
Can Machines Call People? – User Experience While Answering Telephone Calls Initiated by Machine

Adam J Sporka

University of Trento
Via alla Cascata, 56/C
Trento, 38100 Italy
adam.sporka@disi.unitn.it

Jakub Franc

Sun Microsystems Inc.
V Parku 2308/8
Praha 4, 14800 Czech Republic
jakub.franc@sun.com

Giuseppe Riccardi

University of Trento
Via Sommarive, 14
Trento, 38100 Italy
riccardi@disi.unitn.it

Copyright is held by the author/owner(s).
CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.
ACM 978-1-60558-247-4/09/04.

Abstract

Current state-of-the-art spoken dialog systems are aimed at handling telephone calls to automate incoming caller requests. In this paper we explore a scenario which is symmetric to a traditional human-initiated interaction. We report on an initial qualitative study focusing on the opposite type of interaction, i.e. when automated agents place telephone calls to recipients that request interactive dialog from the recipients. The study consisted of 16 telephone calls to participants placed by a simulated agent, followed by debriefing interviews with the participants. The data gained in the study were analyzed to identify factors that influence the acceptance of such calls.

Keywords

Speech modality, telephone services, autonomous call agents, agent-initiated telephone calls, acceptance, qualitative study, wizard of Oz.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces: Voice I/O.

General Terms

Design, Experimentation, Human Factors

Introduction

Despite the increasing pervasiveness of multimodal interfaces, the role of traditional telephone calls is still very important. The number of call centers that use automated telephone services employing both speech input and output is steadily increasing. Their users (callers) can accomplish numerous tasks ranging from simple telephone number queries to train or airline ticket bookings. The call initiated by the human caller is answered by a computerized agent that is capable of understanding caller's speech and producing intelligible speech signal. Depending on the technology employed, the agents can process user input whose form ranges from individual speech commands to natural language [1].

The use of automated telephone services is beneficial for the providers of these services as they can bring significant cost savings as compared to traditional model of human telephone operators. Various aspects of these systems have been evaluated in the HCI literature, including usability [2] and persona design [5].

The automated telephone services are being employed in the reverse direction as well, that is, a telephone agent takes the initiative and places the call to the user (call recipient). Such setup has been traditionally used in somewhat controversial applications, such as telephone marketing or election campaigns. The user acceptance of unsolicited telephone calls is generally very low [4] and many countries have already started to regulate the use of such technology by legislation, such as the Do-Not-Call Improvement Act in the United States (2007).

Nevertheless, there are numerous applications of agent-initiated telephone calls that are legitimate by its nature and are to the benefit of the call recipient: Wake-up calls, informing clients about arrival of the taxi, etc. In these examples, the user would have previously agreed to receive such call and would expect the call to happen.

With increasing autonomy of the telephone agents, it may be expected that computerized telephone services will start placing calls to people previously not involved with the service, as well as we may expect these calls to be multi-turn natural-language conversations, as opposed to simple playback of a single prompt. By the term agent-initiated calls we will refer to this style of communication. An example of agent-initiated call scenario could be a courier service delivering a package to its recipient where the service would make a use of an agent to negotiate the exact time and location of the delivery, while the recipient is not previously aware of any package.

In general, such interaction is necessary and suitable in these conditions:

- A timely response from a user is needed; where e-mail or SMS would be too slow.
- Large number of users would be contacted by the service; where such number of calls would present considerable costs to the provider.
- The users do not have to be informed in advance about such communication and therefore they would not be expecting it.
- Relatively small amount of data (such as an address, telephone number, or meeting confirmation) is to be gathered by the agent.

We believe that it is important to design the interaction with the agent in a way that would facilitate the discrimination between the solicited and unsolicited calls. We also believe that the research of the agent-initiated calls from the user's perspective is an important track of the voice user interfaces research for getting the knowledge needed for designing effective and satisfying interaction.

This note presents a qualitative study of attitude of the users towards the automated agent-initiated calls where an active participation of the users is required. We hope that this note will foster further discussion of this topic.

THE STUDY

We have performed a simple qualitative study whose main goal was to identify the factors that influence the acceptance and overall experience of automated agent-initiated telephone calls that require recipient's active participation in the dialog. The main purpose of this study was to identify problems for further research.

The Participants. 16 experimenters' friends were included in the study. The selected participants represented a very diverse social group and there were very little mutual friends among the participants. The participants represented different professions, different age groups, and different social status. Since the aim of the study was to investigate the genuine behavior when answering previously unannounced telephone calls, the consent to use the data gathered during the call has been solicited from the participants only after the call (see below).

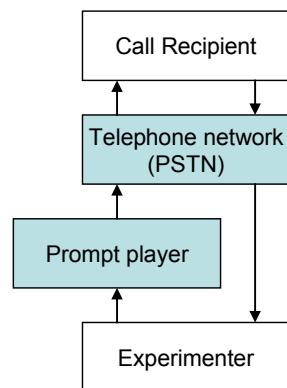


Figure 1: Set-up Overview

Use Case. We have acted the role of the automated agent representing a fictive entity Statistical Research Bureau of the Central Bohemia County. The role of the agent was to remind the user of the oncoming population census and to negotiate with the user the way to deliver the forms needed to be filled out for the purposes of the census.

Telephone Call Scenario. The telephone calls have been conducted in the Czech language. The following phases of the call have been planned:

1. Introduction. Information on the purpose of the call.
2. The agent asks whether the call is not coming in an inappropriate moment. If it does, the agent asks the user when to call again and then terminates the call.
3. The agent asks the user about their experience with the previous census.
4. The agent arranges the delivery of census forms to the participants.
5. Thanks and farewell.

Apparatus. All calls were placed from a fixed telephone line to the participants' mobile telephones in a weekday during regular office hours. The telephone number was visible but unknown to the participants. The functionality of the agent was simulated in a Wizard-of-Oz [3] (see Fig, 1). We have created 28 prompts covering utterance-by-utterance the entire telephone call scenario. These prompts were played back by the experimenters as the dialog was unfolding. The prompts have been designed in a way so that the user would be encouraged to take an active part in the conversation by going beyond simple yes/no questions. The prompts have been synthesized using a state-of-

art Acapela Text-to-Speech synthesizer (Czech voice Eliška).

The Experiment Procedure. Each of the participants went through the following two steps:

- *Call by simulated agent.* We have dialed the telephone number of the participant. If they never picked up, we have cancelled their participation in the study. After they picked up, one of the experimenters played the initial prompt. Depending on the user's response the experimenter played another prompt, and so on, until the goal of the call had been fulfilled or until the participant had hanged up. The duration of the calls varied between 10 and 90 seconds.
- *Debriefing.* A few minutes after the call, one of the experimenters called the participant again from a different telephone number. (We made sure that one experimenter contacted the other experimenter's friends in order to avoid possible bias.)

The debriefing composed of the following:

- The experimenter apologized to the participant, explained the purpose of the experiment, and asked for the permission to use the data gathered during the simulated agent call. All the participants expressed their curiosity about the experiment and gave their permission.
- The experimenter and the participant went through a semi-structured interview covering the following topics: participant's experience throughout the call, previous experience with any automated telephone agents, feeling of being addressed by an automated agent, the trust held towards the agent, and its development during the course of the call.

RESULTS

We have used established procedures for processing of the qualitative data acquired during semi-structured interviews. From the analysis of the interviews we have identified number of factors that were likely to affect the user experience. In the further text, these factors are outlined in individual sections.

Overall Acceptance. 11 call recipients from the total number of 16 answered the call. The reasons for not answering the phone are out of scope of our study. Only 2 out of 11 realized calls went through the entire scenario, i.e. the telephone call resulted in agreement on preferred time and place of census forms delivery.

The main drop-out moments were the following:

- After the agent asks first question, approx. 15 seconds into the call, see #2 in the telephone call scenario. This question required simple yes/no answer. The highest drop-out (6 call recipients) was recorded at this moment.
- After agent asks the first open ended question, 30 – 35 seconds into the call, see #3 in the scenario. Both call recipients quitting dialogue at this moment manifested confusion. They did not hang up the phone immediately after the question was placed, but after a few seconds of "perplexed silence" (term used by a participant during the debriefing).

Relatively low level of acceptance could be explained by various reasons that will be elaborated in the following text. However, the main drop-out points were moments when the dialogue required participants' active cooperation.

Attitude towards the Agent. A broad spectrum of attitudes towards the agent was present in our research sample.

3 participants expressed their preference being called by agents rather than human operators. Their following summarizes their views:

- The agent was perceived as a prevention of wasting human telephone operators' effort and time.
- The call recipient can be sure that the agent uses appropriate language and well-pronounced speech (provided that a high-quality text-to-speech synthesizer is used), as compared to the human operators "whose pronunciation is not always perfect and who sometimes use slang."
- The call recipient can be sure that the agent will manifest stable mood and appropriate behavior, as compared to human operators who "could be moody and not always patient".
- The voice had a pleasant timbre.

4 participants had no preference of human or machine operators. Those people reported being focused only on the content of delivered information and not on its means.

4 participants rejected communication with agents finding it as "superficial", "inhuman", "humiliating", and "deficient" („I would not talk to any tape recording."). These participants showed the quickest drop-out. Most of them told us that they thought it was a "telephone spam".

Some participants who did not perceive agent initiated calls positively mentioned that an agent *answering* their own call is "acceptable". They strictly differentiated

between scenario of "requesting a service" (outgoing calls) where the use of automated agents is considered appropriate, and "being requested" to listen and speak (incoming calls) where an agent is perceived as "derogative" and "not credible".

Previous experience with agents. Most participants reported a previous experience with agent-answered calls, mostly when calling their mobile operators or other customer lines. Only one participant had received an agent-initiated call (a non-interactive presentation of a political candidate). The participant perceived the call quite positively, as something "interesting" and "unexpected".

User Experience of the Call. Various experiences were reported regarding the communication initiated:

- *Curiosity.* Two participants reported that talking to robot was something "novel" and "extremely fun". They were curious how the interview will continue. Other two participants were curious about current state of technologies and that motivated them to maintain the dialogue, at least for some time.
- *Problem of politeness.* One participant reported that she did not feel guilty when having instantly "hung up on a robot". She would not terminate the call with a human, because she would be "shy" and would "feel uncomfortable to treat human operators this way". She had no such feelings for an automated agent.
- *Relevance.* Those who had expressed negative feelings about the call, described their initial reactions as "feeling of being a spam target" or as "being bothered by something irrelevant for them". One participant expressed his fear of being a victim

of a joke and reported that he was not willing to “play the game” and answer the questions.

- *Opening.* The initial 12 seconds of the call spent on agent self-introduction and explanation of the purpose of the call were found as too long by several participants and contributed to the decision to prematurely hang up the call.

Current Situation and Attention of the User. These problems on user side have affected the call:

- *Inconvenient situations* of call placement were reported (together with the perceived irrelevance of the call) as one of the most frequent reasons for dropping out: traveling on subway, being accompanied by a colleague, being in the meeting with colleagues. Two participants hung up because of the inconvenient timing of the call. Both of them were leaving their work and did not feel like to talk to anyone being tired after their day of work.
- *Cognitive overload.* Although the agent introduced himself and explained the purpose of the call, one participant complained about a missing introduction. She reported being so busy figuring out how it could be that she was addressed by robot that she did not concentrate on the content of the agent’s speech.
- *Trustworthiness.* One participant explained his unwillingness to talk to the automated agent as a consequence of his incredulousness.

CONCLUSION

This paper describes an initial qualitative study focusing on the overall user experience of automated agent-initiated calls in which a spoken interaction is requested from the user called. The main aim of the study was to assess the complexity of the problem and to identify topics for further research. The qualitative data

allowed us to identify a number of mutually interacting factors that affected the user experience, on both the service- and user-side. The users reported a broad range of attitudes, ranging from very positive to very negative. Our next research activities will focus on these factors through a series of quantitative studies.

ACKNOWLEDGMENTS

This research has been supported by European Commission Marie Currie Excellence Grant for the ADAMACH project (contract No. 022593).

REFERENCES

- [1] Cohen, M. H., Giangola, J. P., Balogh, J. *Voice User Interface Design*. Addison-Wesley, 2004.
- [2] Hempel, T. (ed.) *Usability of Speech Dialog Systems. Listening to the Target Audience*. Springer-Verlag Berlin Heidelberg, 2008.
- [3] Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Aboobaker, N., and Wang, A. 2000. Suede: A Wizard of Oz Prototyping Tool for Speech User Interfaces. In *Proc. of the 13th Annual ACM Symposium on User Interface Software and Technology* (San Diego, CA, Nov 06–08, 2000). UIST '00. ACM, New York, NY, 1-10, DOI <http://doi.acm.org/10.1145/354401.354406>
- [4] Kolan, P., Dantu, R. Socio-Technical Defense against Voice Spamming. *ACM Trans. Autonom. Adapt. Syst.* 2, 1, Article 2 (March 2007), 44 pages, DOI 10.1145/1216895.1216897.
- [5] Nass, C., Brave, S. *Wired for Speech – How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press, 2005.