

Preference, Perception, and Task Completion of Open, Menu-based, and Directed Prompts for Call Routing: a Case Study

Jason D. Williams, Andrew T. Shaw, Lawrence Piano, Michael Abt

Edify Corporation, 2840 San Tomas Expressway, Santa Clara, CA 95051 USA
{jason.williams, andrew.shaw, lawrence.piano, michael.abt}@edify.com

Abstract

Usability subjects' success with and preference among Open, Menu-based and Directed Strategy dialogs for a call routing application in the consumer retail industry are assessed. Each subject experienced two strategies and was asked for a preference. Task completion was assessed, and subjective feedback was taken through Likert-scale questions. Preference and task completion scores were highest for one of the Directed strategies; another directed strategy was least preferred and the open strategy had the lowest task completion score.

1. Introduction

"How may I help you?"-style interactions are increasingly being used in commercial speech-recognition applications for call routing. This type of "Open" strategy prompt is often preferred assumed to be superior to "Directed" or "Menu" strategies, typically because it is believed to increase caller satisfaction and task completion rates. This paper seeks to explore this assumption in controlled usability testing in a particular domain.

2. Background

Several studies have attempted to assess how prompt strategy affects user satisfaction and task completion scores, but none have applied this specifically to the call routing task.

[1], [2], and [3] give general insights into task completion and caller preference for open vs. directed prompts, but are not applied to the call routing task. [1] and [2] tested a train timetable system of various levels of "open-ness". [1] found that user satisfaction is shown to derive from user perception of task completion, recognition rates, and amounts of barge-in. "No efficiency measures are significant predictors of performance" for real systems, but "in the presence of perfect ASR, efficiency becomes important." [2] found that "Strategies that produce longer tasks but fewer misrecognitions and subsequent corrections are preferred by users." [3] explored a Yellow-pages search task, and found that listing fewer choices over more turns achieved higher preference scores than asking open questions among more choices using fewer turns. "It is interesting to note that even in the case of a search task, where the interaction itself should not matter as much as the information to be retrieved, users did not necessarily prefer the interface that would take them the fastest to the desired result."

[4] and [6] studied user behavior using "open" prompts for call routing. [4] compared a variety of wordings for an "open" prompt, seeking maximal task completion and caller preference scores, but did not compare it with a directed or menu-based approach. [6] assess responses to a human operator's "How can I direct your call?" [7] studied usability

subjects' responses to a banking application's first prompt – a call routing task. The study compared responses to three prompts: "How can I help you?", "Which service do you require", and "Please say *help* or the name of the service you require". Users scored the third strategy highest for "knowing how to select options" (using Likert scores). Subjects expressed no clear preference between the systems, and 69% of subjects said "help" in the first turn in the third experiment.

3. Methodology

3.1. Domain and assumptions

The project was initiated for an Edify customer in the consumer retail domain, referred to here as the Client. The Client sought to route calls based on task area (Table 1) and product category/model ID. The "routing buckets" were ultimately agent queues or self-service automation.

We assumed it would not be possible to reliably capture (or determine absence of) product category/model number *and* task area in one utterance. We did not test this assumption experimentally, and limited ourselves to approaches which obtained the task area and product category/model number in separate utterances. Further, the caller population was composed primarily of first-time or infrequent callers, and we limit ourselves to analyzing the effects of the first encounter; we don't explore the effects of repetitive use.

3.2. Routing strategies

Table 2 shows the three routing approaches we identified. The "Directed" approach sought to glean the task through asking multiple questions; by ordering these questions differently we created three different Directed strategies called Direct-1, Directed-2, and Directed-3. Example interactions with all of the strategies are shown in Figures 1-5.

Each strategy included 2 escalating no-input (i.e., silence) and 2 no-match (i.e., invalid/out-of-grammar) prompts at each interaction point. If a subject triggered a total of more than 2 no-inputs or more than 2 no-matches at a given interaction point, an agent transfer was simulated.

3.3. Experimental design

Each of the strategies above was implemented using a Wizard-of-Oz based system. The telephony interface supported barge-in. The same voice talent was used for all experiments, and the voice coaching attempted to maintain consistency of persona throughout. "Perfect" speech recognition was assumed by the wizard – i.e., all utterances which *could* be classified by the Wizard for a given state were treated as successful recognitions. Thus the system was evaluating utterance content and not utterance quality, the latter of which would impact a deployed ASR system.

Subjects were selected from a pool of the Client’s customers with a representative distribution of age, profession, and gender. Subjects were provided with a cash incentive.

Task area	Description
Technical support	Phone-based product support such as product usage questions and hook-up
Product information	Information about products or accessories that a caller is considering purchasing
New repairs	Finding local repair stores and arranging to mail products to Client for repairs.
Store locations	Finding a local store where a caller can see/buy a particular product
Existing repairs	The status of a repair the Client is currently working on.

Table 1: Five task areas

Strategy	Description
Main-menu	Provide list of task choices as keywords in one menu.
Open	In one interaction, ask open-ended question like “What can I help you with?” possibly supported by examples.
Directed	In multiple interactions, ask a series of constrained, binary questions to determine task area, intermixed with capturing product category.

Table 2: Prompting approaches determining task area

Each session consisted of a short welcome and introduction from a standard script. Subjects were told they would be using two different prototypes of a new system, but not told anything about what was different between them. The subject was presented with 6 tasks to perform with one routing strategy, then a different 6 tasks to perform with a second routing strategy.

Tasks were written and presented one at a time. Example tasks included:

- You’re interested in a 17” TV with the picture in picture feature. You are calling to find out how much they cost.
- This camera was given to you as a gift for your birthday last week. You haven’t been able to get the camera to turn on. You are calling to get help trying to turn it on. (The camera was available to the subject in the room).

A total of 16 sessions were conducted. Each session was video and audio recorded. Because of limited time and subject pool, it was not possible to compare every strategy to every other strategy. Instead, we compared each Directed strategy to the Open or Main-menu strategy, varying which strategy was presented first.

At the end of the session, subjects were asked which system they preferred. Subjects were then read a series of statements about the system they preferred and asked to respond using a 7-point Likert scale, intermixed with several free-response questions.

3.4. Assessing task completion

A task attempt was graded successful if the caller would have been transferred to the correct “routing bucket” for their task

– i.e., both the task and product category/model number had been correctly received by the system. No-inputs (i.e., caller silences) and no-matches (i.e., simulated recognition rejects)

Experiment (strategy)	First in session	Second in session	Total experiments
Main-menu	5	4	9
Open	3	4	7
Directed-1	2	2	4
Directed-2	3	3	6
Directed-3	3	3	6

Table 3: Counts of each experiment conducted, and position (first or second) in the usability session

Experiment (strategy)	Preferred (sessions)	Not preferred (sessions)	No preference (sessions)
Main-menu	5 (56%)	4 (44%)	0
Open	3 (43%)	3 (43%)	1 (14%)
Directed-1	1 (15%)	2 (50%)	1 (25%)
Directed-2	1 (17%)	5 (83%)	0
Directed-3	5 (83%)	1 (17%)	0

Table 4: Preference by strategy

First experiment preferred	Second experiment preferred	No preference expressed
8 (50%)	7 (44%)	1 (6%)

Table 5: Preference for a strategy by position of experiment (first vs. second) in usability session

did not constitute a failed task attempt unless the no-input or no-match limit was reached in one interaction state.

4. Results

Table 3 shows the total number of experiments conducted for each strategy. Table 4 shows preference vs. strategy.

We considered whether the order a caller experienced the strategies could influence their preference; Table 5 shows preference was not correlated with experiment position.

Table 6 shows task completion for all task attempts (across both experiments in all sessions). Table 7 shows the questions scored on the Likert scale, and Table 8 shows the resulting scores for the strategy that callers preferred.

5. Discussion

Preference scores were highest for the Directed-3 strategy. While Directed-3 was not compared directly to Directed-1 and Directed-2 (section 3.3), it seems reasonable to conclude that Directed-3 would be preferred to Directed-2 and Directed-1 given their relative scores vis-à-vis the Open and Menu-based strategies.

From listening to call center agent interactions, we believe that Directed-3 most closely models typical conversation flow, and believe this correlation determined its success – i.e., when a directed human/machine dialog follows human/human norms, it can outperform an open- or menu-based approach; conversely, when a directed human/machine dialog deviates from human/human norms, it is more likely to under-perform open- or menu-based approaches.

Subjects were most successful with Directed-1 and Directed-3; they were least successful with the Open strategy.

Experiment (strategy)	Success (task attempts)	Failed (task attempts)	Undecidable task attempts ¹
Main-menu	48 (89%)	6 (11%)	0
Open	31 (79%)	8 (21%)	3
Directed-1	24 (100%)	0	0
Directed-2	31 (86%)	5 (14%)	0
Directed-3	36 (100%)	0	0

Table 6: Task completion by strategy

ID	Statement (for Likert questionnaires)
1	I felt in control while using the speech system to perform my tasks
2	Overall, this speech system gave me an efficient way to perform my tasks
3	Overall I was satisfied with my experience using the speech system to accomplish my tasks
4	Overall, I was comfortable with my experience calling this speech system.
5	When the system didn't understand me, or if I felt hesitant, it was easy to get back on track and continue with the task
6	Overall, the system was easy to use
7	The choices presented to me by the system were what I expected to complete the tasks

Table 7: Statements used for Likert scale responses

Strategy	N	1	2	3	4	5	6	7	Av
Main-menu	5	5.4	6.2	6.4	6.2	5.5	6.2	5.2	5.9
Open	3	5.7	5.7	6.0	6.3	7.0	7.0	6.0	6.2
Directed-1	1	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0
Directed-2	1	5.0	6.0	5.0	6.0	3.0	5.0	6.0	5.1
Directed-3	5	6.0	6.6	6.8	7.0	6.6	7.0	6.6	6.7

Table 8: Likert scale responses by strategy

In analyzing the subject test scores, we discarded results from Directed-1 and Directed-2 (which were selected only once each).

At a confidence interval of $p \leq 0.05$ (using the Mann-Whitney test [8]), and taking all Likert scores together as a collective measurement of preference, Directed-3 scored higher than Main-menu ($p \leq 0.0052$). Relaxing the confidence interval to $p \leq 0.11$, Directed-3 scored higher than Open. At $p \leq 0.05$, none of the individual comparisons were statistically significant; relaxing the confidence interval to $p \leq 0.07$, users were more satisfied (question 3) with Directed-3 strategy than with the Open strategy. In addition, users were more comfortable (question 4) with Directed-3 than with the Closed strategy, and thought that the Directed-3 strategy was easier to use.

Anecdotal observations for the main-menu strategy included:

- Subjects who did preferred the main-menu strategy cited they liked “knowing all their options.”

¹ One subject exhibited extensive confusion with the Open strategy, and their results are excluded from the percentages reported.

- Subjects who didn't prefer the main-menu strategy had trouble determining which menu option was appropriate.
- Most failures were due to selecting the wrong item from the main menu.

We speculated that, though preferred by some subjects, having a list of choices created a *false sense of control* as the options were interpreted differently by the subjects.

Anecdotal observations for the open strategy included:

- Most subjects using this system did not respond before the pause in the initial prompt.
- Most failures were due to selecting the wrong item from the list that played after the Open prompt.

In addition, subjects who did respond to the open question were more likely to respond to subsequent questions with longer utterances.

We speculated that, when asked an open question, subjects lacked an expectation of the system's abilities and usually waited for more guidance.

Anecdotal observations for the directed strategy included:

- Subjects who preferred the Directed Strategy liked that the system was “taking charge of their problem” and leading them through their choices.
- Subjects who didn't prefer the Directed Strategy preferred having a sense of all their choices.

6. Conclusions and future work

We have shown usability test results for preference and task completion among 3 prompt approaches in the consumer retail domain. A directed approach was found to produce higher task completion and preference scores than an open or a main-menu approach.

Future work would attempt to corroborate these findings through comparing multiple variants of each strategy – recent work has found that wording of open prompts has an appreciable effect on user preference and task completion [4]. In addition, using a “double-blind” experimental set-up with more subjects would give better resolution.

Finally, repeating this test in other domains (e.g., banking, airlines) and with other user populations (e.g., regular users, trained users) would help isolate domain-specific and user-group specific factors determining appropriateness for each approach.

7. Acknowledgements

The authors would like to thank Judee Humberg for providing valuable input on the of the usability test procedure and interview questions.

8. References

- [1] D. J. Litman et al., “Evaluating Response Strategies in a Web-Based Spoken Dialogue Agent.” Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL'98), pp. 780-786, Montreal, Canada, August 1998.

- [2] M. Swerts et al., "Corrections in spoken dialogue systems." Proceedings of the International Conference on Spoken Language Processing (ICSLP 2000), pp. 615-618, Volume II, Beijing, China, 2000.
- [3] W. V. Vanhoucke et al., "Effects of Prompt Style when Navigating through Structured Data" (with accompanying presentation Speech and Search). Proceedings of INTERACT 2001, Eighth IFIP TC.13 Conference on Human Computer Interaction (IOS Press), pp.530-536, Tokyo, Japan, 2001.
- [4] T. Sheeder, "Learning From User Performance: A Laboratory Study." Nuance V-World 2002 (Nuance), Orlando, Florida, 2002.
- [5] F. Stentiford and P. Popay, "The design and evaluation of dialogues for interactive voice response services" BT Technical Journal (British Telecom: United Kingdom). No 1, Vol 7, January 1999. pp 160-171. pp 142-148.
- [6] B. Carpenter and J. Chu-Carroll, "Natural Language Call Routing: A Robust Self-Organizing Approach." ICSLP 98, Sydney, 1998.
- [7] F. McInnes et al., "Effects of prompt style on user responses to an automated banking service using word spotting." BT Technical Journal (BT: United Kingdom). No 1, Vol 7, January 1999. pp 160-171.
- [8] R. Lowry. Concepts and Applications of Inferential Statistics. Vassar College. On-line textbook. <http://faculty.vassar.edu/lowry/webtext.html>.

System: Hello, welcome to <Client>. My name is Johnson, your virtual assistant. Please take note: this service has recently changed. **Now, please tell me which of the following options you'd like: tech support, repairs, product information, or store locations.**

User: Ah, product information

System: Ok, product information. What product are you calling about?

User: It's a television – seventeen inch television

System: What's the model number? It's ok to say, I don't know.

User: Ah, I don't know.

System: No problem. I'm transferring you to an agent who can help you with your question.

Figure 1: Main-menu strategy (sample dialog)

System: Hello, welcome to <Client>. My name is Johnson, your virtual assistant. Please take note: this service has recently changed. **Now, what can I help you with? [2.5 sec pause] You can get tech support, product information...**

User: Ah, yes, I need... I'm thinking about buying a TV

System: Ok, product information. What product are you calling about?

User: It's a television – seventeen inch television

System: What's the model number? It's ok to say I don't know.

User: Ah, I don't know.

System: No problem. I'm transferring you to an agent who can help you with your question.

Figure 2: Open strategy (sample dialog)

System: Hello, welcome to <Client>. My name is Johnson, your virtual assistant. Please take note: this service has recently changed. **To get started, if you've got an event number or work order number, tell me it now. Otherwise, say NEW QUESTION and we'll start from the beginning.**

User: New Question

System: Ok, a new question. What product are you calling about? Or, if it's an accessory you're after, say ACCESSORIES.

User: It's a Television

System: What's the model number? It's ok to say, I DON'T KNOW

User: Ah, I don't know

System: No problem. **Do you own this product now?**

User: No

System: Great. I'm transferring you to an agent who can help you.

Figure 3: Directed-1 strategy (sample dialog)

System: Hello, welcome to <Client>. My name is Johnson, your virtual assistant. Please take note: this service has recently changed. **To get started, do you need help with a product you own now?**

User: No

System: Oh, ok, what product are you calling about?

User: It's a Television

System: What's the model number? It's ok to say, I DON'T KNOW

User: Ah, I don't know

System: No problem. I'm transferring you to an agent who can help you.

Figure 4: Directed-2 Strategy (sample dialog)

System: Hello, welcome to <Client>. My name is Johnson, your virtual assistant. Please take note: this service has recently changed. **To get started, if you've got an event number or work order number, tell me it now. Otherwise, say NEW QUESTION and we'll start from the beginning.**

User: New Question

System: Ok, a new question. **To get started, are you calling about a product you own now?**

User: No

System: Oh, ok. What product are you calling about? Or, if it's an accessory you're after, say ACCESSORIES.

User: It's a Television

System: What's the model number? It's ok to say, I DON'T KNOW

User: Ah, I don't know

System: No problem. I'm transferring you to an agent who can help you.

Figure 5: Directed-3 strategy (sample dialog)