Theophany: Multimodal Speech Augmentation in Instantaneous Privacy Channels

Abhishek Kumar
University of Helsinki
Finland
abhishek.kumar@helsinki.fi

Tristan Braud
HKUST
Hong Kong SAR
braudt@ust.hk

Lik-Hang Lee
KAIST
Republic of Korea
likhang.lee@kaist.ac.kr

Pan Hui
HKUST & University of Helsinki
Hong Kong SAR
panhui@cse.ust.hk

ABSTRACT

Many factors affect speech intelligibility in face-to-face conversations. These factors lead conversation participants to speak louder and more distinctively, exposing the content to potential eavesdroppers. To address these issues, we introduce Theophany, a privacy-preserving framework for augmenting speech. Theophany establishes ad-hoc social networks between conversation participants to exchange contextual information, improving speech intelligibility in real-time. At the core of Theophany, we develop the first privacy perception model that assesses the privacy risk of a face-to-face conversation based on its topic, location, and participants. This framework allows to develop any privacy-preserving application for face-to-face conversation. We implement the framework within a prototype system that augments the speaker's speech with real-life subtitles to overcome the loss of contextual cues brought by mask-wearing and social distancing during the COVID-19 pandemic. We evaluate Theophany through a user survey and a user study on 53 and 17 participants, respectively. Theophany's privacy predictions match the participants' privacy preferences with an accuracy of 71.26%. Users considered Theophany to be useful to protect their privacy (3.88/5), easy to use (4.71/5), and enjoyable to use (4.24/5). We also raise the question of demographic and individual differences in the design of privacy-preserving solutions.

CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; **Privacy protections**; **Usability in security and privacy**; • **Human-centered computing** → **Ubiquitous and mobile computing**; **Mixed / augmented reality**.

KEYWORDS

Assistive technology, Augmented reality, User privacy, Human augmentation, Multi-modal speech augmentation, Speech intelligibility

ACM Reference Format:
Abhishek Kumar, Tristan Braud, Lik-Hang Lee, and Pan Hui. 2021. Theophany: Multimodal Speech Augmentation in Instantaneous Privacy Channels. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3474085.3475507

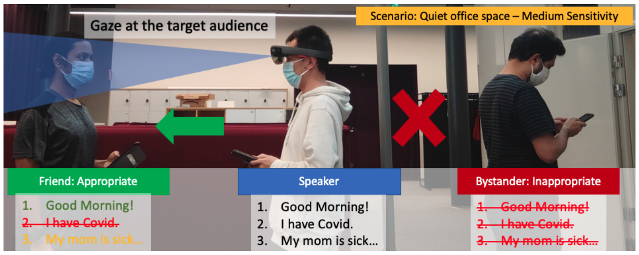


Figure 1: Theophany implementation as a real-life subtitle application. Theophany leverages a multi-modal channel (eye-tracking, cameras, and audios) that recognizes recipient(s) and records the speaker's speech, estimates the privacy sensitivity of each sentence, and dynamically establishes ad-hoc social networks to transmit the textual retranscription to the appropriate recipients' smartphones.

1 INTRODUCTION

Face-to-face conversations are the most effective support of information communication by enabling the full range of visual and auditive cues [6]. However, many factors can affect the transmission channel, whether external (background noise, reverberation, distortion), or internal to the conversation (speech disorders, hearing impairment, absence of contextual cues). These factors significantly reduce speech intelligibility, impeding the naturalness of face-to-face conversations. Speech augmentation in face-to-face conversations can enhance speech intelligibility by transmitting additional information to the conversation participants' personal devices. Speech augmentation raises significant privacy concerns. Bystanders may overhear the conversation, whether intentionally (eavesdroppers), or unintentionally (due to physical proximity). However, by nature, face-to-face conversations are ephemeral, existing only at the time they take place. Speech augmentation threatens this paradigm by recording and transmitting information over digital means. Yet, few studies consider privacy in face-to-face conversations, and most of the proposed solutions reduce speech intelligibility [4, 10, 18, 20, 29].

To address these issues, we introduce Theophany, a privacy-preserving framework for speech augmentation in face-to-face conversations. Theophany operates by deploying privacy respectful ad-hoc social networks on mobile devices. These ad-hoc social networks convey the user's speech content to the conversation

MM '21, October 20–24, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8651-7/21/10...$15.00
https://doi.org/10.1145/3474085.3475507

recipients according to the message's sensitivity and the user's privacy preferences. Theophany relies on the first privacy perception model developed specifically for daily conversations. This model evaluates the user's perception of the privacy risk based on the data sensitivity, the exposure risk due to the location, the location's acoustic property, and the data recipient relevance. Theophany leverages the contextual integrity (CI) framework [21], to minimize the leakage of information disclosure to recipients deemed not appropriate. We integrate this model within a system architecture that can be used to develop any privacy-aware application for face-to-face conversations. This system architecture provides a seamless information–sharing channel that enables social interactions while respecting social distancing, improving social interaction in a privacy-preserving manner without registration process, creating a CI-regulated Augmented Reality Social Network (ARSN). At the time of writing this paper, the social distancing restrictions brought during the COVID-19 pandemic are still in effect, significantly impeding face-to-face conversations. As such, we implement Theophany within a prototype smartphone app that provides privacy-appropriate recipients with real-life subtitles, alleviating the issues of masks and social distancing on face-to-face communication (See Figure 1). During our final user evaluation, we show that Theophany accurately (71.26%) matches the privacy expectations of the participants, while presenting a high technology acceptance.

The main contributions of our paper are as follows:

- We develop a novel **privacy perception model** for daily conversations. The model accounts for the conversation contents' sensitivity, the level of trust in the participants and the environment, and the ambient noise.
- We introduce Theophany, a framework leveraging CI to build privacy-respectful applications for speech augmentation. This framework protects against both voluntary and involuntary overhearing in face-to-face conversations.
- We implement Theophany as an **ad-hoc social network** in which the CI framework governs the flow of information.
- We **evaluate** Theophany through both a user survey and an exploratory user evaluation. Theophany scores closely to users' privacy preferences with *71.26% accuracy*. Participants found Theophany *useful (3.88/5)*, *enjoyable (4.24/5)*, and *easy to use (4.71/5)*. They also expressed that they would adopt Theophany under a lighter form-factor (3.76/5).

2 RELATED WORKS

In this section, we first describe the media richness theory as our justification for designing our privacy-preserved solution. Next, we compare our solution with other privacy-preserving solutions.

Media richness theory: The theory of information richness [5] is a framework that describes a communication medium's capability to reproduce the information delivered through it. This theory is usually employed to rank and assess the efficiency of certain media types in organizational communication [17]. Based on contingency theory and information processing theory [23], we consider the theory of information richness significantly in our solution design, where the richest information channel for person-to-person communication is reserved for the sake of effective information delivery. According to the hierarchy of media richness [1], individuals receive the most information from face-to-face oral communication as most of the communicative cues are available, serving as the signifiers in the user affordance between individuals in the conversations. However, the richness of face-to-face communication deteriorates when some cues are suppressed (e.g., when wearing surgical masks [9]). Additional visuals [15] can serve as a compensation cue for such losses of medium richness. Our paper addresses the loss in media richness during the daily face-to-face conversation and establishes a privacy-conserved channel to present the additional visuals.

User's Privacy Perception: Perez et al. [33] make an attempt to understand user's (sound) privacy perception in terms of acoustic properties of the environment. Kumar et al. [11] propose a framework to quantify user's privacy perception of text data which it generates in real-time and builds a color-code warning system. However, this solution is primarily designed for text data and is not directly applicable to preserve privacy for daily conversations. Liang et al. [18] study the users' perception in an environment in which the audio is recorded continuously. The authors propose that users' privacy perception improves if the quality of the audio is degraded before it is recorded. This result is consistent with other studies showing that slight obfuscation of location data can enhance the users' privacy perception while still enabling the original application [2]. However, these results can not be directly used in the much wider scope of face-to-face conversations.

Speech Augmentation: Recently, the idea of augmenting speech has gained momentum in the wake of the ongoing pandemic. However, its scope can be generalized to many other scenarios, such as hearing impairment or noisy environments. MAScreen [13]converts the user's lip motion to visual content visible to the surrounding audience. Similarly, two smart face masks prototypes [7] address the face occlusion problem: *Mouthy Mask* reproduces the image of the wearer's mouth, while *Smiley Mask* provides symbolic contents. These masks are becoming more useful in public contexts to support short socially-expected rituals, by showing the occluded information from the users' faces, as an alternative form of emoji-style visualization. However, the above studies do not consider privacy conservation. Our paper uniquely considers the visualization of augmented speech with the privacy feature supported by contextual integrity in daily conversation.

3 QUANTIFYING PRIVACY PERCEPTION IN DAILY CONVERSATIONS

In this section, we develop a model for privacy perception in daily face-to-face conversations based on the framework of Contextual Integrity. After summarizing the key elements of this framework, we introduce the primary parameters of privacy perception, from which we derive a model to measure the privacy risk of speech data. Finally, we perform two user studies to tune our model's variables.

3.1 Contextual Integrity

In daily-life conversations, people intuitively adapt the content they share in a conversation, depending on 1) the content, 2) the level of trust they have in their conversation partners and, and 3) the environment [33]. These factors can serve as a basis for understanding the user's perception of privacy and can be used to build privacy-enhancing technologies for daily conversations.

Following this line of reasoning, we design Theophany as a privacy-enhancing technology for daily conversations. We base the design of Theophany on the principles of Contextual Integrity (CI) [12, 21]. The theory of contextual integrity proposes that informational privacy can be achieved by ensuring the appropriateness of information flows in the given context. An appropriate flow is a flow that complies with the norms associated with it. Such norms are determined by five factors: *Sender, Recipient, Subject, Attribute, and Transmission Principle.* For instance, in healthcare, patients (data subject, sender) submit their health-related information (information type) to doctors (recipient) under conditions of strict confidentiality (transmission principle). In this context, the information flow is constrained by the transmission principle of confidentiality, which restricts the information flow to other parties. However, if we replace the doctor with a friend, the transmission principle changes, and thus the appropriateness of the flow.

Theophany uses CI as a system abstraction to prevent the spread of inappropriate information flows when the user speaks during their daily conversations. To achieve this goal, we build a model that detects inappropriate information flows depending on the context (recipients, location, topics). Theophany thus provides the first line of defense against undesired dissemination of the user’s daily conversations by making sure that those conversations are received by the recipient the user actually intended.

3.2 Privacy Perception Parameters

We conduct an online pilot study with 30 participants to determine the parameters that significantly affect users’ privacy perception in daily conversation[1]. We asked users how comfortable they would be in disclosing their personal information (name, age, home address, personal mobile number, personal email address, occupation, blood type, credit card number, medicine, and birthday) in the voice conversation at different circumstances, like bank, office, meetup, medical center. The survey was distributed among the personal connections of the authors. Based on findings from previous research on sound privacy [33] and our observations on user’s discomfort level regarding disclosing information through conversations, we identify five primary parameters that define privacy perception (see Tables 1-5 (Appendix)): Data (Content to be disclosed) sensitivity (S), Exposure risk caused by the physical location in which it is disclosed (E), Acoustic properties of the environment (A), Relevance of the conversation content with the conversation partner (V), and Relatedness of conversation content with the context (R). These parameters influence the privacy risk perceived by the user when they make data disclosure decisions during face-to-face conversations.

3.2.1 Data Sensitivity (S). Each data item D has an inherent sensitivity that does not depend on external factors. For instance, medical information is inherently more sensitive than which school did the user goes to. We define the sensitivity in terms of a given data element in terms of the amount of information (or control) that gets released out of the user’s hand once the data element becomes public/leaked. We adopt three categories for classifying the information in terms of sensitivity: low, medium, and high sensitivity. The reason for choosing only three categories instead of choosing multiple fine-grained categorizations is that the smaller number of categories is more manageable cognitively [22].

3.2.2 Exposure Risk due to Location (E). We define the exposure risk as the risk associated with the physical location of the conversation. The location influences information exposure to outside parties by presenting many bystanders or conditions that force conversation participants to raise their voice, increasing the information transmission distance. A closed office or the user’s home presents a minimal exposure risk, while a mall carries a significantly higher risk. Although places such as concerts and festivals present external noise conditions that reduce the information transmission distance, bystanders’ density increases the exposure risk.

3.2.3 Acoustic Property (A). We define the acoustic property as the amount of noise of the environment. The background noise is related to the risk of privacy leakage since a background noise leads the speaker to speak louder to convey information. Hence, the risk of being heard by an eavesdropper or being recorded by a nearby microphone significantly increases. In prior works, Zarazaga et al. [33] report that subjects are comfortable disclosing sensitive information in locations with higher background noise. We classify the background noise into three categories: 1) low (below 40 dB), 2) medium, (40 to 60 dB), and 3) high (over 60 dB).

3.2.4 Data-Recipient Relevance (V). We define data-recipient relevance in terms of how critical it is for the recipient to receive the data. For example, giving healthcare data to a doctor is critical to the user, no matter how sensitive such data are. A medical doctor has high data-recipient relevance for medical data, leading to a very low perceived privacy risk. On the other hand, there is little relevance if the recipient of healthcare information is a coworker.

3.2.5 Data Context Relatedness (R). We define data-context relatedness in terms of the user’s trust level in the general environment. For example, the user may speak freely on sensitive topics at home without minimal worry about privacy leakage or eavesdropper. On the other hand, at the office, the user may not feel comfortable addressing specific topics, regardless of the actual exposure risk (E). We classify the location into three categories: 1) Private (home), ii) semi-public (office, gym), and iii) public (cafe, mall).

3.3 Risk Perception Model

In this section, we establish the relationship between the privacy risk P of a given data D, and privacy parameters S, E, A, V, and R mentioned in Section 3.2. From Maximilien et al. [19], we establish that the privacy risk P is directly proportional to *Data Sensitivity* S, and *Exposure Risk* E. From Senarath et al. [26], we establish that the privacy risk P of a given data D, is inversely proportional to *Data Context Relatedness* R. In prior empirical research on sound privacy, Zarazaga et al. [33] reported that users were uncomfortable in disclosing sensitive information in locations that have higher background noise, and we also confirm this observation in our pilot study, and therefore, we establish that privacy risk P of the user, is directly proportional to *Acoustic Property* A. From the pilot study, we also find that users perceive a lower privacy risk when disclosing their sensitive data items to their conversation partners who have relevance with the data, and hence from this, we establish

[1] Pilot Study https:\shorturl.at/eDEIQ

that the privacy risk P of a given data D, is inversely proportional to *Data-Recipient Relevance* V. Combining these elements, we consider that the privacy risk P_i of the user i can be obtained by:

$$Risk\_Posed\_By\_flow, P_i \propto \beta \times \frac{S^a \times E^b \times A^c}{V^d \times R^e} \tag{1}$$

Where, β, a, b, c, d, e are real numbers. These variables' values (a, b, c, d, e) correspond to the associated parameter's weight in quantifying the net privacy risk posed by a given flow. For instance, if the value of the variable 'a' is high, it means the contribution of its associated parameter S in determining the risk is also high.

3.4 Model Tuning

We aim to observe how close the relationship proposed in Equation 1 is to the actual privacy risk perceived by users. We adopt the definition of privacy perceived risk from Maximilien et al. [19] and Senarath et al. [26] in the conversation settings. Our definition for the user's perceived privacy risk is as follows: "a measurement that determines the user's feeling of discomfort or reluctance in disclosing a data item in the given conversation context". We follow the protocols presented in [26] as guidelines to conduct studies for quantifying user's privacy perception in daily conversations. Following these guidelines, we conduct two user studies. The first study aims at establishing the general users' privacy perception through an Amazon Mechanical Turk survey. The second study regroups a limited number of privacy experts to perform a fine-grained evaluation of the privacy risk in various scenarios.

3.4.1 Study I: We recruited 196 workers from Amazon Mechanical Turk (AMT) to obtain the data for the dependent variable of our model. The participants were of age from 18 to 65, and 116 participants are women (Detailed demographic information in Tables 7-9 (Appendix)). None of these workers were involved in the pilot study used to identify the privacy perception parameters, or in any later study. In this survey, we define users' perceived privacy risk as their discomfort or reluctance for data disclosure in daily conversations. We ask users about the perceived discomfort when they disclose data during their conversations to measure their perceived privacy risk. We define five settings for daily conversations: 1) Restaurant, 2) Hall, 3) Office Closed, 4) Office Open, 5) Street, and consider the following data items: name, age, address, mobile number, email address, occupation, blood type, credit card number, medicine taken, and birthday. We also extract 45 real excerpts of user conversation of varying sensitive levels on cancer, pet, and family topics, as proposed by Kumar et al [11]. We ask the AMT workers to rate their level of comfort for the following scenarios:

(1) disclosing each of the 10 data items in the five settings.
(2) saying each of the 45 excerpts in the five settings.

The AMT workers rate their feeling of discomfort for disclosure F_i on a five-point Likert scale (very uncomfortable – very comfortable). We aim to determine how close the calculated privacy risk P_i (Equation 1) is from F_i. Through this study, we obtain $F_i \in \{1, 2, 3, 4, 5\}$ for each user in a total of 275 (i.e. 5 x (10 + 45)) scenarios.

3.4.2 Study II: We conduct a second study with a focus group of 8 privacy experts who have experience in designing systems for privacy and security. This study aims to obtain the model's independent variables mentioned in Section 3.2 (sensitivity, exposure, acoustics, recipient relevance, and context relatedness) for the data disclosure scenarios used in the survey. For each of the 275 scenarios, the participants categorize the value of the 5 independent variables into three privacy levels (high – 3, medium – 2, low – 1).

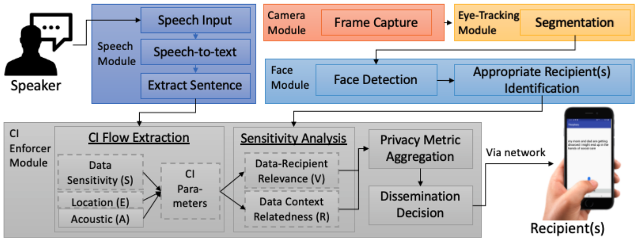


Figure 2: System Architecture. Theophany transforms the user's speech into text and estimates the intended recipients through gaze detection. The CI Enforcer module evaluates the sentences' sensitivity. If the sensitivity meets the speaker's privacy threshold, the sentence is transmitted to the appropriate recipients.

3.4.3 Data Analysis: From the study I, we extract the users' perceived privacy risk for the dependent variable P, while study II evaluates the independent variables (S, E, A, V, R) of our model. We then use curve-fitting on the raw data to determine a, b, c, d, e. As the search space is large, we rely on heuristics used in prior literature [26] on user's privacy perception. This study reports that the relation between perceived privacy risk is in i) direct linear relation with sensitivity (S), ii) direct cubic relation with Exposure (E), and iii) inverse linear relation with Data Context Relatedness (R), which leads to the following values: $a = 1, b = 3, e = 1$. After applying curve fitting with these values, we recover the optimal result for the goodness of fit (R Squared = 0.728, Root Mean Squared Error = 0.412 with 95% confidence interval) under the following form:

$$P_i = \frac{0.56 \times S \times E^3 \times A^{0.8}}{V \times R} \tag{2}$$

So, the parameters of the model are as follows: $\beta = 0.56, a = 1, b = 3, c = 0.8, d = 1, e = 1$. This model gives us the best approximation of user's perceived risk at the core of our CI-based framework.

4 SYSTEM DESIGN

Due to the nature of the transmission medium, information exchanged during face-to-face conversation is propagated to everyone in the vicinity. Theophany aims to minimize the leakage of information disclosure to the recipients who are deemed not appropriate right at the source by relying on the theory of contextual integrity (CI). To achieve this goal, Theophany consists of three modules: *CI flow extraction*, *CI Flow Processing*, and *Privacy Metric Aggregation* as depicted in Figure 2. After extracting the CI parameters from the information flow, the CI flow processing module determines the sensitivity of the flow based on the model presented in Section 3 and the identified recipients of the conversation. Afterward, the *Privacy Metric Aggregation* combines these two components' results into a single metric to determine if the given flow respects/violates

CI norms. If the flow containing the information on the user's sensitive topic is considered appropriate by the CI framework, it is transmitted to the relevant users. if the flow does not contain any information on the user's sensitive topic, the last two modules of Theophany are bypassed. In the rest of this section, we describe these modules and how they integrate the CI theory.

4.1 CI Flow Extraction

This module is responsible for extracting different relevant CI parameters (as stated in Section 3) from the given information flow and existing metadata. These parameters include the actors (sender, subject) and the type of information (attribute). The parameter extractor then maps these parameters onto CI flows. Extracting sender and subject is relatively straightforward since both parameters are the user itself. However, extracting other information like attributes and transmission principle is more challenging. For example, if the user is sensitive about disseminating health-related data (e.g., HIV), the CI Flow Extractor analyses the user's speech to determine whether the conversation contains any HIV-related topics. However, if the user is trying to speak (or send) HIV-related information at home or in a closed office at the hospital, they might be sharing it with a family member or a doctor. Such information flow should be considered admissible even though the user's speech contains sensitive information. In other contexts, such as in the office or at a public place, the flow can be considered non-admissible.

4.2 CI Flow Processing

The *CI Flow Processing module* comprises two components: (1) *Sensitivity Analysis* quantifies the instantaneous sensitivity level of the text (converted from user's speech), and (2) *Eye-Tracking based Recipient Recognition* identifies the recipients in real-time.

4.2.1 *Sensitivity Analysis (S).*

This component is responsible for estimating the sensitivity of the given information flow, i.e. amount of privacy the user's speech may leak. It classifies a given information flow F as sensitive if the *Amount of Information* carried by the flow is beyond the comfort level of the user γ in a user's sensitive context. The amount of information is measured using an information-theoretic metric: pairwise mutual information (PMI) [24]. The PMI of a given flow F in a given context t is given as:

$$PMI(F;t) = -\log_2 \frac{Count(F;t)}{Count(F) \times Count(t)} \quad (3)$$

Where $Count(F;t)$ represents the number of co-occurrences of F and t. If $PMI(F;t) > \gamma$, the flow is sensitive. The problem of sensitivity estimation thus reduces to finding the PMI value for the flow. Calculating this value requires the presence of a corpus. We rely on the solution proposed by Kumar et al. [11] which has a minimal footprint on the mobile systems. Using this approach to determine sensitivity has several advantages over traditional natural language processing (NLP) approaches. Sensitivity estimation is highly dependent on the context, and these contexts tend to evolve outside of the scope covered by the original NLP training dataset.

4.2.2 *Eye-Tracking based Recipient Recognition.*

One of the primary challenges in building privacy-enhancing technology for daily conversation is to locate relevant recipients in real-time. Eye-gaze is an ideal option for this purpose. In both our user study and the Amazon Mechanical Turk (AMT) study, almost every user confirmed that whenever they wish to disclose sensitive information in a conversation, they tend to look at the recipients with very high confidence (avg = 4.82, std = 0.38) on a scale of 1-5 where 5 is the most confident, and 1 the least confident. With the human eye having a high enough resolution only over a few degrees [3], gaze tracking allows us to accurately identify conversation recipients. Using eye gaze to guide information dissemination to other users also solves the problem of the information being heard by an eavesdropper since the eavesdropper would likely not be in the user's line of sight.

4.3 Privacy Risk Aggregation

This module is responsible for aggregating the information received by the CI Flow Processing module and estimating the privacy risk posed by the given flow through equation 2 with the criteria defined in Section 3.2. The privacy risk score calculated by this module can be further used in several ways depending on the privacy framework being used to design a solution. In this work, we use the CI framework, and the privacy risk score is used to determine the transmission principle dynamically. Since the user's privacy perception evolves over time and contexts, the transmission principle needs to be updated accordingly. The Privacy Risk Aggregation module keeps the model up-to-date with the user's privacy perception.

4.4 Enforcing CI in Theophany and Examples

Theophany enforces contextual integrity by identifying the information flow parameters and assessing the flow's appropriateness within the conversation context using three components summarized in Figure 2. The CI parameters are determined as follows: 1) Sender (the user itself), 2) Subject (the user itself), 3) Information Type (determined by *CI Flow Extraction*), 4) Recipient (determined by *CI Flow Processing*), and 5) Transmission Principle (determined by *Privacy Metric Aggregation*) The CI Flow Extraction component preprocesses the user's speech and determines the information type. If the information type is related to sensitive topics, Theophany checks if the input text respects the transmission principle. The user's speech is transmitted (1) if it is not sensitive given the user's privacy leakage tolerance in the given context, and (2) if the recipient is deemed appropriate given the information type, regardless of the sensitivity level of the text. If the CI Flow Extraction determines the information not to be sensitive, the information can be directly transmitted, bypassing the CI Flow Processing module and the Privacy Metric Aggregation module. Otherwise, the CI Flow processing and Privacy Metric Aggregation determine the user's speech's sensitivity after considering multiple aspects: sensitivity of the text, acoustic characteristic, exposure risk, data-recipient relevance, and data-context relatedness (Equation 2).

To illustrate how Theophany determines whether a transmission would violate the user's privacy, we consider the following example: "*my mom and dad are getting divorced I might end up in hands of social care*". We assume that the user considers family-related conversations as moderately sensitive. To determine the transmission principle, Theophany estimates the user's perceived privacy risk for this flow based on Equation 2. Data sensitivity (S) is determined

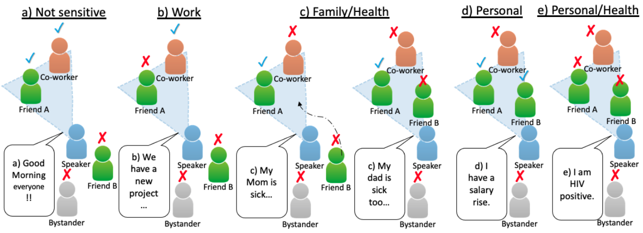


Figure 3: Theophany operation depending on the topic and session. (a) the topic is not sensitive and transmitted to everybody in the user's gaze. (b) the topic is work-sensitive and only transmitted to the coworker. (c) the topic is sensitive and only transmitted to the friend in the user's gaze. A new friend entering the user's gaze only gets the textual transcription once a new session (topic) starts (d). (e) the topic is highly sensitive and nobody gets the textual transcription.

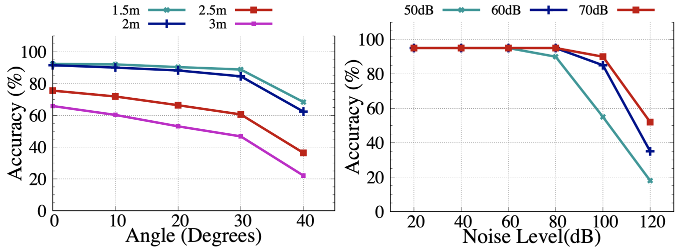
to be 2 following the method in Section 4.2. If the conversation happens in a quiet home environment with a trusted friend, then E, A, V, and R take values of 1, 1, 3, and 3, respectively, for a perceived privacy risk of 0.124. However, if it happens in the festival area with a work colleague, then E, A, V, and R take values of 3, 3, 1, and 1, respectively for a perceived privacy risk of 109.23. Equation 2 ranges between 0.062 and 109.24. To present the privacy score to the user, we thus normalize the score in the range [1-10], leading to 1.005 and 9.87 for the above two scenarios, respectively. A user with a privacy threshold of 6 would consider the transmission in the first scenario appropriate, but not in the second scenario (30 more examples – 6 Speech Excerpt * 5 Scenarios in Table 6 (Appendix)).

4.5 Implementation

We implement Theophany within a real-life prototype system targeted towards improving speech intelligibility with COVID-19 restrictions (social distancing, mask-wearing). We use the user's eye-gaze to infer the intended recipients (see Figure 3). As most smartphones do not have this eye-tracking capability [14], we use an external device that provides eye-tracking in real-time. We thus use Microsoft Hololens 2 for eye tracking while an Android phone (Samsung Galaxy A6 Plus) captures the user's voice and displays the conversation content to the recipients deemed appropriate. Although this setting is appropriate for a demonstration application, it is not feasible for wide-scale distribution. A more realistic approach would consist of integrating all the components within smaller-scale smartglasses, with the added advantage of overlaying subtitles on top of the participants' physical location. However, with the state of current technologies [15], we consider that using a combination of Microsoft Hololens 2 and smartphones represents a good emulation of the setting for demonstration purposes [16].

The implementation includes the following modules: CI Flow Extraction, sensitivity analysis, and eye-tracking module. We implement the first two modules in `Android` (API 16) using `Java 1.15`.

We use Android's SpeechRecognizer API for voice-to-text conversion, and we apply the conversion on 15 seconds-long windows of voice recording. We use the `Apache OpenNLP` 1.9.1 library to perform the primary natural language processing tasks such as stemming, pre-processing, and detecting sentences. For the sensitivity analysis module, we use `JSoup API` (Version 1.12.1), based on the implementation method described in [11]. Eye-tracking is implemented in the C# programming language, using the `Unity 2019.4.11f1 (LTS)` framework. We rely on the components from Microsoft's official HoloToolkit-Unity repository (`MRTK2`). Though eye-gaze serves as an anchor to guide the information dissemination, Theophany still needs to link the eye-gaze data to the identity of the other person (whom the user is looking to), i.e., whether the person is a doctor, spouse, or office worker with who the user is comfortable to share its sensitive information. To do this linking, we extract the circular region of radius 8 cm using eye-gaze as an anchor for the center, and then utilize the face signature proposed by Shu et al. [27]. However, this process needs to take images and perform identity linkage every time the user wishes to speak about a sensitive topic. To alleviate this problem, we introduce a heuristic of the *session*, defined in terms of sensitive topics. The session starts when the user initiates the conversation about a new sensitive topic. The identity linkage step needs to be done only at the beginning. A session associated with the given sensitive topic dies either if the user does not speak about it for five minutes (user can easily adjust) in the same context or if the context changes.



(a) Recognition Accuracy

(b) Audio Accuracy

Figure 4: (a) Recognition Accuracy is high within distance of 2m and angle of 30° from the speaker, (b) Audio accuracy is high for noise at most 40dB higher than the user's voice.

5 EVALUATION

In this section, we seek to establish how well the Theophany framework can protect user privacy in face-to-face conversations, and how effective our prototype system is at augmenting users' speech when typical communication cues are suppressed by COVID-19 restrictions. To this end, we first characterize the accuracy of face and speech modules. Then, we determine the efficiency and acceptance of privacy-enhancing technologies for daily conversations. We first establish the ground truth data on privacy perception. We then discuss the impact of demographic factors on accuracy. We conclude our study with a user evaluation of our prototype system performed on 17 participants from six European countries.

5.1 Recognition and Audio Accuracy

Figure 4a depicts the accuracy of the face module according to the angle and distance between the user and the individual to recognize. The total recognition accuracy declines with both the distance and angle. As long as recipients are within 2 meters and an angle below

30°, the system recognizes the face with high accuracy (over 80%). Figure 4b presents the accuracy of the speech module of the system. As long as the noise level is at most 40 dB higher than the speaker's voice, the audio accuracy is over 80%. It allows users to continue the conversation in their normal voice (60 dB), even in a noisy environment without being forced to speak loudly which might be overheard by bystanders. Overall, our system supports the typical setting of daily conversations, even with socially distancing users.

5.2 Accuracy of Privacy Prediction

The transmission principle (presented in Section 3) should predict the sensitivity of any information flow with high accuracy. Given the lack of ground truth data to assess the accuracy of a privacy model (in the CI enforcer module) in face-to-face conversations, we recruited 53 participants (45 from the Amazon Mechanical Turk and 8 from the local university) from diverse backgrounds to provide such ground truth. There was no overlap between this participant set and the participants who took part in the initial pilot study or in providing data for model tuning in Section 3. Most participants meet less than 20 people a day, with 39% meeting 10 people or less.

We gave the participants a total of 45 affirmations over three topics (pets, family, cancer) of varying levels of sensitivity. These affirmations were taken from real user's conversations on online forums (Online Pet Forum, Reddit). Users on these forums typically mix formal and informal language to express their feelings, resembling voice conversations. We asked the participants to rate how comfortable they would be while saying these affirmations in different locations on a 5-point Likert scale ranging from 1 – very uncomfortable to 5 – very comfortable[2]. The locations are as follows: 1) *Closed Office*, the user is alone or with highly trusted friends, 2) *Closed Office*, the user is in presence of friends or regular acquaintance they may not know outside the professional setting, 3) *Restaurant*, the user is in presence of a limited number of strangers, mild to high background noise, 4) *Hall*, the user encounters a mix of close friends, acquaintances, and strangers, mild to the high background noise, and 5) *Street*, the user is surrounded by a crowd of strangers. Each participant provided their judgment for 225 scenarios, resulting in a total of 11,925 privacy judgments.

After receiving the ground truth information from the test population (53 users), we evaluate the accuracy of the Theophany by calculating the average absolute distance between the participants' answers and Theophany's results for each scenario. On average, participants respond closely to Theophany's results. With an average absolute difference of 1.15 (std=0.46) out of a five-point Likert scale, we can conclude that Theophany deviates from the users' perceived sensitivity by (+) 28.74%, for an average accuracy of 71.26%. This result is slightly skewed by a couple of sentences that Theophany over- or under-estimates. It is important to note that these numbers result from Theophany's generic usage. We expect accuracy to increase with the user-configurable privacy preferences.

5.3 Relation between Theophany Accuracy and Demographics

To understand the relationship between the objective accuracy of Theophany and user demographics, we split our dataset along several directions. The *country of origin* affects the accuracy. Theophany is on average (+) 7.45% more accurate for North American and European participants (75.26% accuracy vs 67.81% for Asians). We also notice a difference between *genders*. Theophany is on average (+) 5.59% more accurate for female participants (74.88% vs 69.29% for male). Both results can be explained by the composition of our user panel for model tuning: the majority of participants were of female gender from European and North American countries. However, our test panel is more varied. Since the majority of the participants belong to age groups 25-35 and 35-45, we evaluate the impact of *age* on the accuracy across two groups: 18 - 35 vs. 35+. We notice a difference of (+) 7.08% in the accuracy between these two age groups (72.04% for 18-35 vs. 64.96% for 35+). Finally, we do not notice any significant impact on the accuracy when we separate the participants on the basis of the *number of people they meet in a day* (<10 vs 10+). Similarly, although one-third of the participants reported that they consider themselves *members of a vulnerable community* in some capacity, we do not notice its impact on the objective accuracy. In conclusion, demographic factors such as age, gender, and cultural background should be given extra attention when designing solutions for daily life conversational privacy.

5.4 User Experiment

Privacy remains subject to individual preferences. As Theophany ultimately leaves the final decision to the user, the user can always bypass the system by speaking loudly at its own risk. It is thus critical to evaluate the user's perception of the system. We evaluate Theophany through an exploratory user experiment.

We invited 17 participants to test the feasibility of our prototype system. These participants were recruited from a local football club regrouping over 100 amateurs of different ages, backgrounds, education levels, and affiliations. All participants are of European origin, with a total of six countries represented (Finland, Sweden, Spain, Italy, Latvia, and Russia). The participants are young (18-34 years old) and educated (higher education). 11 participants are men, and 6 participants are women. These 17 participants were not involved in the initial pilot study for determining the parameters of the user's perception privacy model, in gathering the data (Study I and Study II) for model tuning (in Section 3), and in the objective accuracy evaluation (in Section 5.2). Due to the COVID-19 pandemic, we could not get a more diverse user group, especially old age users.

5.4.1 Privacy Awareness (in Daily Conversations) Survey. We first asked the participants to fill in a survey[3], partly based on Tuunainen et al. [28], divided into three parts: general conversational privacy awareness, measures they take to protect their conversational privacy, and conversational privacy in the context of the pandemic. We asked users to rate several affirmations on a scale ranging from 1 – very unlikely to 5 – very likely.

The general privacy awareness of the users is very high, with an average of 4.17 (+1.17) (std=0.62). Regarding privacy awareness on how they disclose sensitive information in their conversation, users tend to worry about their data privacy and security (avg = 3.75 (+0.75), std = 0.99). Furthermore, they worry about their information getting heard by an eavesdropper (avg = 3.13 (+0.13), std =

[2] Ground Truth Data Collection http://shorturl.at/iDJU9

[3] Technology Acceptance Survey http://shorturl.at/wKO58

1.3), and getting recorded by a microphone that they didn't know (avg = 3.08 (+0.08), std = 1.4). Users tend to rely on adjusting the loudness of their voice to protect privacy in their conversations based on their perception of the environment (avg=3.24 (+0.24), std=1.11). Users tend to think carefully before disclosing any sensitive information in their conversation (avg = 3.92 (+0.92), std = 1.01). However, users had had experiences in the part when they misread the trust level of their conversation, and later the information disclosed during the conversation got exposed to the larger extent for which it was not intended (avg = 2.75 (-0.25), std = 1.36), and this observation is consistent with the phenomenon of immediate gratification studied in prior literature [30] in the domain of online social media, which states that individuals are susceptible to hyperbolic time discounting, i.e. the tendency to increasingly choose a smaller-sooner reward over a larger-later reward. Finally, the COVID-19 pandemic has had a significant effect on people's conversations, especially when sharing some sensitive information. Since the health guidelines mandate the use of face mask, many users (43.49%) have trouble in understanding the speaker due to the loss of media richness caused by face covering. Combined with social distancing guidelines, the speaker is required to take off her/his mask or to speak louder than normal to be understood. The responses to this survey are very weakly correlated with technological literacy (Pearson correlation coefficient lower than 0.42). The privacy awareness of our participants is remarkably high (on average one point higher than the original study from Tuunainen et al. [28]), which usually pairs with high technological literacy.

5.4.2 Technology Acceptance. After explaining the purpose and operational flow of Theophany to our participants, we split them into the group of four participants. One participant takes the role of speaker, while the other three take the role of conversation partners. The speaker wore the Hololens 2 to determine the recipient of the conversation about the given sensitive topic through eye gaze tracking. The recipients used their phones to receive the conversations. Participants were standing at a distance of 1.5 meters to 2 meters from each other, which corresponds to the typical distancing recommendation during the COVID-19 pandemic. Before starting the conversation, the speaker specified which other participant should receive the conversation on the given sensitive topic. The participants were allowed to use the application for a total of ten minutes. We repeated this experiment until each participant in our study has the role of the speaker once. To make the experience more engaging, we randomize the groups so that conversations happen between different participants. Each participant engages in four conversations: one as a speaker, and three as a recipient.

We then asked the participants to fill a technology acceptance survey [32]. We measured the Perceived Usefulness (PU), the Perceived Ease Of Use (PEOU), the Intention Of Use (IOU), and the Perceived Enjoyment (PE) (see Table 10 (Appendix)). Participants found that Theophany enabled them to improve the way they decide to disclose information in their conversation (avg=3.88 (+0.88), std=0.67), to be more aware of the conversation's privacy (avg=4.29, (+1.29), std=0.57), easy to learn (avg=4.82, (+1.82), std=0.32), and plan to use once it gets released with dedicated hardware (avg=3.76 (+0.76), std=0.35). Participants considered using their eye-gaze an accurate approach to recipient detection (avg=4.00 (+1.0), std=0.68). During the feedback session, they also expressed that this solution could be integrated with Glasses, or Specs, to improve their likelihood of adoption. Many participants appreciated the importance of this solution in the context of social distancing. Theophany received particularly high acceptance among the female users.

Despite the objective accuracy of 71.26%, Theophany slightly suffers in terms of perceived accuracy during the experiment, as users report the perceived accuracy of 3.41 (+0.41) (std = 0.49), though still above the average, on the scale of 1-5, where 1 being the least accurate, and 5 being the most accurate, despite the user acknowledging that the usefulness of Theophany in protecting their privacy. This can be explained by the fact that topics of the conversation evolve rapidly during the conversation. During the feedback session, users expressed that if Theophany could be integrated with lightweight Glasses or Specs, they are likely to adopt it for daily communications. Many participants appreciated the importance of this solution in the context of social distancing.

Overall, users who care about their online privacy display a higher acceptance of Theophany. Despite high privacy awareness among many of these users, most of them (88%) are worried about their conversational privacy which might leak their privacy either because of eavesdropper getting their information, or microphones recording their information. Most of these users no longer maintain their social media after the 2016 Facebook-Cambridge Analytica Data Scandal [8], and multiple cases of the European Commission fining Facebook [25] and Google [31] for their over-breaches coming into light. They have now moved to secure applications like Telegram, Signal Messenger, and prefer to interact with many simultaneously via Group Chat on these platforms instead of Instagram or Facebook. But unfortunately, they do not have any solution for speech privacy, and many of them commented that they would be able to adopt the technology if it comes in a more usable form.

Our current user experiment reveals that highly privacy-aware users are likely to adopt a solution that protects their conversation privacy, since they often misread the trust level of their surroundings, resulting in privacy leakage. To confirm these results, it is necessary to perform further evaluation on less technology- and privacy-literate populations, that are the most at-risk.

6 CONCLUSION

This paper presents Theophany, a framework of privacy-preserving augmented speech in face-to-face conversations. Based on the gaze tracking, face signature, and conversation context, the system transmits textual visuals to the appropriate recipients to improve speech intelligibility. In future works, we will move the implementation of Theophany to a head-worn device to reduce the reliance on multiple devices. We will also evaluate other user-centered metrics (e.g., readiness, perceived information loads, user acceptance) on such user interfaces. Finally, we will extend our study to the role of visual cues. We will evaluate how to transmit such cues to conversation recipients and study their impact on perceived privacy.

ACKNOWLEDGMENT

This work has been supported by the Academy of Finland, under Grant 319669, and Grant 325570.

REFERENCES

[1] David G. Allen and Rodger W. Griffeth. 1997. Vertical and Lateral Information Processing: The Effects of Gender, Employee Classification Level, and Media Richness on Communication and Work Outcomes. *Human Relations* 50, 10 (1997), 1239–1260. https://doi.org/10.1177/001872679705001003

[2] A.J. Bernheim Brush, John Krumm, and James Scott. 2010. Exploring End User Preferences for Location Obfuscation, Location-Based Services, and the Value of Location. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (Copenhagen, Denmark) *(UbiComp ’10)*. Association for Computing Machinery, New York, NY, USA, 95–104. https://doi.org/10.1145/1864349.1864381

[3] Isha Chaturvedi, Farshid Hassani Bijarbooneh, Tristan Braud, and Pan Hui. 2019. Peripheral Vision: A New Killer App for Smart Glasses. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI ’19)*. Association for Computing Machinery, New York, NY, USA, 625–636. https://doi.org/10.1145/3301275.3302263

[4] Francine Chen, John Adcock, and Shruti Krishnagiri. 2008. Audio Privacy: Reducing Speech Intelligibility While Preserving Environmental Sounds. In *Proceedings of the 16th ACM International Conference on Multimedia* (Vancouver, British Columbia, Canada) *(MM ’08)*. Association for Computing Machinery, New York, NY, USA, 733–736. https://doi.org/10.1145/1459359.1459472

[5] Richard Daft and Robert Lengel. 1986. Organizational Information Requirements, Media Richness and Structural Design. *Management Science* 32 (05 1986), 554–571. https://doi.org/10.1287/mnsc.32.5.554

[6] Richard L Daft and Robert H Lengel. 1983. *Information richness. A new approach to managerial behavior and organization design.* Technical Report. Texas A and M Univ College Station Coll of Business Administration.

[7] Çağlar Genç, Ashley Colley, Markus Löchtefeld, and Jonna Häkkilä. 2020. Face Mask Design to Mitigate Facial Expression Occlusion. In *Proceedings of the 2020 International Symposium on Wearable Computers* (Virtual Event, Mexico) *(ISWC ’20)*. Association for Computing Machinery, New York, NY, USA, 40–44. https://doi.org/10.1145/3410531.3414303

[8] Kevin Granville. 2018. Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html

[9] Dominic Watt Senior Lecturer in Forensic Speech Science. 2020. The science of how you sound when you talk through a face mask. https://theconversation.com/the-science-of-how-you-sound-when-you-talk-through-a-face-mask-139817

[10] Jin Yong Jeon, Joo Young Hong, Hyung Suk Jang, and Jae Hyeon Kim. 2015. Speech privacy and annoyance considerations in the acoustic environment of passenger cars of high-speed trains. *The Journal of the Acoustical Society of America* 138, 6 (2015), 3976–3984.

[11] Abhishek Kumar, Tristan Braud, Young D. Kwon, and Pan Hui. 2020. Aquilis: Using Contextual Integrity for Privacy Protection on Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 4, 4, Article 137 (December 2020), 28 pages. https://doi.org/10.1145/3432205

[12] Abhishek Kumar, Tristan Braud, Sasu Tarkoma, and Pan Hui. 2020. Trustworthy AI in the Age of Pervasive Computing and Big Data. In *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 1–6. https://doi.org/10.1109/PerComWorkshops48775.2020.9156127

[13] Hyein Lee, Yoonji Kim, and Andrea Bianchi. 2020. MAScreen: Augmenting Speech with Visual Cues of Lip Motions, Facial Expressions, and Text Using a Wearable Display. In *SIGGRAPH Asia 2020 Emerging Technologies*. Association for Computing Machinery, New York, NY, USA, Article 2, 2 pages. https://doi.org/10.1145/3415255.3422886

[14] Lik-Hang Lee, Yiming Zhu, Yui-Pan Yau, Tristan Braud, Xiang Su, and Pan Hui. 2020. One-thumb Text Acquisition on Force-assisted Miniature Interfaces for Mobile Headsets. In *2020 IEEE International Conference on Pervasive Computing and Communications, PerCom 2020, March 23-27, 2020*. IEEE, Austin, TX, USA, 1–10. https://doi.org/10.1109/PerCom45495.2020.9127378

[15] Lik-Hang Lee and Pan Hui. 2018. Interaction Methods for Smart Glasses: A Survey. *IEEE Access* 6 (2018), 28712–28732. https://doi.org/10.1109/ACCESS.2018.2831081

[16] Lik Hang Lee, Yiming Zhu, Yui-Pan Yau, Pan Hui, and Susanna Pirttikangas. 2021. Press-n-Paste: Copy-and-Paste Operations with Pressure-Sensitive Caret Navigation for Miniaturized Surface in Mobile Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 5, EICS, Article 199 (May 2021), 29 pages. https://doi.org/10.1145/3457146

[17] Robert H. Lengel and Richard L. Daft. 1988. The Selection of Communication Media as an Executive Skill. *Academy of Management Executive* 2 (1988), 225–232. https://www.jstor.org/stable/4164833

[18] Dawei Liang, Wenting Song, and Edison Thomaz. 2020. Characterizing the Effect of Audio Degradation on Privacy Perception And Inference Performance in Audio-Based Human Activity Recognition. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI ’20)*. Association for Computing Machinery, New York, NY, USA, Article 32, 10 pages. https://doi.org/10.1145/3379503.3403551

[19] E Michael Maximilien, Tyrone Grandison, Tony Sun, Dwayne Richardson, Sherry Guo, and Kun Liu. 2009. Privacy-as-a-service: Models, algorithms, and results on the facebook platform. In *2009 IEEE Symposium on Security and Privacy Workshops, WEB 2.0 SECURITY AND PRIVACY*. IEEE, Oakland, California, 1–4. http://www.ieee-security.org/TC/W2SP/2009/papers/s4p2.pdf

[20] Markus Müller-Trapet and Bradford N Gover. 2019. Relationship between the privacy index and the speech privacy class. *The Journal of the Acoustical Society of America* 145, 5 (2019), EL435–EL441.

[21] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Washington Law Review* 79, 1 (2004), 119–157. https://heinonline.org/HOL/LandingPage?handle=hein.journals/washlr79&div=16

[22] Marie Caroline Oetzel and Sarah Spiekermann. 2014. A systematic methodology for privacy impact assessments: a design science approach. *European Journal of Information Systems* 23, 2 (2014), 126–150. https://doi.org/10.1057/ejis.2013.18 arXiv:https://doi.org/10.1057/ejis.2013.18

[23] Jeff Reinking. 2012. Contingency Theory in Information Systems Research. In *Information Systems Theory: Explaining and Predicting Our Digital Society, Vol. 1*, Yogesh K. Dwivedi, Michael R. Wade, and Scott L. Schneberger (Eds.). Springer New York, New York, NY, 247–263. https://doi.org/10.1007/978-1-4419-6108-2_13

[24] David Sánchez and Montserrat Batet. 2016. C-Sanitized: A Privacy Model for Document Redaction and Sanitization. *J. Assoc. Inf. Sci. Technol.* 67, 1 (Jan. 2016), 148–163. https://doi.org/10.1002/asi.23363

[25] Mark Scott. 2017. E.U. Fines Facebook $122 Million Over Disclosures in WhatsApp Deal. https://www.nytimes.com/2017/05/18/technology/facebook-european-union-fine-whatsapp.html

[26] Awanthika Senarath, Marthie Grobler, and Nalin A. G. Arachchilage. 2019. A Model for System Developers to Measure the Privacy Risk of Data. In *52nd Hawaii International Conference on System Sciences, HICSS 2019, January 8-11, 2019*. University of Hawaii at Manoa, Grand Wailea, Maui, Hawaii, USA, 6135–6144. https://doi.org/10.24251/HICSS.2019.738

[27] Jiayu Shu, Sokol Kosta, Rui Zheng, and Pan Hui. 2018. Talk2Me: A Framework for Device-to-Device Augmented Reality Social Network. In *2018 IEEE International Conference on Pervasive Computing and Communications, PerCom 2018, March 19-23, 2018*. IEEE Computer Society, Athens, Greece, 1–10. https://doi.org/10.1109/PERCOM.2018.8444578

[28] Virpi Kristiina Tuunainen, Olli Pitkänen, and Marjaana Hovi. 2009. Users’ Awareness of Privacy on Online Social Networking Sites - Case Facebook. In *22nd Bled eConference: eEnablement:Facilitating an Open, Effective and Representative eSociety, June 14-17, 2009*. Association for Information Systems, Bled, Slovenia, 42. http://aisel.aisnet.org/bled2009/42

[29] Petra Virjonen, Jukka Keränen, Riikka Helenius, Jarkko Hakala, and OV Hongisto. 2007. Speech privacy between neighboring workstations in an open office-a laboratory study. *Acta Acustica united with Acustica* 93, 5 (2007), 771–782.

[30] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. “I Regretted the Minute I Pressed Share”: A Qualitative Study of Regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania) *(SOUPS ’11)*. Association for Computing Machinery, New York, NY, USA, Article 10, 16 pages. https://doi.org/10.1145/2078827.2078841

[31] Jeanne Whalen. 2020. Europe fined Google nearly $10 billion for antitrust violations, but little has changed. https://www.washingtonpost.com/technology/2020/11/10/eu-antitrust-probe-google/

[32] Yui-Pan Yau, Lik Hang Lee, Zheng Li, Tristan Braud, Yi-Hsuan Ho, and Pan Hui. 2020. How Subtle Can It Get? A Trimodal Study of Ring-Sized Interfaces for One-Handed Drone Control. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 63 (June 2020), 29 pages. https://doi.org/10.1145/3397319

[33] Pablo Pérez Zarazaga, Sneha Das, Tom Bäckström, Vishnu Vidyadhara Raju Vegesna, and Anil Kumar Vuppala. 2019. Sound Privacy: A Conversational Speech Corpus for Quantifying the Experience of Privacy. In *Interspeech 2019, 20th Annual Conference of the International Speech Communication Association, 15-19 September 2019*, Gernot Kubin and Zdravko Kacic (Eds.). ISCA, Graz, Austria, 3720–3724. https://doi.org/10.21437/Interspeech.2019-1172