Andreas Renz, Thomas Neff, Matthias Baldauf* and Edith Maier

Authentication methods for voice services on smart speakers – a multi-method study on perceived security and ease of use

https://doi.org/10.1515/icom-2022-0039
Received November 16, 2022; accepted February 10, 2023;
published online March 29, 2023

Abstract: With the increasing proliferation of security-critical voice-based services such as voice banking, user authentication on smart speakers is becoming a vital requirement. Prior research on verifying the speaker's identity has been taken a technical perspective predominantly, while respective user-centered research is scarce. To investigate authentication methods for smart speakers from a user's perspective, we conducted a multi-method experiment. In a comprehensive online survey ($n = 696$) and a comparative lab study ($n = 18$) with an advanced functional prototype we studied 6 authentication methods (spoken PIN, biometrics, app with button/voice confirmation, card reader, sound authentication) regarding their perceived security and ease of use. While token-based authentication approaches (in particular an authenticator app on a smartphone) typically are perceived as more secure, they are found inferior when it comes to the ease of use. The currently most frequently used authentication method for smart speakers, the spoken PIN method, seems to represent a compromise between security and ease of use. The sophisticated sound authentication was appreciated for its ease of use, however, was rated worst regarding the perceived security.

Keywords: conversational services; user authentication; voice assistant.

***Corresponding author: Matthias Baldauf**, Eastern Switzerland University of Applied Sciences, Institute for Information and Process Management, St. Gallen, Switzerland, E-mail: matthias.baldauf@ost.ch. https://orcid.org/0000-0002-1876-5082

Andreas Renz, Skyguide Swiss Air Navigation Services Ltd, Wangen bei Dübendorf, Switzerland, E-mail: andreas.renz@gmx.ch

Thomas Neff, Innovation Process Technology AG, Bern, Switzerland, E-mail: thomas.neff@tutanota.com

Edith Maier, Eastern Switzerland University of Applied Sciences, Institute for Information and Process Management, St. Gallen, Switzerland, E-mail: edith.maier@ost.ch

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 [1].

While many popular voice-controlled applications involve non-critical tasks, such as playing music or searching the Web [2], an 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, voice-based banking services for checking one's account balance or even initiating a money transfer [3], and conversational government services for citizens [4]. 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 [5–11]. From a users' perspective, prior work looked at user concerns regarding attacks and threats as well as at mitigation strategies [12]. 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.

[1] Google Nest: https://store.google.com/product/nest_audio.

[2] Amazon Echo: https://www.aboutamazon.com/news/devices/alexa-news-2022.

[3] Apple Homepod: https://www.apple.com/homepod/.

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. Second, we compare users' perceptions of six authentication methods in a multi-method study: We combine a comprehensive online survey ($n = 696$) and a user study with an advanced functional prototype. In detail, we compare knowledge-based (PIN), biometric (voice), and both interactive (authenticator app, card reader) and zero-touch (sound authentication) token-based approaches with regard to their perceived security and ease of use.

This comprehensive multi-method study 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.

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 [13]. 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 [15]. 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 [16]. 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 [17].

In addition, more experimental approaches have been proposed. *Continuous authentication* [7] and *voice resonance* [18] 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 [7]. While continuous authentication concepts are based on wearables such as watches, glasses, earphones, or necklaces [7], vocal resonance depends on a microphone that can be worn on the head, neck, or chest [18]. A further experimental alternative for user verification is *SpeakerSonar* [19] which makes use of inaudible sounds to track the user's direction and compares it with the direction of the received voice command. In a similar vein, *HandLock* by Zhang and Das [20] makes use of built-in microphones and speakers to generate and sense inaudible acoustic signals to detect the presence of a known (i.e., authorized) hand gesture.

2.2 State-of-the-art in securing voice-based services

As a pioneer in the field of conversational commerce [21], *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 [22], Capital One [23], and Ally Bank [3]) as well as from Germany (Sparkasse [24]).

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 [24]. The functions are very similar, including checking account balance, checking recent transaction details, billing due dates, and even transferring money [3].

One approach that addresses the risks related to using a PIN or passcode is two-factor authentication, as, for example, offered by *Futurae* [25]. 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. 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 various threats for voice-based services (cf., [26]). In the following, we summarize user-related threats which originate from the voice interfaces of involved devices and the presence of a speaker emphasizing the need for suitable authentication methods (i.e., in contrast to threats for these connected devices on a network level).

Lei et al. [27] 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, 29]. An example is *DolphinAttack* by Zhang et al. [29]. 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 [30, 31]. Zhang et al. [31] 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 [30]. Countermeasures include word-based and phoneme-based analysis during the publisher's certification process [30] and context-sensitive detectors assessing the impersonation risk [31].

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. We address the following research questions to close this gap:

RQ1: How do users perceive different authentication methods for voice-based services on smart speakers regarding their perceived security?

For the adoption and acceptance of security-critical voice-based services on smart speakers, not only the technical security is relevant. To eventually trust and continuously use such a service, the perceived security of an authentication method plays a crucial role.

RQ2: How do users perceive different authentication methods for voice-based services on smart speakers regarding their ease of use?

To consider the methods' practical acceptability, we contrast the security assessment with the perceived ease of use of the methods. Smart speakers, in particular, promise convenient and natural interactions with digital services; an additional authentication mechanism should not disrupt formerly convenient processes but be easy to use.

To answer these research questions, we conduct a multi-method experiment. First, we aim at a broad view on smart speaker authentication through a large-scale online survey (see Section 4). Then, in a complementary follow-up study we compare selected authentication methods with the help of an advanced functional prototype in a lab environment (see Section 5).

4 Study 1: online survey

To find answers to 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 Study design

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 online survey, 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.

- **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.
- **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.
- **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.
- **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 [32] and the Google Authenticator [33]. 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.
- **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.

4.3 Questionnaire

The first section contained demographic questions (country of residence, year of birth, sex, highest educational degree, employment status) and a question on the participants'

[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.

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 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). To avoid order effects, the five methods were presented in random order.

For each method, we asked participants to rate their agreement with several statements (e.g., “I consider this method prone to errors”) on Likert scales. Reasons for the ratings and additional remarks could be provided in free-text fields. Details on these results of the ratings and assessments per method can be found in [anonymized].

After the participants had become familiar with all five authentication methods, we asked them to rank the methods regarding the perceived security (first rank – highest security; fifth rank – lowest security) and perceived ease of use (first rank – highest ease of use; fifth rank – lowest ease of use). Again, the participants were able to give reasons for their assessments using free-text fields.

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 as interval scales (cf. [34]).

A thematic qualitative analysis [35] 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

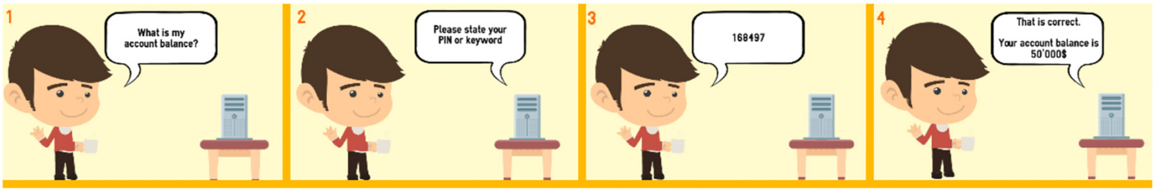
1
What is my account balance?
2
Please state your PIN or keyword
3
168497
4
That is correct.
Your account balance is 50'000$

Figure 1: In the online survey, each authentication method (spoken PIN, biometrics, app with button/voice confirmation, card reader) was explained through a short four-step cartoon within the questionnaire. The example above shows the illustration for the “spoken PIN” method.

from 16 to 75 years ($M = 31.3$; SD = 10.8). Our survey reached broad participant groups: 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 Results

In the following section, we summarize the results of our online survey with regard to participants' perceptions of the perceived security and ease of use of the methods.

4.6.1 Perceived security

Figure 2 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, $F(3.70, 2495.97) = 34.12, p < 0.001$. Pairwise comparisons showed statistically significant differences between the card reader method and the voice, PIN, and biometric method ($p < 0.002$). The biometric method with the lowest rating was significantly worse in terms of perceived security than the four alternatives ($p < 0.001$). Further significant differences were found for button and PIN ($p < 0.001$), voice and both PIN and biometric ($p < 0.019$), as well as PIN and all four alternatives ($p < 0.019$).

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).

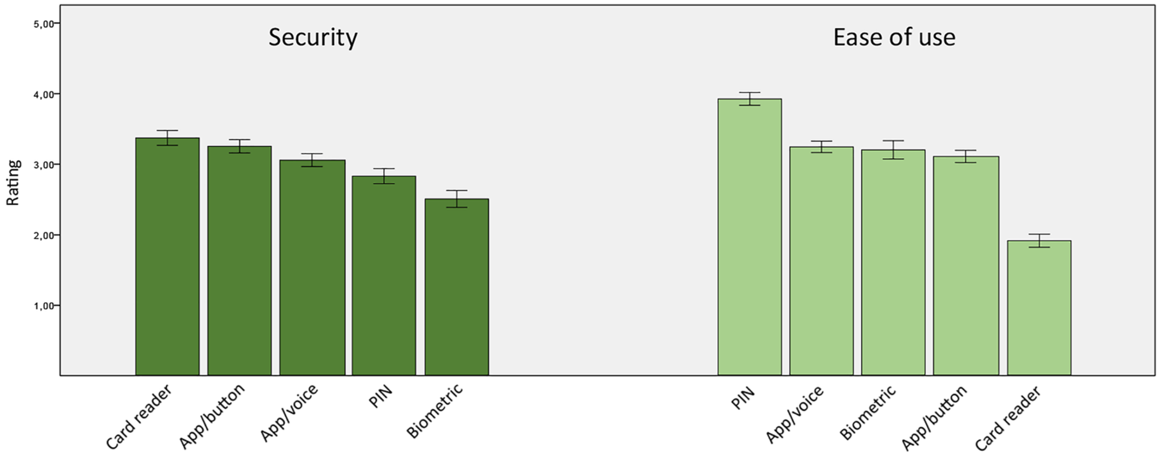

Figure 2: Online survey: the participants' ratings of the perceived security (left) and ease of use (right) of the five authentication methods.

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

4.6.2 Ease of use

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, $F(3.24,2180.867) = 170.17$, $p < 0.001$.

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 ($p < 0.001$). 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 Study 2: comparative lab study

While the online survey provided a first glimpse of users' perceptions of authentication methods for smart speakers, we decided to conduct a follow-up experiment in form of a comparative lab study. The goals of this complementary experiment were (1) validating the survey findings through a prototype and (2) adapting the set of authentication methods according to these first insights.

5.1 Study design

Our comparative study was designed as a within-subject experiment, i.e., each participant tested all of the authentication methods. It was conducted in our university's usability lab and led by one researcher who guided the participants through the procedure. Figure 3 shows the study setup. Overall, each test comprised three phases:

Phase 1: Introduction and exploration. After obtaining the participant's written informed consent, the moderator started the test with questions on the participant's demographics. Furthermore, he asked whether the participant possessed a smart speaker and how often he or she used a voice assistant. Finally, the moderator asked whether the participant likes to test new features of technical devices (5-point Likert scale from 1-very much to 5-not all) to obtain the participant' interest in new technology.

Having completed the introductory questions, the participant was presented the study prototype. The moderator explained the two task types supported by the prototype, checking the account balance and settling an invoice, and demonstrated how to use them. Participants could then experiment with the prototype until they felt comfortable.

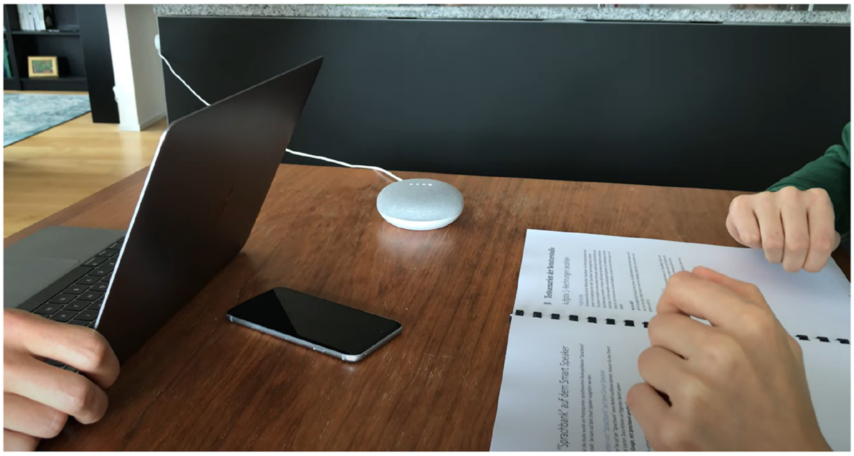

Figure 3: The study setup involved a Google Home Mini smart speaker running the functional prototype and an iPhone 6s with the authenticator apps installed. Participants (right) received written instructions for the tasks. A test assistant (left) guided the participants through the overall test procedure.

In this free exploration phase, no authentication method was applied.

Phase 2: Test scenarios. In this main phase of each test, the participants were asked to complete five tasks, each with a different authentication method. We randomly chose "account balance" and "invoice settlement" tasks in order to make the test procedure varied, yet keep the number of conditions to a manageable size. The order of the authentication methods was systematically varied to avoid sequence effects.

We decided to evaluate four of the authentication methods from the online survey: PIN, Biometrics, and the two app variants with voice and button confirmation. However, instead of the card reader method, which had received worst ratings regarding ease of use and thus seemed ineligible for the smart speaker scenario, we went to include a more advanced method designed for smart speakers: Futurae's smart assistant authentication (cf. Section 2.2).

For each task, the participants received written information on how to set up and use the authentication method. The major steps required for each authentication method were as follows:

- **PIN**: define a custom PIN through the voice command "Setup PIN"; say the PIN when being asked for by the banking service
- **Biometrics**: capture a voice profile through the voice command "Record voice profile"; repeat a randomly selected word when being asked for by the banking service
- **App/voice**: when being asked for by the banking service, open the authenticator app and speak out loud the code displayed
- **App/button**: when being asked for by the banking service, confirm in authenticator app by pushing a button
- **Sound authentication**: for authentication the mobile phone plays a sound which then is detected by the smart speaker

The participants conducted the tasks independently. Still, they could ask the test assistant in case of questions or issues with the prototype.

Phase 3: Comparison and debriefing. Once, a participant had experienced all authentication methods by completing the five tasks, the test assistant conducted a structured final interview involving a brief questionnaire. Participants were asked to compare the five authentication methods and rank them with regard to the perceived security and ease of use. The moderator closed the interview with a final open question on additional remarks and thoughts on the authentication methods experienced and smart speaker authentication, overall.

The tests were conducted in April 2022. Interviews were audio-recorded and, in addition, participants' qualitative feedback was written down by the test assistant during the interviews. Each test session took between 30 and 45 min.

5.2 Study prototype

For gaining results with high ecological validity, we designed and implemented a fake voice-banking application for the study. This functional prototype was supposed to provide a realistic experience to the participants while being configurable for various test conditions. It featured two voice-controlled services, checking the account balance and settling invoices, and five different authentication methods.

Figure 4 shows a high-level architecture overview. We developed the prototype for *Google Home Mini* smart speaker using *Dialogflow Essentials*. This platform facilitates the creation of conversational agents by supporting the definition of dialogues without coding. So-called *Webhooks* enable integrating external services which we used, for example, to set a specific authentication method during execution of the application.

The application could be started with the phrase *"Talk to voice bank"*. The banking application then welcomed the user and asked *"How can I help you')*. By replying *"I'd like to settle an invoice"* or *"What's my account balance"*, the user could start one of the two services:

- **Checking the account balance:** The application returns the static answer *"Your account balance is 17′536.90 Euro."*
- **Settling an invoice:** The application reads each of four invoices (with biller and amount information) and asks the user whether it should be settled.

The authentication methods were implemented in the prototype as follows:

- **User-defined PIN**. This method was implemented as 4-digit variant to resemble its real-life counterparts, e.g., *Amazon's Voice Code.* The prototype supported the definition of an individual PIN by the participants through a custom intent triggered by the command *"Setup PIN".* We made use of a *Dialogflow* parameter of type *number-sequence* to access the spoken PIN and stored it using a self-written Web service (accessed via a Webhook). In a similar manner, we retrieve the PIN during an authentication attempt. If the comparison with the stored PIN fails, the user is asked to try again (without any maximum number of failed attempts). If

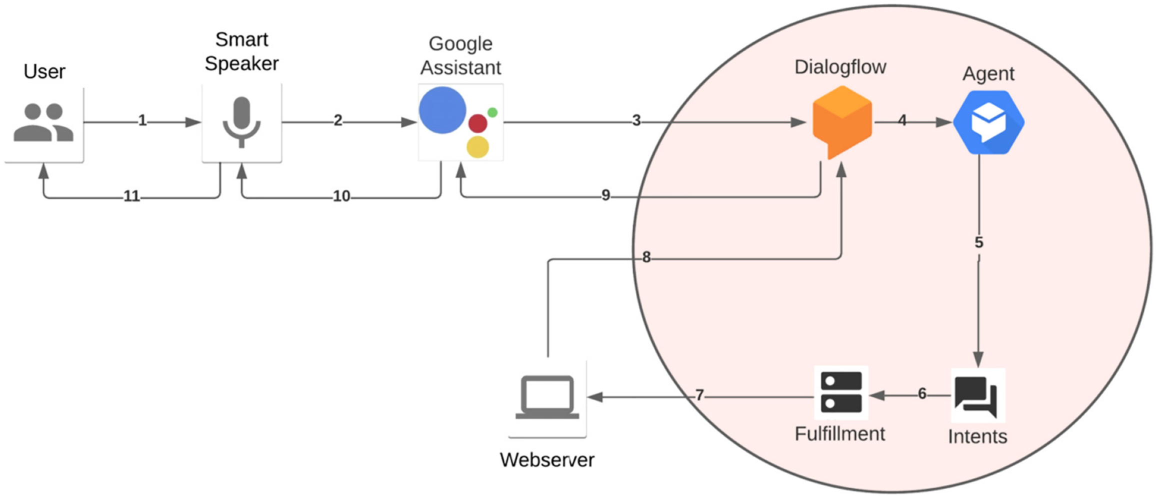

Figure 4: High-level architecture of the study prototype and process steps from a user's input to the smart speaker's response (1–11). Dialogues were modelled in *Dialogflow* and custom logic implemented as services on an external webserver linked through *Webhooks*.

the comparison is successful, the application starts one of the two banking services mentioned above.

- **Biometric authentication**. Similarly to the PIN method, we implemented a fake biometric voice recognition for user authentication. The command *"Record voice profile"* triggers an intent which asks the user to repeat three words. After the user did so, the application pretends to have completed the voice profile and thanks the user. For authentication, the prototype asks to repeat a word (predefined and constant since each participant uses each authentication only once). Again, with the help of a *Dialogflow* parameter, the prototype can access the spoken word and compare it with the predefined one. While this comparison is functional, the prototype obviously does not match voices.
- **Authenticator app with button confirmation**. For implementing this authentication method, we installed *Microsoft Authenticator* on the study smartphone, an *Apple iPhone 6s.* Using *Selenium*,[7] a browser automation tool, we were able to initiate a login in a pre-defined Microsoft account with two-factor-authentication enabled. Thus, the app showed the corresponding confirmation prompt while the user was tricked into believing the voice assistant triggered this action.
- **Authenticator app with voice confirmation**. We installed *Google Authenticator* on the study smartphone to simulate this authentication method. We made use of *Speakeasy*,[8] a one-time passcode generator, to create a valid secret code which we both used in the authenticator app and the voice banking app. Similarly to the PIN method, we were able to access the numeric code spoken by the user via a *Dialogflow* parameter. The codes were compared through *Speakeasy* and, in case of a match, authentication granted.
- **Sound authentication**. For mocking this authentication technique inspired by *Futurae*, we needed to play an audio clip on the study smartphone upon an authentication attempt. To efficiently implement this feature, we made use of *Pushover*,[9] a platform to easily receive and react on push notifications. Through a RESTful API, we were able to trigger a notification with a specific melody. The mobile *Pushover* app was executed in the background, so participants only experienced the melody on the smartphone as soon as the voice banking service requested authentication.

5.3 Participants

For the comparative study, we recruited 18 Swiss participants through convenience sampling within the researchers' acquaintances. We paid attention to compile a well-balanced sample regarding age ($M = 38.6$, $SD = 12.2$) and sex (10 male, 8 female). For the age structure, we aimed at modeling a typical age distribution of online banking usage in [anonymized], i.e., two thirds of the

[7] https://www.selenium.dev/.

[8] https://github.com/speakeasyjs/speakeasy.

[9] https://pushover.net/.

users are under 40 years of age [36]. Table 1 shows the sample's age distribution in detail. Overall, the sample comprised tech-savvy subjects. 16 of the 18 participants saw great potential of smart speakers for everyday use cases and expected widespread usage of the technology. All participants knew PIN-based authentication and 15 participants were familiar with authenticator apps. None of the participants knew about the sound authentication method. Only one participant had used the biometric method, i.e. voice recognition, before. All participants were familiar with Web banking services and use them regularly.

5.4 Data analysis

The data collected through the comparative lab study was analyzed analogous to the one from the online survey (cf. Section 4.4). Participants' order of the authentication methods was mapped to quantitative assessments from 1 (lowest) to 5 (best) during the analysis. Again, SPSS was used to compare methods with a general linear model repeated measures analysis of variance. The participants' responses to open questions were analyzed through a thematic qualitative analysis [35]. Again, two researchers coded the responses to find common themes and patterns. Verbatim quotations (translated to English by the researchers in case of non-English originals) were selected to illustrate relevant themes.

5.5 Results

This section reports on the results of the comparative lab study. We present the quantitative assessments of the participants and summarize their qualitative responses illustrated by representative remarks.

5.5.1 Perceived security

Figure 5 depicts the average scores for each method. With regard to the perceived security, the study participants ranked the authenticator app with voice confirmation best (4.1), followed by the app with button confirmation (3.6). PIN and voice recognition performed similarly (2.7 and 2.5, respectively). The sound authentication method was ranked lowest (2.2).

Table 1: Participant distribution regarding age groups.

Age group	Male	Female
18–29	3	3
30–39	3	3
40–49	2	0
≥50	2	2

Statistical analysis found a main effect of the methods on the perceived security, $F(4,68) = 5.969$; $p < 0.001$. Pairwise comparisons shows significant differences between the highest-ranked app with voice confirmation and the methods PIN, voice recognition and sound authentication ($p < 0.01$). A further significant difference was found between button confirmation and the lowest-ranked sound authentication method ($p < 0.013$).

A closer look at the results revealed that participants experienced the authentication methods' security very differently. This is indicated particularly by the fact that each of the five authentication methods was rated best (and four of them rated worst) at least by one participant. Remarkable is that the authenticator app with voice confirmation was ranked only first or second; the authenticator app with button confirmation was ranked first place like the one with voice confirmation six times, yet more often third or fourth than its app-based competitor. Furthermore it is remarkable that the biometric method was rated fourth and fifth by 6 and 5 participants, respectively; the sound authentication method was rated worst regarding perceived security even by 8 of the 18 participants.

For both of these "audio-based" methods, several participants explained their concerns with the lack of knowledge and transparency on how these advanced approaches work. For the biometrics approach, a few participants had doubts on the method's security and reliability, e.g., when the user had a cough. Two participants suggested combining the biometrics approach with the PIN-based method.

With regard to the sound authentication, several participants articulated their mixed feelings about this method such as *"I've never seen that before. it might be very secure, but it's strange somehow"* (P15) or *"It's strange because it's so unfamiliar"* (P17). P18 also mentioned the novelty of the approach, yet *"[for me] it feels more secure because a second device is involved"*. Three participants associated their perception of security with an active interaction with the system as illustrated by P3: *"This doesn't feel secure to me, I guess I only feel secure when I enter a code."*

5.5.2 Ease of use

With regard to the ease of use of the methods, the study participants ranked sound authentication and biometrics

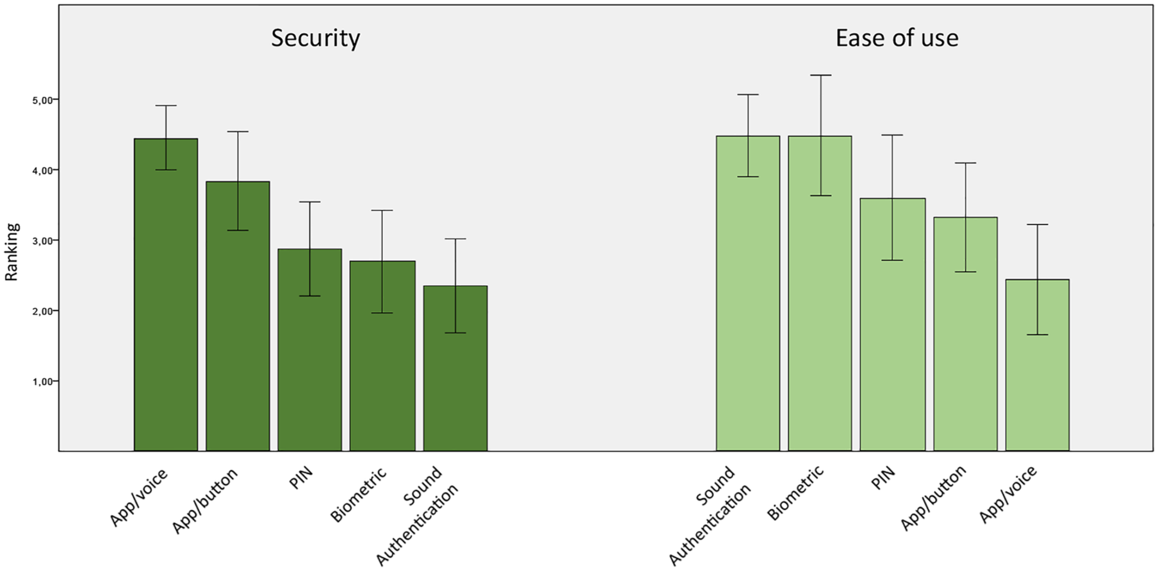

Figure 5: Comparative lab study: the participants' ratings of the perceived security (left) and ease of use (right) of the five authentication methods.

best (3.7), followed by the PIN method (2.9), and the app with button confirmation (2.7). Lowest ranked was the app with voice confirmation (2.0).

We found a main effect of the methods on ease of use, $F(4{,}68) = 4.168$; $p < 0.01$. Pair-wise posthoc tests showed that the two top-ranked methods, sound authentication and biometrics, were significantly better ranked than the app with voice confirmation ($p < 0.002$). The pairwise tests did not show any further significant differences between the methods with regard to the ease of use.

The participants explained their good ratings for sound authentication and biometrics with the fact that no explicit interaction is required (sound authentication) or no additional devices is needed (biometrics). Several participants emphasized the advantage of the sound authentication approach, that *"[they] don't have to take the smartphone in hand"* (P12).

In contrast, for the two app-based methods, several participants criticized the *"media disruption"* (P7), i.e., having to interact actively with a second device and switching their attention from one device (the smart speaker) to another (the smartphone) and back, respectively. They argued that these methods are not suitable for smart speakers: *"Then I can use the voice service on my smartphone right away"* (P15). Still, some participants appreciated the methods' simplicity, in particular of the button-based method, and emphasized that they are familiar with related methods for Web-based online banking.

6 Discussion

In this section, we refer back to our original research questions on the perceived security and ease of use of the methods investigated. We bring together the results of both studies and discuss the findings. Table 2 shows an overview of the results comparing the assessment of the authentication methods in both studies.

6.1 RQ1: perceived security

Regarding the perceived security, our online survey showed a clear preference for token-based methods. The card reader method was perceived best, followed by the methods that involved an authenticator app. We attribute 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.

These first insights were validated by the lab study. Again, the token-based methods, i.e., the two authenticator app variants, were ascribed the best security. These results obtained in our user study with functional prototypes, are in line with findings from a prior scenario-based interview study by Ponticello et al. [12]. Several of their

Table 2: Results overview for all authentication methods investigated in both studies (mean ratings from 1 – worst to 5 – best; ranks from 1 – best to 5 – worst).

Method	Security				Ease of use			
	Survey		Lab study		Survey		Lab study	
	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
App/button	3.3	2	3.6	2	3.0	4	2.7	4
App/voice	3.1	3	4.1	1	3.2	2	2.0	5
Biometric	2.5	5	2.5	4	3.1	3	3.7	1
Card reader	3.4	1	–	–	1.9	5	–	–
PIN	2.8	4	2.7	3	3.8	1	2.9	3
Sound authentication	–	–	2.2	5	–	–	3.7	1

participants also favored token-based authentication methods for security, since “tokens would not be susceptible to the openness of the voice input channel” [12]. Despite its technical sophistication, our participants rated the novel method of sound authentication significantly lower and worst of all the methods investigated.

While the app variants were familiar to most of the participants and did not raise many questions, none of the participants had heard of the sound authentication approach before. It seems, many participants drew their assessment of a method’s security from prior experiences with related technologies and/or knowledge about the technical functionality. The lack of understanding, how the sound authentication works, and its fully automated, non-interactive procedure led to disconcerting moments for several participants which seemed to impact their perception of security. While prior research showed that many users of voice assistants are unsure about the general flow of privacy-related data in such ecosystems [37, 38] and involved (established) authentication processes [12], our study uncovered users’ discomfiture when using novel authentication methods for voice-based services and, subsequently, the lack of trust in the methods, when they do not understand the functionality. Several statements indicate that an active manual task (such as pushing a confirmation button, scanning a fingerprint, typing in a password, etc.) is associated with secure authentication. The absence of such an active task and thus a certain lack of control might have further impacted the respective assessments.

While the lab study confirmed most findings from the online survey with regard to the perceived security, we found the biggest difference in assessments for the voice-based authenticator app. The participants of the lab study rated its security significantly better. As the ratings for the app with button confirmation (a general method many users are familiar with) were similar in the survey and the lab study, we assume that experiencing the prototype of the novel voice-based variant helped participants to better understand and, subsequently, more accurately rate their impression.

6.2 RQ2: ease of use

When it comes to ease of use, the card reader was perceived to be the worst method in our online study. 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.

Our follow-up lab study also showed a preference for approaches that do not require a smartphone or, at least, do not require interaction with it. Our participants (non-significantly) preferred the novel sound authentication and biometric authentication over the PIN method. The authenticator app with voice confirmation was considered too much effort.

Having not to interact with an additional device (i.e., the smartphone) was one of the participants’ most frequently used arguments. The high ratings for the biometric authentication confirm results from prior work [12], which also ascribes the preference to “the natural and effortless interaction”. With regard to the case of smart speakers, which promise seamless convenient interactions with digital services, this is a reasonable argument. Even though the smartphone is a steady companion, picking up this device during a conversation with a smart speaker (or a corresponding conversational application, respectively) breaks the user’s experience.

Even more, since voice assistants and their third-party extensions are available on smartphones as well, the question arises whether the service is not consumed on the smartphone from the very beginning when active usage of a smartphone is required for authentication.

When comparing our results regarding ease of use from the online survey and the lab study, we found major differences for the app with voice confirmation and the PIN method. For these two methods, using a prototype led to strikingly worse scores. We assume that the advantages of the zero-touch authentication methods became much more apparent (and the weaknesses of the alternatives, respectively), as participants practically experienced the methods one after another.

6.3 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. biometrics). Further methods and variants of the considered methods remain subject to future work.

We managed to recruit a large number of participants through social media channels (online survey) and convenience sampling (lab study). 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 (online survey) and the introductory session (lab study). Some user groups that particularly benefit from voice-controlled services, e.g., the elderly and people with impairments (cf. [39]), are probably underrepresented in our sample. Obviously, their requirements need to be taken into account when designing a universally accessible and secure voice service.

7 Conclusion and outlook

In this work, we presented a multi-method study on the perceived security and ease of use of user authentication methods for smart speakers. The number of advanced security-critical voice applications is increasing, yet scientific knowledge on users' perceptions of suitable authentication method is still scarce. 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.

Our multi-method study comprised a comprehensive online survey and a comparative lab study with an advanced functional prototype. In addition to established authentication methods such as authenticator apps or spoken PIN, we investigated a novel sound-based authentication approach for smart speakers, which has not been studied from a user perspective so far. The results of our multi-method study revealed the well-known dilemma of reconciling security and usability: While token-based authentication approaches (in particular an authenticator app on a smartphone) typically are perceived as more secure, they are found inferior when it comes to the ease of use. The currently most frequently used authentication method for smart speakers, the spoken PIN method, seems to represent a compromise between security and ease of use. The modern and sophisticated sound authentication was appreciated for its ease of use, however, was rated worst regarding the perceived security.

We consider the perceived ease of use of an authentication method to be a key criterion for the widespread usage and acceptance of voice-based services that give access to sensitive personal data or trigger security-critical transactions (cf. [12]). Conversational assistants with their manifold services came in with promises to make interaction with digital applications natural, seamless, and convenient. Authentication methods that complicate and prolong this interaction, counteract users' expectations and will not lead to long-term acceptance of the respective services. Thus, we see great potential for zero-touch authentication schemes, such as the biometric method or sound authentication. Particularly in the lab study, where participants tried out and experienced the methods through prototypes, the participants appreciated the ease of use of these two methods. We found that their novelty and non-transparency causes doubts on their security, though. Based upon these findings, we suggest including mechanisms to make novice users familiar with non-standard authentication methods. This could include adding hints and brief tutorials on how authentication works to make users familiar with the method. Furthermore, we recommend establishing (auditory) clues to communicate the security of voice-based services. In graphical user interfaces, we rely on visual hints (such as the lock symbol in a Web browser's navigation bar) to quickly assess the security of a service. Related indicators are currently missing for voice-based services without visual feedback.

These results and observations lead to further research challenges to be addressed in future work. In this work, we focused on authentication methods that were already available in mass-market smart speakers, or their integration into smart speakers could be expected in the near future. However, researchers continuously present novel suitable authentication approaches (e.g., based on wearables), partly particularly designed for smart speakers (cf. Section 2.1). We argue for user-centered investigations of these novel methods beyond technical evaluations and performance assessments. Today, people are mainly familiar with authentication methods which require (at least minimal) user interaction (e.g., confirming in an authenticator app). When it comes to modern zero-touch approaches (such as the sound authentication method in this work), ways to indicate the technology's security to novice users are worth investigating.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This work was funded by Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (no. CRSII5_189955).

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Declaration: This article is a modified and extended version of our previously published conference paper [40]. Whereas the original paper contains several more results from the online survey, we focused this version on the aspects of perceived security and ease of use and complemented the respective former results by a comparative lab study with an advanced functional prototype.

References

1. Statista. *Smart Speakers — Statistics & Facts*, 2022. https://www.statista.com/topics/4748/smart-speakers (accessed Nov 15, 2022).
2. Ammari T., Kaye J., Tsai J. Y., Bentley F. Music, search, and iot: how people (really) use voice assistants. *ACM Trans. Comput. Hum. Interact.* 2019, *26*, 1—28.
3. Ally Bank. *The Ally Skill for Amazon Alexa*, 2019. https://www.ally.com/bank/online-banking/how-to-bank-with-ally/alexa/ (accessed Nov 01, 2022).
4. Baldauf M., Zimmermann H. D. Towards conversational e-government. In *HCI in Business, Government and Organizations*; Springer International Publishing: Copenhagen, Denmark, 2020; pp. 3—14.
5. Anand S. A., Liu J., Wang C., Shirvanian M., Saxena N., Echovib Y. C. 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, ASIA CCS '21*; Association for Computing Machinery: New York, USA, 2021, pp. 67—81.
6. Blue L., Abdullah H., Vargas L., Traynor P. 2ma: verifying voice commands vie two microphone authentication. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security - ASIACCS '18*; Kim J., Ahn G. J., Kim S., Kim Y., Lopez J., Kim T., Eds. ACM Press: New York, USA, 2018, pp. 89—100.
7. Feng H., Fawaz K., Shin K. G. Continuous authentication for voice assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking — MobiCom '17*; van der Merwe K., Greenstein B., Srinivasan K., Eds. ACM Press: New York, USA, 2017, pp. 343—355.
8. Kwak I. Y., Huh J. H., Han S. T., Kim I., Yoon J. Voice presentation attack detection through text-converted voice command analysis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*; Association for Computing Machinery: New York, USA, 2019, pp. 1—12.
9. Sudharsan B., Ali M. I., Corcoran P. 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, 2019.
10. Wang Y., Cai W., Gu T., Shao W., Li Y., Yu Y. Secure your voice: an oral airflow-based continuous liveness detection for voice assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2019, *3*, 1—28.
11. Zhang L., Tan S., Yang J. Hearing your voice is not enough. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security - CCS '17*; Thuraisingham B., Evans D., Malkin T., Xu D., Eds. ACM Press: New York, USA, 2017, pp. 57—71.
12. Ponticello A., Fassl M., Krombholz K. 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, 2021, pp. 475—492.
13. Braz C., Robert J. M. Security and usability: the case of the user authentication methods. In *Proceedings of the 18th Conference on l'Interaction Homme-Machine, IHM '06*; Association for Computing Machinery: New York, USA, 2006, pp. 199—203.
14. Google. *Link Your Voice to Your Google Assistant Device with Voice Match — Android — Google Assistant Help*, 2019. https://support.google.com/assistant/answer/9071681 (accessed Nov 01, 2022).
15. Chang Y. T., Marc D. My voiceprint is my authenticator: a two-layer authentication approach using voiceprint for voice assistants. In *2019 IEEE SmartWorld, editor, Conference: 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*; IEEE: Leicester, England, 2019.
16. Rao V. S., Song T. Biometric Enabled Proximity-based User Authentication. In *Technical Disclosure Commons, (July 09)*, 2018. https://www.tdcommons.org/dpubs_series/1296.
17. Meva D. T., Kumbharana C. K. Comparative study of different fusion techniques in multimodal biometric authentication. In *International Journal of Computer Applications (0975 — 8887)*; Foundation of Computer Science (FCS): New York, USA, 2013, pp. 16—19.
18. Liu R., Cornelius C., Rawassizadeh R., Peterson R., Kotz D. Vocal resonance. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2018, *2*, 1—23.

19. Lee Y., Zhao Y., Zeng J., Lee K., Zhang N., Hossain Shezan F., Tian Y., Chen K., Wang X. F. Using sonar for liveness detection to protect smart speakers against remote attackers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2020, *4*, 1–28.
20. Zhang S., Das A. Handlock: enabling 2-fa for smart home voice assistants using inaudible acoustic signal. In *Proceedings of the 24th International Symposium on Research in Attacks, Intrusions and Defenses, RAID '21*; Association for Computing Machinery: New York, USA, 2021, pp. 251–265.
21. Vassinen R. The rise of conversational commerce: what brands need to know. *J. Brand Strategy* 2018, *7*, 13–22.
22. U.S. Bank. *How U.S. Bank Aims to Shape the Future of Voice Banking*, 2018. https://www.usbank.com/newsroom/stories/new-technology-you-can-bank-on.html (accessed Nov 01, 2022).
23. Collier Z. *The Story of the Capital One Alexa Skill*, 2016. https://developer.amazon.com/blogs/alexa/post/c70e3a9b-405c-4fe1-bc20-bc0519d48c97/the-story-of-the-capital-one-alexa-skill (accessed Nov 01, 2022).
24. Sparkasse. *Sparkasse Banking-App*, 2019. https://www.sparkasse.de/unsere-loesungen/privatkunden/rund-ums-konto/voice-banking/voice-banking-sparkasse.html (accessed Nov 01, 2022).
25. Futurae. *Futurae: Strong Authentication (2fa) and App Security*, 2022. https://www.futurae.com/smart-assistant/ (accessed Nov 01, 2022).
26. Yan C., Ji X., Wang K., Jiang Q., Jin Z., Xu W. A survey on voice assistant security: attacks and countermeasures. *ACM Comput. Surv.* 2022, *55*, 1–36.
27. Lei X., Tu G. H., Liu A. X., Li C. Y., Xie T. The insecurity of home digital voice assistants - vulnerabilities, attacks and countermeasures. In *2018 IEEE Conference on Communications and Network Security (CNS)*; IEEE: Beijing, China, 2018; pp. 1–9.
28. Roy N., Hassanieh H., Choudhury R. R. Backdoor: sounds that a microphone can record, but that humans can't hear. *GetMobile: Mobile Comp. and Comm.* 2018, *21*, 25–29.
29. Zhang G., Yan C., Ji X., Zhang T., Zhang T., Xu W. Dolphinattack: inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS '17*; Association for Computing Machinery: New York, USA, 2017, pp. 103–117.
30. Kumar D., Paccagnella R., Paul M., Hennenfent E., Mason J., Adam B., Bailey M. Skill squatting attacks on amazon alexa. In *Proceedings of the 27th USENIX Conference on Security Symposium, SEC'18*; USENIX Association: Berkeley, CA, USA, 2018, pp. 33–47.
31. Zhang N., Mi X., Feng X., Wang X. F., Tian Y., Qian F. 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, 2019.
32. Microsoft. *Microsoft Authenticator – Apps on Google Play, 2021*. https://play.google.com/store/apps/details?id=com.azure.authenticator, 04/01/2020 06:23:54 (accessed Nov 01, 2022).
33. Google. *Google authenticator – apps on google play*, 2021. https://play.google.com/store/apps/details?id=com.google.android.apps.authenticator2 (accessed Nov 01, 2022).
34. Johnson D. R., Creech J. C. Ordinal measures in multiple indicator models: a simulation study of categorization error. *Am. Socio. Rev.* 1983, *48*, 398.
35. Braun V., Clarke V. Thematic analysis. In *APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological*; American Psychological Association: Washington, DC, 2012, pp. 57–71.
36. Dietrich A., Rey R., Rommel H., Rüesch S. *Wie nutzen Schweizerinnen und Schweizer das E-Banking und Mobile Banking?* 2020. https://hub.hslu.ch/retailbanking/wie-nutzen-schweizerinnen-und-schweizer-das-e-banking-und-mobile-banking/ (accessed Nov 18, 2022).
37. Abdi N., Ramokapane K. M., Such J. M. More than smart speakers: security and privacy perceptions of smart home personal assistants. In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security, SOUPS'19*; USENIX Association: USA, 2019, pp. 451–466.
38. Yao Y., Reed Basdeo J., Mcdonough O. R., Wang Y. Privacy perceptions and designs of bystanders in smart homes. *Proc. ACM Hum.-Comput. Interact.* 2019, *3*, 1–24.
39. Baldauf M., Bösch R., Frei C., Hautle F., Jenny M. 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, MobileHCI '18*; Association for Computing Machinery: New York, USA, 2018, pp. 119–126.
40. Renz A., Baldauf M., Maier E., Alt F. Alexa, it's me! an online survey on the user experience of smart speaker authentication. In *Proceedings of Mensch Und Computer 2022, MuC '22*; Association for Computing Machinery: New York, USA, 2022, pp. 14–24.