Interaction Evaluation of a Mobile Voice Authentication System

Oscar Miguel-Hurtado
School of Engineering and Digital Arts
University of Kent
Canterbury, United Kingdom
O.Miguel-Hurtado-98@kent.ac.uk

Ramon Blanco-Gonzalo
University Group for Identification Technologies (GUTI)
University Carlos III of Madrid
Madrid, Spain
rbgonzal@ing.uc3m.es

Richard Guest
School of Engineering and Digital Arts
University of Kent
Canterbury, United Kingdom
R.M.Guest@kent.ac.uk

Chiara Lunerti
School of Engineering and Digital Arts
University of Kent
Canterbury, United Kingdom
C.Lunerti@kent.ac.uk

***Abstract*— Biometric recognition is nowadays widely used in smartphones, making the users' authentication easier and more transparent than PIN codes or patterns. Starting from this idea, the EU project PIDaaS aims to create a secure authentication system through mobile devices based on voice and face recognition as two of the most reliable and user-accepted modalities. This work introduces the project and the first PIDaaS usability evaluation carried out by means of the well-known HBSI model. In this experiment, participants interact with a mobile device using the PIDaaS system under laboratory conditions: video recorded and assisted by an operator. Our findings suggest variability among sessions in terms of usability and feed the next PIDaaS HCI design.**

Keywords*— *Biometrics, Usability, Evaluation, Mobile

I. INTRODUCTION

Smartphones have become an inseparable companion in our daily lives. In recent years, the introduction of fingerprint sensors as an integral component of devices manufactured by the main vendors has made the use of biometric authentication mechanisms mainstream. This interest in biometric authentication has made many service providers consider incorporating biometric authentication strategies within their mobile applications for securely accessing online services. Market analysis suggests that several finance corporations are considering including voice and face biometric authentication mechanisms to their mobile applications. However, the inherent unconstrained nature of these devices and the wide demographic of potential users have brought new challenges for the biometric community. Mobile phones can be used in many different unconstrained environments which implies uncontrolled biometric sample acquisition. This creates a complex challenge for the analysis of the interaction between human and smartphones.

Within this context, the Private IDentification as a Service (PIDaaS) European project [1] aims to create a multi-factor authentication solution through mobile devices on the cloud that can be easily incorporated to the workflow of third-party applications or services based on three main technologies: biometrics as a main factor to guarantee identity, biometric template protection to ensure user's privacy and a Life Management Platform in order to control the personal data shared with third parties.

The PIDaaS platform will be piloted using three end-user service providers: e-health, e-citizen and e-commerce from three different countries (Spain, Italy and Latvia). In order to ensure the usability and effective interaction with the final PIDaaS Mobile Application (PMA), the Human-Biometric-Sensor Interaction (HBSI) framework for biometric evaluations [2] has been applied to the initial version.

The main target of this evaluation is to analyse how the final users of the platform interact with the PMA on their mobile devices. The experimental assumptions are: i) the final users will be at their work place (quiet or noise office) and; ii) it is the first user's interaction with the PMA, without previous training. This evaluation allows the PIDaaS developer partners to test the PMA user interfaces and ensure the usability and intuitiveness of the implementation. These are key factors for the final users to switch from common authentication mechanisms based on user and password to a more robust and secure mechanism based on biometrics. The results of this evaluation provide valuable feedback to the PIDaaS platform developers to create an enhanced final version of the PMA.

In this work, the novel and extended HBSI methodology applied for evaluation of the PIDaaS voice authentication system is presented. This methodology proposes the integration of mobile analytical tools for logging the user-biometric system interaction and a simplification of the HBSI analysis. Mobile analytical tools log information on how users proceed within the

application, enabling the automatic calculation of HBSI effectiveness and efficiency metrics. Furthermore, the analysis of the timing information can enable the generation of cognitive metrics such as learnability and memorability. Together with the mobile analytical tools, the interaction between the user and the device is also video-recorded, along with logs of the biometric system output at the server side for subsequent analysis and calculation of HBSI metrics related to users' presentations. Finally, the users' perceptions about the use of voice mobile biometrics as an authentication mechanism were collected before and after the experiment, jointly with demographic and previous biometric experience information.

II. Biometric Interaction Analysis

In order to perform such as analysis, the HBSI framework [2] devised at the Purdue University has been applied. The HBSI framework proposes interaction metrics to reach a deeper understanding of commonly used biometric performance metrics such as Failure to Enrol (FTE) or Failure to Acquire (FTA) [3]. Furthermore, the HBSI framework defines metrics related to satisfaction, efficiency, effectiveness (usability), cognitive (learnability and memorability) and physical metrics (ergonomics), sample quality and processing capabilities (signal processing).

The HBSI framework has been tested for different modalities: hand geometry [4], fingerprint [5], face [6] and dynamic signature [7]. The HBSI has also been applied to mobile device implementations for dynamic signature [8]. The work presented in this article applies the HBSI framework to a voice recognition system within mobile environments. The HBSI model is formed by three elements: human (the participants of this experiment), sensor (the mobile device) and biometric system (the PIDaaS platform) and their interactions: ergonomics, usability and sample quality; as depicted in Figure 1:

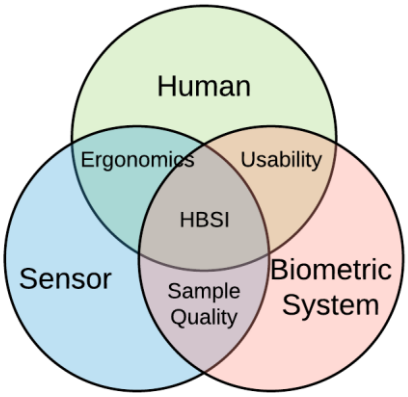

Figure 1 The HBSI model [2]

The human-sensor segment is related to how the users present their biometrics characteristics to the sensor. The sensor in the PMA is the microphone of the mobile device. The analysis of this interaction help us to understand how to better guide users in order to obtain biometrics samples of sufficient quality. The human-biometric system component deals with how users interact with the PIDaaS Platform, mostly through the PMA interface. In this case the evaluation allows to design a better user-centric interface for the final PMA versions. The sensor-system segment is measured through the quality of the biometrics captured samples.

The HBSI presentation metrics are defined by the type of presentations the users make, and their categorisation depends on whether the user makes a correct or incorrect presentation, whether the presentation was detected by the biometric system and whether the presentation was correctly classified by the biometric system [2]. Taking into account these three factors the presentations are classified as unsuccessful interactions: Defective Interaction (DI), Failure to Detect (FTD), Concealed Interaction (CI), Failure to Process (FTP), False Interaction (FI); or as a successful interaction: Successfully Processed Sample (SPS) (see Figure 2).

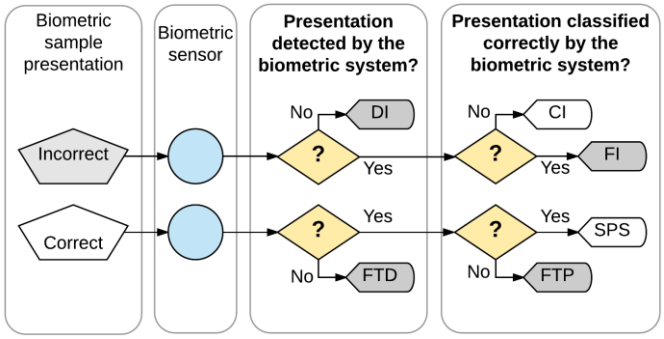

Figure 2 HBSI presentation metrics

This classification of the user's biometric presentation into six types provides a deeper view of how users interact with the biometric system, considering sample presentations that could be overlooked at more common biometric performance evaluation due to the biometric sensors failing to detect the user's interaction (i.e. DI or the FTD presentations). Furthermore, the HBSI framework also defines a set of metrics taken from the usability and the biometric system performance disciplines in order to obtain a holistic view of the user's interaction with the biometric implementation, Figure 3:

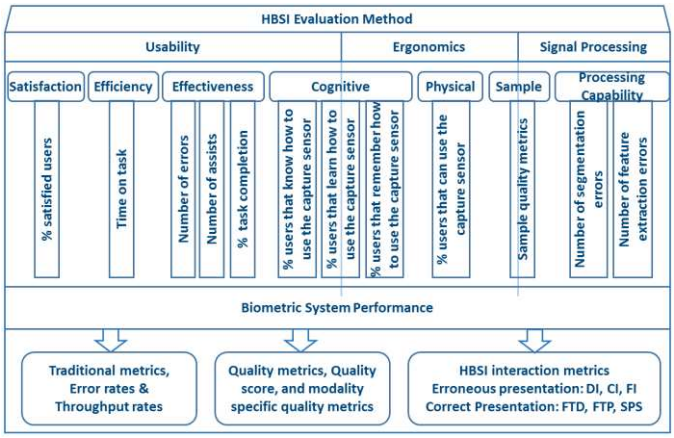

Figure 3 The HBSI evaluation metrics [9]

In this work, we propose an integration methodology for the evaluation of mobile biometrics systems interaction, by using mobile analytical tools and biometric system logs to ensure the sufficient information to enable and facilitate the HBSI analysis. Mobile analytical tools, such as Flurry Analytics [10], are commonly used for mobile applications development providing a deep understanding of the user-application interaction: how users navigate the different options, what events they are conducting. These tools provide powerful visualization and user segmentation capabilities. The information logged by these tools allows for an understanding of what and when the users do within the application, therefore, enabling the calculation of HBSI metrics related to effectiveness and efficiency. Furthermore, the analysis of timing information provided by the mobile analytical tool logs allows cognitive HBSI evaluation such as "how the user learns to use the system" (learnability) or "how the user remembers how to use the system" (memorability). The logged information at the server is related to the sample quality, segmentation, feature extraction errors and comparison scores. The participant's biometric presentations are classified following the HBSI presentation metrics using the operator's logs and the output of the quality module

III. The PIDaaS Platform

The PIDaaS Platform provides an innovative identity management service relying on three main components:

- **Biometrics Template Protection Schemes** (BTPS) component. This component ensure the privacy of the biometric data provided by the users, allowing the creation of multiple biometric pseudo-identities for using in different services providers. This component is split in two modules. The BTPS encoder, implemented within the PMA, which generates the pseudo-identities from the biometric samples. The second module, the BTPS verifier, is located at the PIDaaS server side and performs the comparison between the user's pseudo-identities stored at the server and the one created for a specific authentication request at the mobile side.
- **Life Management Platform** (LMP) allows the PIDaaS users to control their biometric pseudo-identities (renovation, cancelation, expiration dates). Through this component PIDaaS users will be also able to get a historical record of where (which service providers) and when (date and time plus other metadata) their data have been used for authentications.
- **PIDaaS backend,** which provides a gateway to both the PIDaaS Mobile Application and the service providers to access the PIDaaS services located at the LMP component.

A simplified PIDaaS Platform architecture is depicted in Figure 4.

In this experiment, a beta version of the PIDaaS Server has been tested in order to provide feedback to the developer partners before the final version is created. This beta version did not have the BTPS module incorporated. Instead of the BTPS module, a voice biometric algorithm based on MISTRAL [11], an open source software for biometrics applications, has been implemented for the voice biometric authentication.

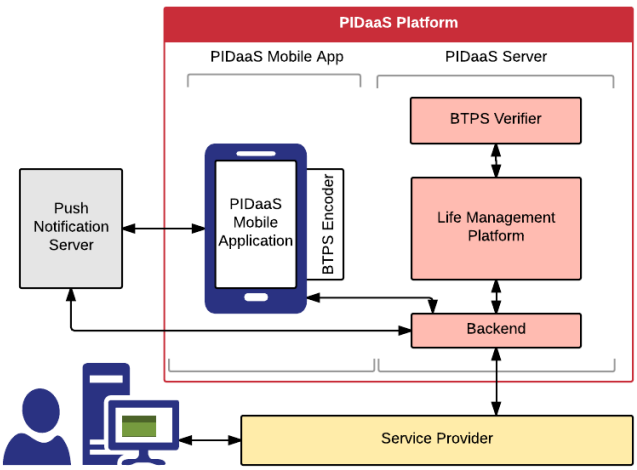

Figure 4 PIDaaS platform architecture

A. *PIDaaS Mobile Application (PMA)*

The PMA is the main user interface for the PIDaaS Platform along with a PIDaaS website where users can control their pseudo-identities. The users need to register into PIDaaS platform using the PMA and provide their voice sample in order to generate the user's voice template (enrolment process). This template is stored and managed by the LMP component. Once the registration is completed, the PIDaaS user will be able to login into different service providers through an authentication delegation. This authentication delegation will trigger a push notification at the user's mobile device registered on PIDaaS. Users will process the authentication request by confirming the authentication request (confirmation interface), providing a new voice sample (voice sample acquisition interface) that will be compared with the user's template (biometric comparison interface). Both the enrolment and the authentication processes are summarized in Figure 5.

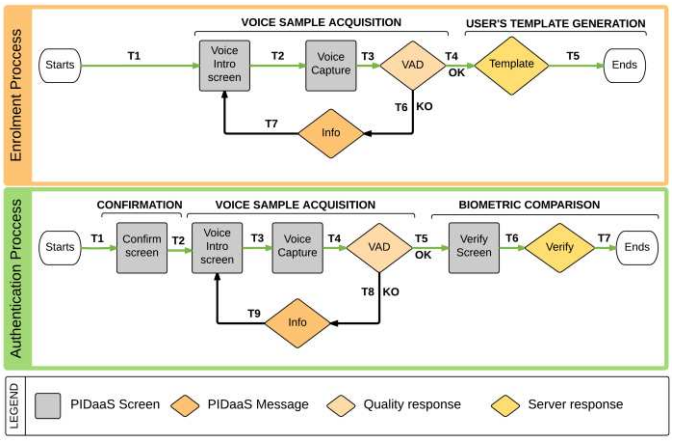

Figure 5 PIDaaS Mobile Application voice flowchart

The voice acquisition interface is shared for both enrolment and authentication processes. The users are prompted with a voice introduction screen where they will be given instruction

on how to provide voice samples. After pressing the "start" button, the user repeats a random sequence of five numbers that appear on the screen. Once the voice is recorded, a quality check module named Voice Activation Detection (VAD) analyses the voice sample and classifies as either correct or incorrect. If the VAD output is correct, a user's template is generated and stored at the LMP module during the enrolment process. For the authentication process, after a VAD correct voice sample acquisition, a new screen is prompted to the user. This screen allows the user to decide whether to proceed or not with the final step of comparing the voice sample against the user's template. In the case that the VAD module classifies the voice samples as incorrect, a message informs the users and redirect them to the voice introduction screen to re-start the voice acquisition.

IV. Experiment Methodology

In this section a complete description of the experimental methodology is provided within the following sub-sections: evaluation crew, scenario settings, participants guidance, and HBSI evaluation metrics used.

A. Evaluation crew

This experiment was conducted with the collaboration of 27 participants from the University of Kent. The only conditions for joining the experiment were to be over 18 years old and be able to speak English fluently. The age range was from 21 to 57 years old, Figures 6 and 7 show the age and nationality distributions:

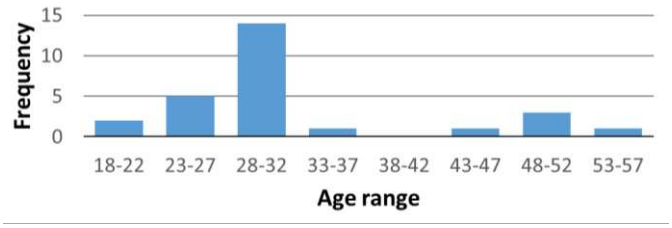

Figure 6 Evaluation crew age histogram

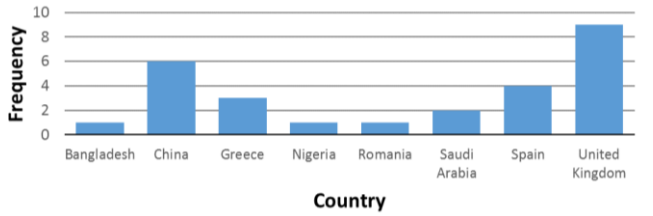

Figure 7 Evaluation crew nationality distribution

Regarding the participant's previous experience with voice biometrics, 66% (18 out of 27) had experience with this modality and 4 of these participants also had previously used voice biometric technology in mobile devices.

Prior to start the evaluation, the participants were informed about the aims of the experiment and how it would be conducted and, if participant agrees, an acceptance sheet was signed.

B. Scenario settings and devices used

The experimental scenario aims to recreate realistically the environments where the PMA is expected to be used: office and home.

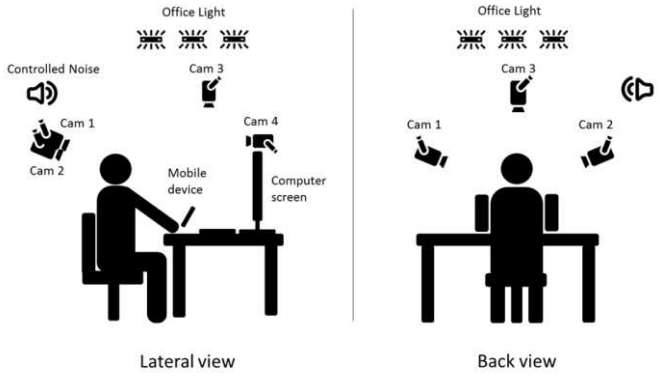

Figure 8 Scenario settings

The environment comprised a desk, a computer, a chair and the mobile device in a well illuminated quiet environment. The office environment has a background noise level ranging from a quiet office (around 40dB) to a large office (around 50dB) [12]. In order to simulate a large office environment, a noise background audio file was played and the volume adjusted to meet 50dB from a speaker connected to the test administrator's computer. The data collection room was windowless with light provided by fluorescent tubes located within the ceiling.

The user's interaction was video-recorded by four cameras. Two Sony PJ410 Handycam video cameras were used to record the user's interaction from left and the right view sides (cam1 and cam2 in Figure 8). Two Microsoft Lifecam Studio Webcams were used to record the user's interaction from top and front view sides (cam3 and cam4 in Figure 8).

The mobile device used for this experiment was an iPhone 5S. This device has a 4" screen with a resolution of 640x1136 pixels. The frontal camera has a 1.2 megapixel sensors with and an aperture of f/2.4. The iPhone 5S had 3 built-in microphones and noise cancelation technology. The microphones are located as depicted in Figure 9.

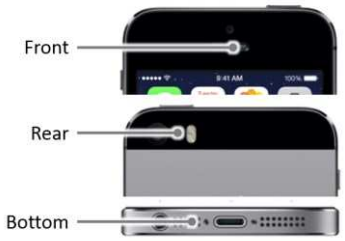

Figure 9 IPhone 5S microphone locations

The users also had to interact with a computer in order to complete different web tasks between verifications (explained below). To do so, a computer screen, keyboard and mouse were located at the participant's desk.

C. User guidance and training

One of the main aims of the experiment was to understand and measure the participant's experience at their first contact with the PMA. The learnability (defined as "*How easy is it for users to accomplish basic tasks the first time they encounter the design?*" [13]) and memorability (when users return to the design after a period of not using it, how easily can they re-establish proficiency? [13]) of the biometric interfaces are key factors for user satisfaction. Therefore, user guidance and training was kept to a minimum being provided mostly by the participant information sheet. Participants were asked to behave as if the operator wasn't at the room and only ask for assist when they don't know how to proceed when interacting with the PMA.

The experiment was split in three sessions in different weeks in order to allow potential voice variations. In the first Session the participants registered for the PIDaaS platform and created their voice templates. After registration in the first Session, the operator generates five authentication requests which are forwarded to the mobile device and the participant execute these requests. Amid the authentication requests, the participants are asked to perform a variety of tasks using a browser on the computer (such as checking the weather). These tasks are intended to distract users between consecutive authentication requests in order to avoid simple repetitions and therefore simulate a more realistic data collection. During the second and the third sessions, participants are asked to login to the PIDaaS platform using the PIDaaS Mobile Applications, execute 5 authentications requests, renew their voice template and carry out another 5 authentications requests (always with web-browsing tasks in between authentications). Three of these last 5 authentications requests are carried out within noisy office environmental conditions.

A brief PMA manual was provided to the participants who decided as and when to consult the document. The manual detailed how the most common PMA actions should be performed, such as: enrolment at the PIDaaS Platform (providing: email, pin, password and voice sample), login, response to authentication requests and renew the voice participant's templates.

Upon the completion of the experimental sessions, the participants filled a post-experiment questionnaire about their biometric technology perceptions.

D. HBSI evaluation metrics

In order to apply the HBSI framework for the biometric-interaction evaluation, the definition of correct and incorrect voice biometric sample presentation was provided. Within the PMA context, a correct voice presentation was defined as when the participant repeats, synchronized with the PMA voice capture interface, the sequence of 5 numbers as they are presented within the mobile screen, in a distinguishable way, with clarity under a reasonable background noise level. An incorrect voice biometric sample presentation was defined as a presentation which doesn't fit the correct voice presentation criteria. An incorrect voice presentation might be due, but not restricted, to the following reasons: a) the user did not repeat all the numbers in the sequence, b) the user did not repeat the same numbers, c) the user repeated the numbers asynchronously, d) the user made an interruption during the sequence, e) the participants occluded the bottom microphone.

Once the correct and incorrect presentations were defined, the 6 HBSI presentation categories (Figure 2) can be mapped within PIDaaS experiment context. Within this context, the experiment assumed that the microphone will always properly record while the PMA shows the sequence of five numbers to the participant. This assumption removes the possibility of DI or FTD.

A CI occurred when the biometric system successfully classifies an incorrect presentation. This classification was made by the VAD module. Due to this misclassification, a CI will be sent to the biometric system for further processing and enrolment or comparison. On the other hand, if the VAD module classified an incorrect presentation as correct, the presentation will fall into the FI type.

If there was a correct presentation and the biometrics system successfully classified as such, the presentation will fall into the Successfully Processed Sample (SPS) and is processed by the biometric system.

Besides the HBSI presentation metrics, the HBSI framework included usability metrics as well: efficiency, effectiveness and satisfaction.

- **Efficiency** was defined as the time spent on performing a task (enrolment or verification) once the users have learned how to proceed. In order to calculate the efficiency of the authentication requests, the time spent for the last 3 authentications on each session was used. Only successful authentication requests without mistakes (i.e. correct voice presentation and successful VAD result) have been taken into account.
- **Effectiveness** refers to the extent to which the product behaves in the way that users expect it to and the ease with which users can use it to do what they intend. This is usually measured quantitatively with the following indicators:
    - Number (%) of errors detected by the test operator (incorrect voice presentations).
    - Number (%) of assists during performing a task.
    - Task completions rate: % of successful voice presentations and correctly classified as such by the VAD module over the total number of presentations at the first attempt.
- **Satisfaction**: measured by means of a questionnaire after the experiment, asking about different aspect of the PMA.

In terms of cognitive metrics, in this work the learnability and memorability as defined in [13] of the PMA have been analysed.

- **Learnability** is related to the % of users that learn how to use the system (i.e. how easy is it for users to accomplish basic tasks the first time they encounter the design?). It was measured by:
    - % of incorrect presentations and,
    - % of successfully completed tasks (without assistance)

at the first attempt during the first Session.

- **Memorability** is related to how the users interact with the application after a period of inactivity. It was measured through the evolution of the learnability metrics at the first attempt during the second and third sessions.

V. Results

A. HBSI presentation metrics

Figure 10 depicts the distribution of HBSI presentation metrics at the authentication task. The categorization of the voice presentation as correct of incorrect was manually done based on the audio-video recordings. Most of the presentations (90%) were SPSs. 3% of the incorrect presentations were wrongly classified as correct by the VAD system (FI) and, therefore, sent also to the biometric server system for comparison, whilst 3% of the correct presentations were wrongly classified as incorrect (FTP), which lead to an overall VAD module error ratio of 7%. Finally, 3% of the presentations were incorrect and appropriately classified by the VAD system, and therefore, not sent to the server for further processing.

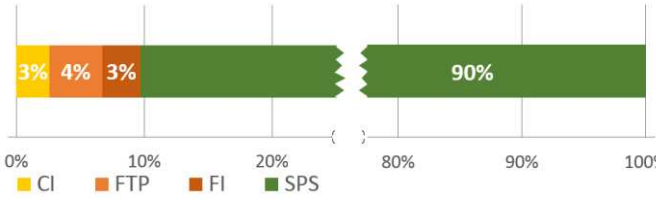

Figure 10 HBSI presentation metrics for voice authentications

B. Usability metrics: efficiency, effectiveness and satisfaction

The efficiency of the enrolment task has been measured over the first Session and for those registrations performed without mistakes (voice correct presentations and successful registration result). The average task-times have been automatically extracted from the logs generated by the mobile analytical tools, Flurry Analytics. The average task-times across sessions are shown at Figure 11 split in voice interface and server times.

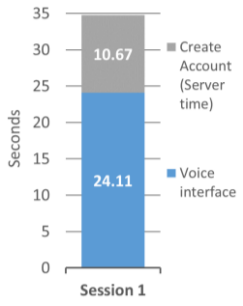

Figure 11 Average enrolment time-task

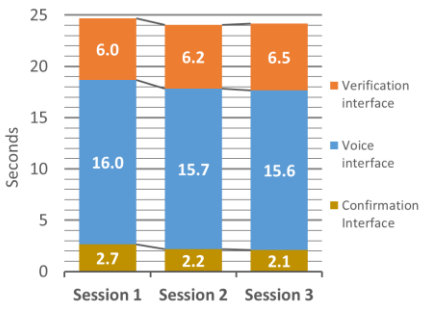

Figure 12 Average verification time-task

The voice acquisition during the enrolment process took an average of 24.1s (±4.9s) and the further processing at the server side took an average of 10.7s (±0.5s). This represents 25% of the whole registration process which took on average 99s (adding email and PIN registrations plus different intermediate screens).

The verification task-time evolution through sessions is shown in Figure 12, wherecan be seen the average times for the 3 verification steps: confirmation, voice acquisition and biometric comparison as detailed in Figure 5. The average time to perform an authentication drops slightly from 24.7s (±2.3s) at the first Session to 24.1s (±1.3s) at the second Session and it stabilises at the third Session. It can be seen how both the confirm interface and voice acquisition times reduce from Session 1 to 2. However, the verify process increased through sessions, most likely due to participant's tiredness.

In terms of effectiveness, three metrics have been analysed: a) number (%) of errors, b) number (%) of assist and c) task completions rate.

The percentage of errors for enrolment and authentication requests are shown in Figures 13 and 14. Figure 13 shows a high rate of incorrect voice presentations at the enrolment process (28.6%). However, after enrolment, and through the 3 different sessions, this rate rapidly decreases as the user learns how to interact with the PMA voice interface. Most of the incorrect presentation at enrolment were due to lack of clarity in the instruction provided at the voice acquisition interface. The most unclear steps have been detected and feedback was provided to the developer team.

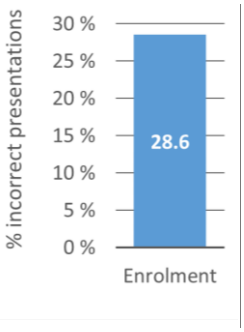

Figure 13 % incorrect presentations at enrolment

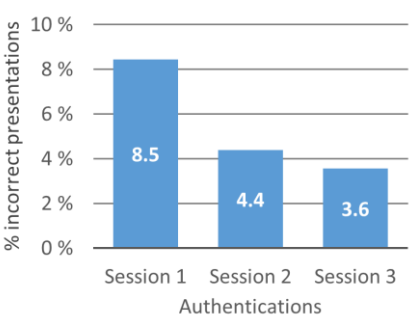

Figure 14 % incorrect presentations at authentication requests

Figures 15 and 16 shows the percentage of assist in the voice enrolment (3%) and the authentication requests (3% for the first Session and none afterwards). It should be taken into account that the participants were asked to not ask the operator. Yet again, the HBSI evaluation allowed to detect where the participants most frequently asked for assist and feedback was provided to the developer team in order to improve the interfaces.

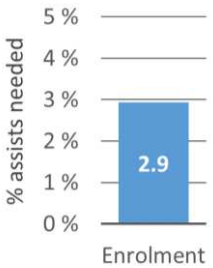

Figure 15 % of assists needed at enrolment

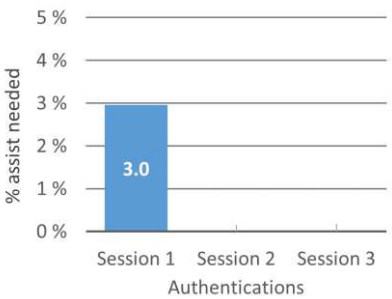

Figure 16 % of assists needed at authentication request

Finally, the percentage of task completion rates are shown in Figure 17 and 18. Only 57% of the enrolments were successful, mostly due to the high rate of incorrect presentations during enrolment and due to incorrect classifications from the VAD. The first Session shows the lowest percentage, 89.2%. In Sessions 2 and 3 the number of incorrect presentations decreases, and therefore the percentage of completion task increases. The slightly lower percentage at Session 3 could be explained by user tiredness.

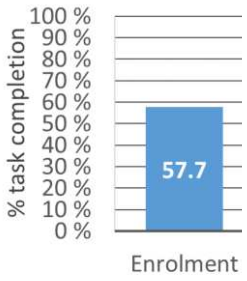

Figure 17 % successful task completion for enrolment

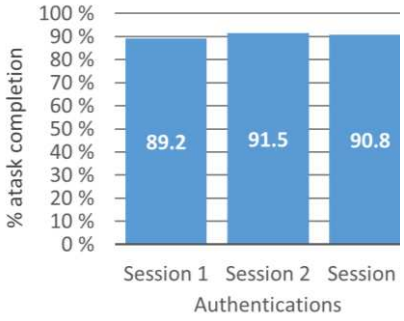

Figure 18 % successful task completion for authentication requests

Regarding satisfaction, the participants were asked to rate from 1 to 5 their satisfaction with the enrolment process, the use of voice biometrics for authentication and their overall experience with the PIDaaS platform. The average satisfaction values are shown in Figure 19:

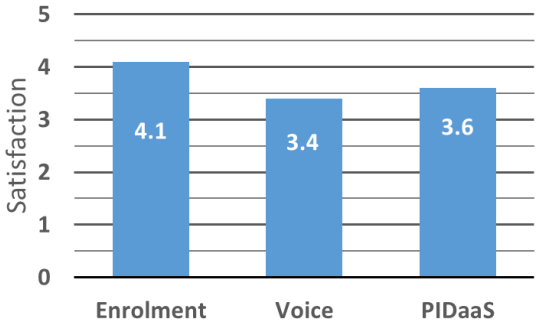

Figure 19 Average satisfaction values

C. Cognitive metrics: learnability and memorability

Learnability has been measured with the number of incorrect presentations, the number of successfully completed tasks (without assistance) in the first attempt during the first Session. These rates are shown in Figure 20-23.

Both the overall enrolment process and the voice enrolment interfaces show poor learnability in terms of incorrect presentations (Figure 20) and successfully completed tasks (Figure 21), especially in Session 1 with less than 40% of participants being able to complete the process successfully at the first attempt without assistance. These results suggest that the interface and the application's user guidance should be reconsidered. Regarding the learnability of the authentications request task, the percentage of incorrect presentations (Figure 22) and the percentage of successfully completed authentication tasks (Figure 23) in Session 1 for the voice interface show acceptable rates. In terms of voice presentation, participants learn how to proceed from the enrolment phase. However, the percentage of successfully completed authentication tasks in Session 1 is again below 50%, due to the number of assists demanded by the participants, which again suggests that the interface and guidance provided within the application for this task should be reconsidered.

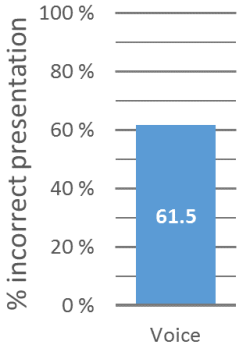

Figure 20 % incorrect presentations for voice at enrolment

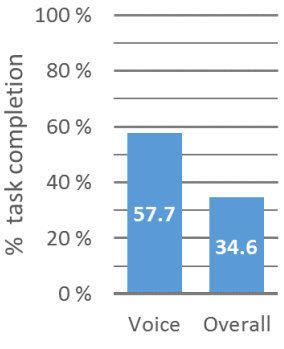

Figure 21 % of successfully completed enrolment which didn't need assistance at enrolment

Memorability assesses how the users interact with the application after a time interval of one week. The memorability has been analysed using the evolution of the learnability metrics through the 3 sessions, in order to measure how the users interact with the applications at their first attempt during Session 2 and 3 in terms of number of incorrect presentations and number of successfully completed authentications. The memorability metrics are shown in Figures 22 and 23. The percentage of incorrect presentations in authentication tasks is the same through the last 3 sessions, which indicates the incorrect presentations doesn't present memorability issues. On the other hand, the percentage of successfully completed authentication tasks improves significantly through sessions, for the overall authentication process between Session 1 and 2, improving slightly in Session 3. The percentage of incorrect presentation in Session 2 and 3, and even more significantly, the increase on the percentages of successfully completed authentication tasks during the last 2 sessions indicates an easily memorable interaction even after a week without using it.

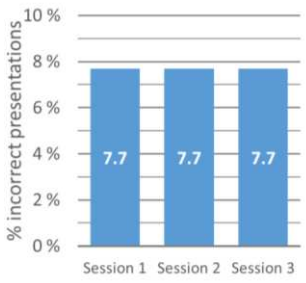

Figure 22 % incorrect presentations for voice during authentication requests

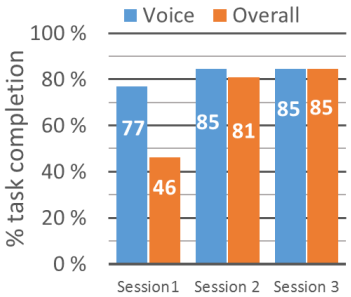

Figure 23 % of successfully completed authentication which did not need assistance at authentication requests

VI. Conclusions

This analysis has provided highly valuable information to improve the next version of the PIDaaS mobile application. The results clearly show that the learnability of the application needs to be improved by better guidance within the PMA to the user through both process, enrolment and authentication. Thus, better user interfaces and participant guidance within the application has been recommended. The improved guidance within the application will avoid user's assistance requests and reduce the user's errors. Hence, it will help to reduce the number of incorrect presentations and raise the rate of successful enrolments.

Another important factor is the enhancement of the VAD module. It has shown a satisfactory accuracy of 93%. However, the enhancement of its accuracy can lead to a higher user's satisfaction, and therefore, a better user experience.

Through this evaluation we have identified the most common user mistakes while presenting the biometric samples related to interface issues. Furthermore, mobile analytical tools have been proven as an adequate tool for data logging and time-task analysis within the HBSI framework. These tools will enable the analysis of HBSI metrics during the PIDaaS service providing large-scale pilots in unconstrained environments.

Last, the presence of the operator in the evaluation room might bias the results of this kind of evaluation. One of the main aims of this experiment was to analyse the first contact of the participants with the PMA. This contact will happen in unsupervised environments during the pilots. Even asking the participants not to ask the operator, participants may tend to do it before really trying to figure out the best way by themselves. As future work, a similar evaluation will be performed without the operator in the evaluation room in order to analyse this issue.

Acknowledgment

This work has been co-funded under the EU ICT Policy Support Programme CIP (Call CIP-ICT-PSP-2013-7, project reference 621021). The authors gratefully thank all the participants that took part of this experiment.

References

[1] Uni. of Kent, CSI-Piamonte, Ricoh, Bantec, Gjovik University College, UAB E-Bros, Eurecat, "Private Identity as a Service (PIDaaS)," *EU CIP*, 2014. [Online]. Available: www.pidaas.eu.

[2] M. Brockly, S. Elliott, R. Guest, and R. B. Gonzalo, "Human-Biometric Sensor Interaction," in *Encyclopedia of Biometrics*, S. Z. Li and A. K. Jain, Eds. Boston, MA: Springer US, 2014, pp. 1–9.

[3] "ISO/IEC 19795-1:2006. Information technology -- Biometric performance testing and reporting -- Part 1: Principles and framework.", ISO/IEC, Geneva, 2006 .

[4] E. Kukula and S. Elliott, "Implementation of hand geometry: an analysis of user perspectives and system performance," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 21, no. 3, pp. 3–9, Mar. 2006.

[5] E. Kukula and S. Elliott, "Implementing Ergonomic Principles in a Biometric System: A Look at the Human Biometric Sensor Interaction (HBSI)," in *Proceedings 40th Annual 2006 International Carnahan Conference on Security Technology*, 2006, pp. 86–91.

[6] E. P. Kukula and S. J. Elliott, "Evaluation of a facial recognition algorithm across three illumination conditions," *Aerosp. Electron. Syst. Mag. IEEE*, vol. 19, no. 9, pp. 19–23, 2004.

[7] M. Brockly, R. Guest, S. Elliott, and J. Scott, "Dynamic Signature Verification and the Human Biometric Sensor Interaction Model," in *IEEE Intern. Carnahan Conf. on Sec. Technology*, 2011, pp. 253–258.

[8] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and E. Bella-Pulgarin, "Automatic usability and stress analysis in mobile biometrics," *Image Vis. Comput.*, vol. 32, no. 12, pp. 1173–1180, Dec. 2014.

[9] S. Elliott, M. Mershon, V. Chandrasekaran, and S. Gupta, "The evolution of the HBSI model with the convergence of performance methodologies," in *2011 Carnahan Conference on Security Technology*, 2011, pp. 1–4.

[10] Yahoo, "Flurry Analitics." [Online]. Available: https://developer.yahoo.com/analytics/. [Accessed: 03-Sep-2015].

[11] E. Charton, A. Larcher, C. Levy, and J.-F. Bonastre, "Mistral: open source biometric platform," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, 2010, pp. 1503–1504.

[12] "Common environmental noise levels." [Online]. Available: http://chchearing.org/noise/common-environmental-noise-levels/. [Accessed: 15-Oct-2015].

[13] J. Nielsen, "An introduction to usability," 2012. [Online]. Available: https://www.nngroup.com/articles/usability-101-introduction-to-usability/. [Accessed: 15-Jun-2015].