al.* as S1 to S9 following the nomenclature used in their paper.

- *Random guessing*: In this scenario, an attacker does not have access to information about the authentication credentials and attempts to guess the appropriate input credential to gain access to the system. Guessing could be random or based on a dictionary. Guessing is one of the more common, and simpler, threats to an authentication system. This is consistent with S3 (Resilient-to-Throttled-Guessing). We do not consider S4 (Resilient-to-Unthrottled-Guessing) as throttling is very easy on a NUI-equipped device and is almost always implemented.
- *Targeted impersonation*: Here an attacker attempts to impersonate a victim by exploiting knowledge about the victim's personal information. This type of attack is applicable to authentication systems that use information from the "what you know" and "what you are" categories. Knowledge-based approaches ("what you know") are vulnerable if users pick weak passwords that are easy to guess. For example, many people use their date of birth as their personal identification number (PIN) [53]. Similarly, authentication using physiological biometrics, like a fingerprint or face ("what you are"), are also vulnerable to targeted impersonation which is commonly known in the biometrics community as a spoofing attack [54]. This issue occurs because many physical biometrics are available either as public information (face), or can be inadvertently left behind (such as fingerprint on a surface). This is consistent with S2 (Resilient-to-Targeted-Impersonation)
- *Physical Observation*: Here an attacker learns an authentication credential by observing (at the moment of entry or later) the credential that is entered into the authentication system, with or without the aid of a digital recording equipment. In the context of mobile devices, this is called the "shoulder-surfing attack". Another example of an observation attack, that is often part of the threat model for biometric authentication, is the so-called "presentation attack" [55]. In this scenario, an attacker has an exact copy (sometimes a digital one) of an authentication credential or interaction that can be re-used ("played back") to gain access to the system. A common instance of this type of attack is to present facial image or video of a person or a fake silicone or gelatin fingerprint to gain illegitimate access to a system. This is consistent with S1 (Resilient-to-Physical-Observation)

As for S10 (Requiring-Explicit-Consent), most authentication schemes listed in this survey do provide this security benefit as users are required to consciously interact with the interface. Exceptions are schemes in which an authentication factor can be captured involuntarily, for example in token-based schemes (which we do not consider in this paper) as well as some of the schemes that use physiological biometrics like traditional face and fingerprint based authentication.

Note that, S5 (Resilient-to-Internal-Observation) refers to the threat where an attacker impersonates a user by intercepting the user's input from inside of the user's device. Examples include acquiring authentication credentials from sensors at the authentication moment or stored information (by installing keylogging malware, gaining root access, etc.) or by decoding authentication credentials from a signal captured from different sensors (side channel attacks). In this paper, a discussion of this threat is omitted as it is more related to security and functionality of the operating system in-use and the implementation of authentication schemes, rather than the design of an authentication scheme itself.

Also, some security criteria listed in Bonneau *et al.*'s framework are not considered here since they are not directly related to user-device authentication. In particular, S6 (Resilient-to-Leaks-from-Other-Verifiers), S7 (Resilient-to-Phishing), S9 (No-Trusted-Third-Party), and S11 (Unlinkable) are not considered since they are more related to web authentication. We also omit S8 (Resilient-to-Theft) as it is related to token-based authentication.

3.2 Usability

Usability is an important property as it relates to how easily users can adapt to the system. The usability traits we cover below are in agreement with the three dimensions of ISO's broad definition of usability: *effectiveness* (accuracy and completeness with which users achieve specified goals), *efficiency* (resources expended in relation to the accuracy and completeness with which users achieve those goals) *and satisfaction* (freedom from discomfort, and a positive attitude toward using the product), and *context of use* (users, tasks, equipment, e.g., hardware, software and materials, and the physical and social environments in which a product is used) [56]. Most of these properties (except for the context of use) are consistent with Bonneau *et al.*'s framework that includes additional properties that do not apply to this survey.

- *Effectiveness*: In user-device authentication context, effectiveness refers to how easily an authorized user can gain access to the device and whether the system can renew a compromised authentication credential. Criteria from Bonneau *et al.*'s framework that fall into this ISO metric are the following:
  - U1 (Memorywise-Effortless), or *Memorability*, refers to the cognitive burden imposed on users in order to use an authentication mechanism. This burden can be compounded when a user needs to remember credentials across multiple systems and devices. Memorability is important to consider whenever an authentication mechanism requires a shared secret between the user and device. In this work, this occurs for interactions that require a "what you know" (knowledge-based) mechanism. For interactions that are purely "what you are", there is little cognitive burden. Memorability can be measured by computing a recall rate by conducting a multi-session experiment where users first create a credential and then recall the same in subsequent sessions on different days.
  - U7 (Infrequent-Errors), or *True acceptance rate*, refers to how often a system grants access when an authorized user performs a correct interaction. This is a particularly important property for interactions that rely on biometric information. In biometrics, a captured biometric trait is never exactly the same as the one enrolled in the system. As a consequence, a trait that is not equal to any of the enrolled traits, but is sufficiently close to one of them (within a threshold) needs to be accepted. This discrepancy results in two types of possible errors: false acceptance and false rejection. While the first relates to security (unauthorized access), the latter relates to convenience and usability as a user may have to provide the input credential multiple times and may even be completely rejected and locked out if the number of attempts exceeds the permitted number due to throttling.
  - U8 (Easy-Recovery-from-Loss) or *Changeability* refers to how easily an authentication credential can be changed or renewed once compromised. In authentication mechanisms that rely on physiological biometric information, e.g., face or fingerprint recognition systems, changeability is an important usability concern that also affects security. Typically, one of two types of biometric information can be compromised: the original biometric trait or a biometric template (the encoding of the original biometric trait) [57]. For example, in fingerprint recognition systems either the original fingerprint image or a fingerprint template can be compromised. Authentication systems that utilize changeable components, e.g., signature or custom gesture, can be renewed in the event of a data breach by enrolling new samples of the credential.
- *Efficiency and satisfaction*: Since there is an implicit economic cost involved in deciding whether or not to use a security mechanism [58], efficiency and satisfaction are important properties to evaluate for any authentication system. Criteria from Bonneau *et al.*'s framework that fall into this ISO metric are the following:
  - U4 (Physically-Effortless) refers to the amount of effort a user spends in order to provide an input credential to the authentication system.
  - U5 (Easy-to-Learn) refers to the amount of time and effort a user spends in order to figure out how to enroll a sample and to provide an input credential to the authentication system.
  - U6 (Efficient-to-Use) refers to the amount of time a user spends in order to provide an input credential to the authentication system.

  The measurements that have been used to quantify effectiveness, efficiency and satisfaction can be divided into two categories: 1) user perception, e.g., the perceived ease of performing an authentication interaction, and the perceived time for successful authentication, and 2) statistical metrics, e.g., the actual time required for successful authentication, and the Failure-To-Enroll rate (FTE). In this regard, the reliability and validity of any usability study are important to consider, particularly when incorporating user perception to evaluate this usability factor [59].
- *Context of use:* Authentication systems that require special user attention or particular environmental conditions would have limited utility if the usage scenario differs from the one it was designed for. Therefore, the context of use is an important criterion to evaluate authentication systems. Below, we identify four context-of-use constraints that may affect the usability of an authentication system.
  - *Eyes-free use:* Authentication systems that require a user to examine information presented on a display obviously violate this constraint. Such authentication systems could potentially pose a limit to the user's freedom of movement [60] or attention to another task being performed while authenticating. For example, recognition-based graphical password systems require attention focused on the image(s) being presented on the screen and hence cannot be performed in an eyes-free mode. Examples of input credentials that can be used for eyes-free authentication are voice, hand gestures, and brain signals.
  - *Hands-free use:* During some activities, such as driving or exercising, users cannot use hands to perform an action needed for authentication. Examples of input credentials that can be used for authentication without the use of hands are eye-gaze, voice, and brain signals.
  - *User state and environmental constraints:* In practice, users may want to authenticate while performing different activities and in different contexts. For example a user may be driving a car or may be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

in a loud bar. Therefore, user state, such as user movement when performing actions required by a given authentication scheme, is an important criterion to determine the usability or applicability of an authentication mechanism. An authentication mechanism that requires a user to be stationary cannot be used while running.

In addition, various environmental factors can negatively affect the quality of a signal used in authentication, thus increasing errors. In voice biometrics, the level of background noise captured by a microphone is one such factor. In image- and video-based biometrics, lighting conditions may have an adverse impact, while in biometrics that require capacitive sensors, like fingerprints, humidity can adversely affect the captured signal. Also, voice-based authentication often needs a quiet location that can also tolerate auditory disturbances. For example, a library is quiet but not tolerant of disturbance caused by voice authentication. However, a park can be quiet and also tolerant of such disturbance.

Since many emerging devices are designed to be used in uncontrolled environments, scenarios and contexts, it is important to consider the above context-of-use factors when evaluating an authentication mechanism.

Note that U2 (Scalable-for-Users) and U3 (Nothing-to-Carry) of Bonneau *et al.*'s framework are not covered here as U2 is more related to memory interference between authentication credentials and not intrinsically related to any authentication system, while U3 is more related to token-based authentication, which is not covered in this paper.

Apart from the security and usability criteria laid out above, there can also be privacy concerns raised when an authentication technique uses physiological biometrics and perhaps even behavioral biometrics. Several issues related to privacy have been raised in the literature for physiological biometric authentication as this information could be viewed as privacy-invasive. For instance, it could be used to match a user record from one application with the record from another application that uses the same biometric modality [61], [62]. Furthermore, it has been shown that biometric samples can possibly reveal other personal information of a user, e.g., gender, age, or even user's genetic or health conditions [61]. Several techniques to address such concerns have been proposed in the literature [63], [64]. In this survey, we do not look at privacy aspects as the solutions that have been proposed do not affect the usability properties and the threat models that we consider.

In the next five sections, we review a multitude of authentication systems organized according to the taxonomy presented in Section 2. Each section focuses on authentication systems that leverage one sensor type, i.e., touch surface, camera, motion sensor, microphone or brain-computer interface. Within each section, methods are grouped by the actuator type (body part used to perform interaction) and input provided to the system (secret, behavioral biometric, physiological biometric or combination thereof). For each sub-group of methods we discuss security considerations and usability constraints; and for each group we discuss avenues for future research.

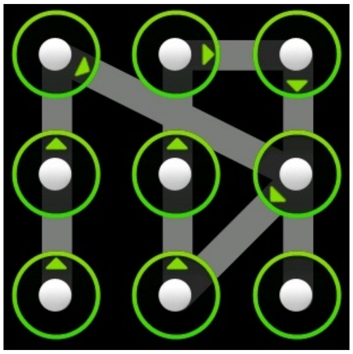


Fig. 3: Android *Pattern Lock* authentication

4 T-∗-∗ Schemes: Touch surface based authentication

A touch interaction produces one or more time series of touch-point coordinates and, for some surfaces, pressure information. Approaches to authentication using either the raw signal or extracted features have been developed.

A wide variety of interactions are possible on a touch surface, e.g., typing on a virtual keyboard, tapping, drawing a shape, or dragging one or more fingers across the surface. On larger surfaces, more sophisticated multi-touch interactions are also possible – surfaces can detect multiple fingers at a time using one-hand or two-hand gestures. There has been some recent work that has also demonstrated the use of pressing body parts, such as an ear, against a touch surface to capture, albeit weak, physiological biometrics.

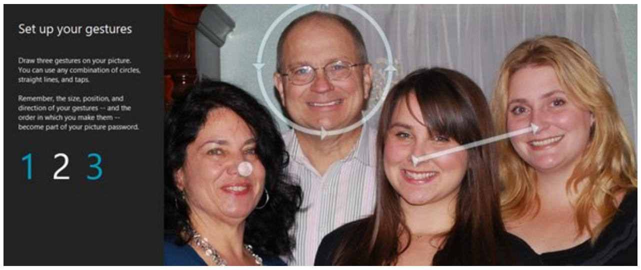
As far as the credential input is concerned, a secret pattern could be entered, behavioral biometrics corresponding to the traits that influence the input can be captured, and physiological characteristics of a body part pressed against a touch surface (e.g., ear shape) can be recorded. Of course, some combination of these inputs can be also used for authentication. Hence with respect to the proposed NUI authentication taxonomy, the categories T-F-∗, T-H-∗, and T-Bo-P are viable.

The rest of the section describes a few representative schemes in each of these taxonomy classes and discusses their strengths and weaknesses.

4.1 T-F-S Schemes: Finger-Touch Techniques Based on a Secret

Perhaps the most widely explored techniques for user authentication on a touch surface involve using one or two fingers. The simplest example of such interaction is when a text password or a PIN is typed on a virtual keyboard [65]. This would be an instance of a T-F-S scheme as the input is a secret. However, studies have shown that typing on a virtual keyboard is slower and harder than on a physical keyboard [5] due to the fat-finger problem [6], user distraction [7], and limited touch-input precision [8], [9]. The Android *Pattern Lock* is another example of a T-F-S scheme. It is a special case of a broader class of "Draw-A-Secret" schemes proposed by Jermyn *et al.* [66] in 1999 for use with a mouse interface, but it is now widely

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science



Set up your gestures

Draw three gestures on your picture. You can use any combination of circles, straight lines, and taps.

Remember, the size, position, and direction of your gestures -- and the order in which you make them -- become part of your picture password.

Fig. 4: Microsoft *Picture Password* authentication

adopted in touch interfaces. In this scheme, a user is asked to create and memorize an exact dot-connected pattern on a 3×3 grid, as shown in Fig. 3.

A T-F-S scheme based not only on text, digits, or patterns, but also on points selected on an underlying picture, called *Passpoints*, was first presented in [67]. The basic motivation for such an approach was that humans are better at remembering visual information as opposed to recalling alphanumeric secrets or even patterns drawn on a grid [68]. The underlying picture serves as a cue for a user to recall the points that represent the password.

There have been many improvements and variations of *Passpoints* proposed in the literature. Most of them can be characterized as T-F-S schemes. Perhaps the most widely used one is Microsoft's *Picture Password* introduced in Windows 8 (Fig. 4). In this case, a user is asked to perform an ordered sequence of at least three gestures, each being either a tap, a line, or a circle, on a background image chosen by the user. A tap gesture is verified if and only if the distance between the enrolled point and the one captured during authentication is less than a predefined threshold. The starting and ending points of a line gesture and the center point of a circle gesture are verified in a similar manner.

T-F-S schemes can also be implemented when a touch-surface area is either very small or not overlaid on a display. For example, gesture-combination lock on Google Glass authenticates employs very simple touch gestures. A user is required to perform an ordered sequence of four touch gestures on a tiny touch surface located on the right of the device. There are eight gestures from which a user can select: swipe forward, two-finger swipe forward, swipe back, two-finger swipe back, hook swipe (swipe forward then back in the same motion or vice versa), two-finger swipe, tap, and two-finger tap [69]. T-F-S schemes on other small devices, e.g., smart watches, can employ a similar approach as used on phones, e.g., PIN or pattern lock. A good survey of authentication schemes on smart watches is provided in [70].

Security: T-F-S schemes have some major drawbacks from the security point of view. Perhaps the most notable one is the weakness of the secret used for authentication, leading to susceptibility to *random guessing*. A study by Bonneau *et al.* [53] showed that some PINs are much more popular than others, e.g., "1234". While blacklisting the most obvious PINs has shown some promise in blocking random-guessing attacks, it is barely effective in preventing *targeted attacks* where date of birth is chosen. An alternative is to assign the password or PIN randomly, but random PINs are difficult to memorize and recall.

Similarly, with respect to "Draw-A-Pattern" schemes, such as the Android *Pattern Lock*, a study by Uellenbeck *et al.* [71] has shown that the 3x3 lock screen used in Android devices is weaker than a three-digit randomly-assigned PIN. For example, most users choose the upper left point as the starting point of their unlock pattern. Furthermore, unlock patterns themselves are very predictable and conform to a small set of familiar shapes. To address this problem, they propose a modification to the grid layout to diversify lock pattern choice. However, the usability of such a technique has not been tested. Siadati *et al.* [72] proposed two other persuasive approaches: using a pattern-strength meter, similar to that of a password-strength meter, and blinking a dot to suggest a starting point during enrollment that would make the first location more uniformly distributed.

However, the entropy of graphical passwords gets affected by the very predictable choice users make given a picture. Effectively, this entropy is much smaller than the theoretical limit due to salient parts of an image that users are more likely to select (hot spots) [73]–[75].

In addition, some T-F-S schemes are vulnerable to *targeted attacks.* For example, it is known that users often choose their date of birth as their PIN. Furthermore, Zhao *et al.* [76] found that users frequently choose family photos, and then, predictably, proceed to perform gestures on their faces. Attackers can mimic this behavior and increase the likelihood of unauthorized access.

Another major drawback of T-F-S schemes is their vulnerability to *physical observation* (or shoulder surfing). PINs, passwords, pattern locks, graphical passwords can all be observed by an attacker. This is especially true for mobile devices that are often used in public. In addition to observing from close proximity, an attacker may also observe someone remotely by using recorded material that was collected intentionally or even unintentionally. For example, unintentional recording could result from a surveillance camera that captured the authentication event. In the case of a smartphone, it has been demonstrated that a PIN entry can be automatically recovered from a video recording made using a Google Glass, hand-held recording devices, or a public surveillance camera [77], [78].

One defense against *physical observation* is to place the attacker at a cognitive disadvantage [79]. For example, when entering a graphical password, the image could be blurred so that an attacker, who has no prior knowledge of the context, is unable to make sense of the image content, but the owner, who has prior knowledge, can potentially recall it and relate it to the secret. Similarly, one proposed solution for PIN entry is to use a randomized keyboard layout as opposed to a fixed one. This could be used in conjunction with *IllusionPIN* where a user and an attacker looking at the screen from different distances see different layouts [80]. This, however, almost doubles the time that a user has to spend typing when compared to a fixed keyboard [77], [78].

Another type of defense suitable for a knowledge-based authentication mechanism is to use a challenge-response approach where a user only presents partial information about a shared secret during authentication [81]. This again results in a significant increase in authentication time and hence much reduced usability.

A security weakness of Android *Pattern Lock* resulting from *physical observation*, in addition to shoulder surfing, is its vulnerability to a “smudge attack”. A 2013 study by Andriotis *et al.* [82] has shown that physical markings on a device, e.g., oily residues or smudges left on the touch screen, along with the knowledge of pattern distributions, can be used to recover 54.54% of lock patterns.

Usability: T-F-S schemes have been widely adopted due to their *effectiveness, efficiency and satisfaction*. T-F-S systems achieve high true-positive rates and yet have high simplicity. They are memory-wise and physically effortless, easy to learn, changeable and take little time to perform. One study [10] indicated significant user preference for the Android pattern lock mechanism as opposed to alphanumeric passwords with 78% preferring the former. This was supported by, another study that investigated the relation between user’s handedness as well as his/her native written language, and user’s perception in terms of usability and effectiveness of a pattern [83]. User studies performed for *Passpoints* by Weidenbeck *et al.* [67] with a small group of users gave positive results but they were completed before the advent of touch surfaces.

However, T-F-S schemes, except possibly gesture-combination lock (on Google Glass), have limitations in terms of *context of use*. They cannot be used in an eyes-free mode or hands-free mode. They also may be difficult to perform while moving at moderate to fast speeds. However, typically they would not be sensitive to ambient conditions such as light and noise. Finally, the usability of each of these touch-surface based authentication techniques may depend on the surface size being used.

4.2 T-F-X Schemes: Finger-Touch Techniques Based on Behavioral Biometrics and a Secret

As mentioned above, the biggest drawbacks of T-F-S techniques are the weakness of the secret and the vulnerability to *physical observation*. The most widely-used approach to address these issues is to capture behavioral biometric traits along with the input. When the input is not a secret, this results in a T-F-B technique. When it is a secret, we get a combination (X) of a secret and a behavioral biometric, resulting in a T-F-X scheme.

A simple example of a T-F-X scheme is when information from a user’s touch-typing style, while entering a password on a virtual keyboard, is used for authentication. In particular, Ben *et al.* [84] have shown that typing characteristics, such as the specific location pressed on a virtual key, the drift from finger press to release, the force and pressure of the touch, and the surface area of the touch, can be used as features to differentiate users. Similarly, another study conducted on 16 subjects demonstrated that typing styles of text messages are unique to individuals and are consistent over time [85]. However, when a password is only a few characters long the authentication error of the system is very high: 32.3% false acceptance rate (FAR) at 4.6% false rejection rate (FRR) for five key presses [84].

To protect Android *Pattern Lock* against *physical observation*, Luca *et al.* [86] studied if a user’s drawing behavior could also be used as an authentication credential. Specifically, the drawing pattern, represented by a time series of $x - y$ coordinates, was compared against an enrolled pattern using a signal-based approach, namely *via* Dynamic Time Warping (DTW). They achieved 77% accuracy with 19% FRR and 21% FAR in a 31-participant study.

One way to incorporate user traits into the ubiquitous PIN-based methods on touch surfaces is to ask the user to draw the PIN instead of typing it. The resulting method checks the correctness of the PIN using a digit recognizer, as well as the drawing style of the digits to authenticate specific users, resulting in a T-F-X technique [87], [88].

Security: The security of T-F-X systems is stronger than that of secret-based authentication as now a biometrics factor provides an additional layer of protection against *random guessing* as well as *targeted impersonation*. Clearly, it is also harder to imitate the behavior of a user learnt from *physical observation*. For example, the *Draw-A-PIN* system was tested against an attacker who wrote the same PIN as the user. A 4.84% EER was reported on a private dataset comprising 20 genuine users and 25 attackers, where the content of a user’s PIN was known to the attackers but not the drawing [88]. An evaluation was also performed where an attacker was provided a recording of user’s PIN drawing and asked to mimic the drawing behavior. In this setting, a 14.11% EER was reported on the same dateset [88].

Usability: T-F-X techniques inherit much of the *effectiveness* (in terms of memorability and changeability) as well as *efficiency and satisfaction* benefits from their original T-F-S counterparts that use only the secret to authenticate users. The second factor, which is the behavioral biometric, is captured without adding a burden as it is derived from the interaction used to enter the secret. But the *effectiveness* of this type of system in terms of True Acceptance Rate could also be negatively impacted due to recognition errors. The authors of [87] argued that PIN drawing could prove to be acceptable and usable by leveraging user familiarity with PIN authentication. However, the drop in authentication performance could also affect usability if the system favors false rejections over false acceptances (security over convenience). Another factor that makes them less usable than T-F-S techniques is that they may not be easy to understand. The user may not have a proper understanding of the “behavior” that is being captured and recorded and may change it (for example, draw a pattern or press a key slower than usual) and be surprised by the negative authentication that results.

Regarding *context of use*, similarly to T-F-S schemes, T-F-X techniques cannot be used in eyes-free or hands-free environments. The ability to authenticate while performing another activity such as walking is further constrained as compared to T-F-S schemes due to the possibility of increased recognition errors. The schemes, however, remain insensitive to ambient conditions such as light and noise.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

4.3 T-F-B Schemes: Finger-Touch Techniques Based on Behavioral Biometrics

Behavioral biometrics, with no underlying secret, can also be used in authentication based on finger touch. One example of a T-F-B technique is drawing a signature on a touch-sensitive display using a fingertip. Since the signature itself is not a secret, authentication is performed purely based on the manner in which the signature is drawn, known as on-line signature recognition. (This scheme can be also viewed as T-F-X if the signature is a secret.) This approach has been a well-established authentication technique for decades although in different contexts.

On-line signatures can be verified using both signal-based and feature-based approaches. Examples of well-known signal-based approaches include DTW [89]–[91] and Hidden Markov Models (HMMs) [92]. The major difficulty for a feature-based approach is identifying a good set of descriptive features that can verify an on-line signature effectively and efficiently. These features include, but are not limited to, the total duration of signature, number of pen-up movements, sign changes of derivatives dx/dt and dy/dt, HMM model parameters [93], and histogram features [93]–[97].

Security: With *random guessing*, where user signatures are not known to the attacker, the authentication performance of a finger-drawn signature on the NYU dataset achieved an average of 3.27% EER when the user template was generated from enrolled samples across two sessions [97]. With a *targeted impersonation* attack where a user signature is known to the attacker, the performance reported on the MCYT dataset was 2.72% EER. Note that, in some work (e.g., [89]), samples of imitated signatures from *physical observation* have been used, where not only the shape of the signature but also timing information is available to attackers. To further enhance security, a more recent study by Buriro *et al.* [98] argued that, when signing on a mobile device, additional features from sensors like accelerometer and magnetometer, that indicate how users hold their devices in three-dimensional space, could help in verifying identity (further strengthening the multi-factor nature of the approach). In their study, a recognition rate of 5.2% FRR at 3.1% FAR was achieved.

Usability: T-F-S techniques inherit most of the usability characteristics of T-F-X techniques. It may be argued that T-F-B techniques are perhaps more *effective* to use in terms of memorability than T-F-S or T-F-X (S+B) techniques in the sense that the need for remembering a secret is eliminated. However, having no secret, security considerations demand low false positives and this leads to higher false negatives, thus negatively impacting *effectiveness* of the system in terms of True Positive Rate. They also take more time to learn and perform and can be considered to have slightly less efficiency and satisfaction than T-F-X techniques. T-F-B schemes are also similar to T-F-X techniques with respect to *context of use*. They can be used potentially under eyes-free constraints, in poor illumination, and in noisy environments. However, false negatives would increase if authentication is attempted while performing an activity, such as walking. Overall, T-F-B techniques provide less usability than T-F-S techniques but do increase the security.

4.4 T-Ha-X Techniques: Multi-Touch Gesture Techniques

On touch surfaces that are larger than what is available on smartphones and wearable devices, touch gestures using the entire palm of one's hand and multiple fingers can be employed. This leads to the category of T-Ha-X techniques. One of the earliest authentication mechanisms using multiple fingers was proposed by Sae-Bae *et al.* [43], [99]. In this approach, a user is authenticated by performing a sequence of multi-touch gestures. This approach uses the gesture pattern and hand geometry information, as well as the gesture style to authenticate users. In particular, a multi-dimensional time series is derived from the time series of 5 touch points, where each touch point is described by $x - y$ coordinates and other touch properties (pressure, size, etc.). Then, a distance between a pair of multi-touch gesture samples is computed using a signal-based approach such as DTW. Another study by Shahzad and Liu [100] has shown that gestures performed by 1-3 fingers, rather than all 5, can also be used for authentication. In yet another study by Yunpeng *et al.* [101], a user is authenticated by swiping multiple fingers in different directions on a touch display. Since multi-touch gesture-based authentication leverages behavioral biometrics and the touch gesture can be a secret, it can be classified as a T-Ha-X (S+B+P) technique.

Security: For a *random guessing attack,* the recognition performance of 5-finger multi-touch gestures was reported at 8.86% EER for an inter-session study (3 sessions) using customized gestures on a private dataset comprising 41 genuine users using a separate threshold for each user [102]. With *physical observation*, the performance reported was 17.17-19.23% EER for 21 types of generic gestures. The system developed by Yunpeng *et al.* [101] achieves EER of 5.84% with 5 training samples for *random guessing*, but improves to 3.02% with 30 training samples. However, EER increases to 3.67% with *targeted impersonation*, where attackers are allowed to watch an animation showing the movements of the targets's fingers on the screen as well as the multi-touch traces left on the screen. EER increases to 4.69% when gestures are performed by a (Lego) robotic hand driven by finger movement patterns constructed from the mean feature vector of genuine samples of all users.

Usability: Initial studies seem to indicate that multi-touch gestures have high *efficiency and satisfaction* [43], [99]. They are easy to learn and understand, and require little memory or physical effort. The same studies also showed that the ease of use, excitement, pleasure and willingness to use correlate well with the authentication performance. This is an encouraging result. Furthermore, users reported a strong level of excitement when they performed their own ("user-defined") gesture. and had the ability to customize the background they drew on, which had the additional benefit of potentially boosting the authentication performance. However, *effectiveness*, in terms of false negatives, and the memorability of the secret, especially when the user has multiple accounts, has not been well studied.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

T-Ha-X techniques provide a broad *context of use*. They can be used in an eyes-free mode and potentially in any ambient conditions with respect to illumination and sound. However, they are likely unusable while performing an activity due to the possibility of increased recognition errors. They may also require a moderately-complicated enrollment process and hence will be potentially harder to understand than T-F-S techniques.

4.5 T-Bo-P Techniques: Body-Print Techniques

In addition to fingers and hand, the now-ubiquitous touch surface has also been proposed to be used with other body parts in order to acquire biometric information. One example of such a technique utilizes the touch surface to capture a user's ear shape [103]. In this specific scheme, the authors capture the set of touch points acquired from a multi-touch surface when a user places his or her phone on the ear while answering a phone call.

Security: An FAR of 0.5% at 7.8% FRR was reported in [103] against a *random guessing* attack. However, the study was conducted on just 12 subjects with each subject providing 12 ear-print samples. More extensive studies are needed to confirm the security of this approach. We note that *targeted impersonation* and *physical observation* attacks are not easy to carry out since the attacker would need (either real or artificial) 3-D models of body parts to spoof the system.

Usability: Although usability studies of body-print techniques have not been performed yet, it seems they would be *effective* and have high *efficiency and satisfaction*. They are simple, quick and memory-wise and physically effortless. They are potentially easy to understand as the enrollment process would be simple. However, they are not changeable. Their *effectiveness* would also be lower than that of T-F-S and T-F-X techniques, as solely relying on a touch pattern match could lead to a decrease in True Positive Rate. They would score very high marks for *context of use*, as the approach would apply in many operational conditions, such as under an eyes-free constraint, in low illumination, under loud noise, and potentially, depending on the body part being used, applicable while performing an activity.

4.6 Avenues for Research

The ubiquity of the touch interface and the attractive simplicity of early T-F-S techniques have led researchers in search of improved security while preserving usability. This has resulted in a variety of techniques clustered around two directions: by considering additional actuators, such as hand and alternate body parts, and by adding physiological and behavioral biometric components. The results have been mixed and clearly a lot more work needs to be done. Admittedly, most of the techniques developed to date alleviate *physical observation* threats to some extent. However, the resulting increase in false negatives, thus negatively affecting usability in terms of *effectiveness*, has been difficult to avoid.

One possible area for future work is to develop classification techniques and training strategies across sessions that improve authentication accuracy thereby enhancing the scheme's usability in terms of *effectiveness* as well as security. For example, a multi-session training strategy can be helpful since many studies (including EEG [104], voice biometric [105], and face [106]) have shown it can be used to model within-user variation more effectively. A long-term study on the consistency of a user's interaction pattern should be considered since the performance may automatically improve over time once a user stabilizes his/her behavior. Another direction for future work is to improve recognition performance by leveraging user interface (UI) elements, such as the background displayed on the screen or visual feedback, to enhance user consistency and memorability. Also, most techniques that have been presented require more extensive security studies against *motivated* forgery under more severe attack models.

In terms of security for multi-touch techniques, the entropy of gestures in particular, and the distinguishability of behavioral biometrics or the question of how many additional bits of security behavioral biometrics provide in general, have not been studied as well. Also, in relation to *random guessing*, the question of how resistant are techniques to dictionary attacks of the type proposed in the recent *MasterPrints* work are left unanswered [19]. Yet another direction would be to quantify the quality of enrolled samples during the enrollment process, similar to a password-strength meter. The system could then prompt users with low-quality samples (which are easy to imitate) to enter new ones [107], [108].

Most T-*-* techniques utilize information from only the touch sensor for the purpose of authentication. However, a lot more information is available in many instances from other sensors such as accelerometer, gyroscope and camera which could be useful for enhancing authentication performance of the system. In body-print for example the hand movement that carries the device to the ear and back could be of value. In multi-touch gestures, taking into account accelerometer and gyroscope data could improve authentication performance when the user is active.

5 C-*-* Schemes: Camera-based Authentication

In this section, we review authentication schemes that use information from interactions captured by a camera (C). These schemes require either static physiological (P) information or dynamic behavioral (B) information to verify identity. We explore four human-body actuators: finger (F), hand (Ha), head (He), and the whole body (Bo).

5.1 C-F-P Schemes: Fingerprint Techniques

While fingerprints are typically captured by capacitive sensors, these devices are specialized and are not designed to be used for general-purpose natural interaction with a device other than authenticating users. Fingerprints can be also acquired by a camera and their image used as a physiological authentication factor [109], [110]. Typically, such approaches rely on minutiae-based features, texture-based features, or a combination of both [111].

Security: Fingerprints have long demonstrated high authentication accuracy. One limitation of a camera-based system is that the quality of a fingerprint image is typically lower than that obtained by capacitive sensor due to such factors as variations of lighting conditions, camera mis-focus, lens distortions, as well as the distance to the camera and pose variations of a user's finger during acquisition. A combination of these factors may significantly reduce authentication performance. In relation to *random guessing*, the performance of fingerprint recognition using a mobile phone camera has been reported at 4.5% EER [109] as compared to around 2% EER when fingerprint impressions are captured by a specialized sensor [112]. C-F-P schemes are immune to attacks that involve *physical observation*, but they are typically not robust to replay attack as biometric traits are static.

Usability: In terms of *effectiveness*, C-F-P techniques are memorywise-effortless as there is nothing to remember. However, true acceptance rate could be lower than its original scheme (e.g., TouchID) due to the lower quality of fingerprint samples. As far as *efficiency and satisfaction* is concerned, user familiarity is one advantage as fingerprints are a well-known biometric trait. However, we are not aware of usability studies to date that address user effort, both perceived and actual, in terms of time and physical action, while performing the task of authentication using camera-based fingerprints. Regarding the *context of use*, there exist disadvantages as the scheme is neither hands-free nor eyes-free. First, at least one hand of the user must be free during authentication to provide fingerprint and another one might be needed to hold a camera, if it is not mounted somewhere. In addition, a user needs to coordinate finger placement so that the fingerprints are captured correctly by the camera. Therefore, this scheme cannot be used when the user's vision is occupied by another task. Another environmental constraint is its sensitivity to lighting conditions.

5.2 C-F-X Schemes: Finger-Motion Techniques

Whereas C-F-P schemes capture physiological information, C-F-X schemes combine (X) behavioral and secret information. One such scheme is the *Leap Password*. Using the Leap Motion sensor [41], Chahar *et al.* [113] proposed to authenticate a user from six in-air one-finger taps performed in front of the device. A set of features are extracted from each of the samples, including the dimensions of the palms and fingers of a user's hand, the timing of the tap sequence, and the motion. Naïve Bayes, artificial neural networks, and random decision forests were used as classifiers, as well as fusion of their results. Since the approach uses a tapping rhythm, finger motion as well as hand geometry information to authenticate users, it can be viewed as an S+B+P multi-factor authentication method (secret S, behavioral B, and physiological P biometric credentials).

Another scheme, *Kinwrite*, proposed by Tian *et al.* [42] authenticates based on "passwords" written with a finger in the air. The system uses the Kinect, a depth camera, to track finger tip motion. 3-D written content as well as writing style are used to authenticate users by calculating the distance between two samples using a DTW algorithm. The scheme is an instance of S+B multi-factor authentication as it leverages both a secret (S) and a behavioral biometric (B).

Security: In terms of the security against *random guessing*, *Leap Password* achieves, at best, 18.83% FRR at 1% FAR on a dataset of tap gestures recorded by 75 users in a single session. Regarding *physical observation*, C-F-X schemes are robust to shoulder surfing attack as they include a biometric factor. They are also moderately robust to replay attacks as it would be hard to replay the finger gestures in three dimensional space. *Kinwrite*, on the other hand, achieves authentication performance of 0% FAR with an average of 1% FRR on a dataset with 35 signatures in total from 18 subjects collected over a period of five months. Each user's signature classifier is trained using only four samples selected randomly from all the samples of that user. This performance is significantly degraded when attackers know the content of user's signature and can observe victims four times. That is, in the experiments performed on four selected signatures – FRR of each user at 0% FAR increases from 0-20% to 10-40%, leading to an average of 25% EER.

Usability: The *effectiveness* of C-F-X schemes in terms of memorability and permanence across a period of time (multiple sessions) has not yet been studied. It is expected that True Positive Rate would be lower when enrolled samples and authenticating samples are collected in different sessions. Memorability of C-F-X schemes would be similar to their touch-based counterparts (T-F-B and T-F-X). However, it would be interesting to investigate how movement along the third dimension affects user cognitive load, recognition performance, and *efficiency and satisfaction* of the scheme. Regarding the *context of use*, similarly to C-F-P schemes, C-F-X schemes cannot be performed when hands are engaged in performing another task. In addition, they require the user to coordinate finger movements to ensure good-quality capture by a camera. Therefore, they cannot be performed in an eyes-free mode. Also, as with other camera-based authentication schemes, they cannot be used in poor lighting conditions.

5.3 C-Ha-* Schemes: Hand-Based Techniques

Camera-based hand authentication can also use physiological features of the hand for authentication. *SignWave Unlock* [114] is a commercial C-Ha-P app using Leap Motion that authenticates users based on hand geometry information captured by stereo cameras and infrared LEDs. In this application, a user simply holds one hand in front of a Leap Motion device. The application then relies on a set of geometric features derived from a user's hand to verify the user. The application has been well-received by users as it has been downloaded more than a million times. Since this approach uses only geometric information from one hand, it is a single-factor authentication method.

In a more intricate approach, not just a hand shape but also hand gesture (hand-pose password) can be captured. In such schemes, authentication credentials comprise a secret (S) and a biometric that could be a combination of physiological (P) and behavioral (B) information resulting in S+P+B multi-factor authentication. One example of such a scheme is the use of the American Sign Language (Fig. 5) to authenticate users. Fong *et al.* proposed using static sign language as a mechanism to enter textual passwords in front of a camera [115]. The proposed method uses a sequence of static images to verify both the user's identity and signalled content. The features used are the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

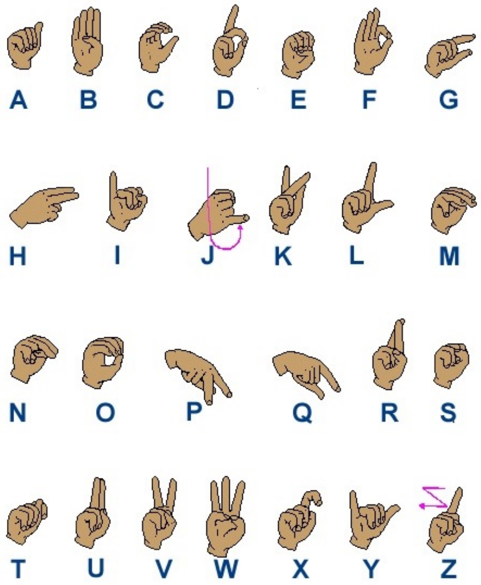


Fig. 5: American Sign Language alphabet that can be used for textual password entry (http://www.iidc.indiana.edu/cedir/kidsweb/amachart.html)

intensity profile, color histogram, as well as hand's dimension and direction extracted from hand-pose images.

Airauth, proposed by Aumi and Kratz [116], is another example of a C-Ha-X (S+B) method using in-air hand gestures captured by a Creative Senz3D depth sensor to verify a secret (3-D hand gesture) and a behavioral biometric (gesture movement pattern). The approach could be viewed as a C-Ha-B method if the gesture is not considered a secret. The system processes each depth frame of a hand gesture to extract 3-D location of all visible fingertips and the hand center. Given a time series of these locations, DTW is used to compute the distance of this series from a template derived from a set of enrolled samples. Wu *et al.* [117] extended this approach by adding physiological attributes (S+P+B). Their method performs gesture matching, between a query and enrolled samples, by comparing covariance matrices of hand's silhouette and depth over a group frames.

Note that, apart from fingerprint and hand recognition, there exist other schemes that make use of body-part physiology to authenticate a user *via* camera, including hand contour [118] and finger knuckles [119], [120].

Security: In terms of security against a *random guessing* attack, it was reported that *SignWave Unlock* can be easily circumvented [121]. Some of the factors that contribute to this are: the accuracy and precision of raw information captured by Leap Motion, effectiveness of the authentication algorithm, and sensitivity of the device to changes in lighting and other environmental conditions. However, none of the studies have reported a performance number so far for the *SignWave Unlock* system. Another study of hand-pose passwords using sign language on a four-user dataset reported a recognition performance of 5.6% and 5.0% FAR at 6.2% FRR when verifying signer and signed content, respectively [115].

As for *Airauth*, the system was tested on user-defined gestures as well as generic gestures. Two user-defined gestures (one's own signature and a custom gesture) and 15 simple generic gestures from 15 participants were used. All gestures were performed with one or multiple (the number based on user's choice) fingers. Each user performed each gesture 10 times with a single finger and 10 times with multiple fingers. For user-defined gestures, 0% EER was obtained in both cases (single and multiple fingers). In relation to *physical observation,* a performance of 2–6% EER was reported for one-finger generic gestures and 1–5% EER for multiple-finger generic gestures. Finally, hand-gesture authentication combining behavioral and physiological attributes [117] resulted in 1.92% EER for 4 generic gestures recorded by 21 individuals in a single session. One advantage of C-Ha-* schemes is their resilience to *physical observation* if biometric information is used. In addition, schemes that use 3-D camera as an interface are more resilient to replay attacks since it is difficult to replay 3-D movement without using very specialized devices.

Usability: In terms of *effectiveness*, the ability to change a gesture in C-Ha-X schemes is a significant advantage over physiological biometrics which cannot be changed if compromised. This is called *changeability* and was discussed in Section 3. In addition, C-Ha-* schemes could be viewed as providing more *efficiency and satisfaction* in terms of physical effort and could be easier to learn than C-F-* techniques as it is more natural to use a hand rather than fingers in front of a camera. In an *Airauth* study it was observed that pleasantness and excitement, which are related to user satisfaction, are well-correlated with the authentication accuracy. However, documented studies of *SignWave Unlock* have not been performed to the best of our knowledge, although this particular instance of C-Ha-P techniques is popular. Regarding the *context of use*, as with C-F-* schemes, a user cannot authenticate using a C-Ha-* technique if hands are not available or one is occupied by other tasks requiring visual attention (the user needs to coordinate hand movement in front of the camera). In terms of constraints on the surroundings and hardware calibration, this would depend on the interface technology. For example, the Leap Motion would require good ambient illumination and hardware re-calibration in case of tracking issues. An additional benefit of C-Ha-* schemes is a touch-free operation. This could be potentially useful in certain situations, e.g., in medical applications.

5.4 C-He-* Schemes: Traditional Face Recognition Techniques

Face recognition is one of the best-known and widely-used authentication mechanisms with a camera interface. Face authentication is provided as an unlocking mechanism on devices supporting Microsoft Windows, Ubuntu, Android and iOS. Face recognition can be based on either a single image, or a video [122]. In addition, advanced face recognition systems have been developed based on a 3-D face model that is reconstructed from 2D images or captured by a 3-D depth sensing camera (e.g., iPhoneX [123]). Fig. 6 depicts 3-D face recognition system used in *FaceID*, an iOS face-based unlocking application on iPhone X. More details on face recognition can be found in [122]. Typically, there are three steps involved in identifying or verifying a person's face: face detection (finding the facial region), feature extraction, and classification. Classification methods can be grouped into three types: holistic approaches, feature-based approaches, and hybrid approaches. Face recognition is a single-factor authentication method based on a physiological biometric (P).

Security: Face recognition has been thoroughly studied in the literature, and its authentication performance has been reported in various contexts. In relation to *random guessing,* 11% and 13% FRR were reported at 0.1% FAR on the *Notre Dame* and *Sandia* datasets, respectively in the 2006 Face Recognition Vendor Test (FVRT) [124], for a single facial image under uncontrolled illumination. However, when the system is applied in the context of mobile devices, authentication performance drops, achieving 10.9% Half Total Error Rate (HTER) [125] due to factors such as occlusions, pose, facial expression, and lighting conditions. This performance drop was also observed in the Android *Face Unlock* mechanism where a study reported that many users fail to log-in or register [126]. Employing the system in an uncontrolled usage scenarios has an adverse impact on FAR and FRR, and consequently the security and usability of the system. A recently-proposed deep convolutional network algorithm, DeepFace, reported far more accurate results on unconstrained datasets [127]. However, this algorithm requires a large number of samples to train a classifier. This could imply that, if face authentication system is used, the performance could be improved over time as the training dataset gets bigger.

In addition to performance issues, the system is vulnerable to several security threats. For example, an attacker may fool the system by presenting a picture or video of the authorized user gathered from *targeted impersonation* or *physical observation* (spoof attack). To counter such attacks, liveness detection mechanisms involving user effort or machine intelligence have been proposed. In terms of user effort, the system may challenge a user to perform some task in order to verify liveness, such as blinking an eye [128], looking at and following a secret icon [129], or moving the handheld device and head simultaneously [130]. In this case, the interaction may be a shared secret ("what-you-know") that can be used as a second factor in the authentication [129]. As for machine intelligence, contextual information from images or video, e.g., texture [131], facial micro-expressions [132], and subtle color variations [133], are utilized to detect spoof attempts. Alternatively, Apple's FaceID utilizes facial structures captured by a 3-D depth sensing camera as an anti-spoofing technique [123].

Usability: Regarding *effectiveness*, while preventive mechanisms against replay attacks can be based on either additional user effort or machine intelligence, this could result in higher FRR (diminished usability) as additional false rejections can be an unintended consequence of liveness detection. In terms of *efficiency and satisfaction*, the scheme is easy to learn and costs little physical and time effort as taking a self-photo is a common activity. With respect to *context of use*, C-He-* techniques can be performed in hands-free mode if the camera is externally mounted. However, in a mobile-device scenario, users are required to hold the device and therefore authentication cannot be performed hands-free. In addition, camera-based authentication needs to be performed in a well-lit environment unless an infrared camera is used (as in Apple's FaceID) to overcome the issue. Furthermore, users are typically required to look directly at the camera so authentication cannot be performed in an eyes-free mode.

Image Credit: Intel Corp

Fig. 6: 3-D facial recognition system) (https://www.counterpointresearch.com/one-billion-smartphones-feature-face-recognition-2020)

5.5 C-Bo-X Schemes: Body-Gesture Based Techniques

One's whole body can also be captured with a camera for authentication. Originally proposed by Lai *et al.* [46], a user can be authenticated by performing a predefined whole-body gesture in front of a Kinect depth sensor. Information from *only* the depth sensor is then processed by extracting either skeletal- or silhouette-based features from the video sequence. These features are then used to compute a matching distance for authentication using DTW (skeletal features) or covariance-matrix distance (silhouette features). Further, a variation of the idea of using gesture-dependent user recognition to learn gesture styles was proposed in [134], [135]. This approach considers both whole-body and hand gestures for authentication using deep convolutional networks. The method uses 3-D body movement pattern, skeletal parameters as well as movement style to authenticate users, and can be considered as a S+P+B multi-factor authentication method.

Security: For a *random guessing attack*, Wu *et al.* [136] reported a performance of 1.24% - 2.78% EER across three predefined datasets. With respect to a *physical observation* threat, similarly to C-Ha-* schemes, the general security advantage of C-Bo-* schemes is their resilience to shoulder surfing as biometric information is used as an authentication credential. In addition, the schemes that use 3-D camera as an interface are resilient to replay attacks as it would be difficult to accurately reproduce three dimensional movement without specialized devices. In terms of spoof attacks, Wu *et al.* [136], performed a study where each user was asked to play the role of an attacker and mimic another user's gesture after briefly practicing with a video recording of the target victim's gesture. The victim for each attacker was selected to be the closest matching user, which is a pessimistic scenario. They found that the performance impact of an impostor imitating a gesture was extremely minimal. In fact, the biometric portion of the gesture seemed resilient to spoof attacks. Specifically, the verification performance only increased to 4.22% - 10.28% EER when dynamic information was available to the attacker.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

Usability: As for *effectiveness*, the impact of users wearing coats and carrying bags, while performing a gesture, on True Positive Rate and False Negative Rate were studied in [137]. This was in addition to studying the memorability of gestures, where they observed how well users were able to recall and replicate their gestures after a week had passed. Under normal conditions in the study, a 1% and 5% EER were reported for two gestures, one selected by the user, and one predefined in a dataset of 40 users, respectively. Wearing coats and carrying bags were found to have a minimal effect on authentication performance causing only a 0.5% and 2.5% increase in EER, while the effect of time caused a more significant degradation increasing EER by 2% and 6.5%, respectively for the above two gestures. Concerning *efficiency and satisfaction*, it would take a comparably longer time and more physical effort to authenticate as compared to C-F-X and C-He-* techniques. However, to date no acceptability studies of C-Bo-X techniques have been carried out, especially when used in public. With respect to the *context of use*, C-Bo-X schemes can be performed in an eyes-free mode but not in a hands-free mode as hands could be used as part of the interaction. In addition, a large and well-lit space is required while users are performing gestures.

5.6 C-E-$*$ Schemes: Eye-Gaze Based Techniques

Authentication using eye-gaze interaction can be performed *via* a regular mobile phone camera or an eye-gaze interface. The authentication factor could be a secret, a physiological biometric or behavioral biometric depending how users interact with the device (e.g., natural gaze or a specifically chosen gaze-gesture).

Iris is rapidly becoming a widely employed physiological biometric to verify the identity of individuals. Although this is mostly done at border crossings and highly-secure locations, there is an increasing effort to use it on mobile devices and other eye-based NUI systems. One such C-E-P scheme is when a user's iris is captured by a mobile-phone camera [138]. In this scheme, iris recognition is performed by applying Adaboost classification to features extracted using a customized version of the Gabor filter in order to cope with users wearing glasses.

In another authentication approach, based on a secret factor (C-E-S), a user is tasked with a specific gaze-gesture, for example serving as a password, and is authenticated based on the correctness of the entered credential. This interaction is typically captured by an eye-gaze interface. One example of this approach is the use of gaze entry in PIN-based authentication proposed by Luca *et al.* [49]. In their work, three approaches for PIN-based authentication were proposed: 1) dwell time – each digit is entered by staring at a specific area for some time, e.g., staring at a particular digit for 800ms, 2) look-and-shoot – each digit is entered by looking at a specific area and simultaneously hitting a button, and 3) gaze-gesture interaction – each digit is entered by performing a specific eye movement pattern, similar to drawing digits with gaze. Another approach to eye-gaze authentication based on a secret factor uses gaze entry to "enter click points" in a cued-recall graphical approach [139]. The user inputs a sequence of locations in the image by simply staring at the secret "click point" locations for a fixed period of time. This sequence is then used to authenticate the user. Bulling *et al.* [140] have demonstrated that the security against random guessing attacks in this approach can be enhanced by using saliency masks to disable the selection of hotspot points. However, the usability trade-offs for this more complex approach have not been studied yet.

In schemes based on a behavioral biometric (C-E-B), there is no specific task for users to accomplish. That is, an arbitrary task can be assigned to users during enrollment and authentication. This approach was first proposed and tested in [50] where two watching tasks were assigned to users in order to capture training and test samples: watching a static text annotation displayed on the screen and watching a comedy video. This approach uses eye movement behavior to authenticate users, and, therefore, can be viewed as single-factor authentication with a behavioral biometric credential. In another scheme, proposed by Sluganovic *et al.* [141], a user is authenticated in a very short period of time by looking at visual stimuli sequences appearing on the screen at random locations. The system then verifies a user's response in terms of gaze pattern as well as the correctness of eye focus to authenticate.

Security: In relation to *random guessing*, the performance in iris authentication on a mobile phone was reported at 0.05% EER for 400 face images captured from 100 persons [138]. For gaze-entry PIN-based authentication, studies have reported a much higher false rejection rate (20.6% - 23.8%) than using gaze-gesture interaction (9.5%) [49]. In gaze-based graphical authentication systems, the authors reported false rejection rates of 27% and 46% when the eye-gaze tolerance area around a "click" point is set to 51x51 and 31x31 pixels, respectively [139]. In [50], eye-gaze interaction using a behavioral biometric was evaluated using a Gaussian mixture model (GMM) applied to histogram features generated from the trajectories of left and right eyes. The reported performance ranged from 28.7% to 47.1% EER depending on the length of training, which ranged from 10 seconds to 9 minutes, and test data, which ranged from 10 seconds to 4 minutes. The best result was obtained when the training lasted 9 minutes and the testing lasted 10 seconds. In the visual stimuli approach [141], the system derives a set of features from the saccade and fixation of gaze response, including the duration, speed, latency, and area. The EER for this approach was reported at 6.2% where samples of the same attackers are also used to train the classification model. The system achieves 7.3% EER when evaluating with five-fold cross-validation for 15 stimuli with median authentication time of 5 seconds.

General security advantages of gaze-based interaction techniques include resilience against *physical observation* (shoulder-surfing and replay attacks) since this type of interaction is hard to observe or record. However, no experiment was performed to assess whether a replay attack can be effective in the case of iris verification. With respect to *targeted impersonation*, the attacker could leverage user information to learn the secret part of authentication factors, e.g., a PIN number. Nevertheless, eye movement can be used as a behavioral biometric [142], [143] to provide a secondary layer of security. Moreover, other biometric traits that could be monitored non-intrusively, like face or iris, could also be used as authentication factors to further strengthen the security and authentication performance of the system.

Usability: Regarding *effectiveness,* a big disadvantage of secret-based approaches, as compared to the behavioral-based ones, is memorability. Another general usability concern in terms of *efficiency and satisfaction* for C-E-* schemes is the time needed for authentication, which is known to be very important to users. With respect to *context of use*, all C-E-* schemes can be performed in a hands-free mode but, obviously, not in an eyes-free mode. Additionally, they require good lighting conditions and stable position of the user. Another disadvantage of gaze-entry PIN-based authentication using dwell time and look-and-shoot is the need for calibration to align the high-resolution eye-tracking system and the space on a display where interaction objects are presented. This is typically performed by the end user.

5.7 Avenues for Research

Future research in camera-based authentication should further investigate usability and security. Usability in terms of *efficiency and satisfaction* can be broadly investigated across-the-board through studying user satisfaction, authentication time and failure-to-acquire rate. In relation to *effectiveness*, robustness to various environmental factors should be investigated. This can be done by evaluating how sensitive a scheme is to environmental setup and adversary factors, e.g., user's distance to a camera, sensitivity to occlusions, lighting, shoulder surfing, etc.

In terms of security against *physical observation*, the vulnerability to replay attacks requires more exploration.

Another research direction could involve validation of *effectiveness* and *security* against *random guessing* on larger datasets collected over a longer period of time (weeks to months). In many cases, studies to date have been performed over a short period of time (days), so the memorability and reproducibility of a scheme over time are not well understood.

Camera-based schemes can benefit by combining physiological, behavioral, and secret information captured from the same user interaction. For example, it is possible to develop eye-gaze based authentication such that the system would simultaneously derive iris, gaze pattern [144], and secret information from user interaction for authentication purposes. Another possibility for camera-based authentication is to explore the use of other contextual information, e.g., eye, face and generic body skeleton model, which could be observed simultaneously by the same camera. One example is SAFE (Secure Authentication with Face and Eyes) where the authentication factors used are a combination of a gaze secret and face biometric [129]. In this system, an authentication phase consists two steps: face verification followed by gaze-secret verification once a user's face is verified. In addition, with the increasing prevalence of 3-D cameras, it is possible to acquire precise depth measurements of the entire body, and utilize it to estimate body-part poses (e.g., hand, head, arm) which may be used as additional information for authentication.

In hand-based authentication, additional physiological information like a hand's texture and appearance can be used as an additional authentication factor to improve authentication accuracy. The *effectiveness* of sign language with regards to memorability for those unfamiliar with it has yet to be explored. While face authentication is a well-studied scheme, advanced spoofing techniques should be explored in order to develop robust countermeasures against *physical observation*. In terms of body gestures, reproducibility can be studied by considering intra-user variations across multiple sessions, as this has proven successful in other biometric modalities, e.g., ECG [145] and speaker authentication [146].

One usability research direction in eye-gaze authentication is to address: *context of use* challenges related to sensor calibration, *effectiveness* in relation to sensitivity or the problem of proper detection of intentional and unintentional gaze, and *efficiency and satisfaction* of this type of authentication interaction in terms of user familiarity with gaze-entry [49], [139].

6 MS-*-X Schemes: Motion-sensor based Authentication

In this family of authentication systems, a user's conscious and intentional movements, e.g., of hands (Ha) or head (He), are *directly* measured by an accelerometer-equipped device such as a Wii remote, a smartphone, or Google Glass, and are matched against enrolled examples of those movements. Although the visual appearance of movements can be *indirectly* recorded by a camera (see Section 5.3), here we focus on gesture interactions measured directly by motion sensors.

6.1 MS-Ha-$*$ Schemes: Hand-Based Techniques

Examples of authentication systems based on hand movements in 3-D space include *uWave* [44] where a user sketches a custom line-drawing (which serves as the password) holding a Wii remote and *in-air signature* [147] where a user "writes" his/her signature in 3-D space while holding a smartphone (each device serves as a "pen") [147]. In both examples, the distances between the time series of accelerometer signals along x, y, and z axes for test and enrolled samples are compared using the well-known DTW algorithm. The system accepts a user if the movement is close enough to the enrolled samples, and rejects otherwise. Both the user-defined hand-movement pattern (secret) as well as the user's hand-movement "style" (behavior/biometric) contribute to the authentication process. Thus, these schemes belong to the family of S+B multi-factor authentication methods that combine both a secret and a behavioral biometric credential. If a task is pre-defined, then a user is not required to memorize any secret and the scheme belongs to the MS-Ha-B category. A study of this approach has also been conducted in *uWave*.

Security: In *uWave*, user-defined gesture samples were collected from 25 users split into 5 equal-sized groups over 5 sessions spread apart in time. In the first session, each user created two custom gestures and performed them once. In the four subsequent sessions, each user performed 5 repetitions of the two gestures. Each user was then attacked at least 20 times by 4 users within the same group. The system achieved a 3% EER on average when the attackers have no visual disclosure (video) of the gesture password, that is, in relation to security against *random guessing*. The EER increased to 10% with visual disclosure of the gesture password, (*physical observation* attack). Similarly, *in-air signature* achieved a 2.5% EER on a single-session dataset of 34 users, where each user contributed 7 samples. In evaluation, 3 samples were chosen at random to construct the template and the remaining 4 samples were used as positive test samples for that user. Negative samples for each user were selected from 3 other users assigned to attack the user. Each attacker studied the in-air signatures by watching a video and attempted to forge those signatures 7 times.

Note that, Ms-Ha-B and Ms-Ha-X schemes provide some security benefits against *physical observation* as they are generally resilient to shoulder surfing since a behavioral biometric is used. In addition, they are reasonably resilient to replay attacks as it is difficult (without a robot) to record the movement and replay it in 3-D space.

Usability: In terms of *effectiveness*, a study of the difficulty of memorizing a *uWave* gesture compared to a text password found statistically significant results indicating that a user-defined gesture is less difficult to memorize than a text password. The findings of the *in-air signature efficiency and satisfaction* study are consistent with those of *uWave*. Users do not have much difficulty in creating and performing an in-air signature and, overall, they rate the scheme's acceptability as high. With respect to *context of use*, Ms-Ha-X schemes can be performed in an eyes-free mode. Obviously, Ms-Ha-X schemes require hand operation. If users are not stationary or stable while performing authentication, signals from motion sensors might be noisy resulting in false negatives and low usability. Lastly, since *uWave* uses a Wii remote as an interface, device calibration may be needed in order to realign the cursor position if it is inaccurately aligned with the pointing direction.

6.2 MS-He-X Schemes: Head-Based Techniques

This is a fairly recent sub-family of authentication systems which is based on music-induced head movements in 3-D space. The prototypical example is *Headbanger* proposed by Sugang *et al.* in 2016 [45]. Here, a user moves his/her head (following a pre-designed movement pattern) in synchrony with beats in a musical cue. Head movements are captured by a system embedded in a head-worn device, e.g., Google Glass. At an algorithmic level, authentication is similar to MS-Ha-X schemes: the distance between two time series of filtered acceleration data from test and enrolled samples is evaluated using DTW. A sufficiently small distance leads to acceptance or, otherwise, rejection. Similarly to MS-Ha-X schemes, both the head-movement pattern (secret) as well as head-movement style (behavior) contribute to authentication. Thus, this too is an S+B multi-factor authentication system that combines a secret and a behavioral biometric credential.

We would like to point out that many active authentication systems using gait biometrics have been proposed [148], [149], where the authentication system validates a user without his/her consent. Such systems are not included in this survey as we focus on point-of-entry user authentication, where an *explicit* authentication interaction is performed by a user. In other words, the activation of authentication system requires an explicit consent from users.

Security: In relation to *physical observation*, similarly to Ms-Ha-X schemes, Ms-He-X approaches provide some resilience to shoulder surfing attacks as a behavioral biometric is used. In addition, they are resilient to replay attacks as it is difficult (without a robot) to record the movement and replay it in 3-D space. In terms of a *random guessing* attack on data collected from 30 users with 40 trials per user (15 users finished all trials in one session and the rest finished in 3 days on average), *Headbanger* attained an EER of 4.43%, 6.14%, and 6.65% for music durations of 10, 6, and 5 seconds, respectively. As for robustness against *physical observation*, where the attacker has access to the video of victims shaking their heads, *Headbanger* attained an FAR of 6.94% on average when 37 users attempted to attack 3 users.

Usability: Regarding the *context of use*, Ms-He-X schemes can be performed in both eyes-free and hands-free modes. However, similarly to Ms-Ha-X methods, if users are not in a stable position the signal from motion sensors might be noisy in Ms-He-X schemes. In terms of other usability properties, we are not aware of any studies completed to date for the MS-Ha-X class of authentication schemes.

6.3 Avenues for Research

No factors that may affect usability of the proposed MS-He-X schemes in terms of *efficiency and satisfaction*, e.g., space requirements and time needed to perform movements, have been explored to date. Also, in current MS-He-X schemes, the music cue and its duration are chosen by the system and a head movement pattern is pre-selected by the user as a secret. What if the music cue is also pre-selected by the user as part of the secret or if the head movement pattern is not pre-selected and is a natural response of a user to the chosen music cue? Further, negative impact on *effectiveness* caused by various impediments such as bulky clothing, backpacks, etc., need to be investigated. Another related research direction would be robustness studies, e.g., robustness against body posture (standing, sitting) and body dynamics (feet movement coupled with head movement).

Regarding security, quantifying the relative value of the movement pattern (secret), anatomy (biometric), and any implicit biases in posture (similar in spirit to the studies conducted in [136] for body gestures) may provide greater insight into which components of MS-*-X schemes have the greatest impact on authentication performance. Finally, one could also consider new systems that simultaneously measure movements of multiple body parts, e.g., hands, feet, head, torso, to improve security against *random guessing* and *physical observation* attacks.

7 M-*-* Schemes: Microphone-Based Authentication

In this interaction, a user is typically authenticated by speaking into a microphone embedded in a device. Voice authentication, also known as speaker recognition, is broadly classified into two types: text-independent and text-dependent [51]. Whisper authentication is a variation of text-independent speaker recognition that aims to enhance acceptability and usability of the system [150]. Yet another possibility while using a microphone for user authentication is to analyze the characteristics of the frequency response created by user's breathing pattern [151] or by skull [152]. In such schemes, users are not required to speak but sound is captured by a microphone in the device held by a user or embedded in a headset.

7.1 M-V-B Schemes: Text-Independent Speaker Authentication

M-V-B schemes perform authentication based on arbitrary spoken content, thus relying on behavioral characteristics of the speaker. They are commonly referred to as text-independent speaker authentication. Recognition techniques used in this context can be broadly divided into statistical models and discrimination-based learning methods [153]. Among common statistical models are the likelihood ratio test and GMM, while examples of learning methods are the Support Vector Machine (SVM) and Artificial Neural Network (ANN). Features often used in these systems include frequency parameters, e.g., Mel Frequency Cepstrum Coefficients (MFCC), and their derivatives, computed at the output of a filter bank or a Linear Predictive Coder (LPC).

A derivative of text-independent authentication, that is less intrusive for users, is whisper-based authentication. One study using a text-independent whispering voice was conducted by Jin *et al.* [150] (single-factor authentication based on a behavioral biometric credential). Based on an algorithm which uses MFCC features in conjunction with a GMM model, they found that whispered speech is less discriminative than normal speech and noise has a large negative impact on recognition performance. In order to improve performance, techniques from regular voice authentication methods are often modified to account for the intrinsic differences between the signal spectra of whispered and normal speech [154]. In particular, Xiaohong and Heming [155] have demonstrated that adaptive fractional Fourier transform cepstral coefficients are a more effective and robust feature set than MFCC coefficients in identifying a speaker from whisper signal when training and test channels are not matched.

Security: A wide range of authentication performance numbers in relation to *random guessing* have been reported for M-V-B schemes due to the variation in devices, environmental factors, and authentication techniques. The best system reports an EER of 3.11% [156] on the I4U's telephone-quality development dataset [157]. In this system, users are prompted to speak a randomly-generated phrase in their normal voice. The system then authenticates users based on voice pitch. This is a single-factor authentication scheme that is resilient to shoulder surfing attack for it uses a behavioral biometric credential. Also, the verification of the content of the spoken phrase can serve as a form of security against replay attacks, as older phrases cannot be reused. Another variation of this scheme is to challenge the users with various questions and use voice responses to extract both context information and voice print to enhance system's performance and increase security against replay attacks [158]. An interesting attack vector for M-V-B schemes is computerized imitation [156] (either by *targeted impersonation* or *physical observation*).

A recognition rate of 32-98% was reported for whisper-based authentication using 6-second tests with 60 speakers (performance variations depended on algorithms and test channels) [155]. However, to the best of our knowledge, no study of whisper-based authentication on devices with an NUI interface has been conducted to date.

Usability: In M-V-B schemes, users are not required to memorize any phrase thus enhancing their *effectiveness*. However, voice-based authentication on personal devices has mixed score on *efficiency and satisfaction* as it takes little time and physical effort to perform tasks and it is easy to learn, but it is considered, in general, insecure and inconvenient [159]. Regarding the *context of use*, all voice-based techniques can be performed without hands and without visual attention. Other issues may impact the security and usability of this scheme and they should be taken into consideration prior to deployment, for example: human factors (e.g., the Lombard effect when a speaker increases vocal effort to compensate for signal-to-noise ratio in noisy conditions [160], emotions, vocal organ illness, aging, and level of attention [156]), device and environmental factors (e.g., level of background noise [161], acoustic disturbances like echo, and microphone frequency response). In terms of whisper-based authentication, we are not aware of any usability studies completed to date.

7.2 M-V-X Schemes: Text-Dependent Speaker Authentication

Unlike in the previous case, M-V-X schemes require that the spoken content of authentication samples be the same as that of the enrolled ones, thus also relying on a secret. Recognition techniques for text-dependent speaker authentication can be broadly divided into dynamic programming approaches, statistical models, and discrimination-based learning methods using frequency parameters at a filter-bank output [162].

Since M-V-X schemes use both spoken content and voice pitch to authenticate users, they belong to S+B multi-factor authentication methods combining a secret and a behavioral biometric credential.

Security: In terms of security against *random guessing* attack, Ram *et al.* [160] have reported text-dependent authentication performance at 7-11% EER on a dataset collected from 48 enrolled users and 40 impostors using a mobile device. This range of EER is due to varying experimental environments, and impostor attacks (speaking the same content). Each of the samples in this experiment was a set of spoken ice-cream flavor phrases. Regarding *physical observation*, a shoulder-surfing attack can compromise the secret text in an M-V-X scheme, but the biometric component provides some resilience. Since the spoken content is static for a particular user, while it is different between users, the authentication performance of text-dependent speaker authentication is usually higher than that of text-independent authentication [163]. However, replay attacks are more feasible in text-dependent authentication as the spoken content is static and the audio can be simply recorded and replayed [164], [165].

Usability: Since the secret phrase needs to be loudly spoken, the *context of use* limits the use of text-dependent speaker authentication to conditions where a user would not be overheard. Also, similarly to M-V-B schemes, M-V-X schemes can be performed without hands and without visual attention. In terms of *effectiveness*, M-V-X schemes, in comparison to M-V-B schemes, have the drawback of memorability for some text needs to be memorized. However, they have the advantage of changeability. Finally, their *efficiency and satisfaction* are similar to those of M-V-B schemes.

7.3 M-He-* Schemes: Non-Vocal Techniques

In authentication *via* bone conduction, users are verified based on an audio signal (e.g., white Gaussian noise), generated in close proximity of the head, that travels through the skull bone to the user's inner ear and is intercepted by a microphone (all components are integrated in a headset) [152]. Essentially, the system authenticates users by analyzing characteristics of the frequency response created by the user's skull. Specifically, MFCC data are extracted from the received audio signal and used for a computationally light-weight 1-NN classifier. This can be considered an M-He-P scheme since the credential is the user's skull structure. *BreathPrint* [151] is another acoustics-based user authentication scheme utilizing a microphone that, instead of user's voice, captures user's breathing sound for authentication. In this scheme, the system authenticates users by performing frequency analysis using Gamma-tone Frequency Cepstral Coefficients (GFCC) as a feature set and applying a GMM-based classifier. This can be considered as an M-He-B[3] scheme where the authentication credential is the breathing pattern.

Security: Regarding *physical observation*, both schemes are secure against shoulder surfing as there is no secret to be revealed. They are also secure against replay attacks, unless specialized, directional, high-quality microphones are used for *targeted impersonation*, since the level of received signal is too low to be captured from a distance. However, the current systems are not very resilient to *random guessing*. Specifically, in the bone-conduction authentication scheme, although the experiment was performed in the best scenario (a controlled setting without any background noise with only 10 participants and one session), the system achieved 6.9% EER, considered too high to be used in practice. Similarly, in *BreathPrint*, the performance of 7-18% FAR at 2% FRR was reported when users performed the same type of breathing (sniff, normal, or deep) and the number of GMM components was 5. This experiment was also conducted in a controlled setting without any background noise with only 10 participants and 3 sessions.

Usability: As this is a pure biometrics-based authentication, there is nothing for users to memorize. These are great usability advantages in terms of *effectiveness*. However, the current bone conduction system with 6.9% EER performance rejects genuine users too often thus reducing its *effectiveness* in practice. Another downside of the bone-conduction scheme with respect to *efficiency and satisfaction* is that users take longer to authenticate (23 seconds) than using *BreathPrint* (0.55-4.8 seconds on average depending on the type of breathing). In addition, if the level of an audio signal being generated in the bone-conduction scheme is too high, it could make users uncomfortable. Clearly, there exists room for improvements to these schemes in terms of usability. Regarding the *context of use,* both schemes involve eyes-free interaction. The bone-conduction scheme is also hands-free since a headset is worn by the user. However, in *BreathPrint* the user may need one hand to hold a microphone in close proximity of the nose. In terms of ambient sound, in bone conduction the audio signal is bypassing the air and traveling directly to user's inner ear thereby reducing the chance of being corrupted by ambient conditions. However, *BreathPrint* experiments show an FRR increase from 0% to 4-37% when ambient noise increases from 50 to 54dB. Both schemes can also be performed in any lighting conditions.

7.4 Avenues for Research

Apart from resolving security and usability issues that exist in both text-independent and text-dependent approaches, one interesting direction for future work is to use a hybrid of these two approaches in order to simultaneously: 1) improve the authentication performance of text-independent systems when only a limited number of training samples from a user are available, thereby enhancing security and usability of the system in terms of *effectiveness*, and 2) enhance the security against replay attacks *(physical observation)* in relation to replay attacks of text-dependent systems. This concept was introduced in a system called "text-constrained" speaker recognition [166]. Another example of this is to use a challenge-response based authentication protocol where the model is created based on a limited number of short-phrases [158]. In whisper authentication, one important question to address is its usability: is it convenient and non-intrusive to use? Also, as many acoustic properties of whisper signals can vary drastically from one's normal voice [167], it would be interesting to develop an effective recognition algorithm that can verify whisper signals using enrolled normal speaking signals. Such an algorithm would allow users to authenticate in the situation when speaking in a normal voice is not appropriate. In addition, as demonstrated by ear-bone conduction authentication, it is possible for a microphone to be used as a sensor to acquire other physiological biometrics, e.g., body fluid and skeleton properties, when embedded in wearable devices that are attached to other parts of the body.

8 BCI-Br-* Schemes: Brain-Computer Interface Based Authentication

Motivated by the advancement of brain-computer interface (BCI) technology, Thorpe *et al.* proposed *pass-thoughts*, an authentication method resistant to physical observation [52]. The method uses a secret thought and the user's brain signal, that is unique among users, to verify identity. Therefore, it is an S+B multi-factor authentication scheme (BCI-Br-X). It can also be considered as BCI-Br-B, if a cognitive task is assigned to a user.

[3] Although breathing characteristics also depend on lung and trachea size, *BreathPrint* measures the sound only at the nose so we consider head (He) to be the actuator.

A BCI-Br-* scheme using a laboratory-grade electrode cap that records brain-wave signals (EEG) was proposed by Marcel and Millán [104]. Another scheme was recently investigated by Chuang *et al.* [168] that used a consumer-grade EEG headset as opposed to a clinical-grade one. Both groups studied performance, usability, and recall of user authentication using brain-wave signals.

Security: In relation to *random guessing*, Marcel and Millán evaluated their system's biometric performance by asking 9 users to perform three brain tasks: imagining repetitive self-paced left-hand movements, imagining right-hand movements, and mentally generating words beginning with the same random letter. For each user, the experiment ran for three days, four sessions per day with 5- to 10-minute breaks between sessions. The method achieved 35.5% HTER when training and testing samples were drawn from different days. While performance degrades over time, it improves to 12.9% HTER when training samples are drawn from two days instead of one.

Chuang *et al.* designed 7 different brain tasks, some consisting of imagining generic activities like breathing, moving fingers, and listening to an audio tone, while others involving the creation and recall of personalized imagined activities like sports, colored-object counting, song or passage recitation, and customized mental thought. Then, a two-session experiment on 15 subjects was performed. In the first session, each user was asked to select four of the seven personalized brain tasks as a secret, and in the second session the user was asked to recall them. In addition, in both sessions, users were asked to perform those seven tasks and answer a series of usability questions. The results demonstrated the feasibility of the scheme using consumer-grade hardware by achieving a good recognition performance on a larger population (1.1-28% HTER depending on the type of brain task, where the recognition threshold is adjusted specifically for each user).

An intrinsic security trait of BCI-Br-X schemes is their resilience to *physical observation* as the interaction signal is not observable [168]. In addition, the scheme is resilient to replay attacks as a brain signal is not recordable from a distance. However, as shown by Chuang *et al.*, the performance of BCI-Br-X schemes is sensitive to the selection of secret tasks.

Usability: In terms of *effectiveness*, studies have been performed to investigate the difficulty in recalling a thought password and user preferences. It was concluded that users have no difficulty recalling thought passwords. Also, the schemes score high marks for *efficiency and satisfaction* as performing brain activity obviously requires no physical effort by the user and takes very little time. Studies show that users have different preferences for mental activity tasks based on the perceived difficulty and enjoyability of the tasks. With respect to the *context of use*, brain tasks can be performed without any physical activity and visual attention. However, a noisy environment could prevent users from concentrating on their thought tasks resulting in a noisy brain signal. In addition, the need for hardware calibration in most of the current BCI systems is a significant usability issue in BCI-based authentication [169].

8.1 Avenues for Research

One direction of research on BCI-based user authentication could be the design of generic brain tasks that are both usable and secure. Also, while an initial investigation of robustness of this scheme against *physical observation* has been conducted [170], the study was limited only to a set of three attackers. Furthermore, the thought processes of the attackers might differ from those of the users due to differences between their native languages and social environments by which they have been surrounded. Future work is needed to address this issue, where the secret is known to attackers.

It may be also possible to use brain activity during another authentication interaction, such as entering a PIN. This would result in one-step multi-factor authentication combining a secret (PIN) with multiple behavioral traits, e.g., keystroke timing, keystroke pressure, brain activity. This would also apply to other authentication interactions, such as voice or 3-D gesture, instead of thought-specific tasks.

Lastly, BCI-based authentication independent of the brain task would go a long way towards natural and effortless interaction with an authentication system as users would not need to memorize a secret. This would be analogous to text-independent speaker authentication (Section 7.1) [153] or learning body-gesture style regardless of the performed action (Section 5.5) [134], [135].

9 Analysis and Conclusions

As is clear from this survey, there exists a rich variety of user authentication techniques proposed for different NUIs. The taxonomy we developed allows their grouping into categories for comparison and analysis. It also reveals modalities for which no authentication techniques have been developed yet. For example, Fig. 2 shows the category MS-Bo-B as blank. That is, there is no technique reported in the literature that uses a behavioral biometric captured using motion sensors attached to the body as input credential. Clearly, one could conceive a technique that involves performing a gesture which is public but yet has enough complexity to provide the ability to discriminate between different users and provide an acceptable false positive rate. However, the usability of such a technique could be an issue depending on the complexity of the gesture. Similarly, other avenues for user authentication could be explored based on the proposed taxonomy elements.

In some cases a missing technique in a given category could stem from the fact that, inherently, it may not provide an attractive authentication mechanism. For example, there is no technique that captures a secret input credential from the entire hand on a touch surface. Although one could conceive such a technique, not capturing a physiological or biological biometric while the input is being recorded would be a waste of the sensor's capabilities. In other cases, the lack of a technique in a particular category could also be a result of the inherent nature of the sensor. A touch sensor cannot be utilized to sense a voice credential and hence the entire corresponding cell in Fig. 2 is blank.

In addition to providing insights on unexplored research avenues, the taxonomy also allows one to select and compare a set of candidate techniques for a given application. Tables 1 and 2 provide a summary of the different interfaces we have considered and the different types of authentication techniques that have been proposed for them. Thus, given an application, one could identify viable techniques from those tables. For example, in the context of Table 1, consider an application where a user has to be authenticated while driving a car. Also, assume that the two sensors already available in the car are a microphone and a camera. The set of techniques listed under M-*-* and C-*-* become potential candidates. However, the best approach depends not only on the available sensors but also on the environmental conditions under which authentication will be performed. If authentication needs to be performed while driving, then only eyes-free techniques may be plausible. A hands-free method may or may not be critical depending whether one hand or two hands are needed and if a country's laws mandate it. Resiliency to motion, while authentication is being performed, may be another desirable characteristic. Armed with these constraints one can narrow down the set of candidates further.

In terms of user state and environmental requirements, one can observe that there is a strong inter-dependence between the user interface and the constraints. Most camera-based schemes would be sensitive to lighting conditions and some would also be sensitive to background appearance. Most voice-based schemes would be sensitive to noise and only be used in locations that have acceptable disturbance level. However, there could be interesting exceptions. For example, skull conduction could be used in a noisy environment as the signal travels inside the user's head and is sent directly to the user's ear bone.

The selection of a technique does not depend only on available sensors and context, but also on the threat model. If shoulder surfing or replay attacks are plausible, then approaches involving only a secret credential become undesirable. For example, text-dependent speaker authentication based on a secret phrase captured by the microphone would not be a candidate for an in-car system, as passengers would be able to hear the phrase. At the same time, if the environment is noisy, a text-independent speaker authentication could lead to difficulties judging by the attainable FAR and FRR rates of state-of-the-art approaches.

Table 2 provides a performance comparison of different approaches. It should be noted that the table should be viewed as a way to compare different approaches as opposed to directly comparing specific techniques. This is due to the fact that the set of techniques presented is not exhaustive but rather representative. More importantly, however, specific techniques have been evaluated independently in the literature with quite different experimental protocols and parameters. While one technique may have been evaluated with 20 users over multiple sessions, another one could have been tested with a few hundred users but over a single session. In one study, the training samples may have been selected randomly and in another case they may have been pre-assigned. Similarly, the thresholds used in various classifiers may have been selected individually for each user or globally. Clearly, these performance measures should be taken only as an indicator of a potential rather than an upper or lower bound.

Tables 1 and 2 together demonstrate the emergence of a variety of user authentication mechanisms using new user interfaces with interaction capabilities. This has created a unique opportunity to redesign old and develop new mechanisms with the goal of providing intrinsic usability and security to the user. However, the tables and the taxonomy also make it clear that despite a variety of possibilities, there is no silver bullet and the choice of a user authentication technique involves making multiple trade-offs. They demonstrate that the selection and design of an authentication mechanism is multidisciplinary in nature and requires experts from different domains to work closely together in order to arrive at a solution with potential for a real-world impact.

As new computing devices equipped with multiple interfaces enter the consumer market, one can envisage new, multi-factor authentication schemes that simultaneously use multiple interfaces to capture several authentication factors [171], not captured in our study. For example, with gaze or eye movement as an authentication interaction, the geometry and appearance of a user's face could be easily captured using embedded 2-D and 3-D cameras as additional authentication factors. This would make it much harder to circumvent the system.

Another interesting direction for future work is to design a mechanism that allows users to enter the same secret in multiple ways depending on the availability of sensors in a device, surrounding environments, and potential threats of the moment. For example, Google Glass users have multiple ways to communicate with the device, including voice control, hand gestures, and tab sliding. The appropriateness of each interface depends on the situation. Having a mechanism that allows a user to operate in multiple ways depending on the surroundings would increase not only the usability but also the security of the system [172]. Another possible framework to enhance the overall accessibility of authentication mechanisms is to move away from one-size-fits-all approaches towards supporting a diverse ecosystem of authentication schemes [173]. In other words, the system could let users select or could assign to them such authentication mechanisms that suit their needs and security level required by a particular application.

ACKNOWLEDGEMENT

This research was supported in part by Thailand Research Fund (MRG 5980078) and by the National Science Foundation (awards CNS-1228842 and CNS-1228869).

REFERENCES

[1] "Garmin's vivoactive 3 smartwatch can now help you pay for things." https://www.engadget.com/, 2017. [Online; accessed 09-Nov-2017].

[2] F. Schaub, R. Deyhle, and M. Weber, "Password entry usability and shoulder surfing susceptibility on different smartphone platforms," in *Proceedings of the 11th international conference on mobile and ubiquitous multimedia*, p. 13, ACM, 2012.

[3] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle, "Improving text passwords through persuasion," in *Proceedings of the 4th symposium on Usable privacy and security*, pp. 1–12, ACM, 2008.

[4] W. Melicher, D. Kurilova, S. M. Segreti, P. Kalvani, R. Shay, B. Ur, L. Bauer, N. Christin, L. F. Cranor, and M. L. Mazurek, "Usability and security of text passwords on mobile devices," in *Proc. of the 2016 Annual ACM Conf. on Human Factors in Computing Systems, CHI*, vol. 16, 2016.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

TABLE 1: Summary of approaches to authentication interaction and their characteristics at the authentication moment († indicates possible taxonomy that is not yet explored in the research literature)

Section	Approach	Interface	Taxonomy	Eyes-free	Hands-free	User state (movement)	Environmental requirement	Average log-in time (seconds)
T-*-* SCHEMES - TOUCH SURFACE BASED AUTHENTICATION								
III.A	Microsoft picture [76]	Large touch display	T-F-S T-F-X (S+B)†	No	No	Relatively smooth	-	4.33 (3 tabs)
III.A	Gesture combination lock [69]	Tiny touch surface	T-F-S T-F-X (S+P)†	Yes	No	Relatively smooth	-	-
III.A-B	Touch typing [65], [84]	Medium-size touch display	T-F-S T-F-X (S+B)	No	No	Smooth	-	1.96 (4-digit)
III.A-B	Android pattern lock [65], [86]	Medium-size touch display	T-F-S T-F-X (S+B)	No	No	Relatively smooth	-	0.91
III.B	PIN drawing [87]	Medium-size touch display	T-F-S† T-F-X (S+B)	Yes	No	Relatively smooth	-	3.7 (4-digit)
III.C	Online signature [97]	Medium-size touch display	T-F-B T-F-X (S+B)	Yes	No	Relatively smooth	-	2.84
III.D	Multi-touch gestures [102]	Large touch display	T-Ha-X (B+P) T-Ha-X (S+B+P)	Yes	No	Relatively smooth	-	1-2 (per gesture)
III.D	Multi-touch swipe [101]	Medium-size touch display	T-Ha-X (B+P) T-Ha-X (S+B+P)†	Yes	No	Relatively smooth	-	0.75
III.E	Bodyprint [103]	Medium-size touch surface	T-Bo-P T-Bo-X (S+P)†	No	No	Smooth	-	-
C-*-* SCHEMES - CAMERA BASED AUTHENTICATION								
IV.A	Fingerprint recognition [109]	2-D camera	C-F-P	No	No	No movement	Good light	-
IV.B	Leap Password [113]	Leap Motion	C-F-X (S+P+B)	No	No	Smooth	Good light	
IV.B	KinWrite [42]	3-D camera (Kinect)	C-F-X (S+B)	Yes	No		Relatively large interactive space	2-12
IV.C	SignWave Unlock [114]	Leap Motion	C-Ha-P	No	No	Smooth	Good light	-
IV.C	Hand pose authentication [115]	2-D camera	C-Ha-S C-Ha-X (S+B)	No	No		Good light, no clutter in background	-
IV.C	Airauth [116]	Creative Senz3D controller	C-Ha-B C-Ha-X (S+B)	No	No	Smooth	Good light	-
IV.C	Hand gesture [117]	3-D camera (Kinect v2)	C-Ha-X (P+B) C-Ha-X (S+P+B)	No	No	Smooth	Good light	-
IV.D	Face recognition [125]	2-D camera	C-He-P	No	Yes	Smooth	Good light	7.5
IV.E	Body gesture [46], [136]	3-D camera (Kinect)	C-Bo-X (B+P) C-Bo-X (S+P+B)	Yes	No		Large interactive space	
IV.F	Iris [138]	2-D Camera	C-E-P	No	Yes	Smooth	Good light	-
IV.F	Gaze-based PIN entry [49]	Eye tracker	C-E-S C-E-X (S+B)†	No	Yes	Smooth	Good light	12-54
IV.F	Gaze-based graphical password [139]	Eye tracker	C-E-S C-E-X (S+B)†	No	Yes	Smooth	Good light	36.7-53.5 (5-gaze-point)
IV.F	Gaze-based task independent [50]	Eye tracker	C-E-B	No	Yes	Smooth	Good light	10
IV.F	Visual stimuli [141]	Eye tracker	C-E-B	No	Yes	Smooth	Good light	5
MS-*-* SCHEMES - MOTION SENSOR BASED AUTHENTICATION								
V.A	uWave [44]	Wii remote controller	Ms-Ha-B Ms-Ha-X (S+B)	Yes	No	Smooth	-	-
V.A	In-air signature [147]	Accelerometer on mobile device	Ms-Ha-X (S+B)	Yes	No	Smooth	-	-
V.B	Headbanger [45]	Accelerometer on Google Glass	Ms-He-X (S+B)	Yes	No	Smooth	-	10
M-*-* SCHEMES MICROPHONE BASED AUTHENTICATION								
VI.A	Text-independent [156]	Microphone	M-V-B	Yes	Yes	-	Quiet	
VI.A	Whisper [150], [154], [155]	Microphone	M-V-B M-V-X (S+B)†	Yes	Yes	-	Quiet	1-20
VI.B	Text-dependent [160]	Microphone	M-V-X (S+B)	Yes	Yes	-	Quiet	-
VI.C	Bone conduction [152]	Microphone	M-He-P	Yes	Yes	-	-	1-10
VI.C	BreathPrint [151]	Microphone	M-He-B	Yes	No	-	Quiet	0.55-4.8
BCI-*-* SCHEMES BRAIN-COMPUTER INTERFACE BASED AUTHENTICATION								
VII	Brainwave [52], [104], [168], [170]	BCI Headset	BCI-Br-B BCI-Br-X (S+B)	Yes	Yes	-	Relatively quiet	5-10

TABLE 2: Recognition performance and experimental setup of authentication approaches (resilience SS refers to shoulder surfing and resilience PA refers to presentation attack). ‡ indicates datasets that are publicly available (for details see [18]).

Approach	Authentication factors		Resilience		Experimental setup			Authentication performance
	Secret (S)	Biometric (P/B)	SS	PA	♯ users	Duration	Remarks	
T-*-* SCHEMES - TOUCH SURFACE BASED AUTHENTICATION								
Microsoft picture [76]	3 gestures of tap, line or circle	-	No	Yes	Not reported			
Gesture combination lock [69]	4 gestures from, 8 predefined	-	No	Yes	On Google Glass Not reported			
Touch typing [65], [84]	Password or PIN	-	No	Yes	32	30 days	4-digit-PIN	3.1% FRR [65]
	-	Touch stroke (B)	Yes	Yes	13	1 sessions	5 keypresses	32.3% FAR at 4.6% FRR [84]
Android pattern lock [65], [86]	Swipe password	-	No	Yes	35	30 days	4-9 strokes	11.5% FRR [65]
	-	Swipe behavior (B)	Yes	Yes	26	21 days	5 strokes	19% FRR at 21% FAR [86]
PIN drawing [88]	-	Drawing pattern (B)	Yes	No	20	10 sessions	PIN attack	4.84% EER
	-	Drawing pattern (B)	Yes	No	20	10 sessions	Imitation attack	14.11% EER
Online signature [97]	User's signature	Signing pattern (B)	Yes	No	180	6 sessions (different days)	-	3.27% EER
	-	Signing pattern (B)	Yes	No	100	5 sessions (same day)	MCYT-100‡ dataset	2.72% EER
Multi-touch gesture [102]	User's gesture	Gesture pattern (P&B)	Yes	No	41	3 sessions	1 gestures	8.86% EER
	-	Gesture pattern (P&B)					1 gestures	17.17-19.23% EER
Multi-touch swipes [101]	-	Gesture pattern (B)	Yes	No	161	6 sessions	Random guessing	3.02% EER
							Statistical attack	4.69% EER
Bodyprint [103]	-	Ear shape (P)	Yes	No	12	1 session	12 samples/user	7.8% FRR at 0.5% FAR
C-*-* SCHEMES - CAMERA BASED AUTHENTICATION								
Fingerprint [109]	-	Fingerprint (P)	Yes	No	22	1 session	-	4.5% EER
Leap Password [113]	5 one-finger tap gestures	Hand geometry/ timing of tap sequence (P&B)	Yes	Yes	75	1 session	-	18.83% FRR at 1% FAR
KinWrite [42]	3-D signature	Signing pattern (B)			18	5 months	Local threshold, /random training/35 signatures	99% FRR at 0% FAR
	-	Signing pattern (B)				5 months	4 signatures	25% FRR at 25%FAR
SignWave [114]	-	Hand geometry (P)	Yes	Yes	Not reported			
Hand pose authentication [115]	Hand sign content	-	Yes	No	4	1 session	-	5.0% FAR at 6.2% FRR
	-	Hand pose (P&B)	Yes	No	4	1 session	-	5.6% FAR at 6.2% FRR
Airauth [116]	3-D gesture	Gesture pattern (B)	Yes	Yes	15	1 session	Customized gestures	0 % EER
	-	Gesture pattern (B)	Yes	Yes	15	1 session	Generic gesture	2-6 % EER
Hand gesture [117]	3-D gesture	Hand geometry / Gesture pattern (P&B)	Yes	Yes	21	1 session	4 generic gestures	1.92% EER
Face recognition [125]	-	Face (P)	Yes	No	160	6 sessions	MOBIO database‡	10.9% HTER
Body gesture [136]	Body gesture	Body build/posture / movement pattern (P&B)	Yes	Yes	40	2 sessions	Imitation attack	1.24-2.78% EER
	Body gesture	Body build/posture (P)	Yes	Yes	40	2 sessions	Imitation attack	4.22-10.28% EER
Iris [138]	-	Iris (P)	Yes	No	100	1 session	4 samples/user	0.05% EER
Gaze-based PIN entry [49]	4-digit PIN	-	Yes	Yes	21	1 session	dwell time look&shoot gaze gesture	23.8% FRR 20.6% FRR 9.5% FRR
Gaze-based graphical password [139]	Graphical password	-	Yes	Yes	45	1 session	-	27%-46% FRR
Gaze-based task independent [50]	-	Gaze movement (B)	Yes	Yes	17	1 session	-	28.7-47.1% EER
Visual stimuli [141]	-	Gaze movement (B)	Yes	Yes	30	1 session	-	6.2-7.3% EER
MS-*-* SCHEMES - MOTION SENSOR BASED AUTHENTICATION								
uWave [44]	3-D gesture	Gesture pattern (B)	Yes	No	25	7 sessions	-	3% EER
	-	Gesture pattern (B)	Yes	No	25	7 sessions	-	10% EER
In-air signature [147]	3-D signature	Signing pattern (B)	Yes	Yes	34	1 session	3 attackers	2.5% EER
Headbanger [45]	Rhythm cue	Head nodding pattern (B)	Yes	Yes	30	3 days	10 secs movement	4.43% EER
M-*-* SCHEMES - MICROPHONE BASED AUTHENTICATION								
Text-independent [156]	-	Acoustic features (B)	Yes	Yes	-	-	NIST 2008‡	3.11% EER
Whisper [155]	-	Whisper pattern (B)	Yes	Yes	60	1 session	Identification	32-98% Accuracy
Text-dependent [160]	Speaking content	Voice print (B)	Yes	No	48	2 sessions	MIT Corpus‡	7-11% EER
Bone conduction [152]	-	Skull structure (P)	Yes	Yes	10	1 session	No background noise	6.9% EER
BreathPrint [151]	-	Breath pattern (B)	Yes	Yes	10	3 session	No background noise	7-18% FAR at 2% FRR
BCI-*-* SCHEMES - BRAIN-COMPUTER INTERFACE BASED AUTHENTICATION								
Brainwave [104], [168]	Brain task	Brain print (B)	Yes	Yes	15	2 sessions	Local threshold	1.1-28.0% HTER [168]
	-	Brain print (B)			9	3 days	Day 1 train, days 2-3 test	35.5% HTER [104]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

[5] L. Findlater, J. O. Wobbrock, and D. Wigdor, "Typing on flat glass: examining ten-finger expert typing patterns on touch surfaces," in *Proc. of the 2011 annual conf. on Human factors in computing systems*, CHI '11, (New York, NY, USA), pp. 2453–2462, ACM, 2011.

[6] D. Vogel and P. Baudisch, "Shift: A technique for operating pen-based interfaces using touch," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, (New York, NY, USA), pp. 657–666, ACM, 2007.

[7] H. Lü and Y. Li, "Gesture avatar: A technique for operating mobile user interfaces using gestures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, (New York, NY, USA), pp. 207–216, ACM, 2011.

[8] D. Wigdor, C. Forlines, P. Baudisch, J. Barnwell, and C. Shen, "Lucid touch: A see-through mobile device," in *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, (New York, NY, USA), pp. 269–278, ACM, 2007.

[9] H.-Y. Chiang and S. Chiasson, "Improving user authentication on mobile devices: A touchscreen graphical password," in *Proc. of the 15th int'l conf. on Human-computer interaction with mobile devices and services*, pp. 251–260, ACM, 2013.

[10] D. Van Bruggen, S. Liu, M. Kajzer, A. Striegel, C. R. Crowell, and J. D'Arcy, "Modifying smartphone user locking behavior," in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, p. 10, ACM, 2013.

[11] "Fast IDentity Online Alliance." https://fidoalliance.org/, 2014. [Online; accessed 25-Sep-2014].

[12] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano, "The quest to replace passwords: A framework for comparative evaluation of web authentication schemes," in *Security and Privacy (SP), 2012 IEEE Symposium on*, pp. 553–567, IEEE, 2012.

[13] R. Biddle, S. Chiasson, and P. C. Van Oorschot, "Graphical passwords: Learning from the first twelve years," *ACM Computing Surveys (CSUR)*, vol. 44, no. 4, p. 19, 2012.

[14] X. Suo, Y. Zhu, and G. S. Owen, "Graphical passwords: A survey," in *Computer security applications conference, 21st annual*, pp. 10–pp, IEEE, 2005.

[15] D. Bhattacharyya, R. Ranjan, F. Alisherov, and M. Choi, "Biometric authentication: A review," *International Journal of u-and e-Service, Science and Technology*, vol. 2, no. 3, pp. 13–28, 2009.

[16] J. Unar, W. C. Seng, and A. Abbasi, "A review of biometric technology along with trends and prospects," *Pattern recognition*, vol. 47, no. 8, pp. 2673–2688, 2014.

[17] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Proc. Mag.*, vol. 33, no. 4, pp. 49–61, 2016.

[18] "Public datasets for user authentication research." http://isis.poly.edu/~benapa/authen_dataset_index.html, 2018. [Online; accessed 18-Dec-2018].

[19] A. Roy, N. Memon, and A. Ross, "Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2013–2025, 2017.

[20] Y. Xu, T. Price, J.-M. Frahm, and F. Monrose, "Virtual u: Defeating face liveness detection by building virtual models from your public photos.," in *USENIX security symposium*, pp. 497–512, 2016.

[21] C. Xu and K. Lyons, "Shimmering smartwatches: Exploring the smartwatch design space," in *Proc. of the Ninth Int'l Conf. on Tangible, Embedded, and Embodied Interaction*, pp. 69–76, ACM, 2015.

[22] P. Peltonen, E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta, and P. Saarikko, "It's mine, don't touch!: Interactions at a large multi-touch display in a city centre," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1285–1294, ACM, 2008.

[23] D. Wigdor, G. Penn, K. Ryall, A. Esenther, and C. Shen, "Living with a tabletop: Analysis and observations of long term office use of a multi-touch table," in *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop on*, pp. 60–67, IEEE, 2007.

[24] O. J. Muensterer, M. Lacher, C. Zoeller, M. Bronstein, and J. Kübler, "Google glass in pediatric surgery: an exploratory study," *International journal of surgery*, vol. 12, no. 4, pp. 281–289, 2014.

[25] Y. Wu and T. S. Huang, "Vision-based gesture recognition: A review," *Urbana*, vol. 51, p. 61801, 1999.

[26] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 52–73, 2007.

[27] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 19, no. 7, pp. 677–695, 1997.

[28] X. Zabulis, H. Baltzakis, and A. Argyros, "Vision-based hand gesture recognition for human-computer interaction," *The Universal Access Handbook. LEA*, 2009.

[29] Y. Duan, H. Deng, and F. Wang, "Depth camera in human-computer interaction: An overview," in *Intelligent Networks and Intelligent Systems (ICINIS), 2012 Fifth Int'l Conf. on*, pp. 25–28, Nov 2012.

[30] "General Motors to fit eye-tracking technology that reveals when a driver is not paying attention to the road." http://www.dailymail.co.uk/sciencetech/article-2740588/, 2014. [Online; accessed 09-Sep-2014].

[31] "Seeing Machines announces Samsung relationship." http://www.everyinvestor.co.uk/news/, 2014. [Online; accessed 09-Sep-2014].

[32] D. Li, D. Winfield, and D. J. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE*, pp. 79–79, IEEE, 2005.

[33] K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz, "Sensing techniques for mobile interaction," in *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pp. 91–100, ACM, 2000.

[34] R. E. Mayagoitia, A. V. Nene, and P. H. Veltink, "Accelerometer and rate gyroscope measurement of kinematics: an inexpensive alternative to optical motion analysis systems," *Journal of biomechanics*, vol. 35, no. 4, pp. 537–542, 2002.

[35] T. Sakaguchi, T. Kanamori, H. Katayose, K. Sato, and S. Inokuchi, "Human motion capture by integrating gyroscopes and accelerometers," in *Multisensor Fusion and Integration for Intelligent Systems, 1996. IEEE/SICE/RSJ International Conference on*, pp. 470–475, IEEE, 1996.

[36] "Honeywell Wi-Fi Smart Thermostat with Voice Control review." http://reviews.cnet.com/smart-home/honeywell-wi-fi-smart/4505-9788_7-35827868.html, 2013. [Online; accessed 04-Dec-2013].

[37] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, T. M. Vaughan, *et al.*, "Brain-computer interface technology: a review of the first international meeting," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 2, pp. 164–173, 2000.

[38] "EEG Hardware Platforms." http://neurosky.com/biosensors/eeg-sensor/biosensors/, 2017. [Online; accessed 26-Nov-2017].

[39] "Intel And Others To Enter Brain-Computer-Interface (BCI) Market, According To Mind Solutions, Inc.." http://online.wsj.com/article/PR-CO-20140114-906425.html, 2014. [Online; accessed 14-Jan-2014].

[40] "Wireless Earbuds Will Record Your EEG, Send Brainwave Data to Your Phone." http://spectrum.ieee.org/the-human-os/biomedical/devices/wireless-earbuds-will-record-your-eeg-send-brainwave-data-to-your-phone, 2016. [Online; accessed 28-May-2016].

[41] "Leap Motion." https://www.leapmotion.com/, 2017. [Online; accessed 3-July-2017].

[42] J. Tian, C. Qu, W. Xu, and S. Wang, "Kinwrite: Handwriting-based authentication using kinect," in *Proceedings of the 20th Annual Network & Distributed System Security Symposium, NDSS*, 2013.

[43] N. Sae-Bae, K. Ahmed, K. Isbister, and N. Memon, "Biometric-rich gestures: a novel approach to authentication on multi-touch devices," in *Proceedings of the 2012 ACM annual conference on human factors in computing systems*, pp. 977–986, ACM, 2012.

[44] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uwave: Accelerometer-based personalized gesture recognition and its applications," *Pervasive and Mobile Computing*, vol. 5, no. 6, pp. 657–675, 2009.

[45] S. Li, A. Ashok, Y. Zhang, C. Xu, J. Lindqvist, and M. Gruteser, "Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns," in *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–9, IEEE, 2016.

[46] K. Lai, J. Konrad, and P. Ishwar, "Towards gesture-based user authentication," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pp. 282–287, IEEE, 2012.

[47] C.-C. Yang and Y.-L. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010.

[48] F. Göbel, K. Klamka, A. Siegel, S. Vogt, S. Stellmach, and R. Dachselt, "Gaze-supported foot interaction in zoomable information spaces," in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pp. 3059–3062, ACM, 2013.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

[49] A. De Luca, R. Weiss, and H. Drewes, "Evaluation of eye-gaze interaction methods for security enhanced pin-entry," in *Proc. of the 19th Australasian Conf. on Computer-Human Interaction: Entertaining User Interfaces*, OZCHI '07, (New York, NY, USA), pp. 199–202, ACM, 2007.

[50] T. Kinnunen, F. Sedlak, and R. Bednarik, "Towards task-independent person authentication using eye movement signals," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 187–190, ACM, 2010.

[51] M. Rogowski, K. Saeed, M. Rybnik, M. Tabedzki, and M. Adamski, "User authentication for mobile devices," in *Computer Information Systems and Industrial Management*, pp. 47–58, Springer, 2013.

[52] J. Thorpe, P. C. van Oorschot, and A. Somayaji, "Pass-thoughts: authenticating with our minds," in *Proceedings of the 2005 workshop on New security paradigms*, pp. 45–56, ACM, 2005.

[53] J. Bonneau, S. Preibusch, and R. Anderson, "A birthday present every eleven wallets? the security of customer-chosen banking pins," in *Financial Cryptography and Data Security*, pp. 25–40, Springer, 2012.

[54] K. A. Nixon, V. Aimale, and R. K. Rowe, "Spoof detection schemes," in *Handbook of biometrics*, pp. 403–423, Springer, 2008.

[55] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*, pp. 1–7, IEEE, 2016.

[56] I. O. for Standardization, *ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs): Part 11: Guidance on Usability*. 1998.

[57] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP Journal on Adv. in Signal Proc.*, vol. 2008, p. 113, 2008.

[58] C. Herley, "So long, and no thanks for the externalities: the rational rejection of security advice by users," in *Proc. of the 2009 workshop on New security paradigms workshop*, pp. 133–144, ACM, 2009.

[59] S. Garfinkel and H. R. Lipford, "Usable security: History, themes, and challenges," *Synthesis Lectures on Information Security, Privacy, and Trust*, vol. 5, no. 2, pp. 1–124, 2014.

[60] L. Gorlenko and R. Merrick, "No wires attached: Usability challenges in the connected mobile world," *IBM Systems Journal*, vol. 42, no. 4, pp. 639–651, 2003.

[61] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE security & privacy*, no. 2, pp. 33–42, 2003.

[62] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[63] C. Rathgeb and A. Uhl, "A survey on biometric cryptosystems and cancelable biometrics," *EURASIP Journal on Information Security*, vol. 2011, no. 1, p. 3, 2011.

[64] V. M. Patel, N. K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 54–65, 2015.

[65] M. Harbach, A. De Luca, and S. Egelman, "The anatomy of smartphone unlocking: A field study of android lock screens," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 4806–4817, ACM, 2016.

[66] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, A. D. Rubin, *et al.*, "The design and analysis of graphical passwords," in *Proceedings of the 8th USENIX Security Symposium*, pp. 1–14, Washington DC, 1999.

[67] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon, "Passpoints: Design and longitudinal evaluation of a graphical password system," *International Journal of Human-Computer Studies*, vol. 63, no. 1, pp. 102–127, 2005.

[68] S. Chiasson, P. C. van Oorschot, and R. Biddle, "Graphical password authentication using cued click points," in *European Symposium on Research in Computer Security*, pp. 359–374, Springer, 2007.

[69] "Google Glass Help." https://support.google.com/glass/answer/4389349?hl=en, 2016. [Online; accessed 28-Nov-2015].

[70] T. Nguyen and N. Memon, "Smartwatches locking methods: A comparative study," in *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[71] S. Uellenbeck, M. Dürmuth, C. Wolf, and T. Holz, "Quantifying the security of graphical passwords: The case of android unlock patterns," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pp. 161–172, ACM, 2013.

[72] H. Siadati, P. Gupta, S. Smith, N. Memon, and M. Ahamad, "Fortifying android patterns using persuasive security framework," *UBICOMM 2015*, p. 81, 2015.

[73] Z. Min, B. Ryan, and S. Atkinson, "The factors affect user behaviour in a picture-based user authentication system: Pixelpin," in *Computer Science and Convergence*, pp. 31–42, Springer, 2012.

[74] Z. Zhao, G.-J. Ahn, and H. Hu, "Picture gesture authentication: Empirical analysis, automated attacks, and scheme evaluation," *ACM Trans. on Info. and System Security (TISSEC)*, vol. 17, no. 4, p. 14, 2015.

[75] A. E. Dirik, N. Memon, and J.-C. Birget, "Modeling user choice in the passpoints graphical password scheme," in *Proceedings of the 3rd symposium on Usable privacy and security*, pp. 20–28, ACM, 2007.

[76] Z. Zhao, G.-J. Ahn, J.-J. Seo, and H. Hu, "On the security of picture gesture authentication," in *Proceedings of the 22nd USENIX conference on Security*, pp. 383–398, USENIX Association, 2013.

[77] D. Shukla, R. Kumar, A. Serwadda, and V. V. Phoha, "Beware, your hands reveal your secrets," in *Proceedings of the 2014 ACM SIGSAC conference on Computer & communications security*, ACM, 2014.

[78] Q. Yue, Z. Ling, X. Fu, B. Liu, W. Yu, and W. Zhao, "My google glass sees your passwords!,"

[79] N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan, "Shoulder surfing defence for recall-based graphical passwords," in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, p. 6, ACM, 2011.

[80] A. Papadopoulos, T. Nguyen, E. Durmus, and N. Memon, "Illusionpin: Shoulder-surfing resistant authentication using hybrid images," *IEEE Transactions on Information Forensics and Security*, 2017.

[81] D. Weinshall, "Cognitive authentication schemes safe against spyware," in *Security and Privacy, 2006 IEEE Symp. on*, pp. 6–pp, IEEE, 2006.

[82] P. Andriotis, T. Tryfonas, G. Oikonomou, and C. Yildiz, "A pilot study on the security of pattern screen-lock methods and soft side channel attacks," in *Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks*, pp. 1–6, ACM, 2013.

[83] P. Andriotis, G. Oikonomou, A. Mylonas, and T. Tryfonas, "A study on usability and security features of the android pattern lock screen," *Information & Computer Security*, vol. 24, no. 1, pp. 53–72, 2016.

[84] B. Draffin, J. Zhu, and J. Zhang, "Keysens: Passive user authentication through micro-behavior modeling of soft keyboard interaction," in *Mobile Computing, Applications, and Services*, pp. 184–201, Springer, 2014.

[85] U. Burgbacher and K. Hinrichs, "An implicit author verification system for text messages based on gesture typing biometrics," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, (New York, NY, USA), pp. 2951–2954, ACM, 2014.

[86] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, "Touch me once and i know it's you!: implicit authentication based on touch screen patterns," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pp. 987–996, ACM, 2012.

[87] T. Van Nguyen, N. Sae-Bae, and N. Memon, "Finger-drawn pin authentication on touch devices," 2014.

[88] T. V. Nguyen, N. Sae-Bae, and N. Memon, "Draw-a-pin," *Computers and Security*, vol. 66, no. C, pp. 115–128, 2017.

[89] A. Kholmatov and B. Yanikoglu, "SUSIG: an on-line signature database, associated protocols and benchmark results," *Pattern Analysis & Applications*, 2008.

[90] M. Faundez-Zanuy, "On-line signature recognition based on vq-dtw," *Pattern Recognition*, vol. 40, no. 3, pp. 981 – 992, 2007.

[91] H. Feng and C. C. Wah, "Online signature verification using a new extreme points warping technique," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 2943 – 2951, 2003.

[92] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro, "Mcyt baseline corpus: a bimodal biometric database," *Vision, Image and Signal Processing, IEE Proceedings -*, vol. 150, pp. 395 – 401, dec. 2003.

[93] E. Argones Rua, E. Maiorana, J. Alba Castro, and P. Campisi, "Biometric template protection using universal background models: An application to online signature," *Information Forensics and Security, IEEE Transactions on*, vol. 7, pp. 269 –282, feb. 2012.

[94] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Pealba, J. Ortega-Garcia, and D. Maltoni, "An on-line signature verification system based on fusion of local and global information," in *Audio- and Video-Based Biometric Person Authentication*, vol. 3546 of *Lecture Notes in Computer Science*, pp. 627–656, Springer Berlin / Heidelberg, 2005.

[95] D. Guru and H. Prakash, "Online signature verification and recognition: An approach based on symbolic representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, pp. 1059 –1073, june 2009.

[96] N. Sae-Bae and N. Memon, “A simple and effective method for online signature verification,” in *Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the*, pp. 1–12, IEEE, 2013.

[97] N. Sae-Bae and N. Memon, “Online signature verification on mobile devices,” *Information Forensics and Security, IEEE Transactions on*, vol. 9, pp. 933–947, June 2014.

[98] A. Buriro, B. Crispo, F. DelFrari, and K. Wrona, “Hold & sign: A novel behavioral biometrics for smartphone user authentication,”

[99] N. Sae-Bae, N. Memon, and K. Isbister, “Investigating multi-touch gestures as a novel biometric modality,” in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pp. 156–161, IEEE, 2012.

[100] M. Shahzad, A. X. Liu, and A. Samuel, “Secure unlocking of mobile touch screen devices by simple gestures: You can see it but you can not do it,” in *Proceedings of the 19th Annual International Conference on Mobile Computing &#38; Networking*, MobiCom ’13, (New York, NY, USA), pp. 39–50, ACM, 2013.

[101] Y. Song, Z. Cai, and Z.-L. Zhang, “Multi-touch authentication using hand geometry and behavioral information,” in *Security and Privacy (SP), 2017 IEEE Symposium on*, pp. 357–372, IEEE, 2017.

[102] N. Sae-Bae, N. Memon, K. Isbister, and K. Ahmed, “Multitouch gesture-based authentication.,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 568–582, 2014.

[103] C. Holz, S. Buthpitiya, and M. Knaust, “Bodyprint: Biometric user identification on mobile devices using the capacitive touchscreen to scan body parts,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3011–3014, ACM, 2015.

[104] S. Marcel and J. d. R. Millán, “Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 743–752, 2007.

[105] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, “Speaker and session variability in gmm-based speaker verification,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 4, pp. 1448–1460, 2007.

[106] R. Wallace, M. McLaren, C. McCool, and S. Marcel, “Inter-session variability modelling and joint factor analysis for face authentication,” in *Biometrics (IJCB), 2011 Inter’l Joint Conf. on*, pp. 1–8, IEEE, 2011.

[107] N. Sae-Bae and N. Memon, “Quality of online signature templates,” in *Identity, Security and Behavior Analysis (ISBA), 2015 IEEE International Conference on*, pp. 1–8, IEEE, 2015.

[108] N. Sae-Bae, N. Memon, and P. Sooraksa, “Distinctiveness, complexity, and repeatability of online signature templates,” *Pattern Recognition*, vol. 84, pp. 332–344, 2018.

[109] M. O. Derawi, B. Yang, and C. Busch, “Fingerprint recognition with embedded cameras on mobile phones,” in *Security and Privacy in Mobile Information and Communication Systems*, pp. 136–147, Springer, 2012.

[110] B. Y. Hiew, A. B. J. Teoh, and O. S. Yin, “A secure digital camera based fingerprint verification system,” *Journal of Visual Communication and Image Representation*, vol. 21, no. 3, pp. 219–231, 2010.

[111] U. Park, S. Pankanti, and A. Jain, “Fingerprint verification using sift features,” in *SPIE Defense and Security Symposium*, pp. 69440K–69440K, International Society for Optics and Photonics, 2008.

[112] D. Gafurov, P. Bours, B. Yang, and C. Busch, “Guc100 multi-scanner fingerprint database for in-house (semi-public) performance and interoperability evaluation,” in *Computational Science and Its Applications (ICCSA), 2010 Int’l Conf. on*, pp. 303–306, IEEE, 2010.

[113] A. Chahar, S. Yadav, I. Nigam, R. Singh, and M. Vatsa, “A leap password based verification system,” in *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, pp. 1–6, IEEE, 2015.

[114] “Battelle SignWave^{TM} Unlock App for Leap Motion Lets You Wave Goodbye to Passwords.” http://www.marketwired.com/press-release/, 2013. [Online; accessed 09-Sep-2013].

[115] S. Fong, Y. Zhuang, and I. Fister, “A biometric authentication model using hand gesture images,” *Biomedical engineering online*, vol. 12, no. 1, p. 111, 2013.

[116] M. T. I. Aumi and S. Kratz, “Airauth: evaluating in-air hand gestures for authentication,” in *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pp. 309–318, ACM, 2014.

[117] J. Wu, J. Christianson, J. Konrad, and P. Ishwar, “Leveraging shape and depth in user authentication from in-air hand gestures,” in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, pp. 3195–3199, Sept. 2015.

[118] D. Schmidt, M. K. Chong, and H. Gellersen, “Handsdown: hand-contour-based user identification for interactive surfaces,” in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pp. 432–441, ACM, 2010.

[119] K. Cheng and A. Kumar, “Contactless finger knuckle identification using smartphones,” in *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proc. of the Int’l Conf. of the*, pp. 1–6, 2012.

[120] L. Zhang, L. Zhang, D. Zhang, and H. Zhu, “Online finger-knuckle-print verification for personal authentication,” *Pattern Recognition*, vol. 43, no. 7, pp. 2560–2571, 2010.

[121] “Hacking Leap Motion apps: Security researchers spoof biometric auto-login system.” http://venturebeat.com/2013/08/13/, 2013. [Online; accessed 09-Sep-2013].

[122] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[123] A. Bud, “Facing the future: the impact of apple faceid,” *Biometric Technology Today*, vol. 2018, no. 1, pp. 5–7, 2018.

[124] P. J. Phillips, W. T. Scruggs, A. J. O’Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, “Frvt 2006 and ice 2006 large-scale experimental results,” *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 32, no. 5, pp. 831–846, 2010.

[125] S. Marcel, C. McCool, P. Matějka, T. Ahonen, J. Černockỳ, S. Chakraborty, V. Balasubramanian, S. Panchanathan, C. H. Chan, J. Kittler, *et al.*, “On the results of the first mobile biometry (mobio) face and speaker verification evaluation,” in *Recognizing Patterns in Signals, Speech, Images and Videos*, pp. 210–225, Springer, 2010.

[126] K. Iacovino, S. M. Kywe, L. F. Cranor, and B. Ur, “Poster: Usability analysis of biometric authentication systems on mobile phones,” in *Proc. of the 10th Symposium on Usable privacy and security*, 2014.

[127] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708, 2014.

[128] G. Pan, L. Sun, Z. Wu, and S. Lao, “Eyeblink-based anti-spoofing in face recognition from a generic webcamera,” in *Computer Vision, 2007. ICCV 2007. IEEE 11th Int’l Conf. on*, pp. 1–8, IEEE, 2007.

[129] A. Boehm, D. Chen, M. Frank, L. Huang, C. Kuo, T. Lolic, I. Martinovic, and D. Song, “Safe: Secure authentication with face and eyes,” in *Privacy and Security in Mobile Systems (PRISMS), 2013 International Conference on*, pp. 1–8, IEEE, 2013.

[130] Y. Li, Y. Li, Q. Yan, H. Kong, and R. H. Deng, “Seeing your face is not enough: An inertial sensor-based liveness detection for face authentication,” in *Proc. of the 22nd ACM SIGSAC Conf. on Computer and Communications Security*, pp. 1558–1569, ACM, 2015.

[131] J. Komulainen, A. Hadid, and M. Pietikainen, “Context based face anti-spoofing,” in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth Int’l Conf. on*, pp. 1–8, IEEE, 2013.

[132] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, “Computationally efficient face spoofing detection with motion magnification,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pp. 105–110, IEEE, 2013.

[133] H.-Y. Wu, M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman, “Eulerian video magnification for revealing subtle changes in the world.,” *ACM Trans. Graph.*, vol. 31, no. 4, p. 65, 2012.

[134] J. Wu, P. Ishwar, and J. Konrad, “Two-stream CNNs for gesture-based verification and identification: Learning user style,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[135] J. Wu, J. Chen, P. Ishwar, and J. Konrad, “Two-stream CNNs for gesture-based verification and identification: Learning user style,” in *Deep Learning for Biometrics* (B. Bhanu and A. Kumar, eds.), ch. 7, pp. 159–182, Springer, 2017.

[136] J. Wu, P. Ishwar, and J. Konrad, “The value of posture, build and dynamics in gesture-based user authentication,” in *Biometrics (IJCB), 2014 IEEE Int’l Joint Conf. on*, pp. 1–8, IEEE, 2014.

[137] J. Wu, J. Konrad, and P. Ishwar, “The value of multiple viewpoints in gesture-based user authentication,” in *The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2014.

[138] K. R. Park, H. Park, B. J. Kang, E. C. Lee, and D. S. Jeong, “A study on iris localization and recognition on mobile phones,” *EURASIP Journal on Advances in Signal Proc.*, vol. 2008, p. 20, 2008.

[139] A. Forget, S. Chiasson, and R. Biddle, “Shoulder-surfing resistance with eye-gaze entry in cued-recall graphical passwords,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1107–1110, ACM, 2010.

[140] A. Bulling, F. Alt, and A. Schmidt, “Increasing the security of gaze-based cued-recall graphical passwords using saliency masks,” in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2019.2893297, IEEE Transactions on Biometrics, Behavior, and Identity Science

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3011–3020, ACM, 2012.

[141] I. Sluganovic, M. Roeschlin, K. B. Rasmussen, and I. Martinovic, "Using reflexive eye movements for fast challenge-response authentication," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1056–1067, ACM, 2016.

[142] O. V. Komogortsev, A. Karpov, L. R. Price, and C. Aragon, "Biometric authentication via oculomotor plant characteristics," in *Biometrics (ICB), 2012 5th IAPR International Conference on*, pp. 413–420, IEEE, 2012.

[143] M. Brooks, C. Aragon, and O. Komogortsev, "Poster: User centered design and evaluation of an eye movement-based biometric authentication system,"

[144] S. Eberz, K. Rasmussen, V. Lenders, and I. Martinovic, "Preventing lunchtime attacks: Fighting insider threats with eye movement biometrics," 2015.

[145] I. Odinaka, P. Lai, A. Kaplan, J. OSullivan, E. Sirevaag, and J. Rohrbaugh, "Ecg biometric recognition: A comparative analysis," 2012.

[146] R. Vogt and S. Sridharan, "Explicit modelling of session variability for speaker verification," *Computer Speech & Language*, vol. 22, no. 1, pp. 17–38, 2008.

[147] J. G. Casanova, C. S. Ávila, A. de Santos Sierra, G. B. del Pozo, and V. J. Vera, "A real-time in-air signature biometric technique using a mobile device embedding an accelerometer," in *Networked Digital Technologies*, pp. 497–503, Springer, 2010.

[148] D. Gafurov, E. Snekkenes, and P. Bours, "Gait authentication and identification using wearable accelerometer sensor," in *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*, pp. 220–225, IEEE, 2007.

[149] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth Int'l Conf. on*, pp. 306–311, IEEE, 2010.

[150] Q. Jin, S.-C. S. Jou, and T. Schultz, "Whispering speaker identification," in *Multimedia and Expo, 2007 IEEE International Conference on*, pp. 1027–1030, IEEE, 2007.

[151] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee, "Breathprint: Breathing acoustics-based user authentication," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 278–291, ACM, 2017.

[152] S. Schneegass, Y. Oualil, and A. Bulling, "Skullconduct: Biometric user identification on eyewear computers using bone conduction through the skull," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1379–1384, ACM, 2016.

[153] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D. Petrovska-Delacrétaz, and D. A. Reynolds, "A tutorial on text-independent speaker verification," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 4, pp. 1–22, 2004.

[154] A. Neustein and H. A. Patil, *Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism.* Springer, 2012.

[155] Q. Xiaohong and Z. Heming, "Adaptive order of fractional fourier transform for whispered speaker identification," in *Automatic Control and Artificial Intelligence (ACAI 2012), International Conference on*, pp. 363–366, IET, 2012.

[156] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech communication*, vol. 52, no. 1, pp. 12–40, 2010.

[157] H. Li, B. Ma, K.-A. Lee, H. Sun, D. Zhu, K. C. Sim, C. You, R. Tong, I. Karkkainen, C.-L. Huang, *et al.*, "The i4u system in nist 2008 speaker recognition evaluation," in *2009 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, pp. 4201–4204, IEEE, 2009.

[158] R. Johnson, T. E. Boult, and W. J. Scheirer, "Voice authentication using short phrases: Examining accuracy, security and privacy issues," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pp. 1–8, IEEE, 2013.

[159] S. Trewin, C. Swart, L. Koved, J. Martino, K. Singh, and S. Ben-David, "Biometric authentication on a mobile device: a study of user effort, error and task disruption," in *Proceedings of the 28th Annual Computer Security Applications Conference*, pp. 159–168, ACM, 2012.

[160] R. H. Woo, A. Park, and T. J. Hazen, "The MIT mobile device speaker verification corpus: data collection and preliminary experiments," in *Speaker and Language Recognition Workshop, 2006. IEEE Odyssey 2006: The*, pp. 1–6, IEEE, 2006.

[161] P. Tresadern, C. McCool, N. Poh, P. Matejka, A. Hadid, C. Levy, T. Cootes, and S. Marcel, "Mobile biometrics (mobio): Joint face and voice verification for a mobile platform," *IEEE Pervasive Computing*, vol. 99, 2012.

[162] A. Larcher, K. A. Lee, B. Ma, and H. Li, "Text-dependent speaker verification: Classifiers, databases and rsr2015," *Speech Communication*, vol. 60, pp. 56–77, 2014.

[163] M. Hébert, "Text-dependent speaker recognition," in *Springer handbook of speech processing*, pp. 743–762, Springer, 2008.

[164] Z. Wu, S. Gao, E. S. Cling, and H. Li, "A study on replay attack and anti-spoofing for text-dependent speaker verification," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, pp. 1–5, IEEE, 2014.

[165] S. Marinov and H. i Skövde, "Text dependent and text independent speaker verification systems. technology and applications," 2003.

[166] D. E. Sturim, D. A. Reynolds, R. B. Dunn, and T. F. Quatieri, "Speaker verification using text-constrained gaussian mixture models," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE Int'l Conf. on*, vol. 1, pp. I–677, IEEE, 2002.

[167] T. Itoh, K. Takeda, and F. Itakura, "Acoustic analysis and recognition of whispered speech," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE Int'l Conf. on*, vol. 1, pp. I–389, IEEE, 2002.

[168] J. Chuang, C. W. Hamilton Nguyen, and B. Johnson, "I think, therefore i am: Usability and security of authentication using brainwaves," in *Proceedings of the Workshop on Usable Security, USEC*, vol. 13, 2013.

[169] J. Grizou, I. Iturrate, L. Montesano, P.-Y. Oudeyer, and M. Lopes, "Calibration-free bci based control.," in *AAAI*, pp. 1213–1220, 2014.

[170] J. Chuang, T. Maillart, and B. Johnson, "My thoughts are not your thoughts," in *Proc. of the Workshop on Usable Privacy & Security for wearable and domestic ubIquitous DEvices (UPSIDE'14)*, vol. 14, 2014.

[171] J. Chuang, "One-step two-factor authentication with wearable biosensors," in *"Who are you?! Adventures in Authentication" (WAY'14), SOUPS'14 Workshop*, pp. 1–4, ACM, 2014.

[172] D. K. Yadav, B. Ionascu, S. V. K. Ongole, A. Roy, and N. Memon, "Design and analysis of shoulder surfing resistant pin based authentication mechanisms on google glass," in *Financial Cryptography and Data Security*, pp. 281–297, Springer, 2015.

[173] A. Forget, S. Chiasson, and R. Biddle, "Towards supporting a diverse ecosystem of authentication schemes," in *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

Napa Sae-Bae is currently a lecturer in Department of Computer Science at Rajamangala University of Technology Suvarnabhumi. She received Bachelor of Engineer in Telecommunication Engineering department and Master of Engineer in Information System Engineering department from King Mongkut's Institute of Technology Ladkrabang, Thailand, and received a PhD in Computer Science from Polytechnic School of Engineering, New York University. Her research interests lie in the area of biometric, authentication, signal processing, pattern recognition, and consumer security. She has published 15 articles in journals and conference proceedings in different venues and holds a patent on multi - touch gesture for user - authentication on touch interface.

Jonathan Wu is currently a senior applied scientist at Amazon. He received his BS/MS in Electrical and Computer Engineering from Carnegie Mellon University and his PhD in Electrical Engineering from Boston University. His research interests lie in deep learning, and computer vision, specifically in the domains of object recognition, and biometric authentication.

Nasir Memon is a professor in the Department of Computer Science and Engineering at NYU Tandon School of Engineering. He has published over 250 articles in journals and conference proceedings and holds a dozen patents in image compression and security. He has won several awards including the Jacobs Excellence in Education award and several best paper awards. He has been on the editorial boards of several journals and was the Editorin-Chief of the IEEE Transactions on Information Security and Forensics. He is an IEEE Fellow and an SPIE Fellow.

Janusz Konrad (M'93–SM'98–F'08) received the Master's degree from Technical University of Szczecin, Poland and the Ph.D. degree from McGill University, Montreal, QC, Canada. He is currently a Professor at Boston University, Boston, MA, USA. He has been actively involved in the IEEE Signal Processing Society; he is currently its Distinguished Lecturer and a Senior Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and recently served as a member-at-large of its Conference Board. He is also an Area Editor of the EURASIP Signal Processing: Image Communications journal. He is a co-recipient of the 2001 IEEE Signal Processing Magazine Award, the 2004-2005 EURASIP Image Communications Best Paper Award and the AVSS-2010 Best Paper Award. He was the Technical Program Co-Chair of ICIP-2000 and AVSS-2010 as well as the General Chair of AVSS-2013. His current research interests include video processing and computer vision, stereoscopic and 3-D imaging and displays, visual sensor networks, human-computer interfaces, and cybersecurity.

Prakash Ishwar received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Bombay in 1996 and the M.S. and Ph.D. degrees in Electrical and Computer Engineering (ECE) from the University of Illinois Urbana-Champaign in 1998 and 2002, respectively. After two years as a post-doctoral researcher in the Electrical Engineering and Computer Sciences Department at the University of California, Berkeley, he joined the Faculty of Boston University where he is currently a Professor in the Department of Electrical and Computer Engineering and the Division of Systems Engineering and an Affiliate Professor with the Department of Computer Science in the College of Arts and Sciences. His current research centers on data science to advance statistical and computational tools for learning and inference problems using both model-based and data-driven methods.

He is a recipient of a 2005 NSF CAREER Award, a co-recipient of the AVSS10 Best Paper Award, a co-winner of the ICPR10 Aerial View Activity Classification Challenge, and co-author of the Student Best Paper Award in the NIPS14 Workshop on Analysis of Rank Data. He has published more than 100 peer-reviewed articles, received numerous grants, served as the Associate Editor of the IEEE Transactions on Signal Processing for two terms, and served as an elected member of the IEEE Signal Processing Theory and Methods (SPTM) Technical Committee and the IEEE Image, Video, and Multidimensional Signal Processing (IMDSP) Technical Committee.