Jacob Leon Kröger*, Leon Gellrich, Sebastian Pape, Saba Rebecca Brause, and Stefan Ullrich

Personal information inference from voice recordings: User awareness and privacy concerns

Abstract: Through voice characteristics and manner of expression, even seemingly benign voice recordings can reveal sensitive attributes about a recorded speaker (e. g., geographical origin, health status, personality). We conducted a nationally representative survey in the UK (n = 683, 18–69 years) to investigate people's awareness about the inferential power of voice and speech analysis. Our results show that – while awareness levels vary between different categories of inferred information – there is generally low awareness across all participant demographics, even among participants with professional experience in computer science, data mining, and IT security. For instance, only 18.7% of participants are at least somewhat aware that physical and mental health information can be inferred from voice recordings. Many participants have rarely (28.4%) or never (42.5%) even thought about the possibility of personal information being inferred from speech data. After a short educational video on the topic, participants express only moderate privacy concern. However, based on an analysis of open text responses, unconcerned reactions seem to be largely explained by knowledge gaps about possible data misuses. Watching the educational video lowered participants' intention to use voice-enabled devices. In discussing the regulatory implications of our findings, we challenge the notion of "informed consent" to data processing. We also argue that inferences about individuals need to be legally recognized as personal data and protected accordingly.

Keywords: privacy, voice recording, speech, microphone, voice assistant, smart speaker, inference attack

DOI 10.2478/popets-2022-0002
Received 2021-05-31; revised 2021-09-15; accepted 2021-09-16.

***Corresponding Author: Jacob Leon Kröger:** Weizenbaum Institute for the Networked Society, Technische Universität Berlin, Germany, E-mail: kroeger@tu-berlin.de
Leon Gellrich: Universität Potsdam, Germany
Sebastian Pape: Goethe Universität, Frankfurt, Germany
Saba Rebecca Brause, Stefan Ullrich: Weizenbaum Institute for the Networked Society, TU Berlin, Germany

1 Introduction

Microphones are found everywhere in today's technology-based society. They are embedded not only into phones, tablets, laptops, electronic toys, cameras, wearables, and car dashboards but also into intercoms, baby monitors, remote controls, smart speakers, and all sorts of other smart home devices. Recordings from these omnipresent microphones, which contain voice commands, memos, and private conversations, are often accessible to a range of different parties. For example, microphones in mobile devices are regularly accessed by device manufacturers, platform providers, installed apps, and even by third-party software components completely invisible to the ordinary user [75]. Private calls and voice messages can, if not properly encrypted, be intercepted by instant messaging providers, network operators, videoconferencing services, and other intermediaries [54, 81, 100]. Similarly, audio data uploaded to a cloud storage system, social networking site, or media-sharing platform can be accessible not only to the audience intended by the user but also to the respective service and/or infrastructure provider [94]. More recently, the soaring popularity of voice-enabled devices [77] and the COVID-19 pandemic with increased rates of virtual meetings and voice/video calls [20] have substantially added to the volume of speech data available to corporations.

Beyond their legitimate processing purposes, these organizations may use personal information extracted from voice recordings for malicious ends or pass it on to other parties with unknown intentions, potentially exposing users to the risk of discrimination, invasive advertising, extortion, identity theft, and other types of fraud and abuse [15, 22]. As shown in recent work, attackers can build a model of a victim's voice using only a limited number of voice samples in order to fool voice-based user authentication systems or to mimic the victim in speech contexts (e. g., leaving fake voice messages or posting fake statements on the Internet) [66]. The use of speech data for unauthorized and unexpected purposes is not limited to hackers and organized crime groups, but also practiced by government agencies [81, 100] and technol-

ogy companies, including major multinational corporations [54, 88].

There is extensive research on the privacy implications of microphone-equipped devices, with numerous studies looking into user perceptions and concerns (cf. Sect. 2.1). However, existing work in this field has almost exclusively focused on linguistic content, i. e., *what* a person says in a voice recording. An equally serious, yet largely neglected, privacy threat is posed by the fact that sensitive personal information can be inferred from *how* a person speaks. In fact, certain patterns and correlations in recorded speech can be much more revealing than the words themselves (cf. Sect. 2.2). As we show in a recent literature review [41], a speaker's voice characteristics and manner of expression can implicitly contain information about his or her geographical origin, age, personality, emotions, level of sleepiness, physical and mental health condition, and more. So far, little is known about users' perceptions of the risks associated with these possible inferences (cf. Sect. 2.3).

Contribution. To examine privacy concerns related to voice recordings, we conducted a survey of 683 Internet users in the UK. Our survey is, to our knowledge, the first to focus specifically on the privacy impacts of audio-based inferences.

- Our results show a widespread lack of awareness about the inferential power of voice and speech analysis, with varying levels of awareness for different types of inferences. (Sect. 5.1)
- While we observe differences in awareness across demographic groups, these differences are not large. Even participants with professional experience in the ICT field score low on awareness. (Sect. 5.2)
- Participants' reactions to audio-based inferences are quite evenly distributed between worried and unworried. An analysis of open text responses offers insight into their reasoning. (Sect. 5.3)
- Participants' intention to use voice-controlled virtual assistants significantly decreased after a short educational video on the topic. (Sect. 5.4)

Outline. The paper proceeds as follows. First, Sect. 2 reviews related literature. Then, we describe our research goals (Sect. 3) and methodology (Sect. 4). In Sect. 5, we present the study results, a discussion of which follows in Sect. 6. We reflect on the limitations of our study in Sect. 7, before we conclude the paper in Sect. 8.

2 Related work

2.1 Privacy perceptions and concerns about audio recording

There are a variety of studies on users' perceptions regarding the privacy impacts of microphone-equipped devices. Aspects that have been investigated include people's mental models for a privacy/utility trade-off [12, 50, 52], concerns about always-listening devices [12, 13, 35, 52, 55, 102], user trust in providers of voice-based services [50, 52], concerns about voice data being shared with third parties or used for other purposes than stated [60, 91], and concerns around microphone-equipped devices getting hacked and other forms of unauthorized access [52, 55, 102].

Studies have also investigated how the context of use affects the acceptability of audio recordings (e. g., at work [35] or in public [64]). Moorthy and Vu [64], for example, investigated the usage of voice-controlled virtual assistants and found that participants preferred to use such devices in private locations to avoid drawing embarrassing attention and being overheard by strangers.

Another line of research focuses on users' awareness and use of privacy controls for microphone-equipped devices, and the willingness to pay for security and privacy features [21, 50, 51, 60, 62]. Recent studies suggest that users are poorly informed about potential privacy risks and therefore not particularly concerned about embedded microphones in their devices [50, 51, 55, 61, 102].

2.2 Sensor-based inference of personal information

It has long been known that sensor data from consumer devices can be analyzed to extrapolate patterns and draw sensitive inferences about the user. The mining of data to illegitimately gain knowledge about a person is referred to as an "inference attack" [47]. Some published review articles summarize data categories that can be inferred from IoT and mobile sensor data [37, 87, 90], video game data [45], accelerometer data [44], and eye tracking data [40, 53].

In a recent literature review, we have examined the wealth of information that can be extracted from voice recordings [41]. Through the lens of advanced data analysis, the voice and manner of expression of a recorded speaker may reveal information about his or her geo-

graphical origin [5, 32], gender[1] and age [34, 83], mental health [4, 73], physical health [17, 18], level of intoxication [9, 86], moods and emotions [36, 95], sleepiness and fatigue [17, 86], and personality traits [78, 85].

For example, researchers have used speech errors and irregularities (e. g., number of false and unintelligible words, interrupts, hesitations) and rhythmicity features for distinguishing alcoholized from non-alcoholized speech [9, 86], used voice hoarseness and sounds like coughs and sniffles for the detection of sore throats and flu infections [18, 33], and used voice pitch variations and speech energy levels for automatic emotion recognition (e. g., anger, compassion, disgust, happiness, surprise) [33, 36, 95]. A large variety of features, including speaking rate, loudness, spectral features and characteristics of linguistic expression, has been applied for the inference of personality traits [78, 85].

While such algorithmic predictions are of course not always correct and can also be significantly impaired by situational factors (e. g., ambient noise, reverberation, microphone quality), remarkable accuracies have already been reported. Polzehl [78], for instance, reached 85.2% accuracy in classifying speakers into ten different personality classes along the Big Five traits (openness to experience, conscientiousness, extraversion, agreeableness, neuroticism). From voice energy features, Sadjadi et al. [83] estimated the age of speakers with a mean absolute error of 4.7 years. In computational paralinguistics challenges, researchers have achieved up to 91% accuracy in speech-based intoxication detection [86]. And, while mental health insights can also be derived from acoustic characteristics (e. g., monotone speech, glottal features) [73], Bedi et al. [4] analyzed semantic coherence and speech complexity to automatically detect emergent psychosis in test subjects, reaching the classification accuracy of traditional clinical interviews.

Findings from such experimental studies usually refer to specific experimental setups and demographic groups (e. g., "prediction of major depression in adolescents" [73]) and are subject to limitations (e. g., laboratory conditions, limited sample size). Therefore, while providing evidence that speech-based inferences in their respective field are possible in principle, their specific findings are not necessarily generalizable to all people and all real-life situations.

It should be noted, on the other hand, that data controllers with access to speech data (e. g., large tech corporations) can be much better equipped in terms of budget, technical expertise and training data than the researchers cited above, meaning that the risk of undesired inferences from audio data is likely bigger in real life than it appears based on published results. Since data analysis methods are often subject to non-disclosure agreements, the most advanced know-how in voice and speech analysis arguably rests within the industry and is not publicly available [41]. It can also be assumed that the variety and effectiveness of audio-based inference attacks will further increase with the rising popularity of voice-enabled devices [77] and continuing advances in computing technologies and audio analysis methods (e. g., feature optimization [9] and deep learning approaches [17]). Therefore, despite the remaining technological challenges and limitations, such attacks pose a real and growing threat to consumer privacy that needs to be taken seriously and thoroughly investigated.

Existing products and features, such as voice-analytics tools for hiring assessment [96], call centers [57] and for illness detection based on smart speaker voice commands [33], indicate that companies intend to use speech data to draw sensitive inferences about users in practice.

2.3 User awareness and perceptions about sensor-based inference attacks

There is little research on users' knowledge of the inferential power of mobile and IoT sensor data. A few previous studies have investigated people's privacy concerns associated with inferences that can be drawn from physiological sensors (e. g., ECG and respiration) [80], smartphone motion sensors [16, 61], in-home sensing applications [14, 103], and IoT systems for companies (e. g., sensors for room occupancy monitoring and energy consumption tracking) [79]. Existing research indicates a widespread lack of awareness about sensor-based inference attacks [16, 61, 103]. Mehrnezhad et al. [61] found that smartphone users are not aware that various mobile sensors can be exploited to infer a personal identification number (PIN) typed on the touchscreen. In general, the researchers observed very low levels of understanding about the existence and functioning of embedded sensors. When presented with examples of personal data inference, users' privacy concerns tend

1 Note that people's personal experience and internal understanding of their gender can vary from the gender socially assigned to them based on reproductive organs, anatomy, chromosomes, etc. Most existing work in the field under investigation focuses on inferring people's self-reported gender.

to increase [16, 61, 80]. However, among the participants surveyed by Crager et al. [16], perceptions about inference attacks significantly varied between different demographic groups. In contrast to our study, none of these research efforts focuses specifically on audio data. And in existing studies on privacy concerns about voice recordings, threats posed by inferential analytics have been largely overlooked [50, 51, 55, 64].

3 Research goals

When the words spoken in a voice recording directly contain private information (e. g., expression of political opinion, credit card details, health-related information), the sensitivity of the recording is obvious. What has not been sufficiently explored in previous research, however, is people's awareness of the wealth of information that can be inferred from a speaker's voice characteristics and manner of expression (cf. Sect. 2). To fill the identified research gap, this study examines users' awareness and concerns regarding audio-based inference attacks, the vulnerability of different demographic groups to this privacy threat, and whether informing people on this issue changes their intentions towards using microphone-equipped devices. In examining these issues, we focus on eight types of audio-based inferences that have emerged from our literature search (cf. Sect. 2.2). For analysis, we group these inferences into three clusters: inferences about demographics (**DEM**), short- and medium-term states (**STATE**), and physical and psychological traits (**TRAIT**). DEM includes geographical origin, gender and age. STATE includes level of intoxication, moods and emotions, sleepiness and fatigue. TRAIT includes mental health, physical health, and personality traits. Note that this paper makes no claim to be exhaustive. For instance, inferences could also be drawn from background sounds in voice recordings (e. g., media, urban or animal sounds) [92] or ultrasonic signals [43], which can be privacy-sensitive but were not included as they do not directly relate to users' voice and speech characteristics. We pose the following research questions:

RQ-1: How aware are people that personal information can be inferred from voice recordings?

Since humans tend to perform quite well in estimating certain speaker attributes based on voice features in everyday life (e. g., age, gender, emotions, intoxication) [49, 65, 84], they may better understand the possibility of automated inferences in these areas (compared to diagnosing a specific mental disorder based on speech features, for example). Human perception of voice identity is particularly related to a speaker's age and gender [6, 48, 65]. Thus, we postulate:

H1: The level of awareness is higher for DEM inferences than for STATE inferences, and lowest for TRAIT inferences.

RQ-2: How does the level of awareness differ across demographic groups?

This question aims to identify at-risk populations by correlating awareness levels with participant demographics. All sorts of domain-related knowledge, technical understanding and privacy experience could assist in understanding the possibilities of modern data analytics. Advanced age, on the other hand, has been associated with lower degrees of ICT literacy [72]. We also explore the relationship between awareness and participants' income, gender, and disposition to value privacy. For lack of clear indications in the literature, these were tested without directional hypothesis. We postulate:

H2.1: Awareness is positively correlated with with previous privacy experience, general privacy awareness, innovativeness (i. e., a person's tendency to be a technology pioneer), general level of education, and with professional experience in data protection law, computer science, data mining, and IT security.

H2.2: Awareness and age are negatively correlated.

H2.3: There are relationships between awareness and income, gender, and disposition to value privacy.

RQ-3: What concerns do people have about the inference of personal information from voice recordings?

Physical and psychological traits are more stable over time and may thus reveal more about a person's life and character than temporary state variables. By contrast, basic demographic information is widely perceived as relatively non-sensitive [63] and is often already entered during sign-up to a digital service. Thus, we postulate:

H3: The level of concern is higher for TRAIT inferences than for STATE inferences, and lowest for DEM inferences.

RQ-4: How do people's usage intentions for voice-enabled devices change when being informed on the topic?

For this question, while audio data can be recorded with any kind of microphone-equipped device, we decided to focus on voice-controlled virtual assistants (VCVAs). This choice was made with regard to the soaring popularity of services like Amazon Alexa and Apple's Siri across the globe [77] and because user voice commands are usually available to the service providers in unencrypted form [54, 88], enabling them to screen the audio for revealing patterns and correlations. Based on previous research suggesting that privacy concerns are an important factor affecting the adoption of voice-enabled devices [29, 64], we postulate:

H4: The educational intervention will have a negative impact on VCVA usage intention.

4 Research methodology

Our four research questions were investigated by means of an online survey. After a brief overview of our study design, this section will provide detailed descriptions of our survey instrument (Sect. 4.1), participant recruitment process (Sect. 4.2), characteristics of our sample (Sect. 4.3) and methods used for data analysis (Sect. 4.4).

Participants' awareness that personal information can be inferred from voice recordings (RQ-1) was studied based on self-reported measures. While alternative approaches exist (cf. Sect. 7), we asked about participants' awareness after showing them a short educational video[2] on the topic, as inspired by Crager et al. [16]. To identify potential at-risk populations with particularly low levels of awareness (RQ-2), we included multiple demographic items in the survey and then correlated the results with participants' reported levels of awareness. To explore participants' concerns about personal information inference from voice recordings (RQ-3), we queried their reactions to the educational video through rating scales and one open text question.

To be able to test whether our video had an effect on participants' interest in using voice-enabled devices (RQ-4), we decided to use two slightly different questionnaires in our study (**Grp-A** and **Grp-B**). In Grp-A, participants were asked about their usage intention twice – once before, and once after the educational video, thus allowing a within-subject comparison to examine the effect of the intervention. However, repeating one question within a questionnaire may introduce a bias: Participants' answers to the repeated question could be influenced by their previous response to the same question. Therefore, to create a control group, participants in Grp-B were asked about their usage intention only once (after the video). By comparing results from the post-intervention question in Grp-A and Grp-B using a Kolmogorov–Smirnov test, we checked whether repeating the question in Grp-A substantially affected participants' responses. Besides underpinning the validity of the within-subject comparison among Grp-A participants, this approach allowed us to conduct a between-subject comparison between the pre-intervention question in Grp-A and the post-intervention question in Grp-B, providing additional insight into the impact of our educational video on participants' intention to use voice-enabled devices.

4.1 Survey instrument

Both questionnaires, Grp-A and Grp-B, consist of one educational video[2] and 53 questions, including three attention checks. The video clip and all questions are exactly the same in Grp-A and Grp-B – only one question is repeated in Grp-A, as will be detailed below. The questionnaires were programmed with the software SoSci Survey (version 3.2.19) [93]. All responses capturing the intensity of feelings or level of agreement were measured on 5-point Likert-type scales. To allow for reproducibility, a copy of all survey items can be found in appendix A. It took participants a median of 8.5 minutes to complete our survey. The questionnaires contained the following:

- **9 privacy demographic items:** Using validated scales directly adapted from Xu et al. [101], participants were asked about their general privacy awareness (PA), disposition to value privacy (DVP), and previous privacy experience (PPE). Results from these items were used in answering RQ-2.
- **2 items on voice-controlled virtual assistants (VCVA):** To ensure a common level of understanding among participants, this section was started with a short textual definition of VCVA, including examples of common features. Participants were

2 Video clip available here: https://youtu.be/Gr22YqS1_VA.

then asked (i) how often they use a VCVA in daily life, and (ii) to what extent they are interested in starting or continuing to use a VCVA. As explained in the beginning of Sect. 4, the position of question (ii) varied between Grp-A and Grp-B for the purpose of inter-group comparison and bias control. In questionnaire Grp-A, the question was posed before *and* after the educational video, in Grp-B it was posed *only once* after the video. Items from this block were used in answering RQ-4.

– **1 educational video[2]:** As a preparation for questions on this topic, participants were presented with a short informational video (1:44 minutes) about audio-based inferences. The video explains that, for certain functions, microphone-equipped devices typically transmit voice recordings to remote company servers, where the audio data can be analyzed to extract various kinds of personal information. Based on previous research (cf. Sect. 2.2), the video lists categories of data that could be derived from a speaker's voice characteristics and manner of expression, namely *geographical origin*, *gender and age*, *mental health*, *physical health*, *level of intoxication*, *moods and emotions*, *sleepiness and fatigue*, and *personality traits*. To ensure that the audio track is audible and the video is watched through to the end, we included a preliminary sound check and displayed a code at the end of the video which was requested on the following page. If the code was not entered correctly, the survey was terminated and the corresponding participant was excluded from analysis.
– **20 items on awareness and concerns regarding audio-based inference attacks:** After the video, the participants were asked (i) whether they had been aware that such inferences are possible, (ii) how concerned they are about the possibility of such inferences, (iii) how often they have consciously thought about this issue before, (iv) how common they think it is for companies to draw such inferences from voice recordings, and (v) how concerned they are about individual categories of inferred information. Items (i) and (iii) were used in answering RQ-1, the other items were used in answering RQ-3.
– **10 technology demographic items:** Adapting a 9-item scale from Parasuraman and Colby [76], this section measures the participants' level of innovativeness (INNO), i. e., the tendency to be a technology pioneer. Then, the participants were asked to select from a list all types of microphone-equipped devices they own. INNO was used in answering RQ-2, the other question was used to provide descriptive sample statistics.
– **10 items on basic demographics and professional experience:** Participants were queried for their age, gender, net income, and level of education. Also, they were asked to specify their level of professional experience in the areas of data protection law, computer science, data mining, and IT security. These items were used in answering RQ-2.

Three attention checks were incorporated in the survey to screen for random responders and potential bot submissions (cf. questions 8, 16, 21 in appendix A). Before the actual online survey was conducted, we administered a pretest to a total of 58 participants using the crowdsourcing platform *Amazon Mechnical Turk* (https://www.mturk.com/). In this way, we were able to test our attention checks and refine the survey instruments, including a clarification of potentially ambiguous wording. Based on the pretest results, there were only minor adjustments.

Our survey instruments and research procedures were approved by the Ethics Committee at Goethe University Frankfurt.

4.2 Participant recruitment

To access a sample of UK adults, we used the services of the online market research firm *respondi AG* (https://www.respondi.com/EN/) which was carefully selected from a list of ten competing panel providers and fulfils the quality management system standards of ISO 20252 [26]. Although crowdsourcing platforms, such as *MTurk* and *Prolific*, offer several benefits in terms of cost efficiency, speed, and flexibility, we favored the option of hiring a panel company for several reasons. Above all, while recent studies have obtained high-quality results from crowdsourced samples [58, 82], there are widespread concerns about generalizability [11, 74, 98]. Significant differences between *MTurk* workers and general population estimates were found in family composition, political attitudes, and religiosity [11], level of education and health behavior [98], social engagement [58], and internet activity [82], to name a few examples.

According to our requirements, the sample for our study was designed to approximate the age and gender distribution of adults (18-69 years) from the latest UK census [24], which also resembles current population es-

Table 1. Participant demographics

Age group	Grp-A (n = 349) male	Grp-A (n = 349) female	Grp-B (n = 334) male	Grp-B (n = 334) female
18-29	39 (11.2%)	40 (11.5%)	39 (11.7%)	40 (12.0%)
30-39	35 (10.0%)	37 (10.6%)	32 (9.6%)	33 (9.9%)
40-49	34 (9.7%)	41 (11.7%)	36 (10.8%)	39 (11.7%)
50-59	33 (9.5%)	30 (8.6%)	31 (9.3%)	30 (9.0%)
60-69	32 (9.2%)	28 (8.0%)	26 (7.8%)	28 (8.4%)
Total	173 (49.6%)	176 (50.4%)	164 (49.1%)	170 (50.9%)

timates from the UK Office for National Statistics [25]. Given a desired power of 95% and an estimated effect size of 0.3, an a priori power analysis revealed a required sample size of around 200 participants. Considering the explorative nature of the study and our available resources, we collected valid responses from n = 683 participants. Survey completers received a small compensation according to the terms of our panel provider. The survey was conducted between June 4 and July 1, 2020.

4.3 Sample characteristics

In total, 1.277 participants signed up for the survey. 588 responses were excluded for being incomplete, either because the participant had closed the questionnaire before answering all survey questions (n=235), or because the participant had failed to pass one of our attention checks (n=353). Additionally, six responses were eliminated due to obvious poor quality of their data, as assessed by independent raters. Our analysis is based on the remaining final sample of 683 participants. The age of participants ranges from 18 to 69 years ($\mu = 42.99$, $\sigma = 14.50$) with 50.7% being females. A breakdown of the age and gender distribution for both test groups is provided in Table 1. 99% of participants report to own at least one microphone-equipped device (95% smartphone, 79% laptop, 54% tablet, 36% smart speaker, 20% voice-enabled remote control, 14% in-vehicle voice control interface, 13% smartwatch). All participants are UK residents.

4.4 Data analysis

Statistical analysis. While all scales in our questionnaire are treated as parametric, we expected – due to the nature of the subject and based on existing literature – that results throughout the survey would be highly skewed (e. g., because related work indicates low awareness levels for sensor-based inference attacks, cf. Sect. 2.3). After visually checking the histograms, Shapiro-Wilk tests confirmed that the survey results for the used scales are not normally distributed ($p < 0.001$). Thus, we used non-parametric tests for comparative analyses.

The Friedman test [23, p. 686ff] was used as a non-parametric alternative to a repeated-measures ANOVA. To test the difference between means of dependent variables, post-hoc Wilcoxon signed-rank tests with Bonferroni-corrected alpha [23, p. 914] were used as a non-parametric alternative to paired t-tests. To test the difference between means of *independent* variables, a Wilcoxon rank-sum test[3] [23, p.655ff] was used as a non-parametric alternative to a two-sample t-test. For correlation analysis, since the commonly used Pearson correlation coefficient (Pearson's r) requires normal distribution of the sample data when attempting to establish whether the correlation coefficient is significant [23, p. 219], we used Spearman's rank correlation coefficient (Spearman's ρ) [23, p. 223ff] instead. To obtain ordinal variables suitable for analysis, the variable income was clustered and the variable education was recoded (e.g., master's degree above bachelor's degree), as shown in the published dataset [38].

We further conducted regression analyses based on the Akaike information criterion (AIC), using forward selection and backward elimination procedures. The software environment R (version 4.0.0) was used for statistical data analysis.

Qualitative thematic analysis. Responses to the open text question (cf. appendix A, № 11) were evaluated using a thematic analysis as proposed by Brown and Clarke [10], which is a method for identifying patterns of meaning within qualitative data.

After familiarizing himself with the data, a first rater systematically assigned descriptive and interpretative codes to all features in the data with potential relevance to the question posed. The resulting codebook was then used by a second rater to independently label and categorize the received responses, adding new codes where deemed appropriate. We used the Cohen's Kappa coefficient to measure inter-rater reliability. All instances of discrepancy were discussed and resolved jointly by the two raters. The assigned codes were then used to identify frequent responses and overarching themes (cf. Sect. 5.3).

3 also known as Mann–Whitney U test

5 Results

We collected n = 683 complete and valid survey responses (n = 349 for Grp-A, n = 334 for Grp-B). In terms of age and gender distribution, Grp-A and Grp-B are approximately identical, both being nationally representative for the UK population between 18 and 69 years. In this section, we analyze the survey responses with respect to the research questions introduced in Sect. 3. We have released an annotated and sanitized dataset containing our results for all participants [38].

5.1 RQ-1. How aware are people that personal information can be inferred from voice recordings?

Our results presented in Fig. 1 indicate widespread unawareness of inferences that can be drawn from voice and speech parameters. Averaged over the eight types of inferences covered in our questionnaire, 67.6% of participants reported to be "not at all" or only "slightly" aware. We observed, however, that the level of awareness strongly differs between the individual inference categories. For example, while 48.2% of participants reported to be "somewhat", "quite" or "very" aware about the possibility of inferring a speaker's gender and age based on a voice recording, this figure drops to 18.7% for the inference of physical and mental health information.

For a statistical analysis of these differences, we compared the three clusters of inferences defined in Sect. 3, namely inferences about demographics (**DEM**), short- and medium-term states (**STATE**), and physical and psychological traits (**TRAIT**). A Friedman test [23, p. 686ff] yielded significant differences in awareness levels between DEM, STATE, and TRAIT ($\chi^2 = 425.53$, $p < 0.001$, Kendall's W = 0.312). Post-hoc Wilcoxon [23, p. 667ff] signed-rank tests with Bonferroni-corrected alpha [23, p. 914] revealed that all three pair-wise comparisons are significant ($p < 0.001$).

In confirmation of hypothesis H1, the test results show that the level of awareness is higher for DEM inferences than for STATE inferences, and lowest for TRAIT inferences. We obtained a moderate effect size for the DEM-STATE comparison (0.342) and large effect sizes for the STATE-TRAIT (0.520) and DEM-TRAIT (0.707) comparisons. Post-hoc power analysis revealed that these tests had a very high power (> 99%).

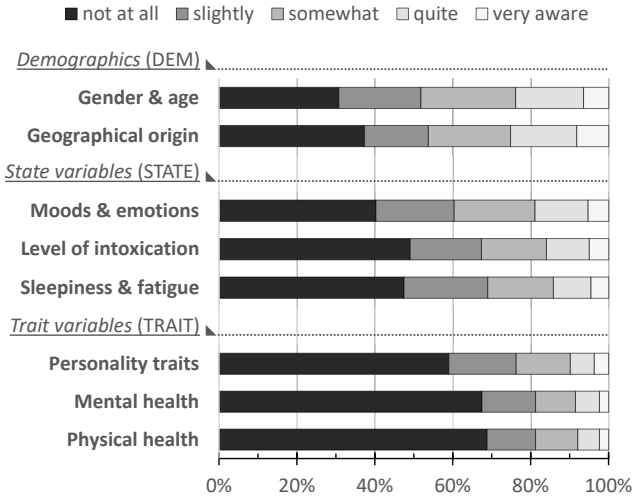

Fig. 1. Distribution of participants' level of awareness dependent on the inferred information

Before taking the questionnaire, a large portion of participants has "never" (42.5%) or "rarely" (28.4%) consciously thought about the possibility of personal information being inferred from speech data. Only a small minority reports to have pondered on this issue "often" (7.0%) or "very often" (2.8%).

5.2 RQ-2. How does the level of awareness differ across demographic groups?

To explore statistical relationships between the awareness levels from RQ-1 and participant demographics, we first calculated three awareness scores for each participant: **AW_DEM** (avg. awareness about DEM inferences), **AW_STATE** (avg. awareness about STATE inferences), and **AW_TRAIT** (avg. awareness about TRAIT inferences).

We then tested for correlations between participant demographic attributes and AW_DEM, AW_STATE, and AW_TRAIT using Spearman's rank correlation coefficient (Spearman's ρ) [23, p. 223ff]. An overview of the correlation results, along with their Bonferroni-corrected significance levels, is provided in Table 6 in appendix B. Note that for all correlations for which we had a directional hypothesis (cf. Sect. 3), the confidence intervals are one-sided, meaning they end at 1.00 or –1.00, respectively.

Supporting our hypotheses H2.1 and H2.2, AW_DEM, AW_STATE, and AW_TRAIT are negatively correlated with participant age and positively cor-

Table 2. Regression results for AW_DEM

Coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.72	0.34	5.15	3.85e-07
age	–0.01	0.00	–2.73	0.00653
gender	–0.30	0.11	–2.67	0.00798
EXP_DM	0.23	0.09	2.70	0.00712
EXP_CS	0.17	0.07	2.54	0.01132
INNO	0.13	0.07	1.97	0.04914
PA	0.11	0.07	1.41	0.15857

Adjusted R^2: 0.1532

Table 3. Regression results for AW_STATE

Coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.62	0.26	2.52	0.01208
EXP_DM	0.17	0.08	2.11	0.03482
EXP_CS	0.16	0.06	2.79	0.00558
PA	0.28	0.08	3.60	0.00036
DVP	–0.12	0.06	–1.90	0.05805
PPE	0.13	0.08	2.11	0.03546

Adjusted R^2: 0.1499

Table 4. Regression results for AW_TRAIT

Coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.58	0.27	2.19	0.02912
age	–0.01	0.00	–1.98	0.04854
EXP_DM	0.25	0.07	3.62	0.00033
EXP_CS	0.14	0.05	2.87	0.00431
PA	0.15	0.06	2.56	0.01070
PPE	0.10	0.06	1.49	0.13707

Adjusted R^2: 0.1954

AW_: average level of awareness about audio-based inference of demographics (DEM) / short- and medium-term states (STATE) / phyisical and psychological traits (TRAIT); **INNO**: innovativeness; **PA**: privacy awareness; **DVP**: disposition to value privacy; **PPE**: previous privacy experience; **EXP_**: professional experience in data mining (DM) / computer science (CS)

related with general level of education, degree of innovativeness (INNO), previous privacy experience (PPE), general privacy awareness (PA), and with professional experience in data protection law (EXP_DP), computer science (EXP_CS), data mining (EXP_DM), and IT security (EXP_IS). Across AW_DEM, AW_STATE, and AW_TRAIT, the most notable correlations were found with EXP_CS, EXP_DM, and EXP_IS. However, using guidelines from Dancey and Reidy [19] to interpret the results, even these are weak correlations.

The correlation test results further indicate that men tend to have slightly higher AW_TRAIT and AW_STATE (but not AW_DEM) than women. No significant correlation was found between participants' awareness and their disposition to value privacy (DVP) or level of income, contradicting hypothesis H2.3.

While the observed tendencies suggest that some population groups are somewhat less informed than others (and thus potentially more prone to unwittingly reveal sensitive information about themselves through speech data), there seems to be little awareness regarding audio-based inferences throughout all population segments. Even among those participants who reported "much" or "very much" EXP_CS (n = 120), EXP_IS (n = 82), or EXP_DM (n = 60), a large portion (49.4%, 45.9%, and 43.1%, respectively) stated to be "not at all" or only "slightly" aware (averaged over the eight types of inferences covered in our questionnaire).

For the continuous demographic variables (INNO, PA, DVP, PPE, EXP_DP, EXP_DM, EXP_CS, EXP_IS, gender, age), we additionally conducted regression analyses based on the Akaike information criterion (AIC) to test their predictive power on awareness. After forward selection and backward elimination procedures, we obtained models for AW_DEM, AW_STATE, and AW_TRAIT, with five to six predictors each and adjusted R^2 ranging from 15% to 19.5%. The results are shown in Tables 2, 3 and 4. For instance, an increase of EXP_DM by 1 point resulted in increased AW_DEM by 0.23, AW_STATE by 0.17, and AW_TRAIT by 0.25. The only predictors that were consistently significant across the three models were EXP_CS and EXP_DM, the other predictors were less relevant and varied between the models. The low R^2 results indicate that only a small portion of the variance in awareness can be explained by the demographic variables from our survey. This supports our above conclusion that all demographic groups under investigation are similarly vulnerable to audio-based inference attacks, i.e. similarly prone to disclose more information about themselves via speech data than they expect.

Regarding the internal consistency of constructs adapted from the literature (cf. Sect. 4.1), we obtained a good or excellent Cronbach's alpha for INNO (0.93), DVP (0.85), and PA (0.81), according to interpretation guidelines provided by George and Mallery [27]. For PPE, we obtained a Cronbach's alpha of 0.66, indicating a questionable internal consistency [27]. The lower Cronbach's alpha for PPE is mainly driven by the second item of the construct (cf. appendix A, № 3) which shows relatively low consistency with the other items of

PPE. Thus, the second item should receive special attention in future uses of the construct and may require improvement and re-validation.

5.3 RQ-3. What concerns do people have about the inference of personal information from voice recordings?

After watching a short educational video on the privacy implications of voice and speech analysis (cf. Sect. 4.1), participants were asked how worried they are about the possibility of personal information inference from speech data. Responses to this question were mixed. While 38.7% of participants reported to be "not at all" or "slightly" worried, a similar proportion (40.7%) stated to be "quite" or "very" worried.

The average participant believes that it is common rather than exceptional for companies to infer personal information from voice recordings. When asked to estimate the prevalence of this practice, 12.6% selected "somewhat" or "very" uncommon, 48.0% were "undecided", and 38.4% selected "somewhat" or "very" common.

In an open text question, we asked the participants to provide a reasoning for their reported level of concern. Except for a handful of cases, all participants offered an intelligible response. The responses were evaluated by two raters using the thematic analysis method proposed by Brown and Clarke [10], as described in Sect. 4.4. The number of assigned codes per response varies between 1 and 6. The coding yielded a Cohen's Kappa of 0.83, indicating a high degree of inter-rater reliability [59]. After instances of discrepancy were discussed and resolved jointly by the two raters, the assigned codes were used to identify overarching themes. Along with the complete dataset of our study, we have released the resulting codebook, including all identified themes [38].

In the following, we will summarize our findings regarding the most salient themes, namely participants' *emotional reactions* (Sect. 5.3.1), *feared data misuses* (Sect. 5.3.2), *perceived benefits* and *inevitability* of voice-based technology (Sect. 5.3.3) as well as participants' *knowledge gap's and misconceptions* (Sect. 5.3.4). Additionally, findings regarding participants' concerns towards specific types of inferences will be presented in Sect. 5.3.5. When quoting responses, we will either state the corresponding number of participants (**Ps**) or, if one individual participant is quoted, state the respective participant ID from the dataset ($\mathbf{P_1}$ **to** $\mathbf{P_{683}}$).

5.3.1 Emotional reactions

Approximately half of the open text responses illustrate or emphasize negative feelings. For instance, the inferential power of voice and speech analysis is perceived as "alarming" (2 Ps), "frightening" (2 Ps), "unnerving" (2 Ps), "unsettling" (2 Ps), "shocking" (2 Ps), "disturbing" (3 Ps), "uncomfortable" (4 Ps), "scary" (8 Ps), "concerning" (8 Ps), "Big Brother[ish]" (12 Ps), "worrying" (14 Ps), and "intrusive" (15 Ps). Partially, the negative reactions are quite strong, showing that confronting people with this issue can elicit "a great sense of helplessness" (P_{361}).

Participants are surprised at the variety of possible inferences, stating that they "hadn't considered" (P_{535}), "didn't realise" (3 Ps), had "no idea" (2 Ps) that "all this information could be revealed [...] by a voice recording" (P_{374}). In some cases, participants even express amazement about the possibilities of modern voice and speech analysis, e. g. describing them as "fascinating" (2 Ps) or "far beyond what I dreamed" (P_{411}). Others express confusion, e. g. by stating, "I don't quite understand how they could possibly get this information from voice alone" (P_{56}). Participant P_{441} concludes with the words: "The world is a lot cleverer than we realise."

On the other hand, there are also participants who state to be completely indifferent about the the privacy implications of voice recordings. "If I like [a technology], I don't care about side effects", says P_{399} and P_{182} claims she "couldn't care less what information people have on me."

5.3.2 Feared data misuses

Participants express concern that microphone-equipped devices may collect more data than required for their functionality, and that the collected data might be used for "unrelated purposes" (P_{460}) without the user's consent or awareness. While companies usually provide some form of privacy policy, it is objected that customers "rarely read them carefully or understand their implications fully" (P_{678}). Feared types of data processing and data misuse include targeted advertising to shape "political views/consumption habits" (P_{524}), data-based discrimination by insurances and employers (12 Ps) as well as "fraud" (2 Ps) and "identity theft" (P_{260}).

Further, there is concern that information extracted from voice recordings could end up in the "wrong hands" (6 Ps) by being passed on or leaked to third

parties, such as affiliate companies, hacker groups, or governmental agencies. Opposition is not only directed against criminal data use and governmental surveillance but also explicitly against "using very personal information for commercial purposes" (P_{39}).

Additional doubts and worries are expressed over the accuracy of inference algorithms. "A lot depends on how this information is interpreted", says P_{671}. Inferences are feared to be "inaccurate and presumptive" (P_{642}) and "taken out of context" (P_{644}), leading to "assumptions being made that aren't actually true for the individual" (P_{26}).

5.3.3 Perceived benefits and inevitability of voice-based technology

Despite their concerns, some participants perceive the disclosure of sensitive personal data as a necessary trade-off for using modern technology. "Unfortunately, I feel this is just the way the world is heading", says P_{559}. Others agree: "companies have been collecting data for years" (P_{476}) and "there is little we can do about it" (P_{664}).

There are also responses specifically focusing on the beneficial uses of microphone-equipped technology, e. g., for creating "convenient products" (P_{129}), supporting "security and crime-fighting services" (P_{137}), targeting "adverts to sell me products/services that may assist" (P_{303}) or "alerting medical services if someone is in danger due to physical or mental health issues" (P_{493}). Voice control is perceived by some as "an evolutionary step in how we and our children will interact with devices" (P_{529}) which will "improve humanity" (P_{413}) and be used "for the greater good" (P_{502}). While they also see potential downsides, optimistic participants are confident that "the benefits far outweigh the negatives at the moment" (P_{610}) and that privacy loss is a "small price to pay for more convenient products" (P_{129}).

5.3.4 Knowledge gaps and misconceptions

It is striking that – although we specifically asked for their reasoning – none of the unworried participants provided a solid justification for their reported lack of privacy concern. Instead, their responses reveal potentially dangerous, yet understandable, misconceptions and false senses of security. For instance, unconvinced by our educational video, some participants do not believe that the presented audio-based inferences are technically feasible at all. While the sources and arguments compiled in Sect. 2.2 suggest otherwise, participants' disbelief in a short educational video on a complex and unfamiliar topic is of course an understandable reaction.

Other participants do not see how data extracted from voice recordings could be used against their interest. "I really don't care as I have no idea how this information could be used to my detriment", states P_{164}. And P_{447} asks: "Why would I be worried? I have nothing to hide." The nothing-to-hide argument, which was put forth by many participants, has been criticized for its narrow view on privacy and for ignoring various threats that can arise from personal data being available to malicious or negligent parties [99].

Some participants explain that they are not worried because they do not own a voice-controlled device, such as a smart speaker. As exemplified in Sect. 1 and illustrated in our educational video, audio data (e. g., voice messages, voice memos, voice calls) can be recorded, analyzed and transmitted to remote servers by a wide variety of devices – not only by voice-controlled devices. Even living entirely without microphone-equipped devices would not guarantee protection against audio-based inference attacks, as a person's voice can – intentionally or unintentionally – be recorded by other people's devices (cf. Sect. 6.2). It should be noted, however, that our video focuses on direct user-device interaction and puts a slight emphasis on voice-controlled devices to prepare participants for questions related to RQ-4, which could be a source of misunderstanding.

Finally, a few participants base their sense of security on the assumption that their data will always be stored securely and only used sparingly and responsibly. For example, they doubt that any information extracted from voice recordings "would be used to identify me personally" (P_{396}), trusting that such data "would be in an anonymous format anyway" (P_{110}), whereas in reality this is often not the case. Others express confidence that their "privacy settings do the job" (P_{211}) and that companies would not use data "negatively against me" (2 Ps), "for truly bad purposes" (P_{449}) or "in any negative ways" (P_{353}). Participant P_{654} states: "Given the [...] general consensus of privacy violations being bad for business, I don't worry too much about inappropriate use." We also received vague and confusing statements along this line, such as "It doesn't trace it back to me personally as they will never meet me" (P_{499}) or "I assume the Internet has protection in place" (P_{427}).

In reality, however, companies can clearly leak or exploit personal data in harmful ways and commonly share such data with a range of third parties (cf. Sect. 1).

In light of the above observations, unworried reactions among participants appear to be largely explained by knowledge gaps.

5.3.5 Concerns towards specific types of inferences

We also asked the participants how concerned they would be if a company used voice recordings to infer specific types of information about them without their awareness. The results are shown in Fig. 2. As can be seen at first glance, the reported level of concern considerably varies between the information categories. For instance, while 60.8% of participants reported to be "quite" or "very" concerned about the disclosure of mental health information, only 31.3% of participants showed the same level of concern about inferences on their gender and age.

For a statistical analysis of these differences, we again compared the clusters defined in Sect. 3, namely inferences about demographics (DEM), short- and medium-term states (STATE), and physical and psychological traits (TRAIT). A Friedman test yielded significant differences in concern levels between these clusters ($\chi^2 = 386.87$, $p < 0.001$, Kendall's $W = 0.283$). Post-hoc Wilcoxon signed-rank tests [23, p. 667ff] with Bonferroni-corrected alpha revealed that all three pairwise comparisons are significant ($p < 0.001$).

The test results confirm hypothesis H3 by showing that the level of concern is lowest for DEM inferences, followed by STATE inferences, and highest for TRAIT inferences. We obtained a moderate effect size for the DEM-STATE comparison (0.409) and large effect sizes for the DEM-TRAIT (0.643) and STATE-TRAIT (0.513) comparisons. Post-hoc power analysis revealed that these tests had a very high power (> 99%).

Even in their response to the open text question (cf. Sect. 5.3), some participants have focused their concern on specific data categories (e. g., P_{519}: "Health [data] isn't really something you want to be shared without consent"), while other types of inferred data, such as age, gender and level of intoxication, were rarely mentioned at all. Analogous to the knowledge gaps noted in Sect. 5.3.4, the variation in concern levels between different types of inferences could indicate a lack of awareness or understanding of how certain data categories can be misused.

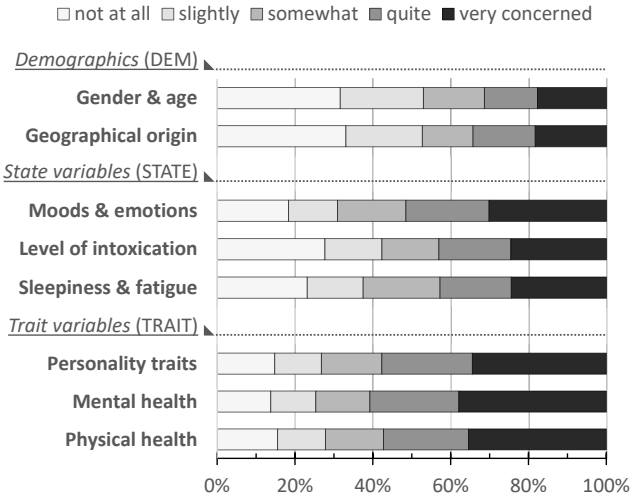

Fig. 2. Distribution of participants' level of concern dependent on the inferred information

5.4 RQ-4. How do people's usage intentions for voice-enabled devices change when being informed on the topic?

To examine whether our educational intervention had an effect on the intention to use a voice-controlled virtual assistant (VCVA), we conducted a between- and a within-subject comparison, as explained in Sect. 4.

First, we compared responses from Grp-A ($n = 349$), where participants were asked about their VCVA usage intention (VCVA_UI) before the intervention, with responses from Grp-B ($n = 334$), where the same question was asked after the intervention. While a Wilcoxon rank-sum test [23, p.655ff] showed a significant difference between VCVA_UI in Grp-A and Grp-B in the direction predicted by hypothesis H4 ($p < 0.05$), i.e. lower VCVA_UI after the intervention, it only yielded a very small effect size (0.086). Given the small effect size, post-hoc power analysis also revealed low power (29%). The slightness of the result is not too surprising, as the expected effect may have been masked by those Grp-A participants who already reported no interest in using a VCVA before the intervention and thus left no room for a reduction in interest.

We then conducted a within-subject comparison in Grp-A. Participants in this group were asked about their VCVA_UI twice, once before and once after the intervention. Of course, repeating a question within a questionnaire may introduce a bias, as noted in Sect. 4. However, a Kolmogorov–Smirnov test revealed no significant difference ($p = 0.52$) between the distribu-

tion of responses to our post-intervention questions on VCVA_UI in Grp-A and Grp-B (despite the large sample sizes), which indicates that the two samples belong to the same distribution and that repeating the question in Grp-A did not substantially affect the distribution. Thus, we proceed with the within-subject comparison.

By comparing results from the pre- and post-intervention items on VCVA_UI within Grp-A, we analyze whether the intervention had a significant effect on VCVA_UI among Grp-A participants. To avoid the masking effect described above, we excluded those participants from the analysis who reported to be "not at all" or only "slightly" interested in using a VCVA before watching the educational video, leaving n = 151 participants for the analysis (Grp-A.1). Within Grp-A.1, a Wilcoxon signed-rank test [23, p. 667ff] not only yielded a significantly decrease in VCVA_UI after the intervention ($p < 0.001$) but also a large effect size (0.590). Here, post-hoc power analysis revealed very high power (> 99%).

These results suggest that for people with a medium or high interest in using such devices, information on the privacy implications of voice and speech analysis can have a strong negative effect on usage intentions, thus supporting hypothesis H4. Further research is required to investigate causes and motivational aspects behind these observations, and to determine whether these changes in reported usage intention sustain beyond the short term and actually translate into shifts in consumption behavior.

Table 5 shows VCVA_UI before and after the intervention, ΔVCVA_UI as well as awareness and concern levels, averaged over Grp-A and Grp-A.1. It can be seen that Grp-A and Grp-A.1 yield similar results for concern and awareness, indicating that Grp-A.1 is a representative subgroup of Grp-A. The similarity further indicates that those participants who reported to be "not at all" or only "slightly" interested in using a VCVA before watching the educational video (and thus excluded in Grp-A.1) showed little interest not because they are more concerned than the rest of Grp-A but, e. g., because they simply have no use for a VCVA.

In Grp-A.1, we additionally tested for correlations between the observed change in VCVA_UI (ΔVCVA_UI) and the participants' level of concern and awareness about the possibility of audio-based inference of demographics (DEM), short- and medium-term states (STATE), and phyisical and psychological traits (TRAIT). We found that concerns are positively correlated with ΔVCVA_UI ($p < 0.001$; Spearman's $\rho = 0.38$ for DEM, 0.42 for STATE, 0.40 for TRAIT). The self-reported level of awareness about audio-based inferences was negatively correlated with ΔVCVA_UI (DEM: $p < 0.01$, Spearman's $\rho = -0.23$; STATE: $p < 0.01$, Spearman's $\rho = -0.21$; TRAIT: $p < 0.05$, Spearman's $\rho = -0.21$). This means that VCVA_UI of participants with higher levels of prior knowledge about audio-based inferences tend to be less affected by our educational intervention. This is not surprising, as the intervention should logically have a higher educational impact on those participants who knew less on the topic beforehand.

Table 5. Mean values for Grp-A and Grp-A.1

	Grp-A (n = 349)	**Grp-A.1 (n = 151)**
VCVA_UI (pre)	2.47	3.97
VCVA_UI (post)	2.11	3.20
ΔVCVA_UI	–0.36 (–14.6%)	–0.77 (–19.4%)
AW_DEM	2.47	2.50
AW_STATE	2.08	2.03
AW_TRAIT	1.60	1.68
CON_DEM	2.62	2.53
CON_STATE	3.18	3.06
CON_TRAIT	3.55	3.40

VCVA_UI: VCVA usage intention before (pre) and after (post) the intervention; **ΔVCVA_UI**: difference between VCVA_UI pre and post intervention; **AW_**: level of awareness about audio-based inference of demographics (DEM) / short- and medium-term states (STATE) / phyisical and psychological traits (TRAIT); **CON_**: level of concern about DEM / STATE / TRAIT inferences

6 Discussion and implications

The results of our survey reveal that there is widespread lack of understanding about the possibilities of voice and speech analysis. For instance, 81.3% of participants are not at all or only slightly aware that physical and mental health information can be inferred from a recorded speaker's voice characteristics and manner of expression. Only 9.8% of participants have often or very often consciously thought about the possibility of personal information being inferred from voice and speech parameters.

Our results, offering a novel contribution by specifically focusing on voice recordings, are consistent with previous findings indicating a lack of awareness about inference attacks based on other types of sensor data [16, 61, 103]. While our analysis shows that awareness varies significantly depending on participants' demographic

attributes, which was previously also shown to be the case for motion sensor-based inference attacks [16], the average level of awareness is low across all demographic groups – even among participants with professional experience in ICT (cf. Sect. 5.2).

In our sample, the degree of worry regarding audio-based inferences is quite evenly distributed between high and low. Results from analyzing open text responses suggest, however, that unconcerned reactions are largely explained by knowledge gaps about the risks that can arise from privacy intrusions (cf. Sect. 5.3.4). While some participants express worry about unauthorized data leakage to third parties, specific types of data misuse, and about being misrepresented by audio-based inferences (cf. Sect. 5.3.2), our results confirm previous findings about people's unwarranted trust in companies' data practices [50, 52, 102, 103] and a widespread nothing-to-hide mentality [50, 51, 91, 99, 103], potentially resulting in a false sense of security.

At the same time, the reported level of concern varies significantly between different categories of inferred data (e. g., high concern about inferred health information vs. low concern about inferred age – cf. Sect. 5.3.5), which may be due to individual preferences but perhaps also indicates a lack of understanding on how certain data categories can be used against the data subject's interests.

Our educational intervention on the privacy implications of voice and speech analysis had a significant negative impact on participants' intention to use voice-controlled virtual assistants. This result aligns with previous findings that users' privacy concerns tend to increase when they are presented with examples of personal data inference [16, 61, 80].

6.1 Consumer education and privacy-enhancing technologies

While Internet-connected microphones have many beneficial applications (e. g., efficient human-computer interaction, assistance for physically disabled people, smart home convenience, driver safety), their increasing ubiquity in modern life calls for a debate on potential social ramifications. Besides the already omnipresent microphones in smartphones, laptops and other mobile devices, the number of installed smart speakers is forecast to reach 640 million globally by 2024 [70]. Nothing is fundamentally wrong with either microphone-equipped devices or speech data mining, but there is clearly a need for appropriate privacy safeguards.

Educating people on existing threats is an important starting point – not only to support informed purchase decisions but also to put critical pressure on the societal actors responsible for protecting consumer privacy in sensing devices.

With regard to data collection transparency in voice-controlled devices, there has been a focus on device recording modes, such as "speech-activated", "manually activated", and "always on" [12, 28]. In the face of recurring security breaches and privacy scandals, users have not only been advised to use the mute feature of their voice-controlled devices but also been encouraged to disconnect power supply or even purposely obfuscate audio signals to protect themselves against corporate and governmental eavesdropping [12].

However, while there are good reasons to be concerned about always-listening devices, it is important to understand that the mentioned safeguards – even if effectively applied in practice – will not prevent audio-based inference attacks (unless, of course, they permanently block the microphone and prevent any recording.) As discussed in this paper, voice and speech characteristics can unexpectedly carry sensitive personal information, which may later be extracted via advanced data analytics (cf. Sect. 2.2). Thus, even if a voice assistant is only consciously unmuted by a user to ask for the weather forecast, for instance, this can already lead to unwanted information leakage (e. g., based on sociolect, accent, intonation, pitch, loudness, or a hoarseness in the user's voice).

To minimize privacy risks, voice recordings should preferably be encrypted before any upload or Internet transfer, and the data processing should take place as much as possible locally on the user's device. In cases where the disclosure of speech data to service providers is unavoidable (e. g., because the data is necessary for service functionality or due to resource constraints of the end device), measures should be taken to prevent the illegitimate inference of personal information.

Some technical approaches that could help to defend against audio-based inference attacks are differential private learning, hashing techniques for speech data, fully homomorphic inference systems, and speaker de-identification by voice transformation [67, 68]. In recent work, for example, Aloufi et al. [2, 3] have proposed privacy-preserving intermediate layers to sanitize user voice input before sharing it with cloud service providers. These approaches, which are based on the automatic identification and obfuscation of sensitive features in speech data, have yielded promising evaluation results for certain use cases, such as protection against

unwanted emotion [2, 3] and gender recognition [3] while maintaining utility of the data for speech and speaker recognition.

Where possible without compromising the required functionality, voice recordings should also be transcribed to text in order to preserve task-relevant information while removing speaking speed, rhythm, voice characteristics, etc. and thus reduce the risk of inference attacks. In their proposed *Preech* system for privacy-preserving speech transcription, Ahmed et al. [1] apply voice transformation and the injection of noise to obfuscate users' voice biometrics and thus prevent unauthorized identification and impersonation.

6.2 Regulatory implications

Considering that (i) existing technical solutions for protecting against sensor-based inference attacks have severe limitations [68, 87] and are still seen as "embryonic research topics" [69], (ii) companies obviously need strong incentives to apply privacy-enhancing technologies [30], (iii) many users are not willing to pay for privacy and their willingness to pay depends on the trust towards the provider of the privacy enhancing technology [31], and (iv) there is – as our study underscores – a very low level of risk awareness among users, adjustments in privacy regulation may be required as well.

To achieve a minimum level of transparency and oversight, inferences should at least be recognized as falling within the scope of data protection law. While the newly introduced California Consumer Privacy Act (CCPA), for example, specifically covers "inferences drawn" as part of its definition of personal information, most other data protection laws – including progressive ones, such as EU's General Data Protection Regulation (GDPR) – do not sufficiently protect individuals against undesired inferences [8]. In a detailed legal analysis, Wachter and Mittelstadt [97] state that the GDPR "focuses primarily on mechanisms to manage the input side of processing. (...) [T]he few mechanisms in European data protection law that address the outputs of processing, including inferred and derived data, profiles, and decisions, are far weaker."

Data protection law could, for example, make it mandatory for companies to provide comprehensive information on all types of inferences that they (attempt to) draw from collected personal data. Given that data mining algorithms are becoming an increasingly accurate and efficient access path to personal information, this could be a sensible measure.

For data subjects, being able to answer the question "who knows what about me?" is a necessary precondition for exercising other data protection rights (e. g., data rectification, erasure, restriction of processing) in an informed manner [39]. The widespread lack of understanding of how personal data can be collected, inferred, and misused calls into question the notion of "informed consent" and may warrant some form of paternalistic government intervention. As we argue in other recent work, people's privacy choices are typically irrational, involuntary and/or easily circumventable [42].

Accordingly, various commentators have proposed a legal shift from the individualistic paradigm of notice and consent ("privacy self-management") towards an increased focus on the ethical and social impacts of personal data use (e. g., [56, 71, 89, 97]). For instance, a general legal prohibition of using certain categories of personal data for ethically indefensible purposes based on the resulting harm potential could be helpful to protect consumers from consequences of their own unawareness.

Another argument against the self-management approach is that it ignores the various externalities that individual privacy choices have on other people and society at large [42]. In today's interconnected world, people often share personal data of other people, giving rise to the notion of "interdependent privacy" [7]. Owners of microphone-equipped devices can become amateur data controllers without the data subject's knowledge or consent. For example, someone might record a phone conversation and share it with a third party. Furthermore, a user's device can record the voice, activities, etc. of persons in the vicinity (e.g., relatives, friends, visitors, bystanders), potentially scaling up the inference problem by a significant factor.

7 Limitations

While surveys are widely used in related empirical studies [16, 61, 79, 80], this form of data collection is subject to several potential limitations.

There is of course the risk of careless or random responding. We incorporated multiple attention checks into our survey to filter out low-quality responses. Only those respondents who passed all quality and attention checks were included in the analysis (cf. Sect. 4.3).

Furthermore, a self-reported survey captures subjective perceptions, which are prone to distortion. In particular, following an approach proposed by Crager

et al. [16], we asked participants for their awareness of audio-based inferences *after* showing them a short educational video on the topic. It may have been difficult for some participants to accurately recall what their level of knowledge was prior to watching the video.

A possible alternative would be to ask participants about awareness before showing the video (e. g., by asking how likely they think different types of inferences are). Even this approach, however, may evoke thoughts that participants would not have by themselves in everyday life. Moreover, participants with low levels of knowledge and skills may have a tendency to overestimate their abilities (a cognitive bias referred to as the Dunning-Kruger effect) [46]. Therefore, future work could build upon this study by using approaches that query the knowledge of participants more implicitly and objectively, instead of using self-reported measures of awareness.

Additionally, learning from our study's limitations, follow-up studies should thoroughly test participants' understanding of educational materials (e. g., in the form of a quiz). It is possible that participants did not understand everything in the video.

It is also possible that participants exaggerated certain responses in an attempt to present themselves in a more positive light, e. g., by stating that they are more familiar with technology than they actually are, by overstating their professional experience in some area, or by falsely claiming to be less interested in using a virtual assistant after our educational intervention. This effect may have been increased by asking participants about privacy attitudes at the beginning of the survey, priming them to think about privacy. We cannot exclude the possibility of a social-desirability bias but believe to have minimized the risk of occurrence through the neutral framing of our educational video and by informing participants in advance that the results of our online survey would be completely anonymous. Asking about privacy attitudes at the end would not have eliminated the issue of priming because, in this case, the privacy focus of the remaining survey may have influenced participants' responses to these questions.

It should also be noted that our findings are only representative for the UK population, which is a typical WEIRD society (Western, Educated, Industrialized, Rich, Democratic). Replication studies in other contexts, such as in Asian or African countries, are required to establish cultural validity.

8 Conclusion

Microphones have become ubiquitous in modern life, embedded into mobile, wearable, and all sorts of smart home devices. While these devices provide useful functions, the increasing availability of private voice recordings to service providers, device manufacturers, app vendors, etc. has also become a major threat to consumer privacy. In this study, focusing on an issue that has received very little research attention to date, we investigated people's awareness and privacy concerns about the wealth of personal information that can be inferred by analyzing a recorded speaker's voice characteristics and manner of expression. Our results indicate a widespread lack of awareness about the possibilities of modern voice and speech analysis. Averaged over the eight types of inferences covered in our questionnaire, most participants reported to be "not at all" (50.0%) or only "slightly" (17.6%) aware. Even participants with professional experience in the ICT field scored low on awareness. Furthermore, while our results for participants' level of concern about audio-based inference attacks do not show a clear tendency, many participants – judging from their text responses – seem to lack the background knowledge required to assess these threats in an informed manner. Overall, the findings of this study underscore that the complexities of modern data processing are beyond the comprehension of ordinary users – which calls into question the notion of "informed consent," a cornerstone of most modern data protection laws, including EU's GDPR. To prevent consent from being used as a loophole to excessively reap data from unwitting individuals, alternative and complementary technical, organizational, and regulatory safeguards urgently need to be developed. At the very least, inferred information relating to an individual should be classified as personal data by law, subject to corresponding protections and transparency rights. Results from our within- and between-subject comparisons suggest that education on data analytics may have an impact on smart device use, the mechanisms and implications of which are an interesting avenue for future research.

9 Acknowledgments

We thank the anonymous reviewers for their constructive feedback. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Shimaa Ahmed, Amrita Roy Chowdhury, Kassem Fawaz, and Parmesh Ramanathan. 2020. Preech: A system for privacy-preserving speech transcription. In *29th USENIX Security Symposium*. 2703–2720.

[2] Ranya Aloufi, Hamed Haddadi, and David Boyle. 2019. Emotionless: privacy-preserving speech analysis for voice assistants. *preprint arXiv:1908.03632* (2019).

[3] Ranya Aloufi, Hamed Haddadi, and David Boyle. 2020. Privacy-preserving Voice Analysis via Disentangled Representations. In *ACM SIGSAC Conference on Cloud Computing Security Workshop*. 1–14.

[4] Gillinder Bedi et al. 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia* 1 (2015), 15030.

[5] Hamid Behravan, Ville Hautamäki, Sabato Marco Siniscalchi, Tomi Kinnunen, and Chin-Hui Lee. 2015. I-Vector modeling of speech attributes for automatic foreign accent recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24, 1 (2015), 29–41.

[6] Pascal Belin, Shirley Fecteau, and Catherine Bedard. 2004. Thinking the voice: neural correlates of voice perception. *Trends in cognitive sciences* 8, 3 (2004), 129–135.

[7] Gergely Biczók and Pern Hui Chia. 2013. Interdependent privacy: Let me share your data. In *Int. Conf. on Financial Cryptography and Data Security*. Springer, 338–353.

[8] Jordan M Blanke. 2020. Protection for 'Inferences Drawn': A Comparison Between the General Data Protection Regulation and the California Consumer Privacy Act. *Global Privacy Law Review* 1, 2 (2020).

[9] Daniel Bone, Ming Li, Matthew P Black, and Shrikanth S Narayanan. 2014. Intoxicated speech detection: A fusion framework with speaker-normalized hierarchical functionals and GMM supervectors. *Computer Speech & Language* 28, 2 (2014), 375–391.

[10] Virginia Braun and Victoria Clarke. 2012. Thematic Analysis. In *APA Handbook of Research Methods in Psychology*. American Psychological Association, Washington, DC.

[11] Jesse Chandler, Cheskie Rosenzweig, Aaron J Moss, Jonathan Robinson, and Leib Litman. 2019. Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior research methods* 51, 5 (2019), 2022–2038.

[12] Varun Chandrasekaran, Kassem Fawaz, Bilge Mutlu, and Suman Banerjee. 2018. Characterizing Privacy Perceptions of Voice Assistants: A Technology Probe Study. *arXiv:1812.00263 [cs]* (2018).

[13] Chola Chhetri and Vivian Genaro Motti. 2019. Eliciting Privacy Concerns for Smart Home Devices from a User Centered Perspective. In *Information in Contemporary Society*, Natalie Greene Taylor et al. (Eds.). Springer, Cham, 91–101.

[14] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly Harrison, Shwetak N. Patel, and Julie A. Kientz. 2012. Investigating Receptiveness to Sensing and Inference in the Home Using Sensor Proxies. In *UbiComp*. ACM, 61–70.

[15] Wolfie Christl. 2017. *How Companies Use Data Against People*. Cracked Labs, Vienna.

[16] Kirsten Crager, Anindya Maiti, Murtuza Jadliwala, and Jibo He. 2017. Information Leakage through Mobile Motion Sensors: User Awareness and Concerns. In *EuroUSEC*. Internet Society, Paris, France.

[17] Nicholas Cummins, Alice Baird, and Bjoern W Schuller. 2018. Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning. *Methods* 151 (2018), 41–54.

[18] Nicholas Cummins, Maximilian Schmitt, Shahin Amiriparian, Jarek Krajewski, and Björn Schuller. 2017. "You sound ill, take the day off": Automatic recognition of speech affected by upper respiratory tract infection. In *Proceedings of the IEEE EMBC Conference*. 3806–3809.

[19] Christine P Dancey and John Reidy. 2007. *Statistics without Maths for Psychology*. Pearson Education.

[20] Evan DeFilippis, Stephen Michael Impink, Madison Singell, Jeffrey T Polzer, and Raffaella Sadun. 2020. *Collaborating during Coronavirus: The impact of COVID-19 on the nature of work*. National Bureau of Economic Research, Cambridge, MA.

[21] Serge Egelman, Raghudeep Kannavara, and Richard Chow. 2015. Is This Thing On? Crowdsourcing Privacy Indicators for Ubiquitous Sensing Platforms. In *Proceedings of the CHI Conference*. ACM, New York, 1669–1678.

[22] Federal Bureau of Investigation. 2020. 2019 Internet Crime Report. https://pdf.ic3.gov/2019_IC3Report.pdf

[23] Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering statistics using R*. Sage Publishing, Newbury Park, CA.

[24] Office for National Statistics. 2013. The National Archives. https://webarchive.nationalarchives.gov.uk/20160110194058/http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-294277 (last accessed on 12 September 2021).

[25] Office for National Statistics. 2019. Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2019. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates

[26] International Organization for Standardization. 2019. ISO 20252:2019. https://www.iso.org/obp/ui/#iso:std:iso:20252:ed-3:v1:en (last accessed on 12 September 2021).

[27] D George and P Mallery. 2003. *Reliability analysis. SPSS for Windows, step by step: a simple guide and reference*. Allyn & Bacon, Boston, MA.

[28] Stacey Gray. 2016. Always On: Privacy Implications of Microphone-Enabled Devices. https://www.ftc.gov/system/files/documents/public_comments/2016/08/00003-128652.pdf

[29] Sangyeal Han and Heetae Yang. 2018. Understanding adoption of intelligent personal assistants. *Industrial Management & Data Systems* 118, 3 (2018), 618–636.

[30] David Harborth, Maren Braun, Akos Grosz, Sebastian Pape, and Kai Rannenberg. 2018. Anreize und Hemmnisse für die Implementierung von Privacy-Enhancing Technologies im Unternehmenskontext. In *Sicherheit 2018*. 29–41.

[31] David Harborth, Xinyuan Cai, and Sebastian Pape. 2019. Why Do People Pay for Privacy?. In *ICT Systems Security and Privacy Protection – 34th IFIP TC 11 International Conference*. 253–267.

[32] Abhinav Jain, Minali Upreti, and Preethi Jyothi. 2018. Improved Accented Speech Recognition Using Accent Embeddings and Multi-task Learning. In *Proc. Interspeech*. 2454–2458.

[33] Huafeng Jin and Shuo Wang. 2018. Voice-based determination of physical and emotional characteristics of users. https://patents.google.com/patent/US10096319B1/en US Patent 10,096,319.

[34] Selen Hande Kabil, Hannah Muckenhirn, et al. 2018. On Learning to Identify Genders from Raw Speech Signal Using CNNs. In *Proc. Interspeech*. 287–291.

[35] Predrag Klasnja, Sunny Consolvo, Tanzeem Choudhury, Richard Beckwith, and Jeffrey Hightower. 2009. Exploring Privacy Concerns about Personal Sensing. In *International Conference on Pervasive Computing*. Springer, 176–183.

[36] Shashidhar G Koolagudi, Sudhamay Maity, Vuppala Anil Kumar, Saswat Chakrabarti, and K Sreenivasa Rao. 2009. IITKGP-SESC: speech database for emotion analysis. In *International Conference on Contemporary Computing*. Springer, 485–492.

[37] Jacob Kröger. 2019. Unexpected Inferences from Sensor Data: A Hidden Privacy Threat in the Internet of Things. In *Internet of Things. Information Processing in an Increasingly Connected World*, Leon Strous and Vinton G. Cerf (Eds.). Springer, Cham, 147–159.

[38] Jacob Leon Kröger, Leon Gellrich, Sebastian Pape, Saba Rebecca Brause, and Stefan Ullrich. 2021. Response data - Survey on privacy impacts of voice & speech analysis. http://dx.doi.org/10.14279/depositonce-12309.2

[39] Jacob Leon Kröger, Jens Lindemann, and Dominik Herrmann. 2020. How do app vendors respond to subject access requests? A longitudinal privacy study on iOS and Android Apps. In *International Conference on Availability, Reliability and Security*. 1–10.

[40] Jacob Leon Kröger, Otto Hans-Martin Lutz, and Florian Müller. 2020. What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking. In *Privacy and Identity Management. Data for Better Living: AI and Privacy*, Samuel Fricker, Michael Friedewald, Stephan Krenn, Eva Lievens, and Melek Önen (Eds.). Springer, Cham, 226–241.

[41] Jacob Leon Kröger, Otto Hans-Martin Lutz, and Philip Raschke. 2020. Privacy Implications of Voice and Speech Analysis – Information Disclosure by Inference. In *Privacy and Identity Management. Data for Better Living: AI and Privacy*, Samuel Fricker et al. (Eds.). Springer, Cham, 242–258.

[42] Jacob Leon Kröger, Otto Hans-Martin Lutz, and Stefan Ullrich. 2021. The myth of individual control: Mapping the limitations of privacy self-management. https://ssrn.com/abstract=3881776. *SSRN* (2021).

[43] Jacob Leon Kröger and Philip Raschke. 2019. Is my phone listening in? On the feasibility and detectability of mobile eavesdropping. In *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 102–120.

[44] Jacob Leon Kröger, Philip Raschke, and Towhidur Rahman Bhuiyan. 2019. Privacy Implications of Accelerometer Data: A Review of Possible Inferences. In *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy (ICCSP)*. ACM, New York, NY, 81–87.

[45] Jacob Leon Kröger, Philip Raschke, Jessica Percy Campbell, and Stefan Ullrich. 2021. Surveilling the Gamers: Privacy Impacts of the Video Game Industry. https://ssrn.com/abstract=3881279. *SSRN* (2021).

[46] Justin Kruger and David Dunning. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77, 6 (1999), 1121.

[47] John Krumm. 2007. Inference Attacks on Location Tracks. In *Pervasive Computing*, Anthony LaMarca et al. (Eds.). Springer, Cham, 127–143.

[48] Norman J Lass, Karen R Hughes, Melanie D Bowyer, Lucille T Waters, and Victoria T Bourne. 1976. Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J. Acoust. Soc. Am.* 59, 3 (1976), 675–678.

[49] Marianne Latinus and Pascal Belin. 2011. Human voice perception. *Current Biology* 21, 4 (2011), R143–R145.

[50] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction* 2 (2018), 1–31.

[51] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. "Alexa, Stop Recording": Mismatches between Smart Speaker Privacy Controls and User Needs. In *Symposium on Usable Privacy and Security*.

[52] Yuting Liao, Jessica Vitak, Priya Kumar, Michael Zimmer, and Katherine Kritikos. 2019. Understanding the Role of Privacy and Trust in Intelligent Personal Assistant Adoption. In *Information in Contemporary Society*, Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin, and Bonnie Nardi (Eds.). Springer, Cham, 102–113.

[53] Daniel J. Liebling and Sören Preibusch. 2014. Privacy Considerations for a Pervasive Eye Tracking World. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1169–1177.

[54] Natasha Lomas. 2019. Microsoft Tweaks Privacy Policy to Admit Humans Can Listen to Skype Translator and Cortana Audio. https://social.techcrunch.com/2019/08/15/microsoft-tweaks-privacy-policy-to-admit-humans-can-listen-to-skype-translator-and-cortana-audio/ (last accessed on 12 September 2021).

[55] Lydia Manikonda, Aditya Deotale, and Subbarao Kambhampati. 2018. What's up with Privacy?: User Preferences and Privacy Concerns in Intelligent Personal Assistants. In *AAAI/ACM Conference on AI, Ethics, and Society*. 229–235.

[56] Alessandro Mantelero. 2018. AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review* 34, 4 (2018), 754–772.

[57] Rob Matheson. 2016. Watch your tone. https://news.mit.edu/2016/startup-cogito-voice-analytics-call-centers-ptsd-0120. (last accessed on 12 September 2021).

[58] Morgan N McCredie and Leslie C Morey. 2019. Who are the Turkers? A characterization of MTurk workers using the personality assessment inventory. *Assessment* 26, 5 (2019), 759–766.

[59] Mary L McHugh. 2012. Interrater Reliability: The Kappa Statistic. *Biochemia Medica* 22, 3 (2012), 276–282.

[60] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. 2017. Toys That Listen: A Study of Parents, Children, and Internet-Connected Toys. In *Proceedings of the CHI Conference*. 5197–5207.

[61] Maryam Mehrnezhad, Ehsan Toreini, Siamak F. Shahandashti, and Feng Hao. 2018. Stealing PINs via Mobile Sensors: Actual Risk versus User Perception. *International Journal of Information Security* 17, 3 (2018), 291–313.

[62] Cristina Mihale-Wilson, Jan Zibuschka, and Oliver Hinz. 2017. About User Preferences and Willingness to Pay for a Secure and Privacy Protective Ubiquitous Personal Assistant. In *Proceedings of the 25th European Conference on Information Systems (ECIS)*.

[63] George R Milne, George Pettinico, Fatima M Hajjat, and Ereni Markos. 2017. Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing. *Journal of Consumer Affairs* 51, 1 (2017), 133–161.

[64] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human-Computer Interaction* 31, 4 (2015), 307–335.

[65] Evelyne Moyse. 2014. Age estimation from faces and voices: a review. *Psychologica Belgica* 54, 3 (2014), 255–265. http://dx.doi.org/10.5334/pb.aq

[66] Dibya Mukhopadhyay, Maliheh Shirvanian, and Nitesh Saxena. 2015. All your voices are belong to us: Stealing voices to fool humans and machines. In *European Symposium on Research in Computer Security*. Springer, 599–621.

[67] Andreas Nautsch et al. 2019. Preserving privacy in speaker and speech characterisation. *Computer Speech & Language* 58 (2019), 441–480. http://dx.doi.org/10.1016/j.csl.2019.06.001

[68] Andreas Nautsch, Catherine Jasserand, Els Kindt, Massimiliano Todisco, Isabel Trancoso, and Nicholas Evans. 2019. The GDPR & Speech Data: Reflections of Legal and Technology Communities, First Steps Towards a Common Understanding. *Proc. Interspeech* (2019), 3695–3699.

[69] Andreas Nautsch, Jose Patino, Natalia Tomashenko, Junichi Yamagishi, Paul-Gauthier Noe, Jean-Francois Bonastre, Massimiliano Todisco, and Nicholas Evans. 2020. The Privacy ZEBRA: Zero Evidence Biometric Recognition Assessment. In *Proc. Interspeech*.

[70] Evan Niu. 2020. Smart-Speaker Volumes Expected to Jump Next Year. https://www.nasdaq.com/articles/smart-speaker-volumes-expected-to-jump-next-year-2020-10-23 (last accessed on 12 September 2021).

[71] Data Ethics Commission of the Federal Government. 2019. *Opinion of the Data Ethics Commission*. German Federal Ministry of Justice and Consumer Protection, Berlin.

[72] Tobias Olsson, Ulli Samuelsson, and Dino Viscovi. 2019. At risk of exclusion? Degrees of ICT access and literacy among senior citizens. *Information, Communication & Society* 22, 1 (2019), 55–72.

[73] Kuan Ee Brian Ooi, Margaret Lech, and Nicholas B Allen. 2012. Multichannel weighted speech classification system for prediction of major depression in adolescents. *IEEE Transactions on Biomedical Engineering* 60, 2 (2012), 497–506. http://dx.doi.org/10.1109/TBME.2012.2228646

[74] Yaakov Ophir, Itay Sisso, Christa SC Asterhan, Refael Tikochinski, and Roi Reichart. 2020. The Turker Blues: Hidden Factors Behind Increased Depression Rates Among Amazon's Mechanical Turkers. *Clinical Psychological Science* 8, 1 (2020), 65–83.

[75] Elleen Pan, Jingjing Ren, Martina Lindorfer, Christo Wilson, and David Chones. 2018. Panoptispy: Characterizing Audio and Video Exfiltration from Android Applications. In *Proceedings on Privacy Enhancing Technologies*. 33–50.

[76] A Parasuraman and Charles L Colby. 2015. An Updated and Streamlined Technology Readiness Index: TRI 2.0. *Journal of Service Research* 18, 1 (2015), 59–74.

[77] Sarah Perez. 2019. Report: Voice assistants in use to triple to 8 billion by 2023. https://techcrunch.com/2019/02/12/report-voice-assistants-in-use-to-triple-to-8-billion-by-2023 (last accessed on 12 September 2021).

[78] Tim Polzehl. 2016. *Personality in Speech: Assessment and Automatic Classification*. Springer, Cham.

[79] Lesandro Ponciano, Pedro Barbosa, Francisco Brasileiro, Andrey Brito, and Nazareno Andrade. 2017. Designing for Pragmatists and Fundamentalists: Privacy Concerns and Attitudes on the Internet of Things. In *Brazilian Symposium on Human Factors in Computing Systems*.

[80] Andrew Raij, Animikh Ghosh, Santosh Kumar, and Mani Srivastava. 2011. Privacy Risks Emerging from the Adoption of Innocuous Wearable Sensors in the Mobile Environment. In *Proceedings of the CHI Conference*. ACM.

[81] Kari Rea. 2013. Glenn Greenwald: Low-Level NSA Analysts Have 'Powerful and Invasive' Search Tool. http://abcnews.go.com/blogs/politics/2013/07/glenn-greenwald-low-level-nsa-analysts-have-powerful-and-invasive-search-tool

[82] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. 2019. How well do my results generalize? Comparing security and privacy survey results from MTurk, web, and telephone samples. In *2019 IEEE Symposium on Security and Privacy*. 1326–1343.

[83] Seyed Omid Sadjadi, Sriram Ganapathy, and Jason W Pelecanos. 2016. Speaker age estimation on conversational telephone speech using senone posterior based i-vectors. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 5040–5044.

[84] Florian Schiel. 2011. Perception of alcoholic intoxication in speech. In *Proc. Interspeech*. 3281–3284.

[85] Björn Schuller, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, Gelareh Mohammadi, and Benjamin Weiss. 2015. A Survey on Perceived Speaker Traits: Personality, Likability, Pathology, and the First Challenge. *Computer Speech & Language* 29, 1 (2015), 100–131.

[86] Björn Schuller, Stefan Steidl, Anton Batliner, Florian Schiel, Jarek Krajewski, Felix Weninger, and Florian Eyben. 2014. Medium-term speaker states — A review on intoxication, sleepiness and the first challenge. *Computer Speech & Language* 28, 2 (2014), 346–374.

[87] Amit Kumar Sikder, Giuseppe Petracca, Hidayet Aksu, Trent Jaeger, and A. Selcuk Uluagac. 2018. A Survey on Sensor-Based Threats to Internet-of-Things (IoT) Devices and Applications. *arXiv:1802.02041 [cs]* (2018).

[88] Dave Smith. 2019. MicrophoneGate: The World's Biggest Tech Companies Were Caught Sending Sensitive Audio from Customers to Human Contractors. Here's Where They Stand Now. https://www.businessinsider.com/amazon-apple-google-microsoft-assistants-sent-audio-contractors-2019-8

[89] Daniel J Solove. 2013. Privacy self-management and the consent dilemma. *Harvard Law Review* 126 (2013), 1880.

[90] Raphael Spreitzer, Veelasha Moonsamy, Thomas Korak, and Stefan Mangard. 2018. Systematic Classification of Side-Channel Attacks: A Case Study for Mobile Devices. *IEEE Communications Surveys & Tutorials* 20, 1 (2018), 465–488.

[91] Anna Ståhlbröst, Annika Sällström, and Danilo Hollosi. 2014. Audio Monitoring in Smart Cities – an Information Privacy Perspective. In *12th International Conference e-Society*. 35–44.

[92] Dan Stowell, Dimitrios Giannoulis, Emmanouil Benetos, Mathieu Lagrange, and Mark D Plumbley. 2015. Detection and classification of acoustic scenes and events. *IEEE Transactions on Multimedia* 17, 10 (2015), 1733–1746.

[93] SoSci Survey. 2020. The Solution for Professional Online Questionnaires. https://www.soscisurvey.de/en/index

[94] Dan Svantesson and Roger Clarke. 2010. Privacy and consumer risks in cloud computing. *Computer Law & Security Review* 26, 4 (2010), 391–397.

[95] Monorama Swain, Aurobinda Routray, and P. Kabisatpathy. 2018. Databases, Features and Classifiers for Speech Emotion Recognition: A Review. *International Journal of Speech Technology* 21, 1 (2018), 93–120.

[96] VoiceSense. 2020. Speech Analysis as a Talent Recruiting and Retention Tool. https://www.voicesense.com/solutions/talent-recruiting-and-retention.

[97] Sandra Wachter and Brent Mittelstadt. 2019. A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Colum. Bus. L. Rev.* (2019), 494–620.

[98] Kelly Walters, Dimitri A Christakis, and Davene R Wright. 2018. Are Mechanical Turk worker samples representative of health status and health behaviors in the US? *PlOS One* 13, 6 (2018), e0198835.

[99] Wikipedia. [n.d.]. Nothing to hide argument. https://en.wikipedia.org/wiki/Nothing_to_hide_argument (last accessed on 12 September 2021).

[100] Wikipedia. [n.d.]. Telephone tapping. https://en.wikipedia.org/wiki/Telephone_tapping (last accessed on 12 September 2021).

[101] Heng Xu, Tamara Dinev, Jeff Smith, and Paul Hart. 2011. Information privacy concerns: Linking individual perceptions with institutional privacy assurances. *Journal of the Association of Information Systems* 12, 12 (2011), 798–824.

[102] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End user security and privacy concerns with smart homes. In *Symposium on Usable Privacy and Security* (Santa Clara, CA, USA). 65–80.

[103] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. 2018. User Perceptions of Smart Home IoT Privacy. *Proceedings of the ACM on Human-Computer Interaction* 2 (2018), 1–20.

A Survey questionnaire

1. What is your gender? (Female/Male/Other)
2. How old are you?
3. Privacy awareness[4]
   - I am aware of the privacy issues and practices in our society.[5]
   - I follow the news and developments about the privacy issues and privacy violations.[5]
   - I keep myself updated about privacy issues and the solutions that companies and the government employ to ensure our privacy.[5]
4. Disposition to value privacy[4]
   - Compared to others, I am more sensitive about the way companies handle my personal information.[5]
   - To me, it is the most important thing to keep my information privacy.[5]
   - Compared to others, I tend to be more concerned about threats to my information privacy.[5]
5. Previous privacy experience[4]
   - How often have you been a victim of what you felt was an improper invasion of privacy?[6]
   - How much have you heard or read during the past year about the use and potential misuse of the information collected from the Internet?[6]
   - How often have you experienced incidents where your personal information was used by a company without your authorization?[6]
6. Voice-controlled virtual assistant[7]

   A virtual assistant is a software agent that can perform tasks based on voice commands, without the requirement for keyboard input. Some examples of commercially available virtual assistants are Apple's Siri, Amazon Alexa, Microsoft's Cortana, and Google Assistant. Among other capabilities, virtual assistants can set reminders, manage con-

4 The constructs Privacy Awareness (PA), Disposition to Value Privacy (DVP), and Previous Privacy Experience (PPE) are adapted from Xu et al. [101]

5 Item was measured on a 5-point Likert scale: Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree

6 Item was measured on a 5-point Likert scale: Never, Rarely, Sometimes, Often, Very often

7 The order of items shown here reflects questionnaire Grp-A. In Grp-B, this component (№ 6, including the two questions) is positioned after the educational video, replacing question № 15. See Sect. 4.1 for exlanation.

tacts, play music, take purchase orders, send messages and calls, provide weather reports, and manage smart home devices.
  - How often do you use a voice-controlled virtual assistant in your daily life?[6]
  - Are you interested in starting or continuing to use a voice-controlled virtual assistant?[8]
7. Please attentively watch this video (1:44 minutes) about the privacy implications of voice data.[9]
8. Please enter the code displayed in the video. (not case-sensitive)
9. Before you watched the video, were you aware that these types of information can be inferred from voice recordings?
  - Geographical origin[10]
  - Gender and age[10]
  - Mental health information[10]
  - Physical health information[10]
  - Level of intoxication[10]
  - Moods and emotions[10]
  - Sleepiness and fatigue[10]
  - Personality traits[10]
10. How worried are you about these possible inferences?[11]
11. Why do you feel this way? Please explain your reasoning in two or more sentences.
12. Prior to this questionnaire, how often have you consciously thought about this issue when using a microphone-equipped device?[6]
13. What do you think, how common is it for companies to draw such inferences from voice recordings?[12]
14. Please rate how concerned you would be if a company used voice recordings to infer personal information about you without your awareness.
  - Geographical origin[13]
  - Gender and age[13]
  - Mental health information[13]
  - Physical health information[13]
  - Level of intoxication[13]
  - Moods and emotions[13]
  - Sleepiness and fatigue[13]
  - Personality traits[13]
15. Previously in this survey, you were asked about your interest in voice-controlled virtual assistants, such as Apple's Siri and Amazon Alexa. You are now asked about this a second time. Please answer based on your current thoughts and feelings, independent from your previous response.
  - Are you interested in starting or continuing to use a voice-controlled virtual assistant?[8]
16. Please indicate the extent to which you agree or disagree with each statement.[14]
  - Other people come to me for advice on new technologies[5]
  - In general, I am among the first in my circle of friends to acquire new technology when it appears[5]
  - I can usually figure out new high-tech products and services without help from others[5]
  - I keep up with the latest technological developments in my areas of interest[5]
  - I enjoy the challenge of figuring out high-tech gadgets[5]
  - I find I have fewer problems than other people in making technology work for me[5]
  - I prefer to use the most advanced technology available[5]
  - Show that you are paying attention by skipping this row without making a tick[5]
  - I find new technologies to be mentally stimulating[5]
  - Learning about technology can be as rewarding as the technology itself[5]
17. Do you own any devices that have a microphone? Select all that apply.
  - Phone/smartphone
  - Laptop
  - Tablet
  - Smartwatch
  - Camera
  - Smart speaker
  - Car with voice control interface
  - Voice-enabled remote control

8 Item was measured on a 5-point Likert scale: Not at all interested, Slightly interested, Somewhat interested, Quite interested, Very interested

9 Video clip available here: https://youtu.be/Gr22YqS1_VA.

10 Item was measured on a 5-point Likert scale: Not at all, Slightly, Somewhat, Quite well, Very well

11 Item was measured on a 5-point Likert scale: Not at all worried, Slightly worried, Somewhat worried, Quite worried, Very worried

12 Item was measured on a 5-point Likert scale: Very uncommon, Somewhat uncommon, Undecided, Somewhat common, very common

13 Item was measured on a 5-point Likert scale: Unconcerned, Slightly concernced, Somewhat concernce, Quite concerned, Very concerned

14 This construct – Innovativeness (INNO) – is adapted from Parasuraman and Colby [76]

  - Other (please specify)
18. What is the highest level of education you have obtained?
  - Finished school with no qualifications
  - Still in secondary school
  - GCSE Level education (e. g., GCSE, O-Levels, Standards)
  - A-Level education (e. g., A, AS, S-Levels, Highers)
  - Some undergraduate education (i. e., university examinations but not completed degree)
  - Degree or Graduate education (e. g., BSc, BA)
  - Post-graduate education (e. g., MSc, MA)
  - Doctorate degree
  - Vocational education (e. g., NVQ, HNC, HND)
  - Other degree or qualification (please specify)
19. Do you have professional experience in the following areas?
  - Data protection law[15]
  - Computer science[15]
  - Data mining[15]
  - Information technology security (IT security)[15]
20. What is your monthly net income? (Net income is defined as your total income after tax and social security deductions.)
  - I do not have a personal income
  - Less than £250
  - £250 up to £500
  - £500 up to £1000
  - £1000 up to £1500
  - £1500 up to £2000
  - £2000 up to £3000
  - £3000 up to £4000
  - £4000 up to £5000
  - £5000 or more
  - Decline to answer
21. To show if we have expressed ourselves clearly enough, please tick the description that best reflects the topic of this study.
  - Health effects of urban air pollution
  - Privacy concerns related to voice recordings
  - Telecommunications in India
  - Professional music production
  - Health concerns about wireless device radiation
  - Landlord and tenant privacy rights

15 Item was measured on a 5-point Likert scale: No experience, Little experience, Some experience, Much experience, Very much experience

B Correlation table

Spearman's rank correlations between participant demographics and paricipants awareness for audio-based inferences are shown in Table 6, along with Bonferroni-corrected significance levels.

Table 6. Spearman's rank correlations between participant demographics and awareness for audio-based inferences

	AW_DEM	95% CI	AW_STATE	95% CI	AW_TRAIT	95% CI
Age	–0.24***	(–1.00 to –0.19)	–0.18***	(–1.00 to –0.12)	–0.19***	(–1.00 to –0.14)
Gender	0.05	(–0.03 to 0.12)	0.11*	(0.03 to 0.19)	0.11*	(0.04 to 0.18)
Income	0.08	(0.00 to 0.17)	0.06	(–0.03 to 0.14)	0.04	(–0.04 to 0.12)
Education	0.29***	(0.23 to 1.00)	0.17***	(0.10 to 1.00)	0.14**	(0.07 to 1.00)
INNO	0.24***	(0.18 to 1.00)	0.18***	(0.12 to 1.00)	0.21***	(0.15 to 1.00)
PA	0.18***	(0.12 to 1.00)	0.20***	(0.14 to 1.00)	0.19***	(0.12 to 1.00)
DVP	0.08	(0.01 to 0.17)	0.05	(–0.02 to 0.13)	0.09	(0.03 to 0.17)
PPE	0.19***	(0.12 to 1.00)	0.18***	(0.11 to 1.00)	0.18***	(0.12 to 1.00)
EXP_DP	0.20***	(0.14 to 1.00)	0.16***	(0.10 to 1.00)	0.14**	(0.07 to 1.00)
EXP_DM	0.28***	(0.22 to 1.00)	0.26***	(0.20 to 1.00)	0.28***	(0.22 to 1.00)
EXP_CS	0.27***	(0.21 to 1.00)	0.22***	(0.16 to 1.00)	0.26***	(0.20 to 1.00)
EXP_IS	0.25***	(0.18 to 1.00)	0.22***	(0.15 to 1.00)	0.22***	(0.14 to 1.00)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

AW_: average level of awareness about audio-based inference of demographics (DEM) / short- and medium-term states (STATE) / phyisical and psychological traits (TRAIT); **INNO**: innovativeness; **PA**: privacy awareness; **DVP**: disposition to value privacy; **PPE**: previous privacy experience; **EXP_**: professional experience in data protection law (DP) / data mining (DM) / computer science (CS) / IT security (IS); **CI**: confidence interval