วารสารวิทยาศาสตร์บูรพา ปีที่ 29 (ฉบับที่ 1) มกราคม – เมษายน พ.ศ. 2567
BURAPHA SCIENCE JOURNAL Volume 29 (No.1) January – April 2024 บทความวิจัย

การยืนยันตัวตนด้วยเสียงแบบอ่านบนโทรศัพท์สมาร์ทโฟน กรณีศึกษาผู้สูงอายุ

Reading-Based Voice Authentication on a Smartphone: A Case Study of Older Adults

เจษฎา บุญสิริ[1*], ทิพยา จินตโกวิท[2] และ นลินภัสร์ บำเพ็ญเพียร[1]

Jedsada Boonsiri[1*], Thippaya Chintakovid[2] and Nalinpat Bhumpenpein[1]

[1]ภาควิชาสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ประเทศไทย

[2] หน่วยปฏิบัติการวิจัยภูมิทัศน์สารสนเทศ ภาควิชาบรรณารักษศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ประเทศไทย

[1]Department of Information Technology, Faculty of Information Technology and Digital Innovation,
King Mongkut's University of Technology North Bangkok, Thailand

[2]Information Landscape Research Unit, Department of Library Science, Faculty of Arts, Chulalongkorn University, Thailand

Received : 26 April 2023, Received in revised form : 22 November 2023, Accepted : 28 November 2023

Available online : 9 January 2024

บทคัดย่อ

วัตถุประสงค์และที่มา : การยืนยันตัวตนเป็นส่วนประกอบสำคัญในการปกป้องสมาร์ทโฟนจากการเข้าถึงโดยไม่ได้รับอนุญาต อย่างไรก็ตาม วิธีการยืนยันตัวตนที่มีอยู่นั้นไม่เหมาะสมกับผู้ใช้สูงอายุ ซึ่งประสบปัญหาความเสื่อมถอยของร่างกาย ส่งผลให้ผู้สูงอายุจำนวนมากมีพฤติกรรมการใช้งานที่ไม่ปลอดภัย รวมถึงปิดระบบการยืนยันตัวตนบนอุปกรณ์ของตน หนึ่งในหนทางแก้ปัญหาดังกล่าวคือการยืนยันตัวตนด้วยเสียง ซึ่งได้รับคำแนะนำให้เป็นตัวเลือกที่เหมาะสมกับผู้สูงอายุมากกว่าวิธีการการยืนยันตัวตนอื่นๆ อย่างไรก็ตาม การยืนยันตัวตนด้วยเสียงมีความเสี่ยงต่อการโจมตีด้วยการเลียนเสียงหรือบันทึกเสียงของผู้ใช้ ในบรรดาแนวคิดเพื่อแก้ไขข้อด้อยดังกล่าว การยืนยันตัวตนด้วยเสียงโดยการอ่านนั้นมีความเรียบง่ายและเหมาะสมกับผู้สูงอายุในทางทฤษฎี กระนั้น ก็ยังมีข้อมูลไม่เพียงพอที่จะยืนยันความเป็นไปได้ รวมถึงความเห็นของผู้สูงอายุและการตอบสนองจากผู้ใช้งานที่เป็นคนไทยต่อระบบ ดังนั้น การวิจัยนี้มีวัตถุประสงค์เพื่อยืนยันความเป็นไปได้ดังกล่าวและทดสอบการยืนยันตัวตนด้วยเสียงโดยการอ่านนั้นเหมาะสมกับผู้สูงอายุจริงหรือไม่

วิธีดำเนินการวิจัย : การยืนยันตัวตนด้วยเสียงโดยการอ่านที่ต่างกันสามรูปแบบได้ถูกพัฒนาขึ้น และทำการทดสอบเพื่อประเมินการใช้งานกับผู้สูงอายุ โดยผู้เข้าร่วมจะถูกแบ่งออกเป็นสามกลุ่มสำหรับแต่ละรูปแบบการการยืนยันตัวตน จากนั้นผู้เข้าร่วมจะใช้ทดสอบระบบโดยการใช้งานฟังก์ชันลงทะเบียนและยืนยันตน ซึ่งผู้วิจัยจะสังเกตและจดบันทึกพฤติกรรมและความเห็นของผู้เข้าร่วม พร้อมด้วยข้อมูลตัวชี้วัดประสิทธิภาพของระบบ (อัตราความสำเร็จ เวลางาน ข้อผิดพลาด)

ผลการวิจัย : ผลการทดสอบเป็นที่น่าพอใจ ในแง่ของประสิทธิภาพ ทั้งการยืนยันตัวตนสามรูปแบบมีอัตรา Task Completion Rate ที่สูง ผู้เข้าร่วมเกือบทั้งหมดสามารถลงทะเบียนและยืนยันตัวตนได้สำเร็จ โดยแต่ละรูปแบบมีความแตกต่างอย่างมีนัยสำคัญเกี่ยวกับเวลาที่ต้องใช้ในกระบวนการการลงทะเบียน อย่างไรก็ตามกระบวนการยืนยันตัวตนนั้นใช้เวลาโดยเฉลี่ยเท่ากัน ข้อผิดพลาดส่วนใหญ่เกิดขึ้นระหว่างการบันทึกเสียง โดยที่ผู้เข้าร่วมลืมกดปุ่มหรือกดผิดจังหวะ สำหรับคะแนนจากแบบสอบถาม SEQ และ SUS นั้น บ่งชี้ว่า การยืนยันตัวตนด้วยเสียงโดยการอ่านทั้งสามรูปแบบนั้นใช้งานง่าย การทดสอบทางสถิติเพื่อเปรียบเทียบผลลัพธ์ของรูปแบบการยืนยันตัวตนทั้งสามรูปแบบ บ่งชี้ว่า การยืนยันตัวตนทั้งสามรูปแบบใช้งานได้ง่าย

ใกล้เคียงกัน และกลไกพิเศษ เช่น การแสดงข้อความรหัสผ่านแบบสุ่ม ไม่ส่งผลในทางลบต่อการใช้งาน อย่างไรก็ตาม ความสามารถในการเลือกข้อความรหัสผ่านได้อย่างอิสระส่งผลเสียต่อการใช้งานโดยผู้สูงอายุ เนื่องจากต้องมีก ารพิมพ์ข้อความ ผู้เข้าร่วมการทดสอบส่วนใหญ่แสดงความคิดเชิงบวกเกี่ยวกับการยืนยันตัวตนด้วยเสียงโดยการอ่าน เช่น ความง่ายในการใช้งานและความสะดวกสบาย

สรุปผลการวิจัย : แม้ว่าผลลัพธ์ที่ได้จะออกมาในเชิงบวก งานวิจัยนี้ยังคงมีข้อจำกัด โดยจำนวนผู้เข้าร่วมการทดลองน้อยกว่าที่ตั้งใจไว้ ซึ่งส่งผลต่ออำนาจทางสถิติของการทดลอง นอกจากนี้ งานวิจัยยังต้องการ การทดลองเพิ่มเติมเพื่อเปรียบเทียบการยืนยันตัวตนด้วยเสียงโดยการอ่านกับวิธีการยืนยันตัวตนอื่นๆ ประเมินความปลอดภัย และทดสอบระบบในเงื่อนไขอื่นๆ เช่น สถานที่กลางแจ้ง เป็นต้น

คำสำคัญ : การยืนยันตัวตน ; การยืนยันตัวตนด้วยเสียง ; โทรศัพท์สมาร์ทโฟน ; ผู้สูงอายุ

Abstract

Background and Objectives : Authentication is a vital component of smartphones to protect devices from unauthorized access. Nevertheless, existing schemes are unsuitable for elderly users due to their age-related difficulties. As a result, many older adults employ insecure practices, including disabling device authentication systems. One promising solution is voice authentication, which has been consistently suggested as a more usable option for older adults. However, voice authentication has a drawback regarding vulnerability to an imitation or recording of an enrolled speaker. Among many solutions for this issue, reading-based voice authentication is relatively simple and theoretically usable for older adults. Still, there is insufficient information to confirm the possibility, including older adults' perceptions and the reactions of Thai subjects toward the system. Therefore, this research intends to confirm that possibility and find whether reading-based voice authentication is usable enough for older adults.

Methodology : Three different styles of reading-based voice authentication were developed, and the testing was conducted to evaluate their usability relative to older adults. Participants were divided into three groups for each authentication style. Then, they would use the systems by enrolling and verifying themselves, where the researcher observed their actions and noted their opinions, along with the systems' performance metrics (Success Rate, Task Time, Error).

Main Results : The results of the test were encouraging. In terms of performance, all three styles achieved high task completion rates; almost all participants successfully enrolled in the system and verified themselves. There were significant differences regarding the time needed to complete enrollment. Nonetheless, the verification process used the same amount of time on average. Most errors occurred during the manual voice recording, where participants either forgot to press a button or pressed at the wrong moment. Both the scores from SEQ and SUS questionnaires indicated that all three styles of reading-based voice authentication were easy to use. The statistical tests to compare the results of all three authentication styles indicated that they were comparably usable, and mechanisms, like random passphrases, could be employed without adverse effects on usability. However, the ability to freely choose a passphrase negatively impacts usability as it requires text input and is thus unsuitable for older adults. Most participants expressed positive thoughts about reading-based voice authentication, like the ease of use and convenience.

Conclusions : Despite the promising results, this research still has limitations. The number of participants was smaller than intended, which affected the study's statistical power. Furthermore, this research needs more experiments to compare reading-based voice authentication with other authentication methods, assess its security aspect, and test the system in other settings, such as outdoor locations.

Keywords : authentication ; voice authentication ; smartphone ; older adult

***Corresponding author. E-mail** : jedsadaboonsiri@gmail.com

Introduction

A smartphone is a valuable device for older adults. Nevertheless, many problems still pose challenges to older adults; one such problem is authentication. Because of their age-related limitations, older adults often struggle to manage authentication and other security measures installed on smartphones. The textual password (or text password) is generally rated low in usability (Yıldırım & Mackie, 2019), especially among older adults, for being tiring and low memorability (Jaspreet Singh & Yvonne Hwei-Syn Kam, 2019). While praised for ease of use and memorability, graphical passwords are outperformed by PINs in input speed and error rates (Grindrod *et al.,* 2016). Also, they are significantly time-consuming compared to other methods (Jaspreet Singh & Yvonne Hwei-Syn Kam, 2019). The biometric-based systems, such as fingerprint and facial recognition, still have many shortcomings, such as imperfect accuracy due to noises in the environment (Haider & Sabahat, 2022) and the degradation or change of biometric traits in older adults (Galbally *et al.*, 2019). Also, biometric-based systems have usage restrictions to obtain biometric samples with acceptable qualities. Due to these shortcomings, a more usable method is necessary.

One promising method is voice authentication. It is a part of voice user interfaces (VUI), which has been consistently suggested as a more usable option for older adults (Schlögl *et al.*, 2013). The benefits of VUIs for older users include hand-free and eye-free interactions (Kowalski *et al.,* 2019) and ease of learning. The novelty of technology can also be overcome once they grow accustomed to it (Stigall *et al.,* 2019). These advantages are significant for the Thai elderly population, where more than half use smartphones (Aranyanak & Charoenporn, 2020). However, many find existing methods (PIN and text password) inconvenient, resulting in refusal to use authentication and subsequent risks. Hence, the more convenient voice authentication can encourage older Thai adults to secure their devices.

Some drawbacks do persist. Aside from perceived security and reliability by users (Renz *et al.,* 2022), the voice recognition system, including commercial services like Microsoft Azure Speaker Verification, can be fooled by an imitation or recording of an enrolled speaker. Therefore, many solutions were proposed to address this issue, such as leveraging a motion sensor (Anand *et al.*, 2021) and an anti-spoofing design of the voice authentication system (Zhao *et al.*, 2021). Nevertheless, there is one solution that does not rely solely on the voice recognition system's accuracy, does not involve special devices or complicated mechanisms, and is still theoretically usable for older users. That solution is reading-based voice authentication.

The key concept is that the system shall display a series of numbers, letters, or regular words as a passphrase for users to read aloud. If the recognized phrase is the same as the provided one and the voice matches the voice profile, then the user's identity claim is valid. For example, Yan and Zhao (Yan & Zhao, 2016) proposed a voice authentication system where users would be asked to read aloud a series of characters they saw on the screen. Bella *et al.* also proposed a similar system (Bella *et al.*, 2020) but using random digits in Indonesian instead of English.

The work of Rehman and Lee (Rehman & Lee, 2019) presented a text-dependent voice-based authentication protocol. However, English words were used instead of a series of characters. Five random English words were selected during enrollment, and the system would randomly display one for users to read aloud in the verification process. Akhtar's work (Akhtar, 2017) resembled the work of Rehman and Lee. The difference was that Akhtar's system was text-independent, giving the system more flexibility due to the broader range of possible passphrases. Nonetheless, the text-independent system was usually outperformed by the text-dependent variance under short-duration scenarios (Tu *et al.,* 2022), along with an increased difficulty in development.

Since they only need to read text displayed on the screen, the required effort from older adults is minimal, as recommended by the guidelines of Usable Security (Garfinkel & Lipford, 2014). Moreover, there is no need to remember secret information, which is helpful for older users whose mental capabilities have deteriorated (Fisk *et al.*, 2009a). Nevertheless, there are still unanswered questions. Are older adults really satisfied with a reading-based voice authentication system? And why? Since previous studies worked with non-Thai people, how do Thai subjects react to reading-based voice authentication? Can the system be incorporated with additional mechanisms, such as random passphrases, to improve security without sacrificing usability? or must the system remain as simple as possible? Moreover, if older adults can choose their passphrase, how does it affect the system's usability? Hence, this research aimed to answer those questions above and find whether reading-based voice authentication is usable enough for older adults.

Methods

An experiment was conducted to achieve the research's goal by testing older adults against three text-dependent reading-based voice authentication styles. The experimental authentication systems consisted of two parts. The front-end side was a mobile application developed with Xamarin, tools to write native Android and iOS applications. The back-end side for voice recognition tasks was built using a Python library named pyAudioAnalysis.

Participants were older adults, defined in this research as any person aged 60 years or more. They also must have prior experience in using smartphones. However, experiences in authentication and the use of VUI were optional. After being recruited, participants were divided into three groups to test three different styles of reading-based voice authentications. The researchers collaborated with a local clinic in Don Mueang District, Bangkok, Thailand, to recruit participants. As older adults might feel uncomfortable and reluctant to participate in an experiment conducted by an unfamiliar person, working with the clinic's staff ensured familiarity and trust, which helped older adults be more willing to participate in the study.

The first step of the experiment was a preliminary questionnaire for general information such as demographic data, smartphone usage, and authentication usage. After completing the questionnaire, participants would receive a tutorial document describing the tasks they needed to do. Participants had one round to practice with the tutorial tasks. If participants did not understand any part of the tutorial, they might request a walkthrough from a researcher. However, participants were excluded from further participation if they could not complete all tasks after the walkthrough.

The experiment consisted of two tasks in total. First, participants needed to enroll their voices into the system. Then, they would verify themselves three times consecutively. At the end of each task, participants would answer the SEQ (Sauro & Lewis, 2016c), which is employed as a post-task questionnaire. They would assess the overall ease of completing a task in the form of level of easiness. When participants completed all tasks given to them, they needed to complete the SUS (System Usability Scale) (Brooke, 1995) for the post-study questionnaire and give their opinions about voice authentication by answering two questions.

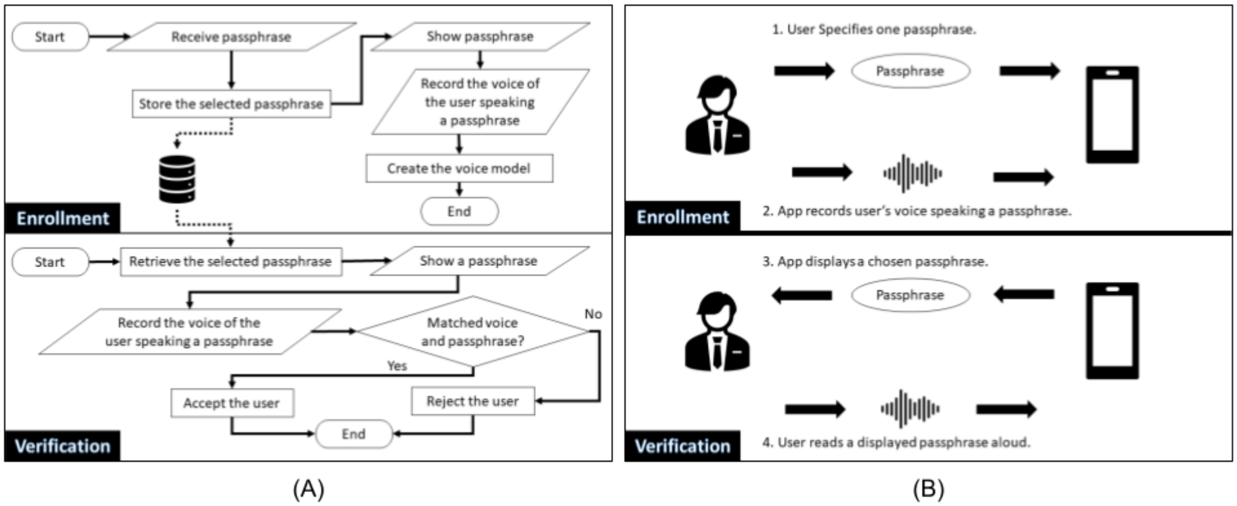

Start
Receive passphrase
Store the selected passphrase
Show passphrase
Record the voice of the user speaking a passphrase
Create the voice model
End
Enrollment
Start
Retrieve the selected passphrase
Show a passphrase
Record the voice of the user speaking a passphrase
Matched voice and passphrase?
No
Yes
Accept the user
Reject the user
End
Verification
1. User Specifies one passphrase.
Passphrase
Enrollment
2. App records user's voice speaking a passphrase.
3. App displays a chosen passphrase.
Passphrase
Verification
4. User reads a displayed passphrase aloud.
(A)
(B)

Figure 1 The first authentication style's process flow (A) and interaction flow (B)

The first authentication style was the base style for the other two styles. Figure 1 (A) shows its process, while Figure 1 (B) illustrates an overall interaction between a user and a system. For a start, a user must specify a Thai word as a passphrase used for verification on a page in Figure 2 (a). The system then stored that phrase and started a sub-process to record the user's voice and create a voice model. In the verification, the system would display a chosen passphrase and have users read it aloud, as shown in Figure 2 (c). If the user's voice and passphrase were matched with stored information, the system would accept the user's identity claim as valid. If the result was a rejection, the user had one more chance to try again. The second unsuccessful attempt, however, would terminate the process. In the real-world scenario, users had to switch to basic authentication methods like PIN or text passwords.

The second and third styles share a similar process with some key differences. In the second style, a user would choose three passphrases instead of one passphrase, as shown in Figure 2 (b), which would be displayed randomly during the verification. For the third style, the system also requested three passphrases from the user. The difference was that During verification, the system randomly displayed one of the participants' chosen passphrases and two decoy phrases, as shown in Figure 2 (d).

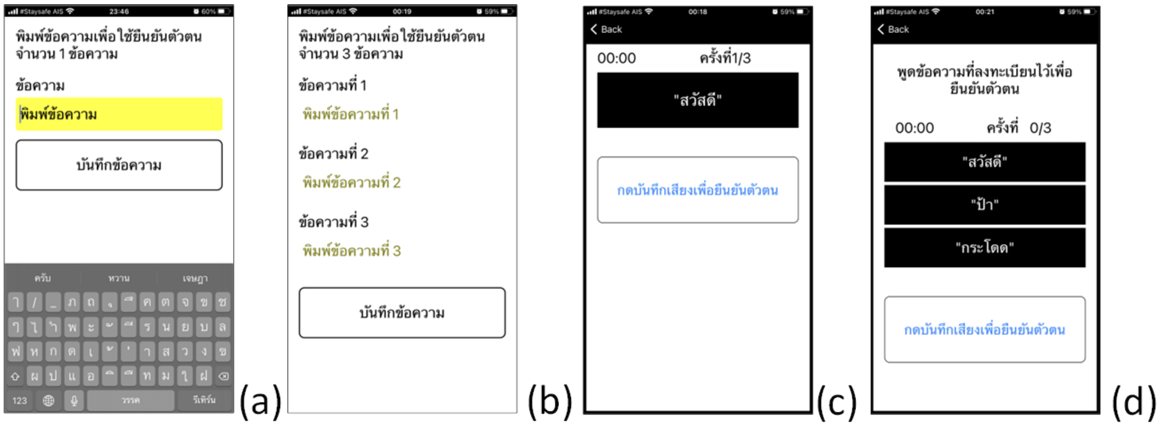


Figure 2 Application's graphical user interfaces

The reason for developing the second and third styles of voice authentication was to investigate if additional security measures (for instance, randomly displaying passphrases) could affect the usability of the system. The concept behind the second style was to create extra obstacles for attackers by using a set of passphrases rather than a single passphrase. For the third style, the correct passphrase was mixed with incorrect ones to confuse attackers. If participants read the wrong passphrase, the system would fail to verify user identity.

Results

The total number of participants was 32, comprised of 13 males and 19 females. 13, 10, and 9 participants experimented with the first authentication, second, and third styles, respectively. The youngest participant was 60, while the oldest participant was 79 years old. The average age of participants was 64.5 years old, and the mode was 63. Nineteen participants worked or used to work in the private sector, e.g., office workers, business owners, and independent contractors. Nine participants worked in government agencies or state enterprises. 2 participants were homemakers, and 1 participant declined to answer this question. For the educational level of participants, 14 participants did not graduate from high school. Nine participants have a high school diploma, and another nine hold a bachelor's degree.

Thirty participants used Android smartphones, and only 2 participants used iOS smartphones. Thirty-one participants had experience using smartphones for more than two years, and one participant declined to answer this question. Twenty-one participants used authentication systems regularly, whereas 11 did not enable them on their smartphones. The textual password was the most used method, followed by fingerprint recognition, face recognition, and pattern-locking. For prior experience using VUI, only 13 participants had used some form of VUI on smartphones.

Performance Metrics

As participants in this experiment had only one chance to complete given tasks, the completion rates of the three styles were calculated by dividing the number of participants completing tasks by the total number of participants in each group. The first task was considered completed if participants successfully enabled the verification function, and in the second task, participants must verify themselves three times to complete the task.

Because of the small sample size, the adjusted Wald (α=0.05) was used to calculate confidence intervals of completion rates in both tasks, shown in Table 1. Although there were differences in the low estimates of completion rates, the Fisher exact test (α=0.05) (Sauro & Lewis, 2016b, p. 5) revealed statistically non-significant differences in the completion rates between the three styles of voice authentication. Researchers, therefore, concluded that older adults' task completion rates were similar across the three styles.

Table 1 Confidence intervals of Rates of Completion in Enrollment and Verification

Task	Style	Low Estimate	High Estimate
Enrollment	First	79.74%	100%
	Second	57.40%	99.99%
	Third	73.70%	100%
Verification	First	79.74%	100%
	Second	75.12%	100%
	Third	54.33%	99.99%

Task completion time was recorded in seconds. The timer started when participants pressed the button to begin a task, and then it stopped when the result of enrollment or verification appeared on the screen. The geometric mean was used for the statistical analysis because of the small sample size (size <=25) (Sauro & Lewis, 2016a; Tullis & Albert, 2013). Table 2 shows geometric means of the length of time for enrollment and authentications of the three styles.

The data of task completion time did not meet two assumptions of the one-way ANOVA. The result of Bartlett's test (α = 0.05) indicated that the enrollment (p-value = 0.0006) and verification (p-value = 0.0005) completion time variances were unequal across the three sample groups. The Kolmogorov-Smirnov Test (α = 0.05) showed that each group's enrollment and verification data were not normally distributed. Therefore, the study employed the Kruskal–Wallis one-way analysis of variance (α = 0.05). The results indicated significant differences in the time participants needed for enrollment between the three interface groups (p-value = 0.0007). On the other hand, there was no significant difference in the time participants needed for verification (p-value = 0.0828).

In conclusion, the amount of time required by older adults to complete the enrollment process significantly differed between the three styles. Older adults who worked with the first style could complete the enrollment process faster than older adults who used the second and third styles. For the verification process, there was no significant difference between the time needed to complete the task in all three styles.

Table 2 Average Task Time in Seconds by Interface Styles

Style	Enrollment Time.	1st Verification Time	2nd Verification Time	3rd Verification Time
First	91.49	16.56	8.82	11.29
Second	160.74	22.55	14.66	13.71
Third	208.77	21.87	14.65	14.59

The errors in the experiment were recorded in simple binary data. For instance, if a participant pressed the wrong location on the screen, this error would be added to the list of errors, and the participant would be marked that he had committed an error. The errors observed during the experiment and the number of participants who committed errors are described in Table 3.

Table 3 The errors during enrollment and verification

Task	Errors committed by participants	Number of participants who committed errors		
		First	Second	Third
Enrollment	Error (1)	3	0	3
	Error (2)	3	2	2
	Error (3)	0	0	1
	Error (4)	1	0	1
Verification	Error (2)	1	5	2

The descriptions of errors are as follows.

(1) **Entered "NULL" as the passphrase.** During passphrase selection, participants must enter phrases into the textbox. Some participants, however, left the textbox blank.

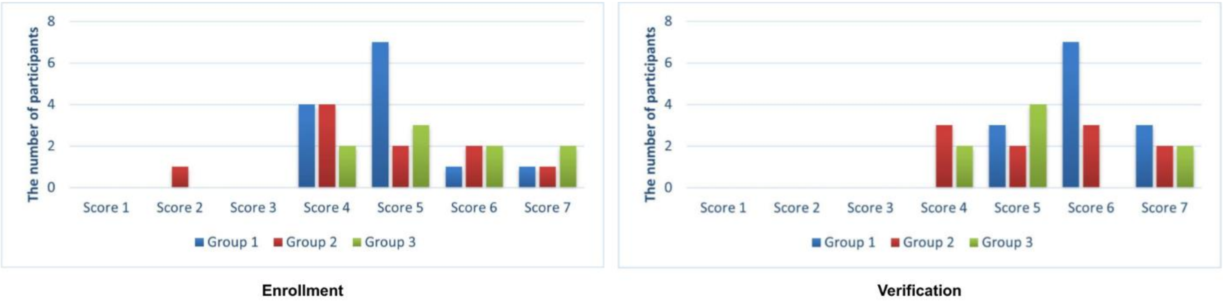
(2) **Errors in pressing voice recording buttons.** Many participants forgot to press buttons or pressed them at the wrong moments.

(3) **Did not complete the process.** Participants did not complete the process for other reasons, such as remaining idle and doing nothing, and required assistance from researchers.

(4) **Missed textboxes.** Participants pressed the wrong spot on the screen and could not select textboxes.

Questionnaires

The results of SEQ questionnaires are represented by the number of participants who chose each score, as illustrated in Figure 3. Score seven means the task is very easy, while score one signifies that the task is very difficult.



Enrollment **Verification**

Figure 3 Result of SEQ for the enrollment and verification tasks

Most participants rated the ease of completing both tasks at a score of four and higher, with only one participant from group 2 rating the ease of Enrollment task at a score of two. Kruskal–Wallis's test was used to determine if there were statistical differences in SEQ scores among the three interface styles. The results indicated a non-significant difference in SEQ scores for enrollment (p-value = 0.3952) and verification (p-value = 0.3278). Participants rated enrollment and verification tasks equally easy to perform with all three authentication styles.

The adjective ratings and SUS acceptability range from the work of Bangor, Kortum, and Miller (Bangor *et al.,* 2009) were used to analyze and interpret the SUS scores. The comparison between SUS scores, acceptability ranges, and adjective ratings is depicted in Figure 4.

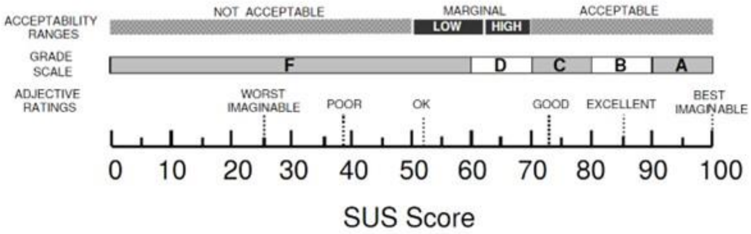


Figure 4 The comparison between SUS scores, adjective ratings, and acceptability ranges

The mean ranks and geometric means of SUS scores for each interface style, including the acceptability and adjective ratings, are described in Table 4.

Table 4 Mean Ranks of SUS Scores by Interface Style

Style	Mean Rank	Geometric Mean	Acceptability	Adj. Rating
First	19.25	64.31	High Marginal	Okay
Second	15.85	54.67	Low Marginal	Okay
Third	18.17	61.86	Low Marginal	Okay

Acceptability ranges and adjective ratings from the results depicted a more favorable impression. Only a few participants gave "failure" scores; the rest rated all three styles as passable and above. The Mean Rank and Geometric Mean show that the first style received the highest score, followed by the third and second styles. Because of the small sample size, the Kruskal–Wallis's Test ($\alpha = 0.05$) was used to determine if there were significant differences between the SUS scores of the three styles. The p-value was 0.5246, indicating a non-significant difference between the SUS scores of the three authentication styles.

Participants' Opinions and Observed Behavior

Two open-ended questions were given to participants to learn about their opinions towards voice authentication. The first question was, "How do you feel about this voice authentication system compared to other authentication schemes?". Nine participants stated that it was easier to use than other methods. Moreover, twelve participants considered this system easier than textual passwords. Nevertheless, seven participants felt it was equal to or worse than other methods.

The second question was, "Do you have any additional comments regarding voice authentication?". Generally, participants had neutral to positive opinions regarding voice authentication, feeling it was convenient and simple. While they acknowledged a considerable effort to learn how to use voice authentication, participants believed they could eventually understand that the method provided sufficient learning time. Even so, many participants expressed negative points about the voice authentication they tested in the experiment. They felt the system required too much effort to complete tasks. Notably, participants disliked that they had to type passphrases during enrollment. Some participants also voiced concerns over privacy and awkwardness if they used the system in public.

During the experiment, participants' behavior and system usage was observed. Overall, participants in each group displayed similar behaviors. Most of them required guidance from the researcher during the tutorial. Many also experienced common problems older adults encounter, such as difficulties in typing and forgetfulness. The minority, however, could learn how to perform the tasks with minimum assistance. They also showed proficiency in using smartphones, such as dexterous typing and recovering from errors.

Discussion

Overall, the results indicate that reading-based voice authentication is usable enough for older adults who view it moderately favorably. Furthermore, security improvement measures implemented in the second and third styles do not negatively affect the system's overall usability. This conclusion is supported by the task success rates and the results from SEQ and SUS scores, indicating that the differences in perceived ease of tasks and usability between the three styles were not statistically significant. Although two more complex styles have longer enrollment task times, speed is not a significant concern for older adults (Grindrod *et al.,* 2016).

Convenience and simplicity are likely primary reasons for positive views toward reading-based voice authentication mentioned in the introduction. Reading passphrases aloud is straightforward and does not involve typing or complex mental abilities like inference; hence, it only requires minimal concentration. Moreover, older adults do not need to memorize passphrases when using the first and second styles. Though the third style does require users to remember their selected passphrases, participants did not specify any significant problems with it.

Unfortunately, the ability to freely choose a passphrase during enrollment negatively impacts the system's usability because older users must provide passphrases via text entry. In addition to being error-prone (Komninos *et al.*, 2018), older adults usually type slower than younger adults (Nicol *et al.,* 2016), lengthening the time to complete the text entry. This difficulty is further emphasized by the lengthy enrollment task times and participants' dissatisfaction with the enrollment process. Therefore, it can be concluded that this approach is unsuitable for older adults.

Another issue is related to the graphic user interface or GUI. Even though the application for the experiment was designed following the guidelines (Fisk *et al.*, 2009b) and recommendations from past work (Balata *et al.*, 2015), many participants still struggled during the experiment. Typical problems of older users with GUIs remained visible, such as trouble reading a passphrase and becoming lost while performing the task. Moreover, the manual voice recording via pressing a button is problematic for many participants and causes errors. Nonetheless, redesigning GUIs and interaction flows may help mitigate these problems.

Based on participants' answers to post-experiment questions, reading-based voice authentication is viewed more favorably than text passwords. Their opinions are consistent with past studies where older adults rate the text passwords as low in usability (Yıldırım & Mackie, 2019) for being tiresome and difficult to memorize (Jaspreet Singh & Yvonne Hwei-Syn Kam, 2019). Nonetheless, text passwords still have advantages in terms of reliability since the results of verification are either true or false. By contrast, verification via voice is not always accurate, and misrecognition can happen.

Due to the lack of experience, participants did not provide opinions comparing reading-based voice authentication with biometric-based methods, such as fingerprint and face authentication. Nevertheless, reading-based voice authentication has some theoretical advantages. It is not affected by water, dirt, or fingerprint quality degradation (Galbally *et al.,* 2019), which impacts fingerprint recognition's accuracy. Additionally, its input method is less restrictive than facial recognition, which limits possible angles and distances between the user's face and the device. Still, older users may prefer fingerprint or face authentication when in public for a few reasons. Aside from the issue of social awkwardness, the performance of reading-based voice authentication will decrease in a noisy environment.

Under the assumption that the voice recognition system is accurate, the vulnerability of reading-based voice authentication primarily involves passphrases. The first style is the most vulnerable since the passphrase is fixed in every verification, making it susceptible to the Replay Attack. The second and third styles can resist such an attack to a degree since more than one passphrase will be displayed randomly during verification. Even if attackers have the records of the target's voice speaking all possible passphrases, they still need to play a correct record matching the displayed passphrase within the time limit.

Conclusions

This study developed a reading-based voice authentication system with different implementations of authentication styles and tested their usability with a sample group of older adults. While the results of the study were encouraging, the research has limitations. Due to the coronavirus pandemic and requirements for participation, the number of participants was smaller than intended, which impacted the study's statistical power. Furthermore, this research has not yet conducted experiments to compare reading-based voice authentication with other authentication methods, assess the method's security, and test the system in outdoor settings with various noise levels. Therefore, additional experiments to address these limitations are necessary, including a long-term study to assess the long-term usability and user acceptance of voice authentication.

Acknowledgments

Researchers are grateful for the clinic staff and the participants who kindly participated in the experiment.

References

Akhtar, M. (2017). *Text Independent Biometric Authentication System Based On Voice Recognition* [Master's Thesis, Masaryk University].

Anand, S. A., Liu, J., Wang, C., Shirvanian, M., Saxena, N., & Chen, Y. (2021). EchoVib: Exploring Voice Authentication via Unique Non-Linear Vibrations of Short Replayed Speech. *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*, 67–81.

Aranyanak, I., & Charoenporn, P. (2020). UX-Based Design of A Mobile Application for Thai Seniors. *Proceedings of the 6th International Conference on Frontiers of Educational Technologies*, 160–163.

Balata, J., Mikovec, Z., & Slavicek, T. (2015). KoalaPhone: Touchscreen mobile phone UI for active seniors. *Journal on Multimodal User Interfaces*, *9*(4), 263–273.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114–123.

Bella, Hendryli, J., & Herwindiati, D. E. (2020). Voice Authentication Model for One-time Password Using Deep Learning Models. *Proceedings of the 2020 2nd International Conference on Big Data Engineering and Technology*, 35–39.

Brooke, J. (1995). SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, *189.*

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009a). Chapter 2 Characteristics of Older Adult Users. In *Designing for Older Adults: Principles and Creative Human Factors Approaches, Second Edition (Human Factors & Aging)* (2nd ed., p. 13). CRC Press.

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009b). Chapter 5 Design of Output and Input Devices. In *Designing for Older Adults: Principles and Creative Human Factors Approaches, Second Edition (Human Factors & Aging)* (2nd ed., p. 61). CRC Press.

Galbally, J., Haraksim, R., & Beslay, L. (2019). A Study of Age and Ageing in Fingerprint Biometrics. *IEEE Transactions on Information Forensics and Security*, *14*(5), 1351–1365.

Garfinkel, S., & Lipford, H. R. (2014). Chapter 4 Lessons Learned. In *Usable Security: History, Themes, and Challenges* (pp. 87–91). Morgan & Claypool Publishers.

Grindrod, K., Hengartner, U., Khan, H., & Vogel, D. (2016, June 22). *Evaluating Smartphone Authentication Schemes with Older Adults*. the 12th Symposium on Usable Privacy and Security, Denver, CO.

Haider, A., & Sabahat, N. (2022). A Usability and Accuracy Measurement of Smartphones Face Recognition. *2022 2nd International Conference on Artificial Intelligence (ICAI)*, 19–25.

Jaspreet Singh & Yvonne Hwei-Syn Kam. (2019). Usable Authentication Methods for Seniors. *International Journal of Recent Technology and Engineering*, *8*(3S), 94–100.

Komninos, A., Dunlop, M., Katsaris, K., & Garofalakis, J. (2018). A glimpse of mobile text entry errors and corrective behaviour in the wild. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, 221–228.

Kowalski, J., Jaskulska, A., Skorupska, K., Abramczuk, K., Biele, C., Kopeć, W., & Marasek, K. (2019). Older Adults and Voice Interaction: A Pilot Study with Google Home. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6.

Nicol, E., Dunlop, M. D., & Komninos, A. (2016). A Participatory Design and Formal Study Investigation into Mobile Text Entry for Older Adults. *International Journal of Mobile Human Computer Interaction*, *8*(2), 20–46.

Rehman, U. U., & Lee, S. (2019). Natural Language Voice Based Authentication Mechanism for Smartphones (Poster). *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 600–601.

Renz, A., Baldauf, M., Maier, E., & Alt, F. (2022). Alexa, It's Me! An Online Survey on the User Experience of Smart Speaker Authentication. *Proceedings of Mensch Und Computer 2022*, 14–24.

Sauro, J., & Lewis, J. R. (2016a). Chapter 2—Quantifying user research. In J. Sauro & J. R. Lewis (Eds.), *Quantifying the User Experience (Second Edition)* (pp. 9–18). Morgan Kaufmann.

Sauro, J., & Lewis, J. R. (2016b). Chapter 5—Is there a statistical difference between designs? In J. Sauro & J. R. Lewis (Eds.), *Quantifying the User Experience (Second Edition)* (pp. 61–102). Morgan Kaufmann.

Sauro, J., & Lewis, J. R. (2016c). Chapter 8—Standardized usability questionnaires. In J. Sauro & J. R. Lewis (Eds.), *Quantifying the User Experience (Second Edition)* (pp. 185–248). Morgan Kaufmann.

Schlögl, S., Chollet, G., Garschall, M., Tscheligi, M., & Legouverneur, G. (2013). Exploring Voice User Interfaces for Seniors. *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, *52*:1-52:2.

Stigall, B., Waycott, J., Baker, S., & Caine, K. (2019). Older Adults' Perception and Use of Voice User Interfaces: A Preliminary Review of the Computing Literature. *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, 423–427.

Tu, Y., Lin, W., & Mak, M.-W. (2022). A Survey on Text-Dependent and Text-Independent Speaker Verification. *IEEE Access*, *10*, 99038–99049.

Tullis, T., & Albert, B. (2013). Chapter 4—Performance Metrics. In T. Tullis & B. Albert (Eds.), *Measuring the User Experience (Second Edition)* (pp. 63–97). Morgan Kaufmann.

Yan, Z., & Zhao, S. (2016). A Usable Authentication System Based on Personal Voice Challenge. *2016 International Conference on Advanced Cloud and Big Data (CBD)*, 194–199.

Yıldırım, M., & Mackie, I. (2019). Encouraging users to improve password security and memorability. *International Journal of Information Security*, *18*(6), 741–759.

Zhao, C., Li, Z., Ding, H., Xi, W., Wang, G., & Zhao, J. (2021). Anti-Spoofing Voice Commands: A Generic Wireless Assisted Design. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*(3), 139:1-139:22.