Lessons from Designing and Studying a Conversational Agent for Group Facilitation

Q. Vera Liao  
IBM Research  
Yorktown Heights, NY, USA  
vera.liao@ibm.com

Dakuo Wang  
IBM Research  
Yorktown Heights, NY, USA  
dakuo.wang@ibm.com

Ameneh Shamekhi  
Northeastern University  
Boston, MA, USA  
ameneh@ccs.neu.edu

Abstract  
In our work, we explore a new paradigm for voice UIs—conversational agents supporting collaborative work, where the agents take social roles (e.g. facilitator, team members) and interact with multiple parties. In this position paper, we introduce our first-stage empirical work for designing a meeting facilitation agent to support group decision-making, including observations of human facilitators, iterative design of the facilitation protocol, a wizard-of-oz user study and focus groups. From this work, we discuss lessons learned for designing voice UIs for agent initiated conversations, group interactions, and decision-support systems.

Author Keywords  
Conversational agent; voice UI; group decision support system.

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

Introduction  
Voice based conversational interfaces are gaining increasing popularity recently, with widely adopted applications such as Amazon Alexa and Google Home.
Voice UIs provide many interaction benefits such as hands-free input, natural information query, and focused navigational paths [2]. However, most existing conversational UIs focus on supporting individual users and handling user-initiated conversations such as question-and-answer (QA) and command execution. In our work, we are interested in exploring the roles that conversational agents can play to support group collaboration, for example, as a group facilitator or a team member. Voice UIs offer several potential advantages for supporting collaborative work. First of all, it may dissolve the boundaries between human-human and human-machine interactions with the same interaction modality. This means that a person can address both his or her human partners and the conversational agent with the same utterances, and that the agent can continuously listen to the conversations happening in the group and actively participate. Users do not need to be interrupted and work with a separate system as with traditional collaboration support tools.

Second, the humanlike conversational interactions and rich space for anthropomorphic designs of agents are a natural fit for providing not only informational support but also social support for collaborative tasks. This is especially valuable when the agent needs to replace some of the effort of a human team member (e.g., facilitation). It could also potentially make interactions that imply social functions or social influence more effective, such as proactive interactions, showing support, soliciting compliance, etc.

Meanwhile, we also foresee many challenges in bringing voice UIs into collaborative processes. Some general challenges for voice UIs may become more severe when there are multiple speakers involved and continuous group activities, such as natural language understanding, turn-taking, interruption, etc [1]. There may also be open questions unique to a group context, such as what kind of social role should the agent play (e.g., exactly like a human facilitator or not) and how to design the voice interactions to convey such a role, how an agent would impact the group dynamics or even social norms, and how to interact with multiple users with different status, opinions and preferences.

In this position paper, we introduce our first-stage design process for a conversational agent that facilitates group meetings to make hiring decisions (i.e., selecting best candidates from resumes reviewing) and report on lessons we learned from a user study.

**System Description**

Our ultimate goal is to develop a conversational agent system named CASSY (collaborative agent for decision support system) to support group decision-making, where a typical task is to evaluate alternatives and select the best option. We chose hiring decision---selecting the best candidate for interviewing---as a use case since it is a common decision for groups in the workplace. We expect a group facilitation agent to provide three kinds of facilitation:

- **Decision making facilitation**: by taking an unbiased position, the agent’s main goal is to enforce a structured decision process to reach a more optimal decision. Examples of these facilitation activities include reviewing and weighing options, eliminating options, and pair-wise comparisons.

- **Meeting facilitation**: the agent should engage in time management and participation management to...
improve meeting efficiency. Examples include: Ice breaking, managing time, agenda description, turn taking management.
• Social support: the agent can also express interest and social support to provide more positive experience. Examples include short verbal phrases to reflect acknowledgement and support.

Typically, a spoken conversational system include four main components: speech to text conversion, dialogue understanding to map raw texts to an intent, response generation based on the intent, and text to speech conversion. The first two steps are often challenging and require a large amount of training data. The problem is more severe in a group context where multiple parties talk. In the first stage of the project, we aimed to create a wizard-of-oz prototype of the agent--- replacing the first two steps with manual control, that is, a human wizard sitting in a different room would choose the intent based on the current conversations following a pre-defined protocol, and the system would then generate and speak out the conversational UIs [3], as a cost-effective way to explore user interactions and collect data to bootstrap from.

Agent Conversation Design
With the Wizard-of-oz setup, the design of CASSY’s conversations focused on two aspects: 1) facilitation activities that CASSY engages in, i.e., what utterances are available in the agent’s conversation space; 2) the protocol for triggering intents and thus speaking of associated utterances. The latter can be shared both as the wizard-of-oz protocol and for the future development of an automated system (e.g., rule based).

Our design process included observational studies of human facilitators for the same hiring task (qualitative analysis of the facilitation strategies and utterances), integration of decision-making and facilitation strategies from literature review, and an iterative design process with multiple pilot studies.

We finalized the conversational design and procedural protocols of CASSY as following:

1. **Introduction**: CASSY initiates the group interactions by greeting meeting participants, introducing herself, and inviting others to introduce themselves.
2. **Agenda setting**: CASSY introduces the hiring decision task and the meeting agenda.
3. **Candidates review**: CASSY asks participants to review candidate resumes and tracks the time. She confirms participants’ readiness to move on to the discussions, otherwise allows one more minute (Example 1).
4. **Criteria discussions**: CASSY suggests the group to start by discussing hiring criteria, and prompts to consider four main criteria (e.g., education, skills) on the resume if not covered.
5. **Eliminating unfavored options**: CASSY invites the group to go through each candidate and eliminate the unqualified ones (Example 2). If participants cannot decide about a candidate after two minutes, CASSY suggests to move to the next candidate. After all candidates are discussed, if the number of elimination is below a threshold, CASSY suggests the group to eliminate more. Finally, CASSY summarizes the remaining candidates, and moves on to the next step.
6. **Decision on the best option**: CASSY asks participants to select the best candidate. If no
Example 2: eliminating unfavored options

CASSY: ok, let’s start narrowing down our choices. Based on the criteria we talked about, can we go through each resume and eliminate ones that are clearly below the bar?

Participant A: OK...hmm...

CASSY: let’s just pick a resume and start with that candidate.

Participant B: I think Candidate A...

... (after 2 minutes)

CASSY: So should we drop this candidate or not?

Participant A: I think we should.

Participant B: I agree.

consensus is reached after 3 minutes, she suggests the group to reflect on their initial voting, or discuss pros and cons of each candidate. When either 30-minutes is up, or the group reaches a consensus, CASSY moves to the last step

7. Exiting the meeting: CASSY summarizes the session, either with a final decision or without a consensus, and thanks the group for participating.

User Study
We conducted a user study where pairs of participants had face-to-face meetings to select the best candidate from a set of five resumes, facilitated by CASSY. 20 groups (40 participants) took part in the study. One of the embodiments of the agent. Therefore, we assigned half of the group to a voice-only agent which spoke out of a Bluetooth speaker, and the other half to the same conversational agent but with the embodiment of an avatar, projected on a Beam (Figure 1).

In this position paper, we focus on reporting general lessons learned for user interactions with the voice UI in the group context for a decision support task. The conclusions for the effect of embodiment was detailed in our upcoming CHI paper [5]. The lessons reported in the next section are drawn from the post-study focus groups, the observations of user interactions, and reflection on our design process. The focus group lasted about 20-30 minutes, following participants’ decision sessions. We started by asking questions about their overall impression about the decision process and the agent’s facilitation. We also inquired about their reactions for different types of agent actions (e.g., meeting facilitation, social behaviors). Lastly, participants were asked about how they would desire the agent to be improved, and new features they would like to add to the system. Discussions were audio recorded, transcribed and analyzed.

Lessons Learned
Lessons for designing (mostly) agent initiative voice UIs
A critical difference of a facilitation agent from popular QA agents such as Amazon Alexa is that the conversations are mostly agent initiated with the primary goal of soliciting compliance to enforce a structured decision process. The agent initiated conversations also happen during continuous collaborative activities, making the issues of floor-taking and interruption critical design challenges. Moreover, different from a QA agent that is usually sufficient to understand (e.g., using an intent classifier) user utterance in one turn, to initiate a facilitation prompt at the right timing requires the agent to continuously process ongoing conversations and infer the group status. While current speech technologies may not be able to handle all the challenges with human-level performance, we think it is useful to reflect on the challenges that our human wizard encountered and the solutions we adopted.

To design the wizard-of-oz protocol, we dedicated significant effort to formulate formal rules for the wizard to trigger utterances that fit the context and at the right timing. Because it is challenging even for a human to keep track of all the conversational content happening in the group, we resorted to mostly simple markers, time tracking and asking for confirmation. For example, the wizard was trained to focus on capturing the mentioning of candidate names during the elimination stage, track the time marker (1 minute), then start asking “so should we keep this candidate” (or variations thereof), if not hearing an elimination
decision made (Example 2). By doing this, the wizard did not have to closely follow every sentences in the discussions.

Still, many participants expressed some level of confusion during the focus group discussions. From the initiation of the interactions (greeting and introduction), participants were quickly used to the pattern of agent initiated conversations, and often waited for the agent to give the next command. At times participants were confused "whether she is listening", "where she is going to talk", and wished to have "a better idea what she is going to do". Adding to the confusion was system latency, for instance, if happening before the second agent utterance in Example 1. One solution came up during the discussion is to have some kind of embodiment to signal listening, interruption and latency (e.g., thinking). For example, Alexa uses flashing lights to signal system listening and processing. Interestingly, we found evidence that having a human face embodiment, while providing familiar modalities (e.g. gaze, facial expression) to signal those social functions, it can backfire and further increase the confusion when latency happened, because the face heightened user attention [5].

Lessons for designing voice UIs in groups
In addition to the main function ---to facilitate a structured decision process, an additional benefit we found of a conversational agent for supporting collaborative tasks is its effective execution of social functions. For example, participants appreciated the “icebreaking” session in the beginning, and generally responded well to the agent’s attempt to balance the discussions by soliciting more input from the less spoken participant. Several participants commented that an important value of a group facilitation agent could be the delegation of social interventions that a human member may not be well positioned to make (e.g. "if my manger...holds up the meeting, I don’t want to be the guy to stand up").

Several challenges of designing conversational UIs also arose due to the group context. In particular, the human wizard reflected on the challenge of accurately assessing individual and group preferences from their conversational interactions, not only because the interactions were continuous and intensive, but also because the expressions were sometimes implicit to avoid face-threatening situations. This kind of challenge may not exist for agents that interact directly with individual users. Moreover, group conversations tend to be less tractable than those of individual users. Even for this structured decision task conducted in a lab environment, we observed diverse trajectories and group behaviors. For example, some groups followed the agent’s command very closely and always waited for the next step, while other groups took more self-initiatives and deviated from the suggested procedure. There were also differences in how groups perceived the agent’s role. While some group commented that they considered the agent to be an active group member working towards the same goal, others perceived it like “a background voice making some suggestions”. To design conversational agents for group tasks, we suggest that one should clearly define the agent’s role and persona, and convey that in the design of conversations and embodiment (e.g., during self-introduction). This may help better constrain the group conversations within the agent’s knowledge base.
Lessons for designing voice UIs for decision support systems (future directions)

Currently, the agent acts as a Level 1 group decision support system (GDSS) according to DeSanctis and Gallupe’s taxonomy of GDSS [4], which mainly targets removing communication barriers. In the focus group discussions, many participants expressed desire for the agent to move up to Level 2 GDSS—also providing informational support to reduce decision uncertainty. For example, some suggested that the agent can provide factual information such as what are the required skills and their statistics for the hiring position. To provide informational support will change the voice UI paradigm to more bi-directional interactions (e.g., QA) instead of agent initiated conversations.

Another future direction that came up during our design process is the integration of voice UI with decision modeling or analytic systems. An example of a decision analytic system is a voting system that tracks not only ranking of candidates but also group consensus and individual preferences, some using advanced electoral methods to counter noises and biases. A voice UI can be an interface for these systems to elicit individual votes. The output of these systems would provide knowledge for the agent to execute nuanced interventions for the group decision-making process. For example, the agent would know on which candidate the group has the lowest consensus, and which group members have the highest disagreement, and focus on encouraging conflict resolution for the controversial candidate or among the disagreeing members, which could help the group reach better understanding and more satisfied decision outcomes.

References