Alexa, It's Me! An Online Survey on the User Experience of Smart Speaker Authentication

Andreas Renz
Eastern Switzerland University of Applied Sciences
St.Gallen, Switzerland

Matthias Baldauf
Eastern Switzerland University of Applied Sciences
St.Gallen, Switzerland

Edith Maier
Eastern Switzerland University of Applied Sciences
St.Gallen, Switzerland

Florian Alt
University of the Bundeswehr Munich
Munich, Germany

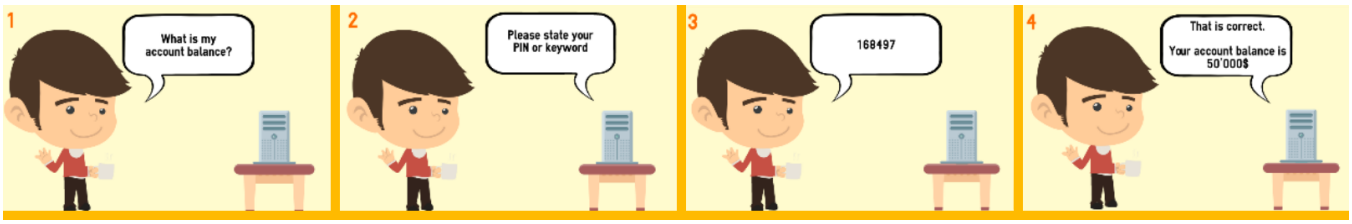

Figure 1: We compare users' experience when authenticating with voice-based services on smart speakers by means of different methods (spoken PIN, biometrics, app with button/voice confirmation, card reader). Each authentication method was explained through a short four-step cartoon within the questionnaire. The example above shows the illustration for the "spoken PIN" method.

ABSTRACT

Verifying the identify of the speaker is a crucial requirement for security-critical voice-based services on smart speakers, such as transferring money or making online purchases. Whilst various studies have explored novel authentication mechanisms for voice-based services, there is little research on the user experience of respective authentication methods. To address this gap, we conducted a comprehensive online survey (n=696). We compared five authentication methods (spoken PIN, biometrics, app with button/voice confirmation, card reader) regarding their perceived efficiency, security, ease of use, and error susceptibility. Additionally, we investigated users' willingness to use security-critical services in banking and government. We found an overall preference to confirm actions triggered by voice by pressing a button on a mobile authentication app followed by PIN-based authentication. In contrast, biometric authentication by voice is considered unreliable, while applying a card reader is regarded secure, yet less convenient.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; • **Security and privacy** → **Usability in security and privacy**.

KEYWORDS

voice assistant, authentication, voice banking

ACM Reference Format:
Andreas Renz, Matthias Baldauf, Edith Maier, and Florian Alt. 2022. Alexa, It's Me! An Online Survey on the User Experience of Smart Speaker Authentication. In *Mensch und Computer 2022 (MuC '22), September 4–7, 2022, Darmstadt, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3543758.3543765

1 INTRODUCTION

Voice assistants have become frequent companions in our homes in the form of so-called smart speakers. Such devices, for example, the *Google Nest Audio*[1], *Amazon Echo*[2], or *Apple HomePod*[3], promise convenient hands-free interaction using natural language with a myriad of apps. Since their first appearance, the market penetration of smart speakers has been steadily growing. By 2021 the number of global smart speaker shipments was projected to reach 186 million, with shipments expected to exceed 200 million annually in 2022 or 2023 [30].

While many popular voice-controlled applications involve non-critical tasks, such as playing music or searching the Web [2], an

MuC '22, September 4–7, 2022, Darmstadt, Germany

ACM ISBN 978-1-4503-9690-5/22/09…$15.00
https://doi.org/10.1145/3543758.3543765

[1] Google Nest: https://store.google.com/product/nest_audio
[2] Amazon Echo: https://www.amazon.com/All-New-Echo-4th-Gen/dp/B07XKF5RM3
[3] Apple Homepod: https://www.apple.com/homepod/

increasing number of services are emerging, which handle confidential information or initiate transactions which might have serious consequences. Examples include triggering online purchases, smart home services for controlling appliances for heating and lighting by voice, or voice-based banking services for checking one's account balance or even initiating a money transfer [1]. For such advanced personalized voice services, user authentication, i.e., proving that the speaker is genuinely the person he/she claims to be, is crucial.

Researchers have developed different approaches to smart speaker authentication [3, 7, 12, 19, 31, 35, 37]. From a users' perspective, prior work looked at user concerns regarding attacks and threats as well as at mitigation strategies [27]. However, knowledge on the users' experience with authentication methods for smart speakers is scarce. Whereas prior research looked at users' privacy and security needs emerging from (negative) user experience when using smart devices, their experience while using different authentication mechanisms has – to our knowledge – not been previously investigated. This is the focus of our work.

Given the smart speakers' promise of convenient, easy-to-use services, an in-depth investigation into the users' perspective is a prerequisite for providing both secure voice-controlled services that are also acceptable to users. First, we provide an overview of authentication methods in general and methods for voice-based services, in particular. We compare these different authentication mechanisms with regard to how users perceive their efficiency, error susceptibility, security and ease of use. Second, on the basis of a literature review, we formulate a series of research questions, which we approach through an online survey (n=696). More specifically, we compare knowledge-based (PIN), biometric (voice), and token-based approaches (authenticator app, card reader) in two contexts (online banking, government services). This comprehensive online survey contributes to the current body of scientific knowledge on users' attitudes towards authentication methods for smart speakers. The results provide insight into the users' subjective perception of crucial factors influencing acceptance and overall preferences.

Our work is meant to inform the design of advanced voice-based services, which are both secure and acceptable for users. In addition, we expect them to provide a solid basis for complementary empirical user studies.

2 RELATED WORK

Our research is based on three strands of previous work: 1) general authentication methods for voice-based services; 2) state-of-the-art in securing voice-based services, and 3) user-related threats concerning voice-based services.

2.1 Authentication Methods for Voice-based Services

Traditional knowledge-based methods include using a code or answering previously defined security question to prove his/her identity [9]. A typical example is a PIN (personal identification number), a four or more digit code that is defined by the user and needs to be uttered at the time of authentication. *Google* and *Amazon* both support this type of authentication for their smart speakers and allow third-party developers to secure their voice services through a user-defined PIN [14]. Although such an approach might be convenient for users, this method has its drawbacks when it comes to smart speakers because other people may be present and overhear the secret code.

Another type of authentication approaches applicable to smart speakers is biometrics. One approach leverages the unique characteristics of the speaker's voice for verifying his or her identity. In its basic form this comprises the generation of a voice print and the comparison of the speaker's voice samples with this registered voice print [11]. Access is granted or a critical request is fulfilled only in case of a match. Another approach of leveraging biometric data is to use the smart speaker in combination with the speaker's smartphone. Its sensors, such as the fingerprint scanner or the camera for facial recognition, can be used to authenticate the user of the smart speaker [33]. Furthermore, it is also possible to combine different biometric methods, such as fingerprint authentication and facial recognition, to perform two separate identification checks and thus enable multi-modal biometric authentication [24].

In addition, more experimental approaches have been proposed. *Continuous authentication* [12] and *voice resonance* [23] follow a similar approach by leveraging body vibrations tracked through wearables (e.g., glasses, chest straps, watches) and transmitted in real time to the smart speaker when a request is made. By checking whether speech samples and the vibrations match, a smart speaker can verify whether a command was issued by the authorized person [12]. While continuous authentication concepts are based on wearables such as watches, glasses, earphones, or necklaces [12], vocal resonance depends on a microphone that can be worn on the head, neck, or chest [23]. A further experimental alternative for user verification is *Speaker-Sonar* [21] which makes use of inaudible sounds to track the user's direction and compares it with the direction of the received voice command.

2.2 State-of-the-Art in Securing Voice-based Services

As a pioneer in the field of conversational commerce [34], *Amazon* introduced a so-called "voice code" for securing voice-triggered purchases via their smart speakers. This voice code is a four-digit code set by the user via the corresponding *Alexa app*. If turned on, the user needs to tell the PIN to confirm purchases through the smart speaker. Furthermore, the PIN can be used to secure and personalize *Alexa skills* (third-party extensions). PIN authentication is a very common authentication method in banking contexts. Many banks that offer voice assistant applications use a PIN or pass code as authentication method. Examples include banks from the United States (U.S. Bank [32], Capital One [10], and Ally Bank [1]) as well as from Germany (Sparkasse [29]).

The methods used generally work in a similar way. To set up the voice app on the smart speaker, users have to initially log in to their banking account on the app or on a computer to set up a 4-6 digit PIN code. After that, the PIN code is active and can be used on the smart speaker [29]. The functions are very similar, including checking account balance, checking recent transaction details, billing due dates, and even transferring money [1].

One approach that addresses the risks related to using a PIN or passcode is two-factor authentication, as, for example, offered by

FUTURAE. FUTURAE developed their own zero-touch, two-factor smart assistant authentication method. It works with Google Assistant and Amazon Alexa and requires a smartphone previously registered to be in the proximity of the smart speaker. When a user prompts the voice assistant to perform a task that requires authentication, a short sound or melody is sent to the smartphone and played automatically. If the smart speaker detects the sound from the smartphone and recognizes it as the correct one, the requested command is executed [13]. Contrary to Voice Match on the Google Assistant, this method uses a second layer to ensure that the user who requests an action is the person who has the right to do so.

2.3 User-related Threats Concerning Voice-based Services

Prior research identified several user-related threats, emphasizing the need for suitable authentication methods.

Lei et al. [22] pointed out general security vulnerabilities due to the fundamental nature of access control, as implemented by current smart speakers. Using *Alexa* as a case study, the researchers criticize that the speaker's identity as well as his/her physical presence are not verified. They describe serious remote attacks (for example, through Bluetooth loud speakers), such as fake online purchases and home burglary, for example, by exploiting smart doors connected to the smart speaker. Lei et al. suggest using a WiFi-based approach to detect the speaker's physical presence.

Specific attacks involve sounds, inaudible to humans, yet recognizable by voice assistants, that trigger respective actions [28, 36]. An example is *DolphinAttack* by Zhang et al. [36]. They demonstrate that voice assistants on today's smart speakers such as *Siri* and *Alexa* react to voice commands on ultrasonic carriers and present a set of both hardware and software-based defense strategies to make these systems resilient against inaudible voice command attacks.

Skill (or voice) quatting refers to exploiting voice commands that either sound similarly or are mispronounced frequently and thus can invoke malicious third-party services unintentionally [18, 38]. Zhang et al. [38] present the example "open capital won" (vs. the original command "open capital one") which might be used to trigger a malicious skill imitating the original one, yet gathering sensitive user information or eavesdropping future conversations. Such attacks can target specific demographic groups [18]. Countermeasures include word-based and phoneme-based analysis during the publisher's certification process [18] and context-sensitive detectors assessing the impersonation risk [38].

3 RESEARCH QUESTIONS

Our review of prior research demonstrates that while many different approaches exist to authentication with smart speakers, an in-depth understanding of users' perception of and experience while interacting with such authentication methods is missing. In detail, we address the following research questions to close this gap:

RQ1: Which security-critical voice-based services in the domains of banking and government are users willing to use?

As foundation, we identify the most favored security-relevant voice-based services in two increasingly relevant domains beyond conversational commerce and smart home scenarios. While comprehensive voice banking services from more and more banks are publicly available on mass-market devices, initial voice-based government services are restricted to information queries (cf. [16]. Yet, advanced applications are increasingly explored in academia and are expected to be publicly available in the near future (cf. [5, 6, 20]).

RQ2: How do users perceive different authentication methods for voice-based services regarding error susceptibility and efficiency?

To maintain convenient and spontaneous interactions also for advanced security-critical services, we investigate how users perceive the efficiency of different authentication methods for smart speakers. We contrast this assessment with another relevant core attribute of Nielsen's system acceptability model [26], the perceived error susceptibility.

RQ3: Which authentication method for voice-based services do users prefer with regard to perceived security and ease of use?

For the adoption and acceptance of security-critical voice-based services on smart speakers the user-perceived security of the authentication methods plays a crucial role. To consider the methods' practical acceptability, we contrast this security assessment with the perceived ease of use of the methods.

RQ4: Which authentication method for voice-based services is preferred by users in general?

Summarizing the impressions of the individual factors, we explore overall preferences for the authentication methods. In addition, we are interested in whether the type of service to be secured (requesting critical information vs. initiating a critical transaction) has an impact on these preferences.

4 METHOD

To answer the above-mentioned research questions, we conducted an online survey. In this section, we present the questionnaire and describe how we collected and analyzed the data.

4.1 Online Survey

We created a questionnaire consisting of 32 questions (multiple choice, single choice, selection and open) using *FindMind*. Details on the content of the questionnaire can be found in Section 4.3. We conducted several pilot tests and iteratively refined the questionnaire. It was publicly available in English and German for five weeks in March and April 2021. Data collection was in line with general national data protection regulations and participants' consent was obtained.

At the beginning of the questionnaire, participants were informed about the objectives of our study and the anonymous collection and processing of the data. They were also made aware of their right to cancel the questionnaire at any time.

We distributed the invitation via social media channels such as *Facebook* and *Reddit* as well as university mailing lists. We also asked participants to forward the survey to their personal contacts. As an incentive, participants had the option to take part in a raffle to win vouchers for an online shop, provided they were prepared to submit their e-mail address.

4.2 Authentication Methods

As outlined in Section 2, various authentication methods for smart speakers have been proposed and investigated in previous work. In our comparative study, we chose authentication methods that were already available in mass-market smart speakers, or their integration into smart speakers could be expected in the near future.

4.2.1 User-defined PIN. In this variant, the user needs to say a predefined six-digit code to confirm a critical action triggered by a voice command. The code is set by the user herself via a corresponding app or Website. Only if the spoken code matches the stored one, the respective action is performed. *Amazon's Voice Code* is a real-life example for this method securing voice-triggered online purchases.

4.2.2 Biometric authentication. This authentication type exploits unique voice characteristics to verify the speaker's identity. First appearances of such biometric authenticaton approaches include Google Assistant's *Voice Match*[4], which is able to differentiate between up to six people's voices for personalizing voice services, and a pilot of a voice-confirmation feature for in-app purchases via Google Play[5]. In the variant of our study, the smart speaker asks the user to repeat a random word to prevent replay attacks through recorded voice samples.

4.2.3 Authenticator app with button confirmation. This method involves a dedicated authenticator app on a smartphone. When a critical action is requested via a smart speaker, a push notification activates the authenticator app. The user can accept or decline the authentication request and critical action, respectively, by pressing a button. This method is a common two-factor authentication approach in online banking, yet has not been implemented for smart speakers.

4.2.4 Authenticator app with voice confirmation. These apps provide time-restricted one-time passwords (OTP) for services registered within the app. Popular examples include Microsoft Authenticator [25] and the Google Authenticator [15]. This method might also be applied to smart speakers: For confirming a critical action requested by voice, the user needs to create an OTP within an authenticator app and must speak it out loud in front of the smart speaker.

4.2.5 Card reader. Another common authentication method for online banking is the use of a card reader[6]. Having inserted her bank card and unlocked it by entering the bank PIN, a user may generate OTPs for online authentication purposes through the device. Given its popularity and relevance in the online banking domain, we envisioned a related method for smart speakers. In analogy to the Web-based version, a six-digit code provided by the smart speaker must be entered in the card reader, and the generated OTP spoken out loud in front of the smart speaker.

[4]Voice Match: https://support.google.com/assistant/answer/9071681
[5]Voice confirmation: https://www.androidpolice.com/2020/05/25/google-assistant-gets-new-confirm-with-voice-match-setting-for-payments
[6]Card reader example: https://www.ubs.com/content/dam/ubs/ch/online_services/documents/anleitung-kartenleser-en.pdf

4.3 Questionnaire

The 32 questions were grouped into four sections. The first section contained demographic questions (country of residence, year of birth, sex, highest educational degree, employment status) and a question on the participants' overall technological savviness ("I like testing the functions of new technical systems."; 6-point Likert scale; 1=strongly agree to 6=strongly disagree). These introductory questions were followed by a definition of smart speakers (including photos of the popular examples Google Home, Amazon Echo, and Apple HomePod) and the description of typical applications such as checking the weather forecast or playing music. Furthermore, the section comprised several questions on participants' experience with smart speakers: whether they currently owned or had ever owned a smart speaker, how often they used which type of smart speaker, and whether they had (privacy) concerns when using a smart speaker.

The second section introduced more advanced and prospective security-critical applications for banking and e-government. We asked participants whether they could imagine using any of the listed voice-based applications, four of which belonged to the banking domain (such as checking the account balance) and five to e-government (such as requesting personal tax information) (6-point Likert scales; 1=strongly agree to 6=strongly disagree).

The main section of the questionnaire presented the five authentication methods (described in section 4.2 and collected the participants' opinions. We used cartoons to illustrate the different authentication methods to ensure a common understanding of the methods and their respective implementation (see Figure 1 for the "spoken PIN" method).

For each method, we asked participants to rate their agreement with the statements "I consider this method prone to errors" and "I consider this method efficient to use" (both on 6-point Likert scales; 1=strongly agree to 6=strongly disagree). Reasons for the rating and additional remarks could be provided in free-text fields. To avoid order effects, the five methods were presented in random order.

Finally, after the participants had become familiar with all five authentication methods, we asked them to rank the methods regarding the personal preference for checking their bank account and for confirming a money transfer (first rank – most likely; fifth rank – least likely).

Furthermore, we wanted to learn about reasons for preferring one authentication method over the other and asked the participants to rank the six factors efficiency, pleasing, few errors, security, time effort, and privacy according to their perceived importance (first rank – most important; sixth rank – least important).

4.4 Data Cleansing and Analysis

After the survey had been closed, the collected data was cleaned. We considered data sets to be invalid and removed them, if the duration spent to complete the questionnaire was below four minutes (i.e., significantly below average times in pre-tests) or the participant's answers showed certain patterns, e.g., the same answer for each question, in particular. Incomplete data sets (which met aforementioned criteria) were reviewed and kept if they contained valid and meaningful qualitative responses. However, a few qualitative

entries were labeled as invalid since they did not answer the corresponding question and therefore were not taken into account in the analysis. Out of a total of 1976 qualitative remarks, 1779 were classified as valid and considered in the further analysis.

Data was analyzed using SPSS. For comparing the methods, we ran a general linear model repeated measures analysis of variance to find main effects and to derive pairwise differences (based on Bonferroni-adjusted p-values). In case of a rejected sphericity assumption, the degrees of freedom were corrected by means of a Greenhouse & Geisser estimate. We assumed continuous concepts for our Likert scales and we treated them them as interval scales (cf. [17]).

A thematic qualitative analysis [8] was conducted to analyze the responses of the participants and to find common themes and patterns. Having familiarized themselves with the data by reading and rereading the questionnaires, two researchers coded the responses for each question using a collaboratively developed codebook. Following an inductive approach, themes were derived from the codes. Constant comparative analysis was performed to iterate the variation between theme occurrences across different participants. We selected verbatim quotations (translated to English by the researchers in case of non-English originals) to illustrate themes relevant for answering the research questions.

4.5 Participants

Overall, 751 participants took part in our survey. In the data cleansing step, we excluded 47 incomplete and eight incorrectly filled-in questionnaires. Our final data set consisted of complete and valid questionnaires from 696 participants (393 female, 303 male). The age of the participants ranged from 16 to 75 years (M=31.3; SD=10.8). Our survey reached broad participant groups worldwide: the major regional groups had their residence in the UK (25.6% of the participants), in the USA (24.0%), and in Germany (20.1%). The remainder was distributed across further 39 countries.

41% of the participants owned (at least) one smart speaker. 3.7% stated to have owned one in the past, but not anymore. In general, the participants considered themselves tech-savvy: 87% of the participants agreed with the statement of openness regarding novel technologies ("I like testing the functions of new technical systems.") with a mean of 4.66 (SD=1.13; 6-point Likert scale; 1=strongly disagree to 6=strongly agree). 74.6% (slightly or strongly) agreed to the statement "I have privacy concerns regarding smart speakers", 26.6% even strongly. The mean on the six-point Likert scale (from 1-strongly disagree to 6-strongly agree) was 4.3 (SD=1.47). Only 4.6% of the participants strongly disagreed.

The voice assistants most often used by the participants turned out to be *Amazon Alexa* (used "frequently" by 35.6% and "sometimes" by 18.7%), *Google Assistant* (22.8% and 17.0%), and *Apple's Siri* (14.3% and 18%). *Microsoft Cortana* (1.5% and 6.6%) and *Samsung Bixby* (3.3% and 4.0%) were used significantly less.

4.6 Limitations

To keep the questionnaire in a manageable size, we had to limit the number of authentication methods investigated. We included available methods for smart speakers (e.g., PIN), available methods not yet applied to smart speakers (e.g., authenticator app), or emerging methods which can be expected soon in mass-market devices (e.g., biometric). Further methods and variants of the considered methods remain subject to future work.

This research was deliberately conducted in the form of an online survey. We aimed to conduct a large-scale study, considering a broad view of authentication for smart speaker services. During the survey, participants were instructed about voice assistants and different authentication methods. When submitting their assessments, they had not experienced the methods one after the other like in a comparative user study. Still, many participants may have had prior experience with some of the available methods (e.g., PIN and authenticator app with button confirmation).

We managed to recruit a large number of participants through social media channels. Thus, a large portion of participants might be considered tech-savvy (which was also indicated by their self-assessment). Not all participants had first-hand experience with voice-based services on smart speakers, but were made familiar with those by the cartoons in the questionnaire. Some user groups that particularly benefit from voice-controlled services, e.g., the elderly and people with impairments (cf. [4]), are probably underrepresented in our sample. Obviously, their requirements need to be taken into account when designing a universally accessible and secure voice service.

5 RESULTS

In the following section, we present the results of our online survey in detail.

5.1 Services

For the banking domain, the use case of checking account balance received the most positive responses overall (Figure 2). 42% of the participants could imagine using a respective service (7% agreed strongly, 21% agreed, 14% agreed slightly). At the same time, 58% were negative with 28% of the participants strongly disagreeing. The use case of checking the details of recent account transactions received the second-most positive responses: 41% of participants were positive about this service (7% agreed strongly, 21% agreed, 14% agreed slightly), but 28% strongly rejected such a service.

Authorizing payments was perceived positively by 27% of the participants (4% agreed strongly, 11% agreed, 12% agreed slightly). The use case with the least positive ratings (22% of the participants) was money transfer (3% agreed strongly, 9% agreed, 10% agreed slightly). In their responses, many participants explained their concerns regarding error-proneness: *"Would have no issue checking my balance but I am afraid of the possibility of my payment or transfer decision not being recorded properly"* (P33).

Overall, none of the four use cases received more positive than negative responses. However, the services to request information (checking the balance of the account and the details of recent transactions) received significantly more positive responses than the security-critical services to trigger transactions (transferring money, authorizing payments).

Of the use cases considered for the e-government domain (Figure 3), the reporting of neighborhood defects through a voice-based service was positively rated by 74% of the participants (18% agreed

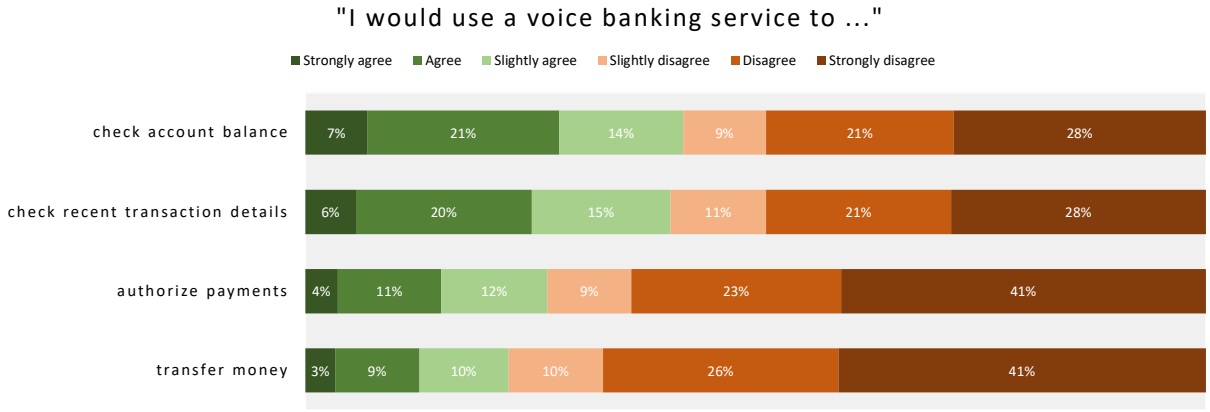

Figure 2: The participants' willingness to use voice-based services in the banking domain.

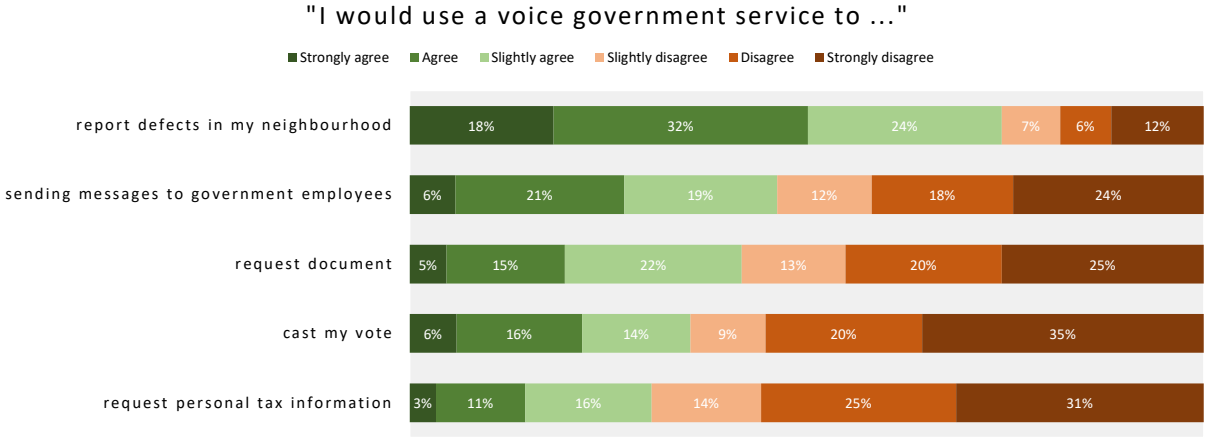

Figure 3: The participants' willingness to use voice-based services in the government domain.

strongly, 32% agreed, 24% agreed slightly). Voice messages to government employees, e.g., to request information, received the second most positive ratings (46% of the participants) received (6% agreed strongly, 21% agreed, 19% agreed slightly). Requesting personal documents via voice was perceived positively by 32% (5% agreed strongly, 15% agreed, 22% agreed slightly).

The use cases of voting and requesting personal tax information received the least positive ratings. 36% of the participants could imagine casting a vote via voice (6% agreed strongly, 16% agreed, 14% agreed slightly). The main reason for rejecting voice voting was mistrust in technology and the severe consequences in the event of technical failure: *"Voting is a very important right to have and one that must be made sure to be untainted by manipulation. If the AI incorrectly recognizes my vote and others, it could cause a lot of problems."* (P536).

30% responded positively regarding requesting personal tax information (3% agreed strongly, 11% agreed, 16% agreed slightly). Only reports of defects in the neighborhood received more positive than negative ratings.

Again, several participants made a distinction between services involving more and less critical data and tended toward using voice services for the latter group. For example, *"I would be willing to use a smart speaker for less confidential transactions such as reporting or sending messages to government officials, but not for more confidential transactions like voting or requesting tax information. I am concerned that information could be stolen from the smart speaker and would not want it to record any confidential information."* (P319)

5.2 Subjective Error Susceptibility and Efficiency

In terms of error susceptibility (Figure 4, left), the button and card reader methods were rated best (i.e., least error-prone) with ratings of 4.0 and 3.8, respectively. The third place rating was voice confirmation with a rating of 3.4. The PIN and biometric method were considered most error-prone, both with ratings of 2.9. Our analysis showed that the method had a significant main effect on the error susceptibility ratings, $F(3.37,2277.74)=102.09$, $p<.001$. Furthermore, all pairwise comparisons, except the last two mentioned, were statistically significant ($p<.006$).

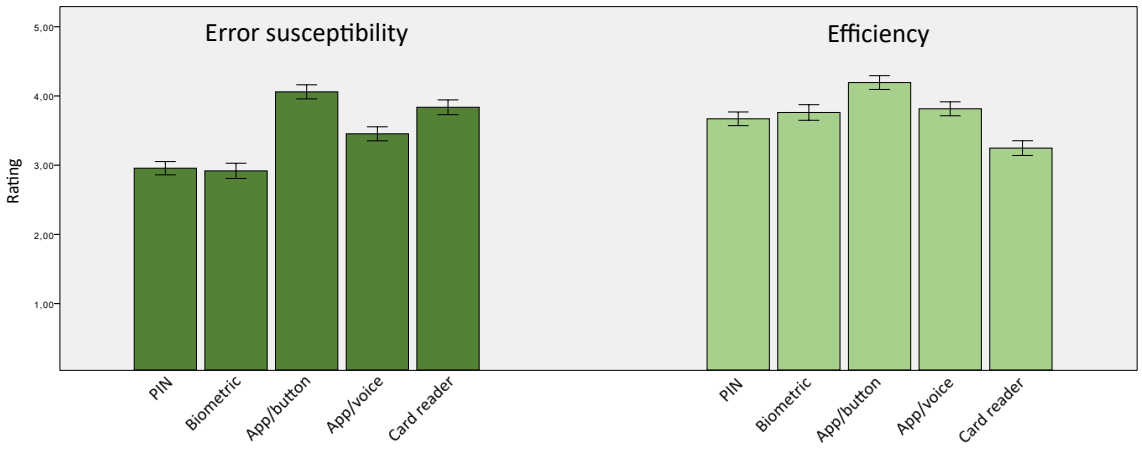

Figure 4: Participant rating of the subjective error susceptibility (left) and the efficiency (right) of the compared authentication methods.

In qualitative responses, a lot of participants had major doubts regarding the reliability of the biometric authentication and often referred to their own experience, e.g. *"I think this method can be prone to errors as a person's voice can at times not be recognised on devices."* (P538). Participants raised questions such as what would happen if a person's voice changes over time naturally or just by having a cold or sore throat. Poor ratings for the PIN method were often due to the fact that the static PIN needs to be spoken out loud, and therefore other methods with non-verbal confirmation were preferred, e.g. *"I like this method [button confirmation] because you still need to verify with your phone, but you do not have to say anything out loud."* (P347).

Regarding efficiency (Figure 4, right), the button method was found to be the authentication scheme with the highest rating with a mean rating of 4.2. Voice, biometric and PIN methods were similarly rated (without significant differences) with efficiency ratings of 3.8, 3.7, and 3.6, respectively. With a mean efficiency rating of 3.2, the card reader method was the lowest rated method. Again, we found that the choice of method had a significant effect on the efficiency ratings, $F(3.34,2258.10)=49.40$, $p<.001$. Pairwise comparisons proved that the button method was significantly better rated in terms of efficiency than the four alternatives ($p<.001$). In contrast, the ratings for the card reader method were significantly lower than those of the other methods. ($p<.001$).

Many participants justified high ratings for the button method by referring to its simplicity. It turned out that having a smartphone involved in the authentication process was not perceived as an obstacle by most participants. E.g., *"This [button confirmation] is a very efficient method and I would be more likely to use this method than saying a PIN."* (P207). However, in 30 comments, the participants considered the requirement of a smartphone inefficient, e.g., *"When this method [voice confirmation] requires the user to get their phone, they may as well be using a mobile app. This method is inefficient."* (P365).

On the other hand, many participants considered this a clear disadvantage of the card reader method, e.g. *"I used this [card reader] personally with a bank I use, but it is slightly inefficient as you always need the keypad or the card reader nearby."* (P633). Furthermore, the reasons for the card reader ratings contained descriptions such as *"slow, outdated, inconvenient, too much effort and not worth the time it takes"*.

5.3 Perceived Security and Ease of Use

Figure 5 shows the participants' ratings of perceived security (left) and ease of use (right) of the five different authentication methods. Our participants ascribed the highest security to the card reader method with a mean rating of 3.4, closely followed by the button method with a mean rating of 3.3. The voice method was rated as 3.1 on average, the PIN method with 2.8. With a score of 2.5, the biometric method received the lowest ratings in terms of perceived security.

The statistical analysis showed that the method had a significant effect on the perceived security, $F(3.70,2495.97)=34.12$, $p<.001$. Pairwise comparisons showed statistically significant differences between the card reader method and the voice, PIN, and biometric method ($p<.002$). The biometric method with the lowest rating was significantly worse in terms of perceived security than the four alternatives ($p<.001$). Further significant differences were found for button and PIN ($p<.001$), voice and both PIN and biometric ($p<.019$), as well as PIN and all four alternatives ($p<.019$).

For the card reader, many participants explained their high ratings regarding security with the involvement of a bank as trustworthy institution. For example, *"This is much more secure than using a mobile phone since the card reader can be sent to you by the bank"* (P636). Furthermore, the requirement to own a physical card was frequently mentioned: *"It seems very secure and reliable since it requires physical possession of a bank card, which is difficult to steal"* (P319).

On the contrary, many participants described their impression that biometric authentication is not secure, often justified by the threat of manipulation. An example is the statement by P319: *"I believe a hacker could imitate my voice if he/she stole enough voice recordings from the smart speaker. It is probably possible to make a*

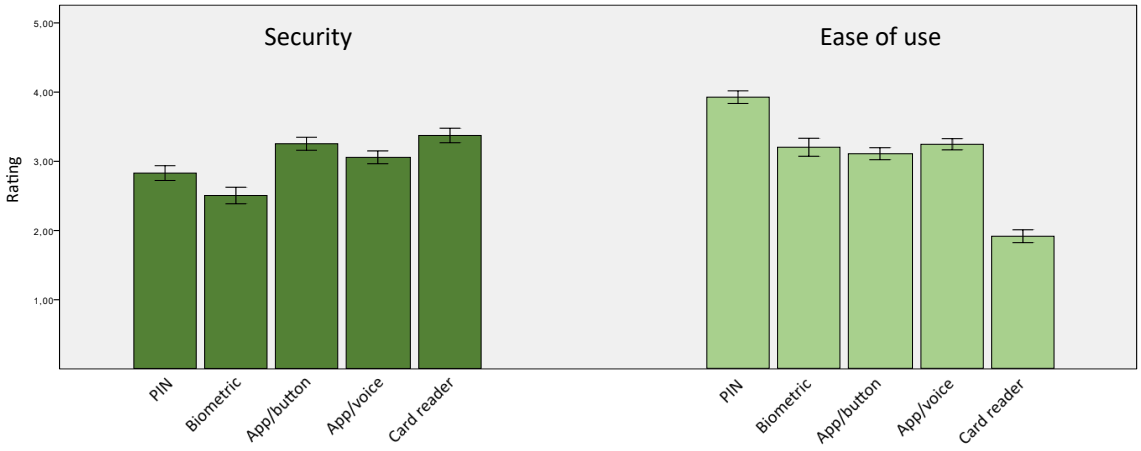

Figure 5: The participants' ratings of the perceived security (left) and ease of use (right) of the five authentication methods.

computer program that imitates voices once you have enough voice recordings. So I am a bit concerned about the security of this method.".

Regarding perceived ease of use, the PIN method was ranked highest with a mean rating of 3.8. The voice, biometric, and button method were rated similarly with mean ratings of 3.2, 3.1, and 3.0, respectively. With a rating of 1.9, the card reader method came out at the bottom. As expected, the statistical analysis proved a significant main effect of the method on the perceived ease of use, $F(3.24,2180.867)=170.17$, $p<.001$.

Pairwise comparisons showed significant differences between the top-rated PIN method and all alternatives, and for the lowest-rated card method and all other methods ($p<.001$). No other pairwise significant differences were found.

The highest-ranking PIN method was frequently described as *"easy to use and quick"* (P556). In contrast, authenticating via the card reader was perceived as tedious since an additional device was needed and the process itself was considered lengthy. For example, P6 described the method as *"[...] too cumbersome these days in my opinion. I don't want to have to use many other devices besides my smartphone. That's why I find smartphone apps to be extremely useful.".*

5.4 Most Important Attributes of Authentication Methods for Voice-based Services

We asked the participants to rank various attributes of authentication methods for voice-based services regarding their subjective importance. Figure 6 depicts how often each factor was ranked 1 (most important) to 6 (least important).

Security was by far the most important factor,: 63% of the participants ranked the importance of security on place 1, 20% on place 2; only 0.5% considered security the least important factor out of the six mentioned. *Privacy* was rated the most important by 13% and the second-most important by 43%, but the least important by 15% of the participants. 18% and 14% of the participants, respectively, considered *efficiency* the most important and the second most important factor in authentication methods for voice-based services (0.2% least important).

Error tolerance was ranked first by 2% and second place by 11% of participants (8% ranked it least important). Both *pleasing* and *time requirement* were low-rated factors. Only 1% and 2%, respectively, ranked these factors on first place, 7% and 4%, respectively, on second place. For 24% and 52% of the participants, these two factors were the least important.

5.5 Overall Preference

When asked for their preferred authentication method to check their account balance, participants rated the button method the best with a mean rating of 3.5. The PIN method was rated second with a mean rating of 3.3, followed by the voice method (3.1) and the card reader (2.8). With a mean rating of 2.4, the biometric method was rated lowest. The statistical analysis showed a significant main effect of the method, $F(4,3355)=72.17$, $p<.001$. All pairwise comparisons except PIN and voice were statistically significant ($p<.003$).

We found the same order of preferences for transferring money: Button was rated best (3.4), followed by PIN (3.3), voice (3.0), card reader (2.9), and the biometric method (2.3). Again, the main effect of the method was significant, $F(4,3355)=72.89$, $p<.001$. Pairwise comparisons showed significant differences between all methods ($p<.001$), except voice and cardreader, and button and PIN. For both tasks, no significant differences were found regarding sex or age.

Figure 7 shows the mean ratings for each method, grouped by the two services investigated. The graph shows the overall consistency of participants' ratings of the methods for the two service types. A significant difference between the two services was found only for the card reader method, which was significantly better at confirming a money transfer ($p<.001$).

The main arguments for the top-rated app/button method are summarized in this response: *"This method is probably the most agreeable to me because it's harder for just anyone to have access to your phone, but it is cumbersome to get out your phone and authenticate. However, it is secure so I would probably check my bank account using this method."* (P608)

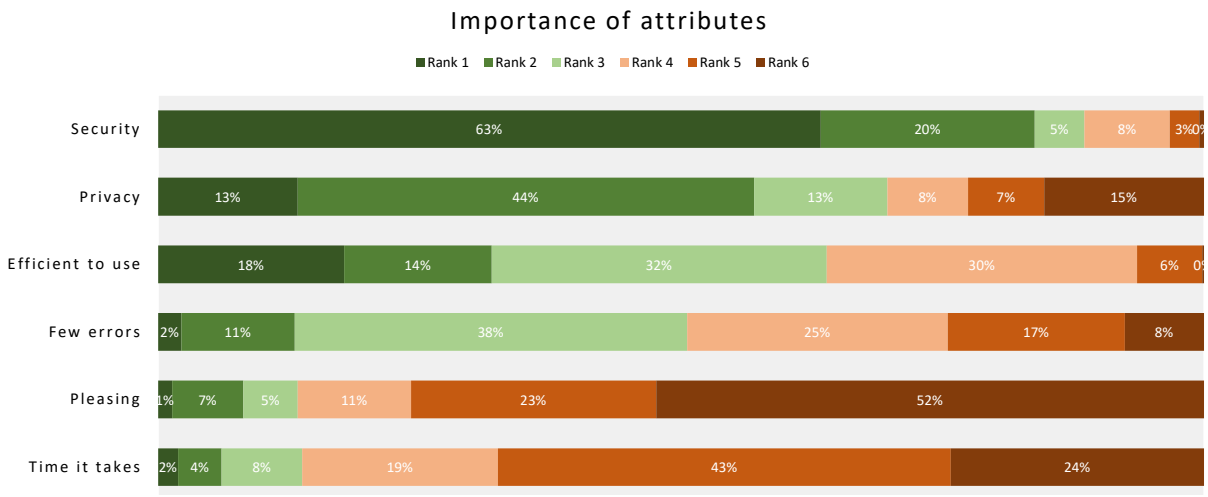

Figure 6: The results of the participants' ranking regarding the importance of different attributes of authentication methods: the top-rated three attributes include security, privacy, and efficient to use.

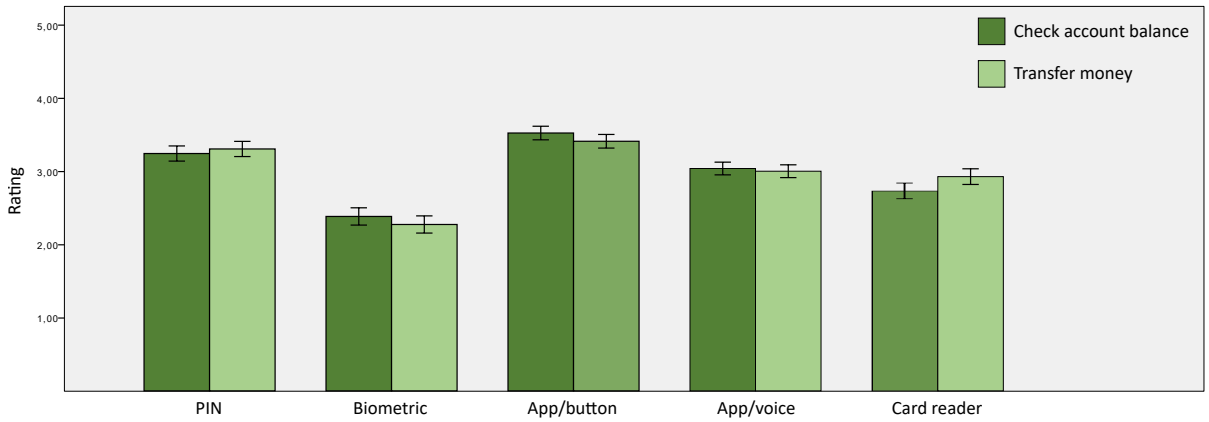

Figure 7: Participants' overall preference of the five authentication methods, regarding two different tasks in the field of voice banking.

6 DISCUSSION

In this section, we refer back to our research questions and provide answers based on the survey results.

RQ1: Which security-critical voice-based services in banking and government are users willing to use?

Overall, we found a preference for information retrieval services over services initiating critical transactions. For the voice banking application, these included checking the account balance and transaction details, for the voice government case reporting defects and sending messages to government representatives. Transferring money and casting votes/requesting tax information were least accepted. We justify this for several reasons. First, we found users' lack of trust in the security and reliability of these devices and several authentication methods. Second, users question the benefit of a voice-based service over existing Web/mobile apps they are familiar with. Third, we assume concerns regarding the complexity of a sophisticated voice-only service. For example, initiating a money transfer comprises several steps, with input errors having critical consequences. Checking and correcting input is considered challenging at a voice-only device with no visual feedback. While our study deliberately focused on smart speaker services, related services seem worth investigating for smart displays with visual support.

It is remarkable that there is general reticence when it comes to voice-based banking services. Checking the account balance was the best-rated service and positively received by 42% of the participants ("slightly agree", "agree", or "strongly agree"), initiating a money transaction via a smart speaker was received positively only by 22% of the participants. Still, for all services investigated, the negative replies outweighed the positive ones.

RQ2: How do users perceive different authentication methods for voice-based services regarding error susceptibility and efficiency?

Four of the methods investigated are perceived as efficient (with ratings of 3.6 and more), only the card reader received significantly worse efficiency ratings. We ascribe this to the necessity of an

additional special device for this method, while the others either do rely on a smartphone or do not require an external device at all.

The button-operated app and the card reader method were perceived best with respect to error susceptibility (i.e., least error-prone). We attribute the top ranking of the authenticator app with button confirmation to its simplicity, which prevents potential errors in generating a code or recognizing a spoken code. Surprisingly, the card reader method was perceived less error-prone than the app variant with voice confirmation, even though both apps rely on the recognition of a spoken code. In addition, the card reader method involves additional steps such as inserting the bank card and unlocking the device. We assume that the users related to previous positive experiences with a card reader for online banking services. On the contrary, we are not aware of an implementation of the voice-confirmed app variant for a publicly available smart speaker service. PIN and biometric authentication are considered more error-prone while in particular the technical reliability of the biometric method is doubted.

RQ3: Which authentication method for voice-based services do users prefer with regard to perceived security and ease of use?

Our study of the perceived security of the methods showed a clear preference for token-based methods. The card reader method was perceived best, followed by the methods involving an authenticator app. We ascribe the card reader's top rank to the users' potential association of the device with its application for online banking services. Similarly, authenticator apps are used by several financial services and have become a de facto standard for security-critical applications in the Web. To our surprise, we found very low confidence in the security of the biometric authentication method. Users perceive this method as immature, in particular it is supposed to be easy to trick.

Contrasting the security assessment, the card reader was perceived to be the worst in terms of ease of use. The process of inserting a bank card as well as generating and providing the code to the smart speaker was perceived as lengthy and cumbersome. While a similar authentication procedure is popular for online banking services, users seem to expect easier-to-use methods for smart speakers which promise convenient and seamless voice-based interactions.

Although the security of the PIN method was rated rather low, it showed its strengths in terms of ease of use. We ascribe this to the knowledge-based approach, which does not require an additional device.

RQ4: Which authentication method for voice-based services is preferred by users in general?

Having been explicitly asked for their preferred authentication method, participants rated the mobile authenticator app with a button best. This is in line with participant's assessments of the method's error susceptibility, efficiency, security, and ease of use. Again, we attribute this overall preference to the familiarity of the method and its availability on smartphones. Note that this top-rated method requires an additional device, in contrast to the PIN method or a biometric method that involves only the smart speaker. It is surprising that the app/button method is perceived as superior to the PIN method, although the latter is the currently predominant method in conversational commerce.

For services with different security requirements, we found a consistent preference for methods. The methods were ranked in same order for requesting information (checking account balance) and initiating a transaction (transferring money). No significant pairwise differences between the service types were found for four methods. Only the card reader was significantly better rated for "transferring money", probably due to the top-ratings for perceived security.

7 CONCLUSION AND FUTURE WORK

We presented a comprehensive online survey on the user experience of authentication methods for voice-based services on smart speakers. While previous work predominantly took a technical perspective on smart speaker security (e.g., by identifying threats and contributing novel authentication schemes), our work contributes scientific knowledge on the users' perception of various authentication mechanisms for smart speakers for the first time.

We studied four crucial acceptance factors (perceived error susceptibility, efficiency, security, and ease of use) and investigated general preferences. We found that the token-based approach of an authenticator app with button confirmation is perceived superior to the PIN method, which is currently the most common authentication method in conversational commerce. A biometric approach exploiting characteristics of the speaker's voice was rated low, in particular regarding perceived security.

The goal of our online survey was a broad view on the user experience of authentication mechanisms for smart speakers. Future work should validate its results through a comparative user study. A respective lab study could use a functional prototype based on an available voice platform (such as *Google Assistant*) or a simple prototype that simulates advanced functionality through a Wizard-of-Oz approach. Furthermore, we consider related questions on novel assistance devices beyond smart speakers worth investigating. Although our study focused on voice-only devices, smart displays (i.e., smart speakers with touch-sensitive displays) provide alternative authentication opportunities.

ACKNOWLEDGMENTS

This work was supported by the projects *conego* (Digital Public Services Switzerland) and *VA-PEPR* (Swiss National Science Foundation, Sinergia, CRSII5_189955).

REFERENCES

[1] Ally Bank. 2019. The Ally Skill for Amazon Alexa. https://www.ally.com/bank/online-banking/how-to-bank-with-ally/alexa/ Accessed: 2022-04-01.

[2] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput.-Hum. Interact.* 26, 3, Article 17 (April 2019), 28 pages. https://doi.org/10.1145/3311956

[3] S. Abhishek Anand, Jian Liu, Chen Wang, Maliheh Shirvanian, Nitesh Saxena, and Yingying Chen. 2021. EchoVib: Exploring Voice Authentication via Unique Non-Linear Vibrations of Short Replayed Speech. In *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security* (Virtual Event, Hong Kong) *(ASIA CCS '21)*. Association for Computing Machinery, New York, NY, USA, 67–81. https://doi.org/10.1145/3433210.3437518

[4] Matthias Baldauf, Raffael Bösch, Christian Frei, Fabian Hautle, and Marc Jenny. 2018. Exploring Requirements and Opportunities of Conversational User Interfaces for the Cognitively Impaired. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, 119–126. https://doi.org/10.1145/3236112.3236128

[5] Matthias Baldauf and Hans-Dieter Zimmermann. 2020. Towards Conversational E-Government. In *HCI in Business, Government and Organizations*. Springer International Publishing, Cham, 3–14. https://doi.org/10.1007/978-3-030-50341-3_1

[6] Matthias Baldauf, Hans-Dieter Zimmermann, and Claudia Pedron. 2021. Exploring Citizens' Attitudes Towards Voice-Based Government Services in Switzerland. In *Human-Computer Interaction. Design and User Experience Case Studies*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 229–238. https://doi.org/10.1007/978-3-030-78468-3_16

[7] Logan Blue, Hadi Abdullah, Luis Vargas, and Patrick Traynor. 2018. 2MA: Verifying Voice Commands vie Two Microphone Authentication. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security - ASIACCS '18*, Jong Kim, Gail-Joon Ahn, Seungjoo Kim, Yongdae Kim, Javier Lopez, and Taesoo Kim (Eds.). ACM Press, New York, New York, USA, 89–100. https://doi.org/10.1145/3196494.3196545

[8] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association, Washington, DC, 57–71. https://doi.org/10.1037/13620-004

[9] Christina Braz and Jean-Marc Robert. 2006. Security and Usability: The Case of the User Authentication Methods. In *Proceedings of the 18th Conference on l'Interaction Homme-Machine* (Montreal, Canada) *(IHM '06)*. Association for Computing Machinery, New York, NY, USA, 199–203. https://doi.org/10.1145/1132736.1132768

[10] Capital One. 2019. Capital One is on Amazon Echo. Questions? Just ask Alexa. https://www.capitalone.com/applications/alexa/ Accessed: 2022-04-01.

[11] Yun-Tai Chang and Dupuis Marc. 2019. My Voicerpint Is My Authenticator: A Two-layer Authentication Approach Using Voiceprint for Voice Assistants. In *Conference: 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation,*, 2019 IEEE SmartWorld (Ed.). IEEE, Leicester, England. https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00243

[12] Huan Feng, Kassem Fawaz, and Kang G. Shin. 2017. Continuous Authentication for Voice Assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking - MobiCom '17*, Kobus van der Merwe, Ben Greenstein, and Kannan Srinivasan (Eds.). ACM Press, New York, New York, USA, 343–355. https://doi.org/10.1145/3117811.3117823

[13] Futurae. 2021. Futurae: Strong Authentication (2FA) and App Security. https://futurae.com/product/iotauth/ Accessed: 2022-04-01.

[14] Google. 2019. Link your voice to your Google Assistant device with Voice Match - Android - Google Assistant Help. https://support.google.com/assistant/answer/9071681 Accessed: 2022-04-01.

[15] Google. 2021. Google Authenticator - Apps on Google Play. https://play.google.com/store/apps/details?id=com.google.android.apps.authenticator2 Accessed: 2022-04-01.

[16] Government Digital Service. 2019. Government uses Alexa and Google Home to make services easier to access. https://bit.ly/32mamSv. Accessed: 2022-04-01.

[17] David Richard Johnson and James C. Creech. 1983. Ordinal Measures in Multiple Indicator Models: A Simulation Study of Categorization Error. *American Sociological Review* 48, 3 (June 1983), 398. https://doi.org/10.2307/2095231

[18] Deepak Kumar, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, and Michael Bailey. 2018. Skill squatting attacks on amazon alexa. In *Proceedings of the 27th USENIX Conference on Security Symposium* (Baltimore, MD, USA, 2018-08-15) *(SEC'18)*. USENIX Association, Berkeley, CA, USA, 33–47.

[19] Il-Youp Kwak, Jun Ho Huh, Seung Taek Han, Iljoo Kim, and Jiwon Yoon. 2019. Voice Presentation Attack Detection through Text-Converted Voice Command Analysis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300828

[20] Sara Lafia, Jingyi Xiao, Thomas Hervey, and Werner Kuhn. 2019. Talk of the Town: Discovering Open Public Data via Voice Assistants (Short Paper). In *14th International Conference on Spatial Information Theory (COSIT 2019) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 142)*, Sabine Timpf, Christoph Schlieder, Markus Kattenbeck, Bernd Ludwig, and Kathleen Stewart (Eds.). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 10:1–10:7. https://doi.org/10.4230/LIPIcs.COSIT.2019.10

[21] Yeonjoon Lee, Yue Zhao, Jiutian Zeng, Kwangwuk Lee, Nan Zhang, Faysal Hossain Shezan, Yuan Tian, Kai Chen, and XiaoFeng Wang. 2020. Using Sonar for Liveness Detection to Protect Smart Speakers against Remote Attackers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 16 (mar 2020), 28 pages. https://doi.org/10.1145/3380991

[22] Xinyu Lei, Guan-Hua Tu, Alex X. Liu, Chi-Yu Li, and Tian Xie. 2018. The Insecurity of Home Digital Voice Assistants - Amazon Alexa as a Case Study. http://arxiv.org/pdf/1712.03327v2

[23] Rui Liu, Cory Cornelius, Reza Rawassizadeh, Ronald Peterson, and David Kotz. 2018. Vocal Resonance. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–23. https://doi.org/10.1145/3191751

[24] Divyakant T. Meva and C.K Kumbharana. 2013. Comparative Study of Different Fusion Techniques in Multimodal Biometric Authentication. In *International Journal of Computer Applications (0975 – 8887)*, IJCA (Ed.). Foundation of Computer Science (FCS), NY, USA, New York, USA, 16–19.

[25] Microsoft. 04/01/2020 06:23:54. Microsoft Authenticator - Apps on Google Play. https://play.google.com/store/apps/details?id=com.azure.authenticator Accessed: 2022-04-01.

[26] Jakob Nielsen. 1994. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[27] Alexander Ponticello, Matthias Fassl, and Katharina Krombholz. 2021. Exploring Authentication for Security-Sensitive Tasks on Smart Home Voice Assistants. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. USENIX Association, Berkeley, CA, USA, 475–492. https://www.usenix.org/conference/soups2021/presentation/ponticello

[28] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. 2018. BackDoor: Sounds That a Microphone Can Record, but That Humans Can't Hear. *GetMobile: Mobile Comp. and Comm.* 21, 4 (Feb. 2018), 25–29. https://doi.org/10.1145/3191789.3191799

[29] Sparkasse. 2019. Sparkasse Banking-App. https://www.sparkasse.de/unsere-loesungen/privatkunden/rund-ums-konto/voice-banking/voice-banking-sparkasse.html Accessed: 2022-04-01.

[30] Statista. 2022. Smart Speakers - Statistics & Facts. https://www.statista.com/topics/4748/smart-speakers Accessed: 2022-06-15.

[31] Bharath Sudharsan, Muhammad Intizar Ali, and Peter Corcoran. 2019. Smart speaker design and implementation with biometric authentication and advanced voice interaction capability. In *27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science*. CEUR-WS.org, Aachen.

[32] U.S. Bank. 2018. How U.S. Bank aims to shape the future of voice banking. https://www.usbank.com/newsroom/stories/new-technology-you-can-bank-on.html Accessed: 2022-04-01.

[33] Varun S. Rao and Teng Song. 2018. Biometric Enabled Proximity-based User Authentication. https://www.tdcommons.org/dpubs_series/1296

[34] Riku Vassinen. 2018. The rise of conversational commerce: What brands need to know. *Journal of Brand Strategy* 7, 1 (2018), 13–22.

[35] Yao Wang, Wandong Cai, Tao Gu, Wei Shao, Yannan Li, and Yong Yu. 2019. Secure Your Voice: An Oral Airflow-Based Continuous Liveness Detection for Voice Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 157 (Dec. 2019), 28 pages. https://doi.org/10.1145/3369811

[36] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. DolphinAttack: Inaudible Voice Commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (Dallas, Texas, USA) *(CCS '17)*. Association for Computing Machinery, New York, NY, USA, 103–117. https://doi.org/10.1145/3133956.3134052

[37] Linghan Zhang, Sheng Tan, and Jie Yang. 2017. Hearing Your Voice is Not Enough. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security - CCS '17*, Bhavani Thuraisingham, David Evans, Tal Malkin, and Dongyan Xu (Eds.). ACM Press, New York, New York, USA, 57–71. https://doi.org/10.1145/3133956.3133962

[38] Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian, and Feng Qian. 2019. Dangerous Skills: Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, Los Alamitos, CA, USA. https://doi.org/10.1109/sp.2019.00016