Hey Google, Can I Ask You Something in Private?

The Effects of Modality and Device in Sensitive Health Information Acquisition from Voice Assistants

Eugene Cho
Donald P. Bellisario College of Communications
Penn State University
University Park, PA, USA
exc75@psu.edu

ABSTRACT

Modern day voice-activated virtual assistants allow users to share and ask for information that could be considered as personal through different input modalities and devices. Using Google Assistant, this study examined if the differences in modality (i.e., voice vs. text) and device (i.e., smartphone vs. smart home device) affect user perceptions when users attempt to retrieve sensitive health information from voice assistants. Major findings from this study suggest that voice (vs. text) interaction significantly enhanced perceived social presence of the voice assistant, but only when the users solicited less sensitive health-related information. Furthermore, when individuals reported less privacy concerns, voice (vs. text) interaction elicited positive attitudes toward the voice assistant via increased social presence, but only in the low (vs. high) information sensitivity condition. Contrary to modality, the device difference did not exert any significant impact on the attitudes toward the voice assistant regardless of the sensitivity level of the health information being asked or the level of individuals' privacy concerns.

CCS CONCEPTS

• Human-centered computing → **Interaction devices;** *Sound-based input / output*

KEYWORDS:

Conversational agent(s); Information sensitivity; Modality; Privacy concerns; Social presence; Virtual assistant(s), Voice assistant(s)

ACM Reference format:

Eugene Cho. 2019. Hey Google, Can I Ask You Something in Private? The Effects of Modality and Device in Sensitive Health Information Acquisition from Voice Assistants. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glagsow, Scotland, UK*. ACM, New York, NY, USA. Paper 258, **9** pages. https://doi.org/10.1145/3290605.3300488

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK.

ACM ISBN 978-1-4503-5970-2/19/05...$15.00.
DOI: https://doi.org/10.1145/3290605.3300488

1 INTRODUCTION

The rising popularity of voice-initiated virtual assistants such as Amazon's Alexa, Apple's Siri, and Google Assistant is driving competition among technology giants. These systems differentiate themselves from other virtual agents by mainly communicating via "voice" without specific non-verbal cues (e.g., facial expression) that are embedded in some virtually embodied conversational agents. While they are generally used for news, entertainment, and other practical information updates (e.g., weather, traffic), the increasing prevalence of voice assistants in individual households allows them to serve as a cost- and time-effective source to obtain day-to-day health information. In addition to being easily accessible at home, the capability of voice assistants to engage in conversations with users in a private yet natural manner is increasing voice assistants' value as healthcare assistants [33, 37]. Driven by such distinctive qualities of voice assistants, for example, Alexa is seen as one of the forefront technologies that could be used to promote personal health check-ups for rather sensitive health matters such as breast cancer [30].

Yet, the value of the intuitive and seamless voice interactions between users and voice assistants in health communication has not grasped much scholarly attention. Thus, this study aims to fill the gap by examining the modality effects of voice assistants in the context of health information acquisition. In one exceptional study, Miner et al. [17] have explored how different voice assistant systems (i.e., Siri, Google Now, S Voice, and Cortana) respond to health questions related to mental health, interpersonal violence, and physical health. Unfortunately, they arrived at a conclusion that the voice assistants in general were not yet capable of offering consistent and complete answers. However, not only are voice assistants getting smarter

through algorithmic learning, more research on user perceptions (beyond voice assistants' reactions) merits attention to better design voice assistants for the purpose of healthcare.

2 VOICE AS A HUMAN-LIKE FEATURE IN VOICE ASSISTANT INTERACTIONS

According to Sundar, Jia, Waddell, and Huang [34], modality is a powerful affordance that has significant impact on the interaction between users and digital media, in turn affecting user perceptions associated with the content as well as the media platform itself. Especially, voice could serve as a powerful modality affordance among audio-activated assistants, considering that many systems (e.g., Apple's Siri, Google Assistant, Microsoft's Cortana) allow users to choose input modality (e.g., voice vs. text), which generally corresponds to the output modality.

Earlier research in synthesized computer speech has documented how imbuing certain interpersonal communication attributes in human-computer interactions could enhance user evaluations toward computers and virtual agents based on the computers are social actors paradigm (CASA) [24, 25]. The CASA framework suggests that when virtual beings reveal human-like qualities, people will respond to non-human agents as they do to human actors [24, 25], "even when they know that machines do not possess feelings, intentions, 'selves,' or human motivations" (p. 325) [18]. For instance, previous findings show that computer agents are evaluated more positively by users when the agents are emotionally-adaptive (vs. non-adaptive) [15], match [19, 23] or complement [13] the personality of the users (e.g., extroversion, dominant characteristics), show reciprocity [18], or simply deliver friendly facial expressions (e.g., smiling vs. neutral) [6, 26].

When it comes to audio cues signaled by virtual agents, studies tended to focus on particular variations in voice output (e.g., emotional tone of voice [21], vocal pitch [6]), with stronger interests in embodiment and non-verbal cues beyond voice (e.g., facial expression), to explore how human-like virtual agents can be perceived by users [9]. On the other hand, the pure voice effects have gained less attention. However, we should note that simply conversing with a virtual agent through voice itself (compared to other means such as text messaging) could deliver a powerful anthropomorphic impression of machines to users [7]. Some may argue that such tendency would be too obvious to be tested, since speech is one of the most unique human capabilities that can be incorporated in computer agents [27]. Yet, we should not forget that texting (vs. traditional telephony) could also be considered as personal and intimate means to interact with other counterparts among modern day users [14]. Also, it is a good time to directly test if audio (vs. textual) interaction indeed affords more human-like experiences to users, considering that many of the recent voice assistant technologies offer various modality options (e.g., voice vs. text), albeit focused on voice exchanges by design.

Moreover, the highlight of modern-day voice assistants is that they are capable of having natural conversations with users even "without" human-like non-verbal or embodied cues. This necessitates the shift from putting more weight on embodiment or non-verbal behaviors of virtual agents to validating the simple but fundamental voice effects. While mostly tested in the context of human-to-human communication, Walther [38] proposed that individuals could rely on verbal cues to compensate for missing nonverbal components in technology-mediated interactions, which may also apply to human-to-computer conversations. In fact, Nass and Gong [22] found that people felt more comfortable sharing their personal information when a human-recorded (more human-like) voice was projected from a voice assistant with no face (i.e., disembodied), compared to when the human-like voice was heard from a virtually embodied agent with synthetic facial features. As virtual voices resemble those of humans more and more, speech from machines can serve as an impactful cue that signals human-like qualities of virtual agents, especially when other nonverbal cues are absent.

Admittedly, there have been several attempts to disentangle the effects of audio (vs. textual) interactions with (virtually embodied) conversational agents. Berry, Butler, and de Rosis [4] compared how exposure to (a) text-only information, (b) voice from no-face agents, and (c) voice from agents with face could result in disparate psychological consequences. The results indicated that text-only health information was easier to understand compared to voice messages from agents with or without face. However, in their study, the text message was shown to participants in the form of a word document, which may have not been acknowledged as being sourced from a virtual agent. Another study on avatar-mediated interactions revealed that text chat is inferior to other real-time interaction modalities including audio, video, and avatar-mediated communication in increasing emotional closeness toward the interactant [3]. In this case, however, participants were interacting with a person, not a computer agent, which may signal different relational meanings. Qui and Benbasat [31] also found that when a web-based

product recommendation agent spoke with a human voice (vs. TTS voice, text), social presence of the agent increased. One limitation of this study would be that participants were simply exposed to a web-based output with different modalities without personally interacting with the agent.

To fill the above gaps in research, we expect to better disentangle the effects of audio vs. textual interactions with voice assistants by manipulating the modality of the "dialogue" between the user and the virtual agent from the same virtual "source". In particular, this study hypothesizes that voice (vs. text) interaction will heighten feelings of anthropomorphic interactions in the form of elevated social presence (i.e., "the sense that other intelligent beings co-exist and interact with you, even if those beings are non-human and only seem intelligent;" pp. 289-290 [12]), which in turn will elicit positive evaluations toward virtual beings [13, 29].

3 MODERATING ROLES OF INFORMATION SENSITIVITY AND PRIVACY CONCERNS

When demanding information from smart devices through voice commands, the context of interaction as well as the characteristics of information could also influence how modality alters user perceptions. This is due to the possible intrusiveness associated with audio input/output, especially in public settings. For example, previous research on modality and mobile devices revealed that users tend to actively engage in voice interactions in private settings (e.g., home), compared to public settings (e.g., park, subway) where they seem to adopt more gesture- and touch-based interactions [32]. Similarly, gesture-based interaction with a mobile device was found to be preferred over voice-based interaction among users around strangers [40]. Specifically focused on voice assistants, Moorthy and Vu [20] found that users felt the use of mobile applications such as Siri more comfortable in private (vs. public) settings and for non-private (vs. private) information inquiries compared to a keyboard-based search. Their findings reflect how voice-based interaction could be subject to individuals' perception of privacy and sensitivity of the context.

Especially, for health information acquisition, privacy matters could become more crucial. Relevant to this point, Bansal, Zahedi, and Gefen [2] found that perceived sensitivity of health information and privacy concerns negatively affected users' intention to disclose health information on websites. When applied to voice assistants, it is reasonable to postulate that voice (vs. text) interactions may not always result in better evaluations toward the voice assistants, especially when users are in a socially/publicly sensitive context. For example, if a person is trying to acquire sensitive health information from a voice assistant out in public, s/he is likely to prefer interaction via text over voice. Certain personal traits, such as holding high concerns over online privacy breach, can also drive users to avoid voice interactions. On the other hand, text messaging can be further preferred among users in retrieving health information from mobile devices due to its convenience [1, 39]. In sum, while voice can heighten social presence of, thus leading to positive evaluations toward, voice assistants, it may not always result in positive outcomes especially when users solicit sensitive health information, or report high privacy concerns. Thus, this study attempts to test such possibilities by examining the moderating roles of the (a) sensitivity level of health information being solicited, and (b) user's level of general privacy concerns, in the effects of modality on social presence and user attitudes toward the voice assistant.

However, predicting the moderating effects of information sensitivity and privacy concerns may not be so simple, since reluctance toward voice interactions in "highly sensitive" contexts would not necessarily compromise perceived "social presence" of the voice assistant. That is, even when users feel uncomfortable speaking to voice assistants regarding certain health matters, it may not make the voice assistant less human-like. In contrast, in "low-sensitive" settings, voice (vs. text) is expected to enhance the social presence level, since users are not inhibited by sensitivity of the information or privacy concerns. Thus, the following hypotheses are proposed based on the prediction that voice will increase social presence, and in turn induce positive attitudes toward the voice assistant, but only in "low-sensitive" contexts in terms of (a) the (manipulated) sensitivity level of the retrieved health information and (b) the (self-reported) level of user's privacy concerns as an individual difference.

H1: Voice (vs. text) interaction will increase perceived social presence of the voice assistant, but only when the sensitivity of the requested information is low (vs. high).

H2: Voice (vs. text) interaction will increase perceived social presence of the voice assistant, but only among users with low (vs. high) privacy concerns.

H3: Voice (vs. text) interaction will indirectly increase positive attitudes toward the voice assistant, mediated by the levels of perceived social presence.

4 DEVICE EFFECTS IN VOICE ASSISTANT INTERACTIONS

In addition to modality options, some systems let users interact with voice assistants through multiple devices. For example, users can talk to Microsoft's Cortana via mobile phones as well as laptops. Google Assistant also offers services through both smart home devices (e.g., Google Home) and smartphones. Such device difference can also have impact on user perceptions. Especially, for smart home devices, users become solely reliant on vocal cues compared to smartphones due to the absence of screen information. In the context of multimedia education, the cognitive load theory suggests that humans have limited amount of cognitive space to process verbal and visual information simultaneously [36], thus informing educators to recognize the cognitive load in multimedia learning environments [16]. If that is the case, only receiving audio feedback may enhance user experiences compared to obtaining identical verbal information through two different modalities (e.g., text + voice). Conversely, some may argue that the voice output accompanied by text information on the mobile screen may produce accumulative effects. This argument resonates with the dual channeling theory which posits that verbal information can be processed more effectively with complementary visual aids [28].

However, studying the multi-modality effects with voice assistants is not always clear for several reasons. First, the smartphone screen offers a complete replica of the vocal response simply in a text version, rather than offering a visual/pictorial representation of information. Second, device difference itself can elicit disparate psychological outcomes that does not solely derive from the modality combinations (e.g., due to higher perceived attachment to smartphones). Third, it is unclear how the device difference and the associated modality difference will affect users' attitudes toward the voice assistant beyond recall or learning outcomes. For the aforementioned reasons, the following research questions (in lieu of directional hypotheses) are proposed to explore the device effects on social presence and attitudes toward the voice assistant as well as the moderating effects of information sensitivity and privacy concerns:

RQ1: Will the level of perceived social presence alter by device difference (i.e., smart home device vs. smartphone)?

RQ2: Will the effects of device difference (i.e., smart home device vs. smartphone) on perceived social presence of the voice assistant be moderated by (a) information sensitivity and/or (b) privacy concerns?

RQ3: Will the device difference (i.e., smart home device vs. smartphone) indirectly increase positive attitudes toward the voice assistant, mediated by the levels of perceived social presence?

5 METHOD

5.1 Participants and Study Design

Fifty-three undergraduate students from an undergraduate course at a southeastern university were recruited for extra credit. The sample was slightly female-dominant (18 men, 35 women) with the age ranging from 19 to 23 (M = 20.15, SD = 0.93). A 3 modality/device (i.e., smart home device + voice, smartphone + voice, smartphone + text) X 2 information sensitivity (i.e., low vs. high) mixed factorial experimental design was employed. The modality and device conditions were combined since smart home devices are not capable of offering text output. In addition, while the modality/device assignment was conditioned as a between-subjects factor, information sensitivity was assigned to participants as a within-subjects factor. A particular voice assistant system developed by Google (i.e., Google Assistant) was adopted for this study. The rationale behind the decision was that not only did Google Assistant offer services through mobile devices via both voice and text inputs/outputs, but also through smart home devices without having a strong brand attachment to a certain device at the time of study (e.g., Amazon's Alexa/Echo is known more for its service through smart home devices).

5.2 Manipulated Conditions

After giving consent, each of the participants was randomly assigned to one of the three device/modality conditions (smart home device + voice: N = 18; smartphone + voice: N = 18, smartphone + text: N = 17), and entered a room with a table that had the assigned device on top. In terms of the device/modality manipulation, an Android mobile phone (i.e., Motorola G5) was used for the smartphone condition, and Google Home device for the smart home device condition. Participants were asked to use either audio input for the voice condition or textual input for the text condition. When users offered text input to the smartphone, text results appeared on the mobile screen (see Figure 1). When users prompted Google Assistant using their voice via the smartphone, identical screen information was given with the voice assistant reading the textual content at the same time. For the smart home device condition, only audio output was provided.

Afterwards, participants were specifically instructed to ask two types of health questions (i.e., less vs. more sensitive health questions) to Google Assistant in a randomized order. For the low sensitivity task, participants received a paper that had a list of 7 questions related to allergies, cold, and flu (e.g., "Can headache medication cause headaches?," "Could oatmeal baths help you with itchy skin?," "How long do cold symptoms last?"). For the high sensitivity information condition, participants were given 7 health questions related to sexual health (e.g., "Do you have to have sex to get an STD?," "What are the benefits of masturbation?," "Do condoms affect orgasms?"). After they went through all of the 7 health questions for each of the conditions, participants completed an online questionnaire containing the measured variables of interest using an iPad device.

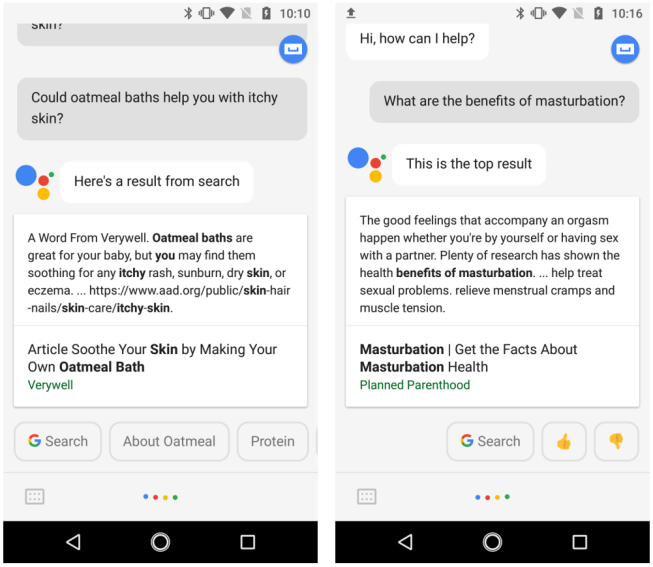

Figure 1: Smartphone screen output for low (left) and high (right) information sensitivity conditions.

5.3 Measured Variables

5.3.1 Manipulation Check for Information Sensitivity. Two items were created to test the effectiveness of the manipulation for information sensitivity: whether the set of questions asked and answers received from Google Assistant was (a) "too sensitive" or (b) "too private" (1 = strongly disagree, 7 = strongly agree). The 2 items were combined since they showed high reliability for both low (r = .81, p < .001; M = 1.92, SD = 1.10) and high (r = .91, p < .001; M = 2.83, SD = 1.58) information sensitivity conditions.

5.3.2 Privacy Concerns. Four items from Dinev and Hart [5] were used to measure the individuals' reported level of privacy concerns (e.g., "I am concerned that the information I submit online could be misused," "I am concerned about submitting information online because it could be used in a way I did not foresee"; α = .91, M = 4.35, SD = 1.58; 1 = strongly disagree, 7 = strongly agree).

5.3.3 Perceived Social Presence. Nine items from previous social presence scales [8, 12] were modified to fit the study context (e.g., "There was a sense of human warmth during the interaction," "While I was using Google Assistant, I felt as if she was talking to me," "I paid a lot of attention to what she said"; 1 = strongly disagree, 7 = strongly agree). The scale was reliable for both low (α = .94, M = 3.90, SD = 1.35) and high (α = .94, M = 4.08, SD = 1.31) information sensitivity conditions.

5.3.4 Attitudes toward the Voice Assistant. The attitudes measure was based on 21 items that were constructed to evaluate certain usability characteristics of online services [11, 35] (e.g. Good, Useful, Cool, Interesting, Smart; 1 = describes very poorly, 7 = describes very well), which also showed high reliability for both less (α = .97, M = 5.47, SD = 1.13) and more sensitive (α = .97, M = 5.47, SD = 1.08) health questions.

5.3.5 Control Variables. Gender and prior usage of voice assistants were included as control variables. Previous experience with voice assistants was measured in two levels, for both Google Assistant (M = 2.23, SD = 1.35), and other systems such as Siri and Alexa (M = 4.19, SD = 1.72) (1 = never heard of it; 2 = heard of it but have not used it; 3 = barely use it; 4 = have some experience with it; 5 = use it occasionally; 6 = use it often; 7 = use it all the time).

6 RESULTS

Before testing the main hypotheses, the effectiveness of the information sensitivity manipulation was confirmed. As expected, sexual health questions were perceived as more sensitive compared to general health questions related to allergies, cold, and flu (t(52) = 4.33, p < .001; M_{low} = 1.92, SE_{low} = .15; M_{high} = 2.82, SE_{high} = .22).

After confirming that the manipulation of information sensitivity was successful, the main analyses were run. In particular, to examine if voice would have stronger effects than text on perceived social presence of the voice assistant, but only when users inquired less sensitive information (*H1*), and among those who reported low levels of privacy concerns (*H2*), a 2 (modality: voice vs. text) X 2 (information sensitivity: low vs. high) mixed model repeated measures analysis of variance (ANOVA) was employed. To do so, the between-subjects effects of modality (i.e., voice vs. text), privacy concerns, and the interaction term (i.e., modality X

privacy concerns) on social presence were tested under the within-subjects contrast of information sensitivity (i.e., low vs. high). In addition, gender, prior history of using voice assistants, and the device condition (i.e., smart home device vs. smartphone) were included as control variables in the model.

As a result, a significant two-way interaction effect emerged between modality and information sensitivity on perceived social presence ($F(1, 45) = 5.69$, $p = .02$, partial $\eta^2 = .11$). As expected by *H1*, voice (vs. text) interactions significantly enhanced perceived social presence toward the voice assistant, but only in the low information sensitivity condition ($M_{text} = 3.50$, $SE_{text} = .37$; $M_{voice} = 4.15$, $SE_{voice} = .24$) (see Figure 2). In the high information sensitivity condition, the voice assistant was perceived fairly socially present regardless of the modality ($M_{text} = 4.10$, $SE_{text} = .37$; $M_{voice} = 4.11$, $SE_{voice} = .24$). For the moderating effects of privacy concerns (*H2*), a marginally significant effect appeared ($F(1, 45) = 3.32$, $p = .08$, partial $\eta^2 = .07$), which will be explained in more detail in the mediation analysis (*H3*).

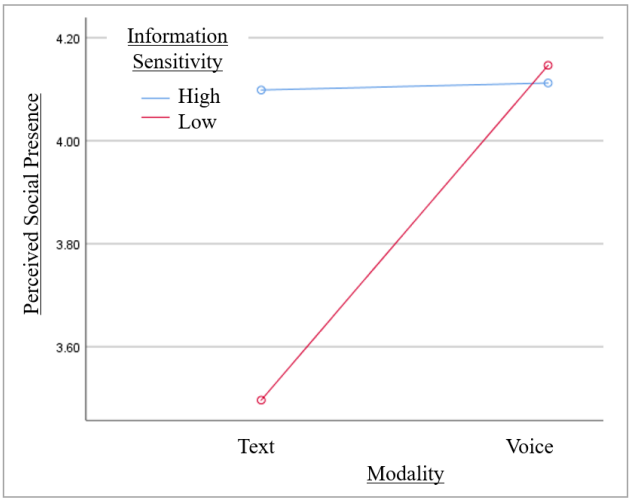

Figure 2: Interaction between modality and information sensitivity on perceived social presence.

To test the mediating effects of perceived social presence on the relationship between modality and attitudes toward the voice assistant in its entirety (*H3*), a moderated mediation analysis was run through Process (Model 7) [10] for low and high information sensitivity conditions respectively, with privacy concerns serving as a moderator between the independent variable and the mediator. Due to the insignificant effects of the control variables (i.e., gender, previous experience with Google Assistant and other voice assistants) on perceived social presence (*p*s > .85), the control variables were excluded from the moderated mediation analyses except for the device difference (i.e., smart home device vs. smartphone).

First, for the low information sensitivity condition, the entire moderated mediation model appeared to be significant (index = -.35, 95% biased-corrected 10,000 bootstrap CI[-0.6304, -0.1096]). Privacy concerns had a significant moderation effect with modality on social presence ($b = -0.58$, $t = -2.18$, $p = .03$), which in turn positively affected user attitudes ($b = 0.59$, $t = 6.90$, $p < .001$) (see Figure 4). More important, the indirect effects of modality on user attitudes via social presence appeared significant, but only among users who reported low privacy concerns (*M*–1*SD*), $b = .94$, 95% biased-corrected 10,000 bootstrap CI[0.2123, 1.9195], lending support to both *H2* and *H3*. On the other hand, no significant indirect effects of modality on attitudes through social presence were seen for users with moderate (*M*), $b = .39$, 95% CI[-0.1587, 1.0729], or high (*M*–1*SD*), $b = -.15$, 95% CI[-.7946, .6923] levels of privacy concerns. When the moderating effect of privacy concerns was decomposed, the pattern suggested that voice (vs. text) interaction led users to perceive the voice assistant more human-like (socially present), but only when they had little concerns over privacy (*M*–1*SD*), $b = 1.56$, $t = 2.55$, $p = .014$ (see Figure 3). Again, for participants who reported medium (*M*; $b = 0.64$, $t = 1.43$, $p = .16$) and high (*M*+1*SD*; $b = -0.28$, $t = -0.45$, $p = .66$) levels of privacy concerns, modality did not alter perceived social presence.

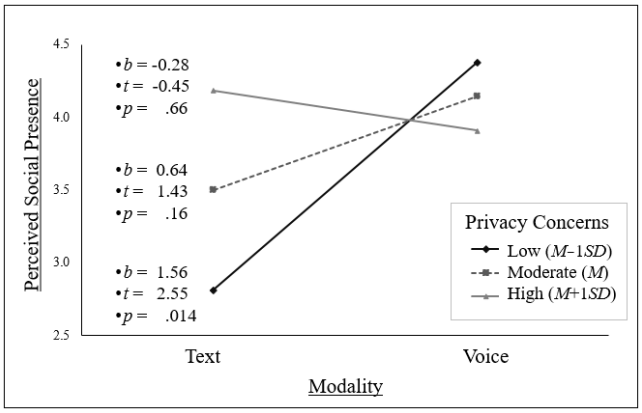

Figure 3: Interaction between modality and privacy concerns on perceived social presence.

Second, for the high information sensitivity condition, the effects of moderated mediation was not significant, index = -.26, 95% biased-corrected 10,000 bootstrap CI[-0.5805, 0.0167] (see Figure 5). That is, no significant mediating effects of social presence emerged between modality and user attitudes, regardless of users' level of privacy concerns: $b = .40$, 95% CI [-0.3232, 1.2195] for low, b < .001, 95% CI[-0.5854, 0.5602] for medium, $b = -.40$, 95% CI[-1.0926, 0.3099] for high levels of privacy concerns. The interaction effect of modality and privacy concerns on social presence was also non-significant ($b = -0.44$, $t = -1.65$, $p = .11$).

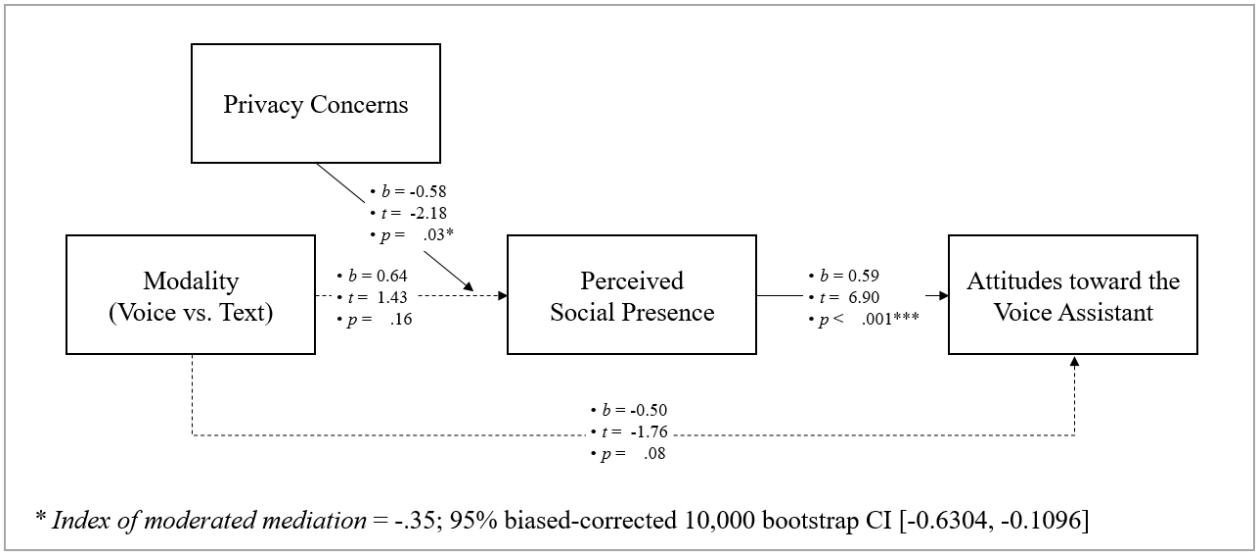

Figure 4. Conditional indirect effects of modality on attitudes toward the VA in the low information sensitivity condition.

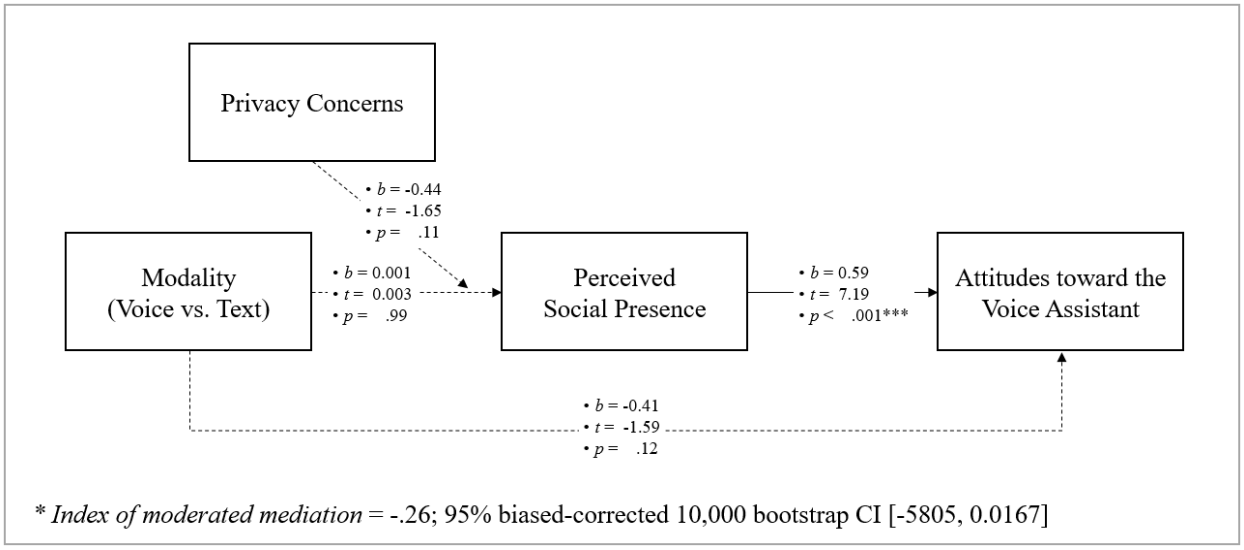

Figure 5. Conditional indirect effects of modality on attitudes toward the VA in the high information sensitivity condition.

When the device effects were tested, using the smart home device (vs. smartphone) did not show a significant main effect on perceived social presence (*RQ1*; $F(1, 45) = 0.001$, $p = .98$, partial $\eta^2 < .001$), even after controlling for modality and other control variables. Furthermore, the device difference was not moderated by information sensitivity (*RQ2a*; $F(1, 45) = 0.93$, $p = .34$, partial $\eta^2 = .02$) nor privacy concerns (*RQ2b*; $F(1, 45) = 0.25$, $p = .62$, partial $\eta^2 = .005$) to alter social presence. In turn, the mediating effect of social presence on the relationship between modality and user attitudes was not further tested due to the above non-significant results (*RQ3*).

In sum, the findings indicate that voice (vs. text) interactions increased social presence of the voice assistant, but only when the information sensitivity was low (*H1*), and the users reported less privacy concerns (*H2*). In turn, increased social presence led to more positive attitudes toward the voice assistant (*H3*), confirming all of the proposed hypotheses. On the other hand, when the information sensitivity and individual's privacy concerns were high, voice did not work better than text to alter social presence. In fact, text was as impactful as voice to induce human-like perceptions of the voice assistant in those cases. In addition, device difference did not show significant effects on social presence (*RQ1~3*).

7 DISCUSSION

Due to the drastic enhancement in naturalness of synthetic voices, voice assistants are now able to resemble some aspects of human-to-human interactions in terms of their speech capabilities. The findings suggest that voice (vs. text) interactions elevated the feelings of having a social conversation between the user and the voice assistant, which further led to positive evaluation toward the agent. However, those patterns occurred only when the information being asked was less sensitive in nature, and while individuals reported low levels of privacy concerns.

Notwithstanding the significant moderation effects consistent with the main hypotheses for the low information sensitivity condition (*H1*) among users less concerned over privacy (*H2*), the specific interaction patterns deserve attention. Instead of voice diminishing the social perceptions of the voice assistant in highly sensitive contexts (both in terms of information sensitivity and privacy concerns), text showed strong effects on perceived social presence as much as voice. It is likely that, in sensitive settings, interactions with a voice assistant become more engaging and stimulating to users to heighten social presence of the assistant regardless of the modality. In contrast, in the condition where users asked less sensitive questions, the effect of voice (vs. text) appeared significant, especially among users with low privacy concerns. It turns out that modality effects become more prominent in less (vs. more) sensitive settings, which offers noteworthy design implications. In particular, it seems that designers could benefit from focusing on voice-associated features when delivering general (non-sensitive) health information via voice assistants. Similarly, designers could also personalize modality options based on individuals' level of privacy concerns to improve user experiences. For instance, encouraging audio interactions among people with low concerns over privacy could increase perceived social presence of, in turn enhancing attitudes toward, the voice assistant.

On the other hand, the device difference did not exert powerful psychological impact on users. It seems that when voice-based applications are used, smartphones (vs. smart home devices) do not elicit more or less attention to visual outputs added to the audio responses. By the same token, smart home devices (vs. smartphones) did not encourage users to pay more attention to the solely offered audio cues without screen information. Perhaps, difference in input modality (i.e., voice vs. text) matters more than the output modality in affecting user perception toward voice assistants. Alternatively, it could be that for voice input in particular, there are no accumulative effects deriving from combination of modalities (i.e., text + voice) as long as the output modality has an audio component to it. Despite the non-significant effects of device difference on user perception and attitudes, the findings still offer some design implications considering that many smart home devices only offer voice interaction options. It seems that building systems that only allow audio input (without the textual input option) could deliver stable user outcomes regardless of the sensitivity level of the information being exchanged or individuals' privacy concerns, which reflects the advantages of using solely audio-interactive smart home devices. On the other hand, with the recent development of new smart home devices that incorporate screen features (e.g., Amazon's Echo Show), again, designers should consider the disparate effects of modality as a function of information sensitivity and privacy concerns. That is, while offering both voice and text input options would not make a difference in highly sensitive settings, designers should be more careful with adopting text interactions in less sensitive contexts.

Despite the possible theoretical and practical implications of the study, few limitations merit note. One limitation was associated with the nature of the experiment conducted in a controlled lab setting which restricts the generalizability of the findings. Furthermore, users' choice to retrieve sensitive (vs. non-sensitive) information is contingent upon various social and spatial contexts (e.g., public settings) in reality. Thus, studies comparing the modality effects among various contexts in natural environments may aid future researchers to extend the understandings of user psychology in interactions with voice-initiated conversational agents.

ACKNOWLEDGMENTS

The author would like to thank the anonymous referees for their valuable comments.

REFERENCES

[1] Adrian Aguilera and Clara Berridge. 2014. Qualitative feedback from a text messaging intervention for depression: benefits, drawbacks, and cultural differences. JMIR Mental Uhealth 2, 4: e46. http://doi.acm.org/ 10.2196/mhealth.3660

[2] Gaurav Bansal, Fatemeh M. Zahedi, and David Gefen. 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decision Support System 49, 2: 138-150.

[3] Gary Bente, Sabine Rüggenberg, Nicole C. Krämer, and Felix Eschenburg. 2008. Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. Human Communication Research 34, 2: 287–318.

[4] Dianne C. Berry, Laurie T. Butler, and Fiorella de Rosis. 2005. Evaluating a realistic agent in an advice-giving task. Itnl Journal of Human-Computer Studies 63, 3: 304-327.

[5] Tamera Dinev and Paul Hart. 2005. Internet privacy concerns and social awareness as determinants of intention to transact. Itnl Journal of Electronic Commerce 10, 2: 7-29.

[6] Aaron C. Elkins and Douglas C. Derrick. 2013. The sound of trust: Voice as a measurement of trust during interactions with embodied conversational agents. Group Decision and Negotiation 22, 5: 897-913.

[7] Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and humanrobot interaction. In Social Robotics. ICSR 2012. Lecture Notes in Computer Science, vol 7621, Shuzhi S. Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons, Mary-Anne Williams (eds.). Springer, Berlin, Heidelberg, 199-208.

[8] David Gefen and Detmar S. Straub. 2003. Managing user trust in B2B e-services. e-Service Journal 2, 2: 7-24.

[9] Jonathan Gratch, Jeff Rickel, Elisabeth Andre, Justine Cassell, Eric Petajan, and Norman Badler. 2002. Creating interactive virtual humans: some assembly required. IEEE Intelligent Systems 17, 4: 54-63

[10] Andrew F. Hayes. 2013. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York, NY: The Guilford Press.

[11] Sriram Kalyanaraman and Shyam S. Sundar. 2006. The psychological appeal of personalized online content in Web portals: Does customization affect attitudes and behavior? Journal of Communication 56, 1: 110-132.

[12] Kwan Min Lee and Clifford Nass. 2003. Designing social presence of social actors in human computer interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03). 289-296. http://doi.acm.org/10.1145/642611.642662

[13] Kwan Min Lee, Wei Peng, Seung-A Jin, Chang Yan. 2006. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in Human–Robot interaction. Journal of Communication 56, 4: 754-772.

[14] Rich Ling. 2004. The mobile connection: The cell phone's impact on society. San Francisco, CA: Morgan Kaufmann Publisher.

[15] Syaheerah L. Lutfi, Fernando Fernández-Martínez, Jaime Lorenzo-Trueba, Roberto Barra-Chicote, and Juan M. Montero. 2013. I feel you: The design and evaluation of a domotic affect-sensitive spoken conversational agent. Sensors 13, 8: 10519-10538.

[16] Richard E. Mayer & Roxana Moreno. 2003. Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist 38, 1: 43-52

[17] Adam S. Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Elini Linos. 2016. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Internal Medicine 176, 5: 619-625.

[18] Youngme Moon. 2000. Intimate exchanges: Using computers to elicit self-disclosure from consumers. Journal of Consumer Research 26, 4: 323-339.

[19] Youngme Moon and Clifford Nass. 1996. How "real" are computer personalities: Psychological responses to personality types in human-computer interaction. Communication Research 23, 6: 651-674.

[20] Aarthi E. Moorthy amd Kim-Phoung Vu. 2015. Privacy concerns for use of voice activated personal assistant in the public space. International Journal of Human–Computer Interaction 31, 4: 307-335.

[21] Christos N. Moridis amd Anastasios A. Economides. 2012. Affective learning: Empathetic agents with emotional facial and tone of voice expressions. IEEE Transactions on Affective Computing 3: 260-272. http://doi.ieeecomputersociety.org/10.1109/T-AFFC.2012.6

[22] Clifford Nass and Li Gong. 1999. Maximized modality or constrained consistency? In Proceedings of the AVSP 99 Conference. Retreived from http://www.isca-speech.org/archive_open/avsp99/av99_001.html

[23] Clifford Nass and Kwan Min Lee. 2000. Does computer-generated speech manifest personality? An experimental test of similarity-attraction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00), 329-336. http://doi.acm.org/10.1145/332040.332452

[24] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. Journal of Social Issues 56, 1: 81-103.

[25] Clifford Nass, Jonathan Steuer, Ellen R. Tauber. 1994. Computers are social actors. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94), 72-78. https://dl.acm.org/citation.cfm?id=191703

[26] Jay F. Nunamaker, Douglas C. Derrick, Aaron C. Elkins, Judee K. Burgoon, and Mark W. Patton. 2011. Embodied conversational agent-based kiosk for automated interviewing. Journal of Management Information Systems 28, 1: 17-48.

[27] Joseph P. Olive. 1997. The talking computer: Text to speech synthesis. In Hal's legacy: 2001's computer as dream and reality, David G. Stork (ed.). The MIT Press, Cambridge, MA, 101-130.

[28] Allan Paivio. 1986. Mental representations: A dual coding approach. Oxford, England: Oxford University Press.

[29] Eun Kyung Park and S. Shyam Sundar. 2014. Can synchronicity and visual modality enhance social presence in mobile messaging? Computers in Human Behavior 45: 121-128.

[30] Judith Potts. 2018. New technology is rapidly improving cancer care, and Alexa is at the vanguard. The Telegraph. Retrieved from: https://www.telegraph.co.uk/health-fitness/body/new-technology-rapidly-improving-cancer-care-alexa-vanguard/

[31] Lingyun Qiu and Izak Benbasat. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. Journal of Management Information Systems 25, 4: 145–182.

[32] Tiago Reism, Marco de Sá, and Luís Carriço. 2008. Multimodal interaction: Real context studies on mobile digital artefacts. In Haptic and audio interaction design, Antti Pirhonen and Stephen Brewster (eds.), Springer, Berlin, Germany, 60–69.

[33] Laura Stevens. 2017. 'Alexa, can you prevent suicide?' How Amazon trains its AI to handle the most personal questions imaginable. The Wall Street Journal. Retrieved from: https://www.wsj.com/articles/alexa-can-you-prevent-suicide-1508762311?mod=e2fb

[34] S. Shyam Sundar, Haiyan Jia, Franklin T. Waddell, and Yan Huang. 2015. Towards a theory of interactive media effects (TIME). In The handbook of the psychology of communication technology, S. Shyam Sundar (ed), John Wiley & Sons, Chichester, UK, 47-86.

[35] S. Shyam Sundar, Qian Xu, Saraswathi Bellur, Jeeyun Oh, and Haiyan Jia. 2011. Beyond pointing and clicking: How do newer interaction modalities affect user engagement? In Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA'11), 1477-1482. http://doi.acm.org/10.1145/1979742.1979794

[36] John Sweller. 1999. Instructional design in technical areas. Camberwell, Australia: ACER Press.

[37] Pragati Verma. 2018. Why voice assistants are gaining traction in healthcare. Samsung NEXT. Retrieved from: http://samsungnext.com/whats-next/voice-assistants-aihealthcare/

[38] Joseph B. Walther. 1992. Interpersonal effects in computer-mediated interaction: A relational perspective. Communication Research 19, 1: 52–90.

[39] Tyler Watson, Scot Simpson, and Christine Hughes. 2016. Text messaging interventions for individuals with mental health disorders including substance use: A systematic review. Psychiatry Research 243, 30: 255-262

[40] Julie R. Williamson. 2012. User experience, performance, and social acceptability: usable multimodal mobile interaction. Ph.D Dissertation. University of Glasgow, Glasgow, Scotland.