

Using Design to Increase Disclosure:

A Study of the Effects of Modality and Gender of Prompt on Willingness to Disclose

Erica Robles

Symbolic Systems Dept.
Stanford University
ewoka@stanfordalumni.org

Hilary Bienstock

Dept. of Communication
Stanford University
hilaryb@stanfordalumni.org

Marissa Treinen

Computer Science Dept.
Stanford University
marissat@cs.stanford.edu

Charles Heenan

History Department
Stanford University
heenan@stanfordalumni.org

Clifford Nass

Dept. of Communication
Stanford University
nass@leland.stanford.edu

Contact Information:

Dept. of Communication
Stanford University
Stanford, CA 94305-2050
Phone: 1-650-723-5499
Fax: 1-650-725-2472

ABSTRACT

Disclosure of personal information is valuable to individuals, governments, and corporations. This experiment explores the role interface design plays in maximizing disclosure. Participants ($N=100$) were asked to disclose personal information to a telephone-based voice user interface (VUI) in a 2 (recorded speech vs. synthesized speech) by 2 (gender of voice) by 2 (gender of participant) between-participants experiment; with male and female participant graphical user interface (GUI) control groups. GUI participants disclosed more than VUI participants. Recorded voice elicited more disclosure than a synthesized voice. There was a significant interaction between modality and gender, with greater disclosure for a female recorded voice or a male synthesized voice, showing a cross-over effect for gender of modality output.

KEYWORDS

Disclosure, Voice User Interface (VUI), Graphical User Interface (GUI), Synthetic Voice, Text-to-Speech (TTS), Recorded Voice, Personal Information, modality.

INTRODUCTION

Why are companies so eager to pay people to fill out questionnaires? Why are residents legally required to fill out the United States Census? Why are we willing to tell our close friends secrets, but not total strangers?

Personal information is a currency people trade to establish trust.[2] The degree of information exchange between individuals is a primary marker of the strength of their relationship.[1] For example, close friendships or relationships are often marked in terms of one's readiness to disclose information to another person.¹ Yet, this mutual disclosure arises as much from self-interest as from altruism: A person who discloses an item of personal information will expect something in return. This reciprocal "value" can take the form of returned friendship, an equally desirable piece of personal information, or a personal favor.

In the public and corporate spheres, personal information is no less valuable as a unit of exchange. The amount of information traded between an individual and the government or between an individual and corporations is also a marker of trust in that relationship. As with individuals, this information trade is driven by mutual self-interest. In the case of the U.S. Census, the government

¹ Disclosure is defined as the willingness to divulge personal information to another.

wants personal information about its population so that it can operate more efficiently, allocating resources where most needed. Similarly, corporations want personal information because it allows them to sell to customers more efficiently and to attract advertisers.

However, neither the government nor corporations can trade friendship for information. In these relationships, payment takes a more quantifiable form. Citizens are willing to disclose their personal information to the Census because they receive, as payment, influence over how government resources are allocated. Individuals are willing to divulge personal information to corporations in return for services, discounts, or actual cash payments. In a very real sense, an individual's personal information has a cost that can be defined in terms of what another individual, government, or corporation must "pay" in order to induce its voluntary disclosure.

In this paper, we demonstrate that *design* can reduce the "cost" of acquiring information from individuals. Specifically, we explore how the modality and design decisions of an interface affects users' willingness to disclose information.

Currently, information is gathered primarily through the use of textual surveys, both in paper format and over the Internet. Over the years, graphical user interfaces (GUIs) have been employed primarily because they were the most available and cost-efficient option. The advent of voice technologies, however, heralds the arrival of exciting new possibilities: Surveys conducted via voice may provide a useful alternative to the current text-based questionnaires in terms of eliciting disclosure.

MODALITY EFFECTS ON DISCLOSURE

GUI vs. VUI

The difference between GUIs and voice user interfaces (VUIs) is more than just a simple design choice. Voice response is more ephemeral than written response: A spoken answer to a question is psychologically more fleeting than a visual mark.[4] If participants are more willing to divulge information in an ephemeral situation, a VUI should induce more disclosure than a GUI. This result would suggest that perhaps the default interface for inducing disclosure should be vocal rather than visual.

On the other hand, voice tends to increase feelings of social presence [4, 8]. If people are more willing to disclose to a medium that will suggest indifference, then a standard GUI interface may be the appropriate design choice, because standard text questionnaires create a neutral environment for revealing information. Thus, a GUI may be the psychologically "safer" option that would elicit higher levels of disclosure.

Characteristics of VUIs

Recorded vs. Synthesized Speech

The most important distinction in VUIs is between those that use recorded speech and those that use synthesized speech or text-to-speech (TTS). A synthesized voice sounds unnatural and inherently unfamiliar, lacking the clarity and prosody of normal human speech. The robotic quality of the speech may emphasize its non-human origin, thus putting the user "on guard." As a result, respondents may be more willing to disclose to recorded voices that give the illusion of a more familiar social interaction.

Conversely, it may be that people are more willing to disclose to a synthetic voice precisely because of its inhuman qualities. Quite often, individuals are cautious about disclosing information which they may find embarrassing. People may be more comfortable divulging information to an entity that seems neutral and clinical, much like divulging information to a psychiatrist. They may connect the recorded voice to an interaction with a nosy stranger, but respond to the synthetic voice as a data collector that does not have a personal stake in the information.

Gender

Research has shown that both recorded speech and synthesized speech are assigned a gender by listeners [4, 8, 9]. These assignments can cue social stereotypes that influence user behavior [5]. For example, both genders will reveal more to a male voice or to a female voice regardless of the type of voice, responding to the stereotypes of non-critical acceptance and unshockability, respectively [9]. Conversely, people may feel more comfortable discussing certain personal issues with a voice of their own gender, whether recorded or synthesized, reflecting a tendency for social identification.[5].

EXPERIMENT

To separate out the various effects of GUI vs. VUI, gender of prompt, and gender of participant on level of disclosure, we created a ten-condition experiment. Because of the practical import of disclosure and the concern that the presence of an experimenter would have an independent and uncontrolled impact on disclosure, we opted for a field experiment rather than a laboratory experiment. Corporations and government entities do not often have the liberty of providing a small, intimate setting when seeking personal information. Laboratory experiments undermine the utility of studies of disclosure because they provide an artificially intimate environment. A field investigation provides the possibility that could directly apply the study's results with a higher probability of achieving a predictable outcome.

The experiment was framed as a simple survey. Participants first read a website with a consent form, and then, after they indicated their agreement, proceeded to the next page. Then, depending on condition, participants either completed the GUI version of the questionnaire or were directed to call a phone number to hear the voice version of the questionnaire. Except for the different interfaces, the questions in each of the conditions were identical and were asked in the same sequence. Following the initial questionnaire, every student used a standard GUI to complete an attitudinal questionnaire.

Method

Participants

All participants were university students, primarily between 18 and 24 years old. They received course credit for answering the survey. Participants were all native English speakers. Participants were randomly assigned to one of the five modality conditions (see below). Ten males and 10 females were assigned to each condition. All participants gave informed consent. Students were debriefed after the experiment during a class session.

Procedure

The experiment compared three different presentations of a questionnaire: a standard GUI, recorded speech, and synthesized speech. Within the voice interfaces, equal numbers of males and females heard either a male or female voice. Thus, the experiment was a 2 (presentation modality: synthesized speech vs. recorded speech) by 2 (voice gender) by 2 (participant gender) design, with two extra cells: males and females responding to a genderless GUI interface. The experiment had a balanced, between-participants design.

Participants received email notification that they had been selected, and were given a URL to access the experiment. They were also given a one-digit condition number which they inputted several times throughout the course of the experiment. Participants were able to complete the experiment anywhere they wished, and were instructed to do so in a room where they had telephone and Internet access, and where they could be alone.

The assigned URL for each subject was the starting page for either the text condition or a voice condition. Participants were taken to different pages depending on their assigned condition. Participants in the text condition went directly to the 63-item questionnaire, which they completed online, selecting their condition number from a pull-down box at the top of the page. Participants in the voice conditions went to a web page with instructions for calling the phone number of the voice experiment. When they called the specified number,

they immediately entered their condition number. After a brief introduction, they began the questionnaire.

The questionnaire itself consisted of sixty-three questions, all of the form, "Have you ever ...?" The questions inquired about the participant's sex life, drug and alcohol use, and other behaviors such as lying and cheating. The questions were modeled on those found in the 500-question Purity Test.²

To ensure a variety of invasiveness levels, the experimenters first rated questions from the original 500 question Purity Test on an invasiveness scale of 1-5, where 1 designated a banal subject and 5 was highly invasive. The questionnaire was designed to include questions from all five different invasiveness ratings, with an emphasis on questions with a rating of 3, 4, and 5. The set of questions was also chosen to cover a wide range of activities in which college students might engage. (A copy of the questionnaire can be found at www.ewoka.com/disclosure).

Banal questions included "Have you ever slow danced?" while a slightly more invasive question asked, "Have you ever forgotten events that occurred while you were drunk?" A still more invasive question was, "Have you ever read magazines or books to enhance your sexual technique?" The most invasive items on the questionnaire covered topics such as whether participants had ever had a sexually-transmitted disease.

Participants responding to the questionnaire were instructed to answer each question with the words "yes," "no," or "no answer." Students in the text condition answered by filling in the appropriate radio button next to each question, while students in the voice condition spoke their answer into the telephone receiver. If the voice-recognition software failed to recognize a student's answer, the student would hear an apology and the question would be repeated. In all conditions, the student was required to provide an answer to all questions before continuing to the next part of the experiment. No answers were associated in any way with a participant's identity and every effort was made to reassure participants of this fact.

After completing the survey, text participants clicked the "Proceed" button at the bottom of the web page. Voice

² The 'Purity Test' is an online questionnaire about a person's sex life, drug and alcohol use, and other behaviors. The test gives takers purity percentages based on how many behaviors they have not engaged in out of the total number of behaviors the test asks about. Taking this 'test' is common among college students in the United States, who may or may not share their results with others. The 500 question version of the test can be found online at www.ewoka.com/disclosure.

participants hung up the phone and clicked the “Proceed” button on the instruction page. Each button took the participants to a web-based attitudinal questionnaire, in which participants selected their condition number, and then were asked about their experiences completing the questionnaire.

Manipulation

The telephone interface was controlled by the CSLU Toolkit. The recorded speech conditions were obtained from a male and female graduate student. The synthesized speech was presented by the default male and female voices in the Toolkit. The Toolkit performed the recognition of the spoken utterances as well as a brief introduction and error messages. Pre-tests were used to ensure that the synthesized voices were intelligible and were clearly marked as male and female. [5]

Measures

We developed two behavioral measures based on the responses to the 63-item survey.

Refusal to Answer was an index based on a count of the number of “no answer” selections made by each participant. Higher levels of refusal to answer are associated with lower levels of disclosure.

Willingness to Admit was an index based on the number of “yes” answers selected by each participant. Although some participants showed their lack of disclosure by refusing to answer the questions, others exhibited a lack of disclosure by giving the socially desirable “no” response rather than the (possibly) truthful “yes” response. This is a standard strategy in exploring disclosure [7]. Thus, participants with more “yes” answers were considered to be more disclosive.

Because attitudes and behaviors are sometimes not aligned [1], we also determined how invasive participants perceived the questionnaire to be. *Perceived Invasiveness* was an index created by averaging participants’ answers to five questions from the attitudinal questionnaire, answered on ten-point Likert scales: How invasive the questionnaire seemed and how embarrassed, shocked, stressed, and relaxed (reverse-coded) the participants felt. The index was reliable (Cronbach’s $\alpha = .59$).

RESULTS

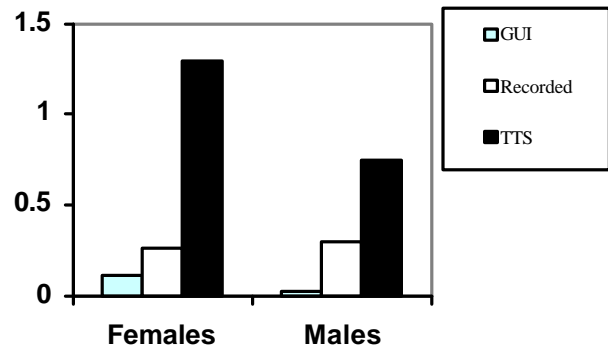
Modality and Gender of Participant

To assess the effects of modality on disclosure, we performed a two-way ANOVA with modality and gender of participant as the between-participants factors.

Modality had a significant effect on participants’ refusal to answer, $F(2,92) = 3.66, p < .03$ (see Figure 1). GUI participants were significantly more willing to disclose

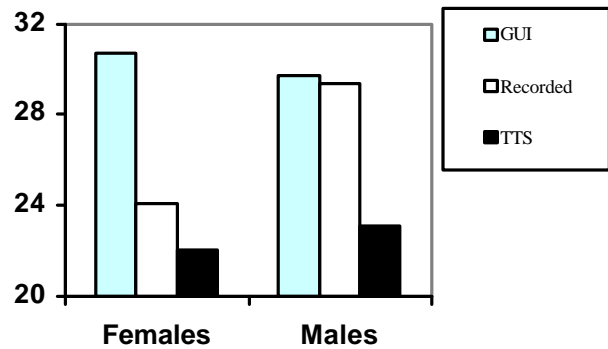
than TTS participants; there was no difference between recorded speech and the other two conditions. There was no effect for gender of participant, and no interaction.

Figure 1: Refusal to Answer: Modality by Gender



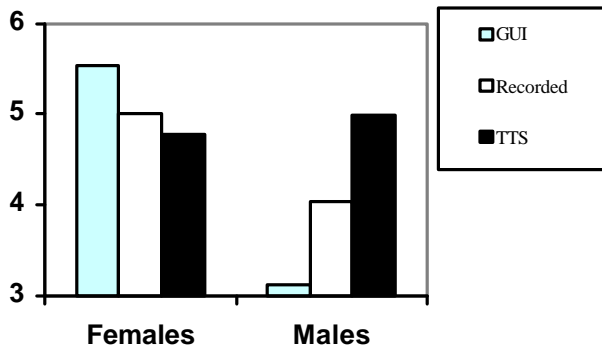
A similar pattern emerged for willingness to admit (see Figure 2). Once again, GUI participants were significantly more willing to disclose than were TTS participants, $F(2,92) = 3.05, p < .05$, while recorded speech participants were intermediate and not significantly different than the other two groups. Gender and the gender-by-modality interaction were not significant.

Figure 2: Willingness to Admit: Modality by Gender



The results for perceived invasiveness were quite different than would be suggested by the disclosure behavior (see Figure 3). In this case, there was a significant gender of participant by modality interaction, $F(2,92) = 4.42, p < .02$. Post-hoc tests demonstrate that the interaction was being driven by the GUI and TTS conditions, such that females perceive the GUI to be more invasive than TTS, while males felt the opposite. There was also a main effect for gender, such that females perceive the survey as more invasive than do males, $F(1,92) = 9.43, p < .001$.

Figure 3: Perceived Invasiveness: Modality by Gender

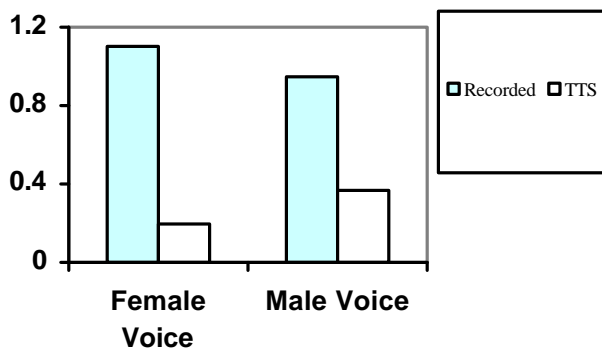


Gender of Voice and Gender of Participant

To analyze the effects of gender of voice, we must exclude the GUI conditions and analyze the data as a 2 (recorded speech vs. TTS) by 2 (female vs. male participant) by 2 (female vs. male voice) ANOVA.

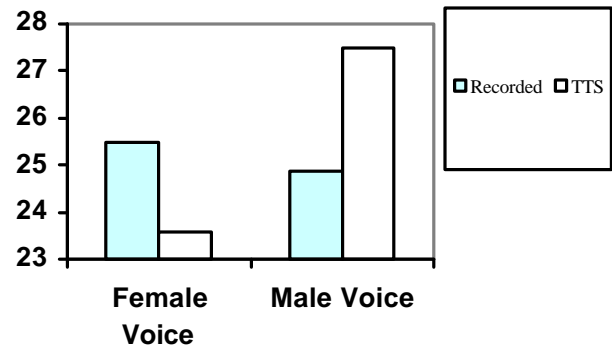
Synthetic speech participants refused to answer significantly more than did recorded speech participants (see Figure 4), $F(1,72) = 4.05, p < .05$, in contrast to the non-significant results in the previous analysis (a result of decreased power). There were no other significant effects.

Figure 4: Refusal to Answer: Type of Voice by Gender of Voice by Gender of Participant



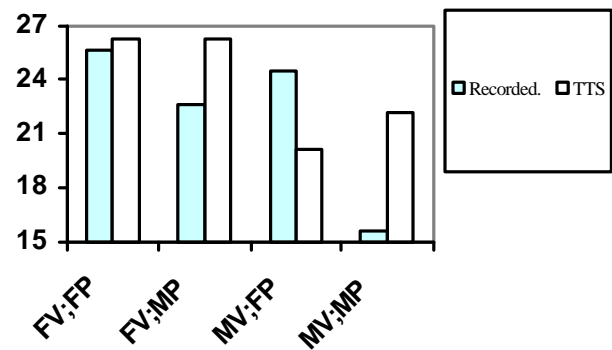
For willingness to admit, there was a significant interaction between type of voice and gender of voice, $F(1,72) = 3.96, p < .05$ (see Figure 5). For female voices, participants were more willing to admit inappropriate behaviors to synthesized speech as compared to recorded speech, while for male voices, participants were more willing to admit to recorded speech than synthesized speech. There were no other significant effects.

Figure 5: Willingness to Admit: Type of Voice by Gender of Voice by Gender of Participant



There was a significant interaction between the type of speech and the gender of the participants, $F(1,72) = 4.44, p < .04$, with respect to perceived invasiveness (see Figure 6). Males felt that TTS was more invasive than recorded speech, while females felt that recorded speech was more invasive. There was also a main effect for gender of the voice, $F(1,72) = 7.86, p < .01$, with the female voice seeming more invasive than the male voice.

Figure 6: Perceived Invasiveness: Type of Voice by Gender of Voice by Gender of Participant



DISCUSSION

The present results demonstrate that disclosure, traditionally believed to be affected primarily by reinforcement, can be manipulated by design. Specifically, the level of disclosure was affected by the choice between a GUI and a TTS VUI, between a synthetic voice and a recorded voice, and between a male voice and a female voice. Design proves to be an economical opportunity to increase the probability of obtaining data and to improve the quality of the data that is obtained.

As described above, a major objective of e-commerce companies and other entities both on and off the Internet is the collection of personal information. As technology advances, designers continue the search for ever-more

effective ways to accomplish this goal. This study suggests a number of considerations that designers should take into account when employing the new technology of voice interfaces to gather information about individuals.

The above results indicate that a standard GUI is an effective means of gathering information from users, and is actually more effective than TTS. However, GUIs are not universally better: Females perceive them as more invasive. There is no difference between GUIs and recorded speech, but given the greater expense and complexity of the latter, and the fact that there was a consistent (though not significant) pattern favoring GUIs over recorded speech, GUIs are clearly the preferred methodology, except when there are issues of literacy or ubiquity (telephone systems are more widespread than web or computer access).

The gender of the user does affect people's *feelings* about the invasiveness of surveys, but does not seem to affect disclosure *behavior*. Thus, if the primary concern is the quality and quantity of information, then one need not worry about the gender of participants.

If designers choose to implement voice interfaces, they should take special care in casting the voice. The results suggests that users are more likely to divulge information to a female recorded voice than a male recorded voice. Users are also more disclosive to a male synthesized voice than to a female synthesized voice. It is clear from these results that voice matters. The question of why still remains, and should be resolved by later research.[8]

This study tested situations in which the questionnaire was presented and answered in the same modality: text with text or voice with voice. Studying GUI presentation with voice response or auditory presentation with text response may also reveal interesting effects on disclosure. Such results would be invaluable to designers concerned with accessibility for the disabled or those taking advantage of alternative and novel technologies.

In addition to modality, researchers might study how different output technologies could affect disclosure. This study used a personal computer for the GUI condition, and a telephone for the VUI condition. Other media could be employed, however, such as a voice which is presented by a personal computer. A telephone is generally regarded as a "dumb" medium that cannot store answers; however, it is also often used to reinforce relationships through the disclosure of information. Computers have different associations for users, and the difference in users' mental models could affect disclosure as well. Other media, such as hand-held devices, may also be used to manipulate levels of disclosure.

Designers can also consider how a various combinations of modalities may manipulate disclosure. For example, one could study the effects of synthetic faces [6] and embodied conversation agents [3] on disclosure. There are other variants as well: Participants watching themselves disclose information, hearing a playback of the supplied information in their own voice or in another's, hearing different voices asking different questions, having famous voices (e.g., James Earl Jones) ask the questions, comic chat, in which an individual is *represented* by a computer character, or personalizing their questionnaire experience by choosing the presenter.

This study shows that disclosure is affected by obviously gendered voices presenting identically worded questions. Research suggests that men and women ask questions differently [9]; researchers should explore the interaction between gender of voice and the *way* that questions are asked.

One limitation of this study is that the only permitted answers were "yes" or "no" (or "no comment"). In many cases, individuals must disclose richer content than can be acquired by a simple yes/no construct. Future research should examine whether the effects observed here are stronger or weaker when users must elaborate rather than provide a single syllable.

While concerns about privacy might make manipulation of disclosure via design morally suspect, methods for increasing disclosure have many pro-social aspects. For example, it is possible that more males would be more comfortable with, and hence complete, the United States' Census if they were given the option to complete it with a synthetic voice on the telephone. Doctors may receive more complete and accurate information from patients if they present their on-line medical questionnaires by using the face of a doctor.

Increasing the quantity and quality of user information has generally proven to be a frustrating experience in which one must constantly increase the rewards to an ever-more-jaded user community. One of the mantras of the design community is "know your users"; the present research demonstrates that proper design can facilitate this process.

REFERENCES

1. Bandura, A. 1997. *Self-efficacy: The exercise of control*. San Francisco: W. H. Freeman and Co.
2. Blau, P. 1986. *Exchange and power in social life*. New York: Transactions Publishers.
3. Cassells, J., Sullivan, J., Prevost, S., & E. Churchill (Eds.). 2000. *Embodied conversational agents*. Cambridge, MA: MIT Press. Clark, H. 1996. *Using language*. New York: Cambridge University Press.

4. Clark, H. 1996. *Using language*. New York: Cambridge University Press.
5. Lee, E. J., Nass, C., & Brave, S. 2000. Can computer-generated speech have gender? An experimental test of gender stereotypes. *CHI 2000 Extended Abstracts* (pp. 289-290), Amsterdam, Netherlands.
6. Massaro, D. W. 1997. *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
7. Moon, Y. 1998. Impression management in computer-based interviews: The effects of input modality, output modality, and distance. *Public Opinion Quarterly*, 62(4), 610-622.
8. Nass, C. & Gong, L. (2000). Social aspects of speech interfaces from an evolutionary perspective: Experimental research and design implications. *Communications of the ACM*.
9. Tannen, D. 1996. *Gender and discourse*. New York: Oxford: Oxford University Press.