Supplemental Material: SPA: Verbal Interactions between Agents and Avatars in Shared Virtual Environments using Propositional Planning

Andrew Best*
Dept. of Computer Science
UNC Chapel Hill

Sahil Narang†
Dept. of Computer Science
UNC Chapel Hill
http://gamma.cs.unc.edu/pedvr/

Dinesh Manocha‡
Dept. of Computer Science
University of Maryland

1 Introduction

A Additional Method Details

Figure 1 details our two-stage planning pipeline. Figure 2 details our natural-language communication pipeline in an example dialog.

Sample Lexion Entry: The lexicon used to generate the results presented in section 6 of the main document consists of a small subset of English words annotated with sample sentences and reference hints for the parser. The word "location" as it appears in the lexicon is tagged with the hint "where" and "InSpace" which are other forms seen in our domain descriptions. As a predicate, It is assigned several template sentences with annotated natural-language intentions. One such sentence with the label *predicate answer* is

"the [PREDICATE:NAME] of [PREDICATE-ENTITY:DEF-ARTICLE-NAME] is [PREDICATE-ENTITY:DEF-ARTICLE-NAME:GALLERY]. "

This template provides parameters for binding an entity with a definite article, e.g. the statue, to an entity with the type specifier gallery. A sample binding of the sentence would be

"The location of the Venus de Milo is Gallery B".

B Additional Performance Results

B.1 Implementation and Performance Benchmarks

Our experiments were conducted on a desktop pc with an Intel Xeon E5 CPU, NVIDIA TitanX GPU and 16gb of RAM. We coupled our propositional planner with Rasa NLU [2] for semantic parsing. User utterances were captured via microphone and automated speech recognition. Our algorithm was implemented in python, and our VR experiments were performed with the Occulus Rift HMD. We couple our approach with the 3D animation system described in [1].

In addition to the results reported in section 6 of the main document, we evaluated the algorithm's performance as a function of domain size and number of agents. Consistent with prior propositional planning approaches, our algorithm scales linearly in the number of agents and exponentially in the size of the problem domain. Figure 3 details our experimental results. Table 1 provides additional details about the number of agents, desires, and verbal interactions in our benchmarks.

B.2 User Evaluation

This section provides a formal description of the user study we conducted to evaluate the plausibility of agent-avatar interactions and the overall simulation generated as a result of our algorithm. In addition, we provide the complete set of participant responses and additional response analysis.

Experiment Goals & Expectations: We hypothesize that verbal communication between agents and avatars will enhance the perceived plausibility of the simulation, and generate positive impressions as compared to the control conditions.

B.2.1 Comparison Conditions

- **No Agents**: In the no agents case, a user avatar explores a virtual environment without any virtual agents present.
- **No Communication**: In the no communication case, a user avatar explores a virtual environment with agents who could not interact using natural-language communication.

B.2.2 Experimental Design

This study was conducted based on a within-subjects, paired-comparison design. Each scenario was displayed with a text-based prompt to provide the appropriate context. Participants were shown two pre-recorded videos of a subject interacting with the system in a side-by-side comparison of our method and one of the comparison methods. They were then asked to answer a short questionnaire before moving on to the next scenario. The order of scenario and the positioning of the methods was counterbalanced.

B.2.3 Environments

The multi-agent tradeshow scenario and multi-agent museum were used in this study. Three confederates were recruited to participate as the avatar in the environments. In trials using our method, the confederate was allowed to interact with the agents using natural-language communication. In each case, the avatar was piloted from a first-person view. Their interactions were recorded via screen capture and a microphone.

Tradeshow: The avatar was instructed to find the "registration booth". They were shown a picture of the booth before beginning their task but were not told its location. In the SPA case, virtual agents in the environment were able to interact and provide the location of the booth to the avatar. We refer the reader to the main document for visual examples of the benchmarks.

Museum: The avatar was instructed to find a specific statue in the museum but was not told the location of the statue. The statue in question was Lucy, courtesy of the Stanford University Computer Graphics Laboratory. In the SPA case, a virtual agent near the avatar's starting position was provided knowledge of the location. The avatar was able to ask this agent the location of the statue. In addition, two agents were placed along the path to the goal who would interrupt the avatar's progress and ask the avatar for the locations of other statues as they passed.

B.2.4 Metrics

Participants were asked a set of common questions for both comparison methods, with specific additions for each comparison method.

Common Metrics: Participants were asked to indicate which simulation more closely reflected a real-world scenario on a Likert scale with 1 indicating strong preference for the method presented on the left, 7 indicating strong preference for the method presented on the right, and 4 indicating no preference. They were then asked the impact of the following items on their preference: the presence of

*e-mail: best@cs.unc.edu
†e-mail: sahil@cs.unc.edu
‡e-mail: dm@cs.umd.edu

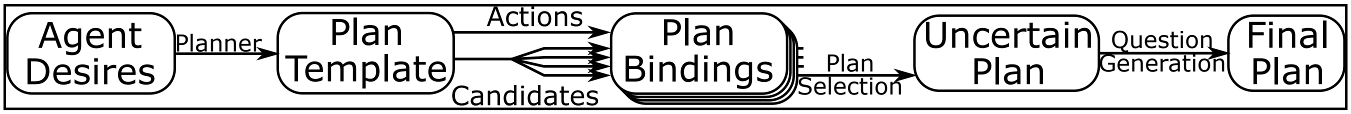


Figure 1: **Two-stage Action Planning with Incomplete Information**: Each agent is given a desire to achieve during simulation. We propose a two-stage planner which generates action plans despite uncertainties in the agent's knowledge. The first stage generates a plan template and a set of candidates for each argument in the plan. The second stage generates a set of candidate bindings. The algorithm selects the plan with the least uncertainty and generates an action plan from the bindings. If any uncertain information is present, an uncertainty resolution action is created, yielding the final action plan which may include asking questions, or exploring the environment.

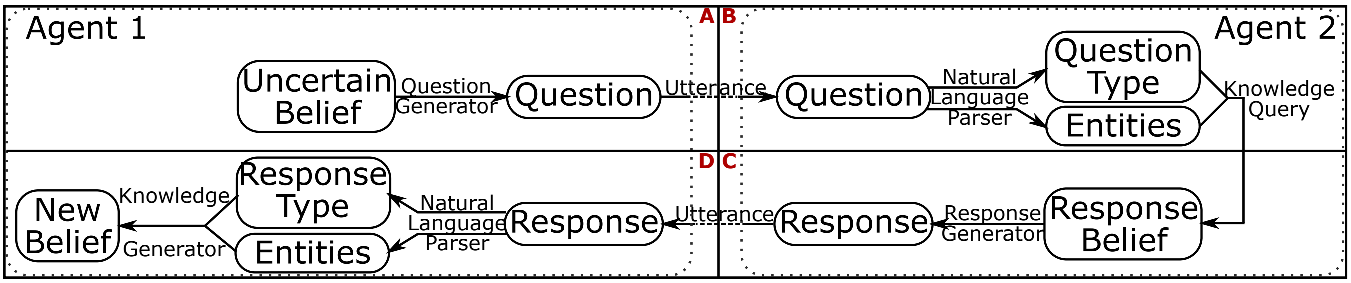


Figure 2: **Natural Language Communication for Virtual Agents**: This figure illustrates a sample interaction between two agents using our natural-language approach (clockwise from top). **(A)** Agent 1's plan yields an uncertain belief. The agent generates a question from the belief template. The question is communicated as a natural language utterance. **(B)** Agent 2 receives the utterance and parses it into the relevant question type and entities. The agent queries its knowledge-base for an answer to the question, yielding a response predicate. **(C)** Agent 2 uses our approach to generate a response utterance. **(D)** Agent 1 receives the utterance, generates the appropriate response type and entities, and processes these into a new belief which is stored in the knowledge-base.

natural language communication, the quality of the verbal responses from the agents, and the quality of the animation. These were answered on a Likert scale with 1 indicating strong negative impact, 7 indicating strong positive impact, and 4 indicating no impact.

No Agent Metrics: Participants were additionally asked what impact the presence of the virtual agents had on their preference.

No Communication Metrics: Participants were additionally asked which of the methods demonstrated more plausible interactions, in which simulation did the agents appear to benefit more from their interactions with the avatar, and in which simulation did the avatar appear to benefit more from their interactions with the virtual agents.

B.2.5 Results

The study was taken by 14 participants. We normalized the data for comparative questions such that a response of 1 indicates strong preference for our method. We collapsed the common metrics across trials as well as plausibility of interactions question for the No Communication metric and the presence of virtual agents from the No Agents metric. We performed a one-sample t-test comparing the mean of each question with a hypothetical mean of 4 (no preference or no impact). We limit our discussion below to questions which directly deal with natural-language interaction and preference for the methods. Table 2 gives complete details of the participant responses collected for our user evaluation.

We found the question "Which simulation more closely reflects a real-world scenario" significant in both the no agents condition ($t(27) = -6.204, p < 0.000$), and the no communication condition ($t(27) = -7.887, p < 0.000$). We found the question "What impact did the presence of natural language interaction have on your answer" significant in both the no agents condition ($t(27) = 10.200, p < 0.000$), and the no communication condition ($t(27) = 14.925, p < 0.000$). We found the question "What impact did the quality of the verbal responses from the agents have on your answer" significant in both the no agents condition ($t(27) = 5.473, p < 0.000$), and the no communication condition ($t(27) = 9.218, p < 0.000$). We found the question "In which simulation did the interactions between the user and the agents seem more plausible" significant in the no communication condition ($t(27) = -6.765, p < 0.000$). It was not asked of the no agents condition. We found the question "What impact did the presence of virtual agents have on your answer" significant in the no agents condition ($t(27) = 13.478, p < 0.000$). It was not asked of the no communication condition. perception of the impact of natural language interactions.

Analysis: As detailed in section 6 of the main document, participant responses demonstrate the benefits of our algorithm in terms of generating plausible agent-avatar interactions (2.46 ± 1.20). In both comparative conditions, either without agents or without communication, participants found our method to generate simulations and interactions with better reflect real-world scenarios (2.29 ± 1.15 and 2.29 ± 1.46).

Overall, Participants preferred our approach in 84% of responses. Of those responses, 84% were strong preferences ($r \leq 2$). Figure 4 and Figure 5 provide additional details about our method's advantages over prior approaches.

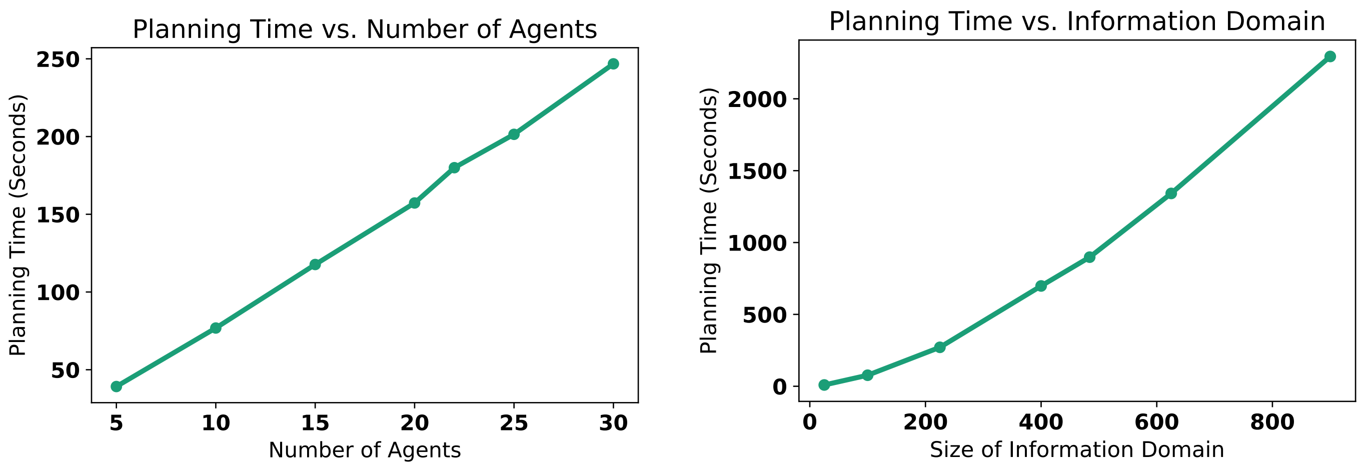
References

[1] S. Narang, A. Best, and D. Manocha. Simulating movement interactions between avatars & agents in virtual worlds using human motion constraints. *Proc. of IEEE VR*, 2018.

[2] Rasa.ai. Language understanding with rasa nlu, 2017.

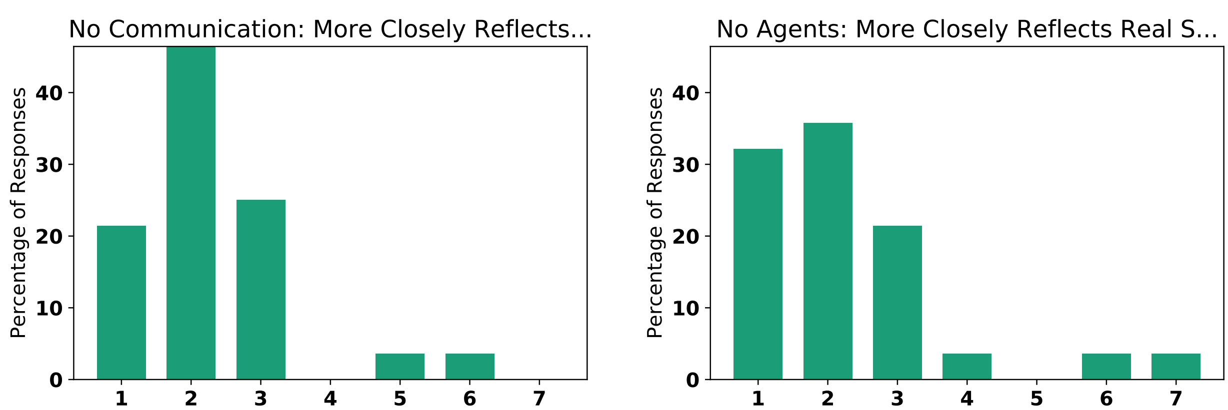
Table 1: **Performance Benchmark Details. We detail number of agents, desires, and the NL-I details of the benchmark scenarios including how many questions were asked, statements made, facts overheard by agents, and parser failures. We observe, as expected, that as the number of agents and desires increases, the amount of information gained from overhearing nearby agents increases.**

Scene	Agents	Desires	Statements	Questions	Facts Overheard	Parser failures
Anti-podal Circle	10	30	14	5	28	0
Evacuation	11	10	4	0	20	0
Museum	5	9	20	13	3	0
Trade Show	4	1	3	2	0	0



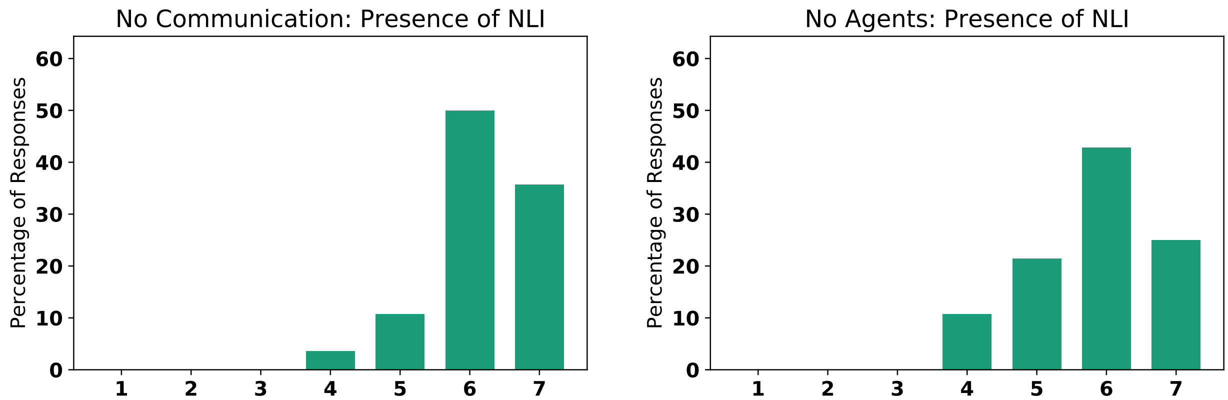
(A) (B)

Figure 3: **Performance Results Varying Number of Agents and Problem Size**: **(A) Varying agents on a fixed domain size (100 predicates):** We observe that our algorithmic approach's performance scales linearly in the number of agents. **(B) Varying domain size for a fixed number of agents (10 agents):** We observe that our algorithm's performance scales exponentially in the size of the problem domain. This is consistent with propositional approaches. However, our two stage planning approach enables rapid replanning after the initial planning step, reducing overall planning time.



(A) **(B)**

Figure 4: **Histogram data of user responses for Which scenario better reflects real-world scenarios**: Participants in our evaluation found simulations using SPA significantly more plausible compared to simulations with a prior approach (**A**) and simulations with no agents (**B**). This indicates that the presence of agents has a positive impact on plausibility and that our agents behave sufficiently well to increase plausibility with respect to agents lacking SPA.



(A) (B)

Figure 5: **Histogram data of user responses for impact of natural language interactions** Participants in our evaluation found the presence and quality of the natural language interactions had a significant impact on their preference for our approach to simulations with agents lacking SPA **A** and simulations without agents **b**. In addition, the preference for the quality of the natural language interactions generated with SPA is stronger when compared to agents not able to communicate.

Table 2: **Frequency of Responses in User Evaluation. This table shows the frequency of participant responses in the user evaluation, as well as the means and p-value for a one-sample t-test with a hypothetical mean of 4. For comparative questions, responses less than 4 indicate preference for our agents. For impact questions, responses greater than 4 indicate positive impacts. We found participant responses to all question significant.**

Question	1	2	3	4	5	6	7	mean	std	p-value
NL-I Agents vs Non-Interactive Agents										
Comparative Questions (NL-I Agents left)										
More closely reflects real scenario	6	13	7	0	1	1	0	2.29	± 1.15	< 0.000
Agents benefit more from interaction	11	4	1	11	0	1	0	2.57	± 1.53	< 0.000
User benefits more from interaction	17	10	0	0	0	1	0	1.54	± 1.00	< 0.000
More plausible interactions	5	13	5	2	3	0	0	2.46	± 1.20	< 0.000
Impact Questions										
Presence of natural Language	0	0	0	1	3	14	10	6.18	± 0.77	< 0.000
Quality of the verbal interactions	0	0	2	1	3	18	4	5.75	± 1.00	< 0.000
Animation of the virtual agents	0	0	4	13	3	3	5	4.74	± 1.36	0.010
NL-I Agents vs No Agents										
Comparative Questions (NL-I Agents left)										
More closely reflects real scenario	9	10	6	1	0	1	1	2.29	± 1.46	< 0.000
Impact Questions										
Presence of the virtual agents	0	0	0	0	8	10	10	6.07	± 0.81	< 0.000
Actions of the virtual agents	0	0	0	6	9	10	3	5.36	± 0.95	< 0.000
Presence of natural Language	0	0	0	3	6	12	7	5.82	± 0.94	< 0.000
Quality of the verbal interactions	0	1	2	3	7	12	3	5.29	± 1.24	< 0.000
Animation of the virtual agents	0	0	3	11	10	2	2	4.61	± 1.03	0.004