

Playing with Words: From intuition to evaluation of game dialogue interfaces

Serdar Sali^{1,a}, Noah Wardrip-Fruin^{1,a}, Steven Dow^{3,c}, Michael Mateas^{1,a}, Sri Kurniawan^{1,b},
Aaron A. Reed^{2,a}, Ronald Liu^{1,a}

¹{sali, nwf, michaelm, srikur, rliu}@soe.ucsc.edu

²aaareed@ucsc.edu

³spdow@stanford.edu

^a Computer Science Department

UC Santa Cruz

1156 High Street, Santa Cruz, CA

^b Computer Engineering Department

UC Santa Cruz

1156 High Street, Santa Cruz, CA

^c HCI Group

Stanford University

353 Serra, Stanford CA



Fig 1. (a) Star Wars: Knights of the Old Republic, (b) Façade, and (c) Mass Effect present three different dialogue interface approaches. Knights of the Old Republic displays complete sentences of dialogue in its menus, Façade uses natural language understanding, and Mass Effect displays short, abstract representations of the actual line of dialogue.

ABSTRACT

Dialogue systems are central to role-playing games, adventure games, interactive fictions, and some forms of interactive drama and cinema—but we have little empirical evidence about how, or even whether, the design of dialogue system interfaces shapes gameplay experience. In this paper we present the results of a study directly comparing three different dialogue system interfaces implemented over the complete course of a dramatic, story-focused game. We find that, holding the rest of the game steady, changing the dialogue interface produces significant changes in gameplay experience. Further, these changes shape perceptions of the system well beyond the interface and its operation. The changes are also sometimes resonant with, and sometimes at odds with, conventional wisdom about these interface options in the game design and writing communities. Our study compared three interfaces: a sentence selection interface (which appears to maximize story involvement), an abstract response menu interface (which maximized reasoning about the underlying game structures), and a natural language

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

FDG 2010, June 19-21, Monterey, CA, USA

Copyright 2010 ACM 978-1-60558-937-4/10/06... \$10.00

understanding interface (which maximized a sense of presence and engagement with the overall experience). Our full results provide valuable future guidance for those seeking a dialogue interface that is resonant with their gameplay experience goals.

Categories and Subject Descriptors

K.8.0 [Personal Computing]: Games

General Terms

Game design, human computer interaction

Keywords

Game design, Dialogue, Natural Language Understanding, User Interfaces

1. INTRODUCTION

Resisting the Dark Side's temptations in Knights of the Old Republic, grilling a murder suspect in Indigo Prophecy, or flirting shamelessly in Façade—choosing what to say is an important part of many games. In this paper we present a direct comparison study of dialogue interfaces, investigating the impact of how dialogue choices are made available to players. Our results provide important future guidance for game design, especially as the underlying structures of character and story become more complex.

Our comparative study uses a version of Michael Mateas and Andrew Stern's Façade, instrumented to support three different

screen-based dialogue interfaces for complete runs of the game. We investigate three dialogue interfaces: free-form text entry (as in the original *Façade* and most interactive fiction), a dialogue menu that specifies exact language to be spoken by the audience-controlled character (as in games like *Planescape: Torment* and *Star Wars: Knights of the Old Republic*), and relatively abstract specification of responses (as in games like *Indigo Prophecy* and *Mass Effect*). Previous studies have not addressed how different dialogue interfaces change the user experience of an otherwise identical game.

Prior to our study, the only guidance available to dialogue interface designers has been their own intuition and the assertions, sometimes contradictory, of other game designers and writers. We review some of this prior guidance, later drawing out how the results of our study help situate earlier received wisdom. We find that sentence selection interfaces most strongly motivate story involvement, and though the time required to read them is considered their major drawback in industry, few participants reported being bothered by this. We also find that abstract response interfaces, though commonly praised in industry for preserving natural conversation flow, are actually most effective at facilitating a sense of game-level control—and feel least conversationally natural to players. We find that natural language understanding interfaces, though certainly somewhat error-prone (as the conventional wisdom holds), are surprisingly engaging and enjoyable for players.

2. RELATED WORK

Though we know of no direct comparison studies of different dialogue interface approaches carried out within game development groups, industry certainly engages in iterative refinement of interfaces, sometimes driven by playtesting. Perhaps the best-known recent example of such work is BioWare's on the dialogue interface for *Mass Effect*. This reportedly went through 10-12 iterations, internally, aimed primarily at speeding up the player's ability to choose responses (to preserve conversation flow) and secondarily at allowing the player character to perform lines without repeating aloud something the player had just read silently [10].

A concern with conversational flow is one of the common objections to the dialogue interface that BioWare used for games prior to *Mass Effect*: sentence selection interfaces such as that of *Star Wars: Knights of the Old Republic* (or the more recent *Dragon Age: Origins*). As Brent Ellison argues, “reading all the possible responses takes time and brings conversation flow to a halt” [4]. Similarly, Lee Sheldon observes that “While [a sentence selection interface] gives the writer even more opportunity for character revelation, especially of the player character, it adds more text to read—one reason it only occasionally shows up in console games, and why designers are forever trying to find ways to shorthand it” [12].

Sheldon's observation that sentence selection interfaces provide an opportunity for revealing information about the player character touches on two further pieces of design wisdom about such interfaces. First, as Sheldon points out, game writers can make player characters “far more witty, articulate, and wise (or boring, tongue-tied, and stupid!) than the player himself.... We give him the chance to stand up in a conversation with Albert Einstein or Dorothy Parker or Dennis Miller and hold his own.” However, the fact that sentence selection interfaces can offer players enticing things to say (and enticing roles to play) points in the opposite

direction of the second piece of design wisdom about such interfaces. This is an outgrowth of the conventional “player character as transparent cipher” viewpoint that has led even story-rich games such as the *Half-Life* series to choose the odd conceit of a silent player character. As Richard Rouse puts it:

[W]hen players want to play games, often they want to play themselves. If the character they are controlling has a very strong personality, there is a distancing effect, reminding players that the game is largely predetermined and making them feel like they are not truly in control of what happens in the game. Particularly frustrating are adventure games that feature strongly characterized player characters who keep speaking irritating lines of dialogue. [11]

In addition, Sheldon argues that sentence selection interfaces can, through well-written sentences, provide multiple dimensions of information about the choices players are making. This can include the topic, the approach to the topic, and how forcefully the character presents her case. As discussed further in our results section below, we found that sentence selection interfaces could create an illusion of greater nuance, without being strongest at communicating information about gameplay impact. They were, in addition, particularly strong for characterization—not only of the player's character, but of all characters in the conversation.

As noted above, the most common alternative to sentence selection interfaces is a different kind of menu—one that displays shorter, less fully realized conversational options. The display may be of discourse acts, topics, tones of response, partial responses, more diverse conversational actions, and so on. With games like *Mass Effect* and *Indigo Prophecy* adopting menus that present a combination of these, we can recognize this hybrid form as a prominent alternative to sentence selection. In this paper we refer to these as abstract response menu interfaces.

In addition to abstract response menu interfaces being discussed as faster to view (with the expectation that this helps preserve conversational flow) another common observation is that they aren't as predictable for players as sentence selection interfaces. Ellison writes of sentence selection: “There is no ambiguity in the player's decision.” Sheldon says that, on the other hand, abstract menu interfaces “can interject an immersion-harming game played between designer and player—What is my player-character going to say next?” Within the gaming community, the early level *Mass Effect* “renegade” response (on Eden Prime) that resulted in physically hitting another character, without this being indicated in the interface, may have helped cement the impression that in this approach the mapping between selection and performance can be unpredictable for players. But such dramatic divergences are relatively unusual, and our findings run in the opposite direction: the abstract response menu was found comparatively easy for strategic control, laying bare the impact on game state rather than the specifics of enunciation, but this also had the result of shifting the interaction's feel from one of conversation to limited-option gameplay.

This brings us to the third type of interface in our study, one rarely employed in current commercial games: natural language understanding (NLU). By allowing (and attempting to interpret) free form textual input from players, NLU interfaces potentially enable a much greater range of player response than any single-depth menu could display. This type of interface is common in the independent game design/writing community of interactive fiction practitioners, who create games in which most actions are

specified textually and interpreted by a parser. The community has also built up significant discussion of the issues [13].

The mainstream game community, however, has not explored this option. Sheldon writes, “We’re nowhere near ready to turn over conversations with major characters to AI” and characterizes this interface option as “outside the scope” of his book. Within the game industry, the best-known contemporary experiment with this interface is *Façade* (e.g., well-known industry commentator Ernest Adams writes, “*Façade* is one of the most important games ever created” [1]). This is part of the reason we selected *Façade* for our study, along with its depth of conversational interaction and the fact that its underlying structure was amenable to supporting multiple dialogue interface options. Ellison argues:

[NLU interfaces] are rare in modern games for two reasons. The first is that the freedom they provide is extremely time-consuming to produce. The system needs hundreds of potential responses to accurately simulate a single, short conversation.

*The second reason is that even the most robust parsers frequently misinterpret the player’s input. In *Façade*, an innocent inquiry can send the NPCs [non-player characters] into shock, horrified by what they thought the player just said. These misunderstandings ruin virtual relationships and frustrate the player, while at the same time exposing the program’s failings and distracting the user from the interaction.*

Anecdotal evidence and past work have certainly shown that *Façade*’s NLU systems make errors, but Mehta et al’s earlier study found that “the narrative cues offered by an interactive drama, coupled with believable character performance, can allow players to interpretively bridge system limitations and avoid experiencing a conversation breakdown” [8]. In fact, our results find this interface option, though critiqued by many players, was also the one they found most enjoyable .

The work on *AR Façade* by Dow et al [3] is perhaps the closest work to our experiment in the game studies literature. *AR Façade* is an augmented reality version of *Façade* in which the player interacts with the couple Trip and Grace in a physical space decorated identically to the couple’s apartment, and Trip and Grace are superimposed on this physical space through a head-mounted display. Dow et al compared three different versions of *Façade*: *AR Façade* (the augmented reality version, where players use speech and physical gestures to interact), the original desktop version of *Façade* in which the player types to speak, and a speech input version in which the player speaks to a microphone. A wizard operator typed player statements into the system in the augmented reality and speech input versions. The results of this study revealed that “increased presence does not necessarily lead to more engagement” and that “mediation may be necessary for some players to fully engage with certain interactive media experiences.” In our study, although not as pronounced, we also found similar effects with the NLU version.

Another relevant study on *Façade* is Milam et al’s phenomenological study of interactive narrative experiences. The goal of Milam et al’s study was to understand “how players understand and internalize their interactive experiences” [9]. Sixteen prominent themes that defined how people experienced *Façade* emerged from their data. We noticed similar themes and patterns in our data, however we were more interested in studying these themes in terms of commonly discussed game design parameters such as immersion, engagement, presence and agency.

3. EXPERIMENT DESIGN

We believe *Façade* is an ideal platform for experiments such as ours. While most games interleave conversational interaction with other forms of gameplay, *Façade* is an interactive drama in which almost all player activity is conversational. This means that player responses to non-conversational elements of play do not have the potential to color our results—and that a 15-minute play session includes an amount of conversation that would take much longer to achieve (and evaluate) in a game that also included combat, world exploration, and so on. In addition, precisely because most player actions take place through conversation, the specifics of the dialogue interface are likely to have a noticeable impact on gameplay. At the same time, it is also worth noting that our results may be influenced by the specifics of *Façade*, which remains a very unusual game, though one that may represent a potential future for games in which dialogue is a central element.

Specifically, for our experiment we built two additional versions of *Façade* (Fig. 3) that employ dialogue interfaces similar to those described above. In the sentence selection version the user selects the actual line of dialogue spoken from a pre-scripted list. In the abstract-response version the user selects actions and short phrases which are abstract representations of an exact line of pre-scripted dialogue to be spoken by the player character, which is not visible beforehand. In both versions, the game pauses indefinitely for player input. The sentence-selection version and the abstract-response version were identical except for the options in the menu, since we wanted the two experiences to feel as similar as possible. Although inevitably the menu-based versions were more linear, we tried to make sure that all major and many minor plot variations were still explorable through the menus. Recognized interactive fiction author Aaron A. Reed was the lead writer for the menu-based versions. The third version is the original *Façade*, which allows free-form text entry and has an NLU module that parses the input and maps it onto discourse acts. In the NLU version, conversation is continuous: the player can interrupt the characters at any time and the NPCs don’t indefinitely pause to wait for player input.

- Q1.** Would you like to play this game again? Why\why not? Which version did you enjoy the least? Which one did you enjoy the most? Why?
- Q5.** How engaged were you in different versions of the game? Can you rank them in terms of engagement?
- Q8.** Which interface variation made you most motivated to move the story forward? Why?
- Q13.** How much influence did you feel over the story using the different versions?
- Q15.** How did you form strategies and make decisions? How easy or hard was it to execute your strategy?

Fig 2. Sample questions from our interview.

We conducted within-subject controlled experiments. Each session, taken together with the interview, lasted between 120 and 150 minutes. We recruited 42 participants, most of whom were from an introductory game design class taught at our university. Participants were compensated with extra credit in the class. Seven experiments had to be thrown out due to issues in the

testing sessions, leaving us with thirty-five participants. The class was open to all majors, although most subjects were Computer Science students. We only recruited native English speakers with gaming experience so that language ability and familiarity with game interfaces were not influencing factors. We also randomized the play order to account for learning effects.



(a)



(b)



(c)

Fig 3. Different versions of Façade. (a) the sentence selection version, (b) abstract-response version, (c) the NLU version.

At the end of the play sessions, we conducted semi-structured interviews designed to explore how the different dialogue interfaces enable presence, control, engagement, agency and

enjoyment, which are all frequently discussed concepts in the game design community, and are important potential design goals when designing interactive experiences. We were also able to draw quantitative data from our interviews using participants' preference ranks between different versions along these dimensions. Some example questions from our interviews can be found in Fig. 2. Results were analyzed using Grounded Theory [6]. While we believe deeper discussion of these issues is beyond the scope of this paper, we define presence as the sense of "being there in the mediated environment" [5]. Presence is shown to be positively related to enjoyment in previous studies [15].

Engagement is also an important property of all interactive experiences and a necessary precondition for presence [2]. We are using a more complete formulation of agency proposed by Wardrip-Fruin et al [14], which also takes into account how system understanding shapes players' perception of agency, resulting in players attempting to take actions that are supported and that have meaningful game impact.

4. RESULTS AND DISCUSSION

Engagement

Despite the dismissal of this interface option by the mainstream game community, slightly more than half of our participants (54.3%) reported being most engaged in the natural language understanding (NLU) version. 45.7% reported that the menu-based versions were more engaging. Of those, half felt most engaged in the sentence-selection version, while the other half found the abstract-response version more engaging (Fig. 4).

Most Engaging

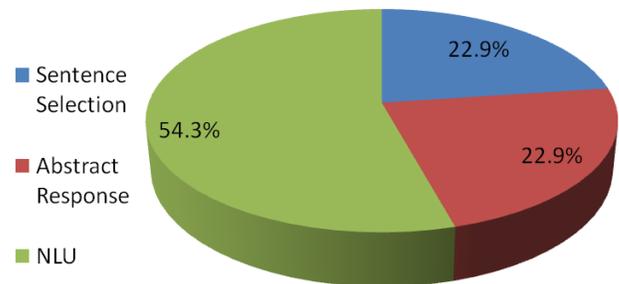


Fig 4. Engagement results. 54.3% of participants found the NLU version the most engaging.

The participants noted that the NLU version was engaging because it allowed them to say whatever they wanted to say, even though they had difficulty making their statements understood by the game. One participant said he found the NLU version to be the most engaging one because:

"I was able to actually talk and give my own words and [I didn't have] to deal with random dialogue that someone else had generated. I felt like I was actually controlling the character more instead of just 'here's five thoughts, pick one.'"

Another reason that participants thought the NLU version felt more engaging was it required considerably more attention than the other versions. Since the players had to figure out when to interject and how, they had to maintain a constant level of focus and stay alert. One participant noted:

“[The NLU version is more engaging] because you are focused on trying to get the right answer more, and you are focused on what different possibilities do I have, and [it requires] more critical thinking. You just have to think more when you have more options.”

While the original *Façade* managed to be engaging by establishing a high sense of attachment between the player character and the player and by requiring constant attention, the menu-based versions achieved engagement by involving the players more in the dramatic situation conveyed by the conversation through more accurate interpretation of what they chose to say. One participant noted:

“The menu system had more conversational engagement with me because I could choose from more explicit options in which way the conversation would go [...], instead of just taking a shot in the dark and saying placating things like you would do in an everyday social situation. I felt myself saying more things that represented what I was really feeling about the characters and the situation, so I got more wrapped up in the dramatic aspects of the interaction between Grace and Trip and myself.”

As we will discuss further below, the difference between the two menu-based versions was more defined by the target of control. Some participants enjoyed having control over what effect their actions would have, whereas others preferred having control over what specifically the player character would say. Knowing the expected outcome of a line of dialogue made the game feel easier, whereas knowing what the player character would say allowed a better association with the character, resulting in a better sense of engagement. When asked which version was the most engaging, a participant noted:

“The [sentence selection] version afforded many more options in terms of how you wanted to [play a specific way], because you had lots of possible dialogue to choose from. [In the NLU version] I felt like the limitations of the computer program in turn limited the gameplay aspect, so I couldn’t really utilize the freedom of speech as much as I felt would be possible. And the [abstract response version] was the most limiting of them all, because it was just giving you categories of dialogue instead of specific sentences [that allow you] to try and hit a specific emotion.”

Another participant noted the following regarding the other versions:

“For the [abstract response version] the difference [from the sentence-selection version] was the text was really unnatural [...] and it made it too obvious what was going on in the background. And for the [NLU version] it seemed like everything I said had no effect whatsoever—like they just ignored everything I said pretty much, unless it was ‘I agree’, ‘I disagree.’”

Participants who felt the abstract-response version was more engaging thought knowing the outcome of their conversation choices made moving through the experience easier. As a result, they felt more engaged in the story. One participant offered this comparison between the abstract-response version and the other versions:

“[In the abstract-response version] it was just easier to figure out how you are going to have an effect on the story. In the [NLU] version I felt like they can’t understand what I was saying. I was limited to certain things. In the [sentence-selection version] it’s not certain what you are going to say means exactly what to them.”

Challenges with different interfaces

According to 71.4% of our participants, the free-form text entry version was the most challenging to learn and use (Fig. 5). As one would predict from conventional game design wisdom, the participants frequently complained that the game didn’t understand them. They struggled to figure out how to phrase their responses so that the game would correctly interpret and react to them. One participant said the NLU version had “too much freedom” and “it was too difficult to know when you can actually say something and what you are supposed to say.” Another felt that the freedom was actually an inhabiting factor: “You can’t make snap judgments. It disrupts the flow.” Another participant stated:

“The most frustrating was typing in my own responses. I guess I’d type in something that had a keyword in it, so [Trip and Grace] would take that keyword and use it how they were programmed to respond to that keyword but not in the context of my sentence.”

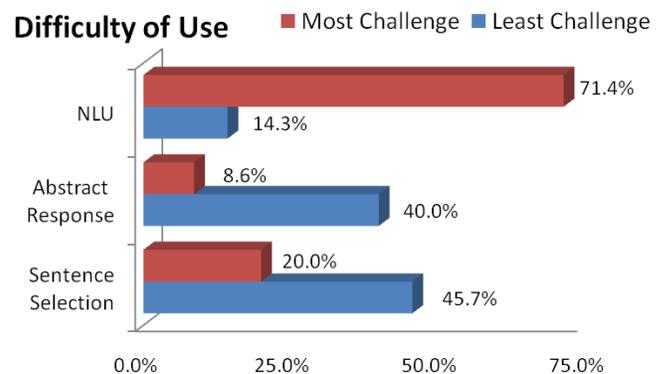


Fig 5. Challenge level. A significant majority of our participants found the NLU interface the most challenging.

As the NLU system in *Façade* maps user utterances onto a limited number of discourse acts, the sense of freedom inevitably broke down in some instances. Figuring out what utterances the NLU system supports was an essential step towards using the NLU version efficiently. Participants who were able to overcome this challenge were more likely to enjoy the system, whereas some participants weren’t able to make the transition from the constraints and affordances of real-life conversation to the modeling of conversation in the game. One participant reported feeling as if she was “behind a glass wall” the entire time she was using the NLU version, because Trip and Grace didn’t understand and didn’t respond to her. Another participant noted that while it was fun to watch Trip and Grace interact with each other, it felt like they didn’t want to interact with him.

For some participants, unconstrained typing also suggested informal communication strategies that were inappropriate. One participant noted that he tried to use slang and everyday college language, while another tried abbreviations such as “u” instead of “you.” Both styles of discourse were not forms understood by the NLU system.

The NLU version also does little to suggest when and what player actions are appropriate. Most of the participants stated that the game should more clearly indicate to them when their input is expected—the suggestions ranged from implementing subtle prompts that signal to the players when their input is expected to the game pausing entirely to wait for input. Some participants

expressed a need for a tutorial, or better feedback on how their input was processed and understood by the system. Although one of Mateas and Stern’s goals when developing the NLU system for *Façade* was to make sure the system never says, “I don’t understand” [7], a few participants actually wanted the system to somehow inform them that it didn’t understand, so that they wouldn’t miss the opportunity to interact.

Despite the fact that it most approached the natural flow of conversation presentation, pace and timing were also found to be significant drawbacks with the NLU version. The participants felt that the time required for deciding on a response, then formulating that response so that the game would understand it, and finally actually typing in that response, was too long. By the time the player was done responding to Trip or Grace, the characters would have already moved on to another topic, and the player input would no longer make sense in the original context: the opportunity to interact would be lost.

“The prompts definitely helped, not so much with what I needed to say. [...] but just the fact that it let me know when I could respond to what I was supposed to. [...] In the [NLU] version, you are trying to type in a response, and you are trying to think of something to say, and they have already moved beyond the question. It’s over before you can get anything out.”

Participants reported developing several strategies to cope with these problems. One participant tried to anticipate events and have a response typed in so that he could press enter to submit it at the correct moment. Another reported only typing in very short and simple phrases and sentences that he was sure the game would understand. While these are valid strategies, they also defeat possible goals of having a realistic conversation system, as player utterances are reduced to simple keywords that cannot possibly capture the nuances of real-life dialogue, and pacing becomes an inhibiting, rather than enabling, factor for presence.

A relatively minor number of participants reported issues using the menu-based versions. The most significant challenge that our participants experienced with the sentence selection version was that the options didn’t give the player a clear sense of what the outcome of speaking that line would be. Despite the fact that the time consuming nature of reading full responses is seen as the major drawback of this interface in the game design community, only two out of 34 participants complained about having to scroll through and read all the options.

We also asked our participants what improvements should be made to the interface. A more accurate NLU system and a prompt that informs the player when his or her input is expected were the most popular answers for the typing version. Participants wanted to see more options on the screen in the menu-based versions, reducing the need for scrolling, which is in fact one of the additional advantages of the interface introduced by *Mass Effect*, which arranges menu options radially..

Sense of Control

The three interfaces in this experiment offer different paths to giving the player a sense of control over the game world. The abstract-response version gives the player more control over the ultimate outcome whereas the NLU version gives the player more direct control over the avatar. This may explain some seemingly-contradictory statements about agency and control in the related literature. The sentence-selection version, on the other hand, may be seen as aiming for a more balanced approach by allowing some

degree of both authorial control and player freedom, allowing the player to choose among authored responses.

Our participants reported having the strongest sense of control in the abstract-response version (Fig. 6). Knowing the outcome of speaking a certain line of dialogue made players feel more influential in the game world.

While participants still enjoyed having total control over what the player character said, ultimately the difficulty of the interface coupled with the interpretation problems inherent to NLU resulted in a loss of control. In fact, 65.7% of our participants reported that they felt least in control using the NLU version.

Looking at our results, as predicted by our earlier work on agency, it seems players experience a greater sense of control when interfaces make the outcomes of their actions clear, rather than offering an illusion of greater control that isn’t entirely supported by the underlying system. Some participants even reported feeling overwhelmed by the freedom offered by the NLU version. They felt that so many things seemed possible that they had no idea how to choose a particular thing to say.

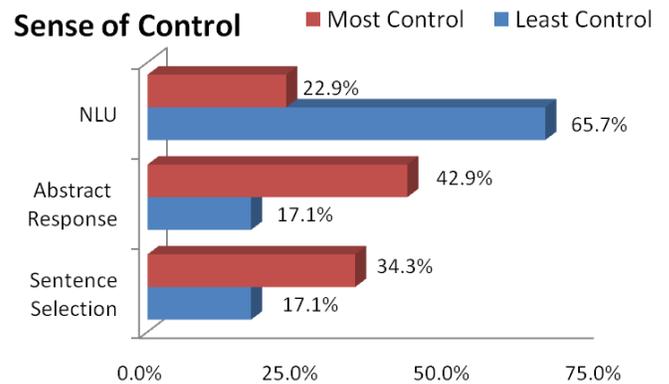


Figure 6. Sense of control. Participants felt most influential using the abstract-response version, and least influential using the NLU version.

Realism & Presence

During our interviews, we also tried to gain further insight on how a conversation system can feel realistic and natural (given this is a stated goal for many game designers) and how well our three different systems can support and maintain a sense of realism and naturalness. Participants who found the NLU version unnatural mostly complained of interpretation errors and felt that despite *Façade*’s attempts to the contrary, the limitations of the program were still very visible. One participant noted that “interacting with [Trip and Grace] still felt like interacting with software.” Another said “in [the NLU version] you can still tell the program is trying to hit keywords in a database.”

Most participants thought that the abstract-response version was an unrealistic model of conversation, because it strayed too far away from the realities of day-to-day conversation. Players are accustomed to conversation at the level of words and sentences, not discourse acts. Another factor that decreased presence in the abstract-response version was, as one would predict from prior game design wisdom, the mismatch between player’s intent when selecting a short, abstract response and the actual line of dialogue spoken by the player character as a result. One of our participants said:

“[In the sentence-selection version] you can get more into the character’s shoes. [In the abstract-response version] if you agree with [either Trip or Grace] your character might end up saying something [...] rude to the other person and that might not have been what you intended. [For example] if you disagree, [the spoken line] might make it sound more rude than you’d actually say.”

The main facilitator of presence was control over the player character’s statements. Participants noted having a stronger sense of control over what the player character will say in the NLU version, which potentially allowed them to be themselves in the world of Façade.

This increased sense of presence came with some trade-offs. Some participants felt that the increased sense of presence resulted in too much responsibility. They felt that the fate of Trip and Grace’s marriage was entirely in their hands and as a result the experience stopped being enjoyable.

Interestingly, we also observed that our participants felt they were more bound by social norms and conventions when using the NLU version. A participant noted that while she was trying to play a more difficult character in the abstract-response version and the sentence selection version, she definitely tried to be nicer in the NLU version. She stated:

“... [the NLU version] was more like a social situation than a multiple choice test. I was less inclined to say those things [I said in the other versions] that I wouldn’t normally say.”

Although on the surface free-form text entry seems like the most natural model of conversation, games present a system that players expect to be able to understand and influence more directly than real-life interactions. Some participants felt that with the NLU version, formulating and executing a plan was almost impossible, which resulted in a loss of control over the experience.

Story involvement

Our participants reported that they felt significantly more motivated to move the story forward using the sentence-selection version (Fig.7). One participant stated that “[the menu] was already there for me, it was easier for me to see what I wanted to do.” Another participant said:

“[in the sentence selection version] you had a lot more variety, a lot more range [compared to abstract response version], you had a lot more leeway, you can somewhat agree or somewhat disagree, whereas in the [abstract response] version you had to either go with this person or that person ... In the [NLU] one I was just so out of control that I just felt stranded a lot.”

Interestingly, this perception of more range and leeway was just an illusion: the sentence-selection and abstract-response versions were identical except for what was displayed in the selection menus. Still, this participant felt that with the sentence-selection version, she could relate more to the options and not feel trapped into taking a certain path in the game—having the actual lines of dialogue entirely written out gave the participant an illusion of more range and variety even though they were mapped to the same discourse acts as in the abstract-response version.

The immediate reaction from one of our participants when he saw the option to flirt with Trip in the very first menu was “Now I’m tempted to try this!” Some participants wanted to go back to the game after finishing and try other options to see how the

characters would react. Many participants stated that with the sentence-selection version they felt more like a character within the plot, although that character wasn’t them. None complained of being forced to play a particular character, though this is one of the concerns about this interface option expressed in the game design community. One of the participants noted that having prescribed options made even Grace and Trip feel more fleshed out and well-developed, and as a result made him more involved in the game world with a higher sense of purpose. Participants felt the menu-based versions better placed their character in context with the game’s dramatic events and the network of social relations between Trip, Grace and the player character. While free-form text entry, coupled with the game’s relative lack of back story for the player character, provided a blank canvas for the player to reflect his/her personality on, some participants didn’t feel that they had enough compelling reasons to care for Grace and Trip and their marriage. As a result, their actions felt meaningless.

Story Involvement

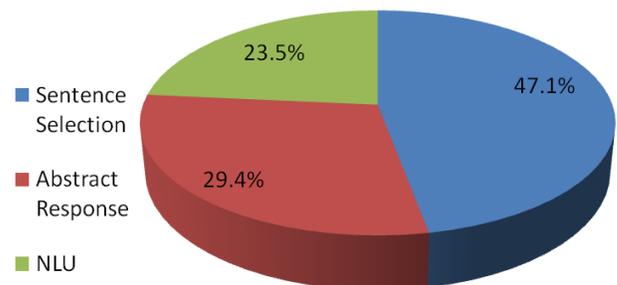


Figure 7. Story involvement. Participants were more involved in and more motivated to move the story forward in the sentence-selection version.

Enjoyment

Despite all the significant drawbacks that they mentioned, and flying in the face of conventional wisdom, more than half of our participants still reported enjoying the NLU version the most (Fig. 8). The participants particularly enjoyed being able to say whatever they wanted and interrupt the characters at any time, in contrast to the discrete and limited interaction possibilities offered by the menu-based versions. When asked which version she liked the most, one participant noted:

“The [NLU version] for sure. You had a lot more freedom in what you could say. If you want to put in your opinion about something while they are talking, it felt like you could do it then rather than just waiting for this [menu] to pop up with limited choices of what you want to say. Maybe it’s not something you want to say but you don’t have a choice. I guess it’s the freedom and the real-time thing too. [The NLU version] is not like wait-go-wait-go.”

Participants also noted that while the NLU version had its problems, it was a very fulfilling experience when it actually worked. As noted by a participant:

“[I enjoyed the NLU version the most] when they threw me out. I told Trip to shut up ... I didn’t really expect [to be thrown out]. Seeing the reaction kind of quickly made [the NLU version] engaging.”

A participant even noted that the NLU version gave him the most sense of control, because it enabled him to take a back seat and let Trip and Grace work out their problems on their own. When playing the NLU version, he just answered enough questions so he didn't get thrown out. He stated:

"[In the NLU version] I thought maybe if I just listened they would figure things out themselves. It was clear from the first two [playthroughs] that what I said wasn't helping them [In the typing version] I had more control, even though I didn't control anything because I didn't influence them at all I felt like I was more in control and more able to help them ... I influenced them by not doing anything and just listening, and letting them work things out themselves."

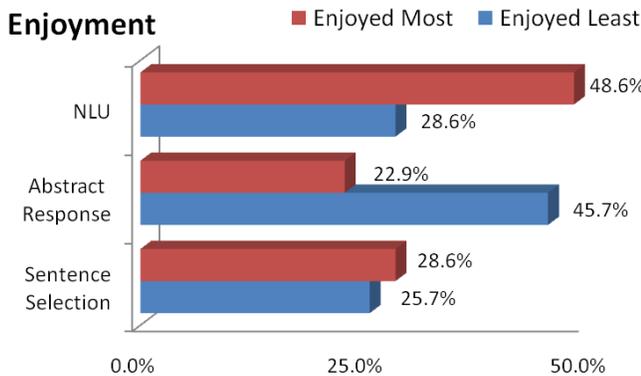


Fig 8. Enjoyment. Participants noted enjoying the NLU version the most, and the abstract-response version the least.

Our subjects noted that although they realized the limitations of the system, the NLU version still gave them an illusion of freedom that is absent from other versions; as one participant noted, when using the NLU version it “felt like everything was possible.” Another said that his “imagination just died in the [menu-based] ones.”

This illusion of freedom, however, did not translate to enjoyment for all participants. Some felt it was too difficult to figure out what to say since too many things seemed possible at any given moment. As a result, players’ sense of agency suffered. Participants who enjoyed the sentence-selection and abstract-response versions usually noted that the game allowed them to think, make decisions, and execute their plans, in contrast to the NLU version which suffered from misinterpretation and timing problems, along with too much unguided freedom which made it difficult to figure out what actions to take to move the story forward.

5. CONCLUSION

Our direct comparison of game dialogue interfaces has some surprising findings. In particular, players enjoy most the interface approach—natural language understanding—that also makes them feel least in control and often produces frustrating errors. However, this does not necessarily indicate that game designers should choose NLU dialogue interfaces. Sentence selection interfaces were found to be most effective for producing a sense of story involvement. Common critiques (that sentence selection takes too long to read and puts words in the player’s mouth) are not supported by our findings. Abstract response menu interfaces, in contrast, which have been praised for preserving natural conversation flow, are found most unnatural in our results, though

most effective at producing a sense of control relative to the game system. Perhaps most fundamentally, our study demonstrates that game dialogue interfaces have a profound impact on the experience of gameplay, even when all other aspects of the game are held steady, something of which all designers should be aware.

REFERENCES

- [1] Adams, E., “You Must Play Façade, Now!” Gamasutra, July 28, 2005. http://www.designersnotebook.com/Columns/075_You_Must_Play_Facade_Now/075_you_must_play_facade_now.htm
- [2] Brown, E., Cairns, P., “A Grounded Investigation of Game Immersion.” CHI '04, Extended Abstracts on Human factors in Computing Systems, 2004.
- [3] Dow, S., Mehta, M., Harmon, E., MacIntyre, B., Mateas, M., “Presence and engagement in an interactive drama.” Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2007.
- [4] Ellison, B. Defining Dialogue Systems. Gamasutra, July 8, 2008. http://www.gamasutra.com/view/feature/3719/defining_dialogue_systems.php
- [5] IJsselsteijn, W.A., Riva, G., “Being There: The experience of presence in mediated environments.” 2003.
- [6] Martin, P.Y., Turner, B.A. “Grounded Theory and Organizational Research.” The Journal of Applied Behavioral Science, 22(2), 141., 1986
- [7] Mateas, M., “Games and Natural Language Understanding.” Grand Text Auto, September 17, 2004. <http://grandtextauto.org/2004/09/17/games-and-natural-language-understanding/>
- [8] Mehta, M., Dow, S., Michael M., and MacIntyre, B. “Evaluating a Conversation-centered Interactive Drama.” Autonomous Agents and Multi-Agent Systems (AAMAS'07), 2007.
- [9] Milam, D., Seif El-Nasr, M., and Wakkary, R., “Looking at the Interactive Narrative Experience through the Eyes of the Participants,” Proceedings of the 1st Joint International Conference on Interactive Digital Storytelling: Interactive Storytelling, Vol 5334, 96 – 107, 2008.
- [10] Nutt, C. AGDC: BioWare Charts Writing For Mass Effect. Gamasutra, September 7, 2007. http://www.gamasutra.com/php-bin/news_index.php?story=15406
- [11] Rouse, R. Game Design Theory & Practice. Wordware Publishing, Inc., 2001
- [12] Sheldon, L. Character Development and Storytelling for Games. Course Technology PTR, 2004
- [13] Short, E. Conversation. Emily Short’s Interactive Storytelling. <http://emshort.wordpress.com/how-to-play/writing-if/my-articles/conversation/>
- [14] Wardrip-Fruin, N., Mateas, M., Dow, S., Sali, S., “Agency Reconsidered.” Digital Games Research Association (DiGRA), 2009.
- [15] Weibel, D., Wissmath, B., Groner, R. “Presence vs. Flow in the Context of Computer Games.” International Society for Presence Research (ISPR), 2007.