When the Tissue Box Says "Bless You!": Using Speech to Build Socially Interactive Objects

Haiyan Jia
Media Effects Research Lab
College of Communications
Pennsylvania State University
University Park, PA 16802, USA
hjia@psu.edu

Mu Wu
Media Effects Research Lab
College of Communications
Pennsylvania State University
University Park, PA 16802, USA
mxw5142@psu.edu

Eunhwa Jung
Media Effects Research Lab
College of Communications
Pennsylvania State University
University Park, PA 16802, USA
eoj5032@psu.edu

Alice Shapiro
Learning and Performance Systems
Pennsylvania State University
University Park, PA 16802, USA
ars301@psu.edu

S. Shyam Sundar
Media Effects Research Laboratory
College of Communications
Pennsylvania State University
University Park, PA 16802, USA
&
Department of Interaction Science
Sungkyunkwan University
Seoul 110-745, Korea
sss12@psu.edu

CHI 2013 Extended Abstracts, April 27 – May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

Abstract
From the Internet of Things to ubiquitous computing, smart objects are everywhere and have become a significant part of the information supply chain. However, these objects remain invisible to end-users mostly because they do not interact with them. Our project is devoted to brainstorming different design possibilities for building interfaces for these smart objects. This paper explores one such possibility—outfitting the object with a speech interface. Study participants (N = 63) witnessed the experimenter sneezing, followed by a "Bless You" from either a nearby tissue box, a robot in the room, or a person in the room. Surprisingly, users found the speaking tissue box to be as social and agentic as a humanoid robot and a human. We also found significant moderating effects of users' preference for consistency, parasocial tendency and power usage. Participants who scored high on these traits were more likely to regard the study object as intelligent and likeable. Users also tended to show the same non-verbal reactions to the tissue box as they would to a human or a robot.

Author Keywords
Smart object; robotics; speech; agency; socialness.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human factors; design; measurement; experimentation.

Introduction

As smart objects become part of our daily lives, it is important to ensure that they interact with human users rather than simply operate autonomously. Studies have revealed that the lack of direct methods of interaction with smart objects is detrimental to user experience [2]. There is an urgent need to design socially adaptable objects and robots, for social situations and to cater to users' need for emotional support. We propose speech-based interfaces as a solution. Speech is a natural form of interaction that occurs in social situations, and can be an effective addition to the design of robotic objects

Speaking is Social

According to "Computer as Social Actors" (CASA) research [7], the tendency to apply social rules and expectations to mediated presentations is heightened when machines or objects are designed with cues signaling social characteristics. In fact, human-robot interaction (HRI) scholars have suggested adding verbal and/or nonverbal cues of expressions and emotions in order to imbue a sense of socialness into robots [1]. If speaking is indeed an influential social cue, the addition of speech to an object that otherwise lacks social cues is likely to result in positive user perceptions of the object, in terms of socialness, friendliness, and related evaluations, thus making it equivalent to a social robot (*Hypothesis 1*).

Speaking is Agentic

Of course, most social robots have the advantage of an anthropomorphic morphology. Although appearance may compete with speech as a social cue for determining agency [5], adding a speech interface could be an effective method of solving the agency negotiation between a human user and a smart object. HRI literature has shown that speech alone may give rise to users' attribution of "sourceness" and agency for robots (as opposed to the programmer, or other entity controlling the robot) [8]. In fact, CASA studies have shown that different voices coming from the same computer box are treated as distinct agents [6]. Therefore, even when the object does not possess human morphological characteristics, the addition of a speech interface by itself can imbue agency to the object. It is likely that users will perceive a high level of humanness and intelligence in their interaction with the smart objects (*Hypothesis 2*).

Smart objects that are unable to communicate directly with users may be perceived as agentic, despite a relative lack of social cues, because of their ability to automatically capture, store, and transmit data among one another [3]. However, such automaticity may be perceived as a threat to user privacy and agency due to limited user control over personal data and the data transaction process. This creates an agency-negotiation paradox between the convenience brought by automaticity and the need for human-user agency. Previously studied methods to counterbalance negative effects of agency in interface design include adopting social cues that highlight user involvement and user control. These cues have been shown to reduce the perceived assertiveness and dominance of the robot/machine [2]. Instead, a speaking voice is socially

Figure 1: Laboratory set-up in pretest. The robot, the tissue box, the human confederate (female) are shown. The person standing is the experimenter who sneezed. Participant would be seated in front of the laptop.

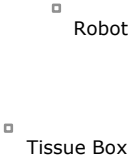

Figure 2: Due to confusion regarding source of the sound, the robot and the tissue box were placed at opposite sides of the room in the actual experiment. The robot was to the left of the participant while the tissue box was to the right.

relatable so that users will perceive a high level of reciprocity in the interaction. Therefore, the addition of speech interfaces to smart objects will not only provide agency to those objects but also to users (*Hypothesis 3*).

To test our hypotheses, we conducted an exploratory experiment to empirically test the speech affordance in (HOI). Our broad goals are to determine first-cut how user perceptions and user experience.

Research Design

We conducted a between-subjects experiment wherein study participants (63 undergraduate students) were randomly assigned to one of three conditions: 1. robot, 2. tissue box, and 3. human interactant (control condition), with 21 participants in each condition. The majority (67%) of the participants were female, with an average age of 20 years (SD = 2.60).

As a cover story, participants were told that the study was about cognitive games and were asked to visit the research laboratory one person per experimental session. Upon arrival, the participant would see a 13.5-interactant for Condition 3) situated at noticeable places in the laboratory (see Figures 1 and 2), while the experimenter greeted them and introduced the study.

Participants first filled out an online pre-test questionnaire. Then, the experimenter started to explain the rules of an online cognitive game (Tower of Hanoi), interrupted by a sneeze. The sneeze, fabricated but believable, triggered consolation ("Bless you!" and "Here, have a tissue.") from the robot in one condition, the tissue box in another condition, or the research assistant in the control condition. Once the experimenter took a tissue, it/she went on to say, "Take care!" while the participants observed. In practice, both the robot and the tissue box were equipped with a blue-tooth speaker that was operated by a third researcher monitoring the experiment through a one-way mirror. In the human interactant condition, the research assistant would utter the same words as the robot and the tissue box. The experimenter would complete the game introduction and time the game session. After the participants finished the game or went on for five minutes, they were asked to complete the post-test questionnaire. Upon completion of the questionnaire, they were debriefed and thanked by the experimenter. Each experimental session took about 30 minutes to complete.

To ensure consistency across the three conditions, all greetings were recorded or said in a female human voice. The research assistant avoided any verbal or non-verbal interaction with the participants throughout the experimental session. Participants' reactions towards the social interactions between the experimenter and the interactant were video-recorded.

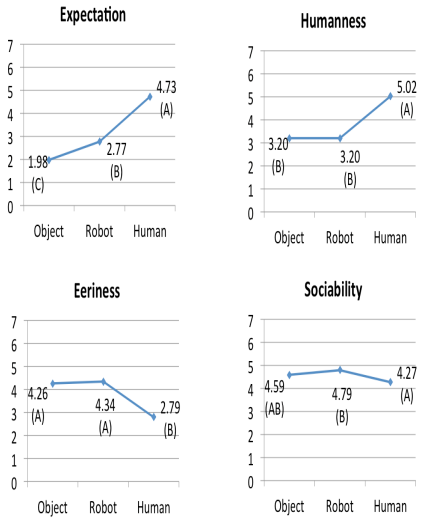

Figure 3. Cross-condition comparison: Condition means that do not share a letter differ according to Tukey HSD test, at $p < .10$.

Measurement

The pre-questionnaire measured user characteristics that may alter their perceptions, including *usage of technologies*, *power usage*, *preference for consistency*, *anthropomorphic tendency*, *socialness*, *parasocial tendency*, *curiosity*, *attitude towards automated communication*, *immersive tendency* and *demographics*. As a manipulation check, participants were asked to identify the source of the speech (SOS). The post-test questionnaire started with filler questions asking about participants' enjoyment and immersion while playing the cognitive game. This was followed by open-ended questions asking about the participants' experience during the experimental session in terms of recall, evaluation of lab equipment, and overall experience. A series of mediating variables, including *attitude toward the experimenter*, perception of the interactant being *sociable*, *attractive*, *friendly, humanlike*, *intelligent*, *uncanny*, *reciprocal*, and *easy to use* (applicable only to object and robot conditions), were measured on a 7-point Likert-type scale, ranging from 1 being "strongly disagree" to 7 being "strongly agree." Dependent variables, including *unexpectedness*, *attitude toward the interaction with the SOS*, and *behavioral intention toward SOS*, were also measured using Likert-type scales. Videotaped user reactions were coded as behavioral measures.

Results

Data from this exploratory study revealed interesting findings. Observational data showed source attribution behaviors in all three conditions: participants oriented towards SOS, some even smiling and/or nodding, upon hearing the greetings. Participants in the object and the robot conditions showed more attention (facial expressions such as eye opening, eyebrow raising, mouth opening; verbal comments such as "That is cool!" and "Where did that come from?") than those in the human condition.

One-way analyses of variance with self-reported user-perception data showed that the speaking tissue box was perceived as friendly, attractive, intelligent, reciprocal, and nearly as sociable as a robotic or human interactant. The human condition only differed from the other two in terms of humanness, unexpectedness and eeriness (being uncanny), while the difference between the other two conditions lacked statistical significance (see significant and near-significant results in Figure 3). In general, these results lend support to our hypotheses.

When user characteristics were taken into consideration, preference for consistency turned out to be a significant moderator of user perceptions of the interactant, such as attractiveness, humanness and intelligence. More specifically, participants who generally prefer consistency found the talking tissue box to be quite unexpected, but at the same time, more attractive, humanlike and more willing to interact with it, compared to those who rated low on preference for consistency (see Figure 4 for the significant interaction results).

Perceived intelligence of the interactant was significantly influenced by individual characteristics including a preference for consistency, parasocial interaction tendency, and power usage. A significant interaction effect showed that participants higher in parasocial tendency regarded the tissue box as being much more intelligent than their peers who are lower on parasocial tendency. Similarly, power users rated

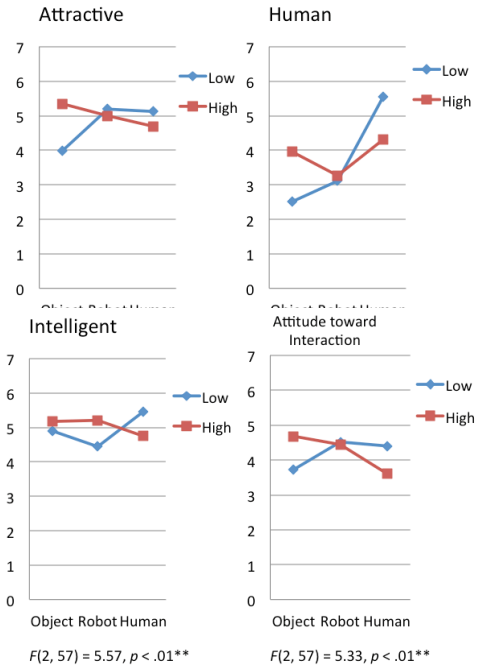

Figure 4. Interaction effects of preference for consistency and interactant type. For the purpose of illustration only, preference for consistency is coded as a dichotomous variable based on median split.

the tissue box and robot as more intelligent than non-power users.

Since participants were only observers of the interaction and had no pre-existing relationship with the experimenter who was consoled, it is possible that they were not personally involved in the social scenario. To test this possibility, participants' attitude towards the experimenter was treated as a moderating variable. Interaction effects showed that, in the object condition, participants who reported higher liking of the experimenter also found the tissue box more social ($p < .01$), attractive ($p < .01$), intelligent ($p < .01$), and even more humanlike ($p < .01$), and they were more willing to interact with it, given the opportunity ($p < .01$), than those who reported lower liking of the experimenter.

Discussion and Implications

From our preliminary study, comparisons of different interactants that ontologically differed in humanness and anthropomorphic attributes but which shared a similar speech affordance did not yield differential results, in line with our expectations.

The CASA predictions are supported by the findings. Videotaped recordings show that people applied human-human social rules in their stance toward the objects. The positive response would be amplified if an individual were highly involved in the social situation (i.e., high level of parasocial tendency, liking for the experimenter). This shows promise that, with a direct or prolonged relationship between the individual and the smart object, such social interactions would lead to more positive results, cognitively, emotionally and behaviorally. Findings confirm theories and research that suggest that humans are "voice-activated," responding intuitively and socially to speaking objects as if they are actual human beings.

The positive effects may result partially from the novelty effect, given that people who found the situation surprising also showed positive attitudes and behavioral intentions. However, novelty may result from the social nature of the interaction rather than the speech affordance. This suggests a design principle for smart objects and robots in a social context: speech should be implemented as an interface feature with a view to boosting the perceived socialness of objects. Instead of using modalities such as text, visual cues and regular auditory alerts that are generally less social ways of communicating with users, incorporating meaningful speech may be an ideal replacement or addition in design. Such a rule might also be applicable to other similar social, anthropomorphic cues, which could also trigger innate social responses from users as well as improved user perceptions and attitudes. Moreover, future design and evaluations of HRI and HOI should also take into account key perceptual and psychological factors discovered in this study, such as robotic/object agency, relatedness, and socialness.

A noteworthy finding is that power usage positively moderated the effect, indicating that, for users familiar with such technical novelty, the effect of socialness in this interaction is even stronger. Still, it will be interesting to test, possibly in a longitudinal setting, when the novelty effect wears off, how people would react to speech affordance in a social interaction with objects and machines. Uttering the appropriate social expressions for specific situations and scenarios promises not only to attract attention of individuals

preferring consistency, but also to command positive reception from power users, thus showing the potential for long-lasting effects after the innovation becomes widely accepted.

Our study is highly exploratory. It tests a novel, unconventional method of interaction between humans and objects, and shows intriguing results. It broadens the CASA model into interactions with everyday objects and provides both theoretical and design implications for HRI and HOI. As the technology of Internet of Things (IoT) becomes pervasive, we will likely see a proliferation of smart objects that are not only equipped with relevant information but also receptive to human input. The HCI community would do well to address the interaction potential of these objects by designing appropriate affordances that contribute to satisfactory, indeed enjoyable, user experiences with them.

Ongoing work at our lab is focused on comparing the speech affordance with other modality options, including visual cues. Also, the social scenario in this study was a rather simple one. To test the holistic effects of dialogue capabilities, a variety of social settings need to be constructed, which would involve more complex language processing.

Acknowledgements

The first three authors were supported by Summer Research Grants awarded by the College of Communications at Penn State University. The last author was supported by a grant (R31-2008-000-10062-0) from the World-Class University program of the Korean Ministry of Education, Science and Technology, awarded to Sungkyunkwan University (where he is serving as a visiting professor of interaction science).

References

[1] Holzapfel, H. A dialogue manager for multimodal human-robot interaction and learning of a humanoid robot. *Industrial Robot: An International Journal 35,* 6 (2008), 528–535.

[2] Jia, H., Wu, M., Jung, E., Shapiro, A. and Sundar, S. S. Balancing human agency and object agency: An in-depth interview study of the Internet of Things. Proceedings of the *International Workshop on Digital Object Memories for the Internet of Things, DOMe-IoT 2012 at the 14th International Conference on Ubiquitous Computing* (Ubicomp 2012), Pittsburgh, PA.

[3] Joinson, A. N. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anatomy. *European Journal of Social Psychology 31* (2001), 177-192

[4] Mayer, R. E. and Moreno, R. Aids to computer-based multimedia learning. *Learning and Instruction 12*, 1(2002), 107–119.

[5] Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M. and MacDorman, K. F. A mismatch in the human realism of face and voice produces an uncanny valley. Perception *2*, 1 (2011), 10-12.

[6] Nass, C. and Steuer, J. Voices, boxes, and sources of messages. *Human Communication Research 19* (1993), 504–527.

[7] Reeves, B. and Nass, C. *The media equation: How people treat computers, televisions, and new media like real people and places*. Cambridge University Press, 1996.

[8] Sundar, S. S. and Nass, C. Source orientation in human-computer interaction: Programmer, networker, or independent social actor? *Communication Research 27,* 6(2000), 683-703.