Two Studies of Open vs. Directed Dialog Strategies in Spoken Dialog Systems

Silke M Witt, Jason D Williams

Edify Corporation, 2840 San Tomas Expressway, Santa Clara, CA 95051 USA
`{switt,Jason.Williams}@edify.com`

Abstract

This paper analyzes the behavior of callers responding to a speech recognition system when prompted either with an open or a directed dialog strategy. The results of two usability studies with different caller populations are presented. Differences between the results from the two studies are analyzed and are shown to arise from the differences in the domains. It is shown that it depends on the caller population whether an open or a directed dialog strategy is preferred. In addition, we examine the effect of additional informational prompts on the routability of caller utterances.

1. Introduction

"How may I help you?"-style prompts are increasingly being used in commercial speech-recognition applications for call routing. This type of *Open Strategy* prompt is often preferred to a *Directed Strategy* prompt such as "Main Menu – please say 'check balances, transfer funds or order checks" because it is widely believed to increase caller satisfaction, decrease task completion times, and increase task completion rates. This paper seeks to investigate whether these assumptions hold true for different commercial spoken dialog systems with different caller populations and for different types of open strategy prompts.

This paper describes and discusses two usability studies that were conducted by Edify's Natural Language Solutions Group as part of two different speech system development projects. Study A was done for a large consumer electronics company and tested different dialog design approaches for a technical support hotline (see Table 2 for a listing of task areas). Study B was done for a large US tele-communications company to automate human resources queries from employees, for examples callers wanting to update their direct deposit or check their vacation balances. In the case of Study A, the caller population consisted of external, infrequent callers with no clear expectation of the task, whereas the caller population for Study B was an internal population with some familiarity of the application.

2. Background

While it is a known fact that the choice of prompt wording has a large influence on the performance of a speech recognition system, so far few studies have assessed the effects of prompts type on different caller populations and task completion. The results reported here are for commercial systems optimized first for task completion and second for task completion time.

Several previous studies have looked at related questions regarding the effect of prompting design on caller responses.

[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. [2] found that "Strategies that produce longer tasks but fewer recognition failures and subsequent corrections are preferred by users." [3] explored a Yellow-pages search task. The study found that explicitly listing fewer choices using more turns was found to be more preferred by users 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 [5] studied user behavior in open prompts used for call routing. [4] compared a variety of wordings for an "Open" prompt to find an approach that maximized task completion and caller preference in a call routing task, but did not compare it with a directed or menu-based approach. [5] assessed responses to a human operator's "How can I direct your call?"

[6] studied usability subjects' responses to a banking application's first prompt – a call routing task, same as in the studies presented here. Three prompt strategies were compared, Open, Mid and Directed, where the Mid strategy refers to an open prompt that has a more specific wording than an entirely open prompt, for example "Main menu – which service do you require?". Users reported significantly higher rates for "knowing how to select options" for the *Directed* style than the others. However, overall, no clear preference between the systems was reported. 69% of subjects asked for "help" in the first turn in the *Directed style* experiment.

[6] differs from the present studies in three ways: (1) the Study A makes a distinction between a *Directed* and *Menu* strategy, assessing the effect of number of turns vs. options at each turn; (2) "routability" was assessed in [5] assuming keyword spotting rather than assessing information content and assuming an SLM, and most importantly, (3) in [6], none of the tested prompts explicitly listed choices; the results from [6] demonstrate that when callers know how to get a list of choices, the vast majority will request them. Therefore, the present experiments include prompt variants which list the routing choices.

3. Experimental Setups

Study A was implemented using a Wizard-of-Oz-based system. A wizard controlled what was "recognized" at each state. The telephony interface supported barge-in throughout all prompts, and the speech detection was done by the telephony card (not the wizard). Study B on the other hand was based on a speech recognition system containing only a single dialog stage, again barge-in was enabled and the telephony card made the speech/no speech decisions. The decision as to whether an utterance was routable (assuming an SLM-type of recognition grammar) was done by a human listening to the utterances. In both cases live calls were diverted from an existing hot-line.

3.1. Study A Experimental Setup

We identified three components of possible introductory greetings and two routing strategies, open and directed. The wording for each of these prompts can be found in Table 1. These prompts were combined to create 6 experiments, shown in Table 2.

Prompt segment	Prompt text
Hello	"Hello, welcome to ACME"
Earcon	*[earcon]*
Persona & recently changed notice	My name is Justin, your virtual assistant. Please take note, this service has recently changed."
Directed	"Please tell me which of the following options you'd like: tech support, repairs, product information, or store locations."
Open	"What can I help you with? *[2.5 sec pause]* You can get tech support, product information, repair information, or store locations. Which would you like?"

Table 1: Prompt types and wordings for Study A

Experiment	Prompt construction
1	Hello + Directed
2	Hello + Open
3	Hello + Earcon + Directed
4	Hello + Earcon + Open
5	Hello + Earcon + Persona & recently changed notice + Directed
6	Hello + Earcon + Persona & recently changed notice + Open

Table 2: Experiments of Study A

3.2. Study B Experimental Setup

The Study B experiments were conducted with the results of Study A in mind. Thus the experiments were designed to focus on questions not yet answered by Study A. These questions were:

a) Whether the preference for open or directed prompting depends on the domain of the application, and

b) Whether it is more effective to use open prompt followed by delayed help or rather preceded by examples.

By 'Delayed help,' we mean playing the help message if there is more than 2 seconds of pause. The advantage of this design is that those callers who are in need of additional information will get to hear this information if they are hesitating with their answer while for those callers who know what to say the pause provides a natural opportunity to interrupt the system.

Table 3 lists the wording of the prompts used in study B and Table 4 defines the experiment setups.

Prompt Segment	Prompt text
Hello	Hello. Thanks for calling our HR Self Service and Information Line. My name is Jeanine and this is our new speech recognition system.
Open	*what would you like to do?*
Mid	*please tell me the reason for your call*
Delayed help	*For example, you can say: I NEED TO RESET MY WEB PASSWORD or I'D LIKE TO TALK TO SOMEONE ABOUT SAVINGS PLANS*
Preceding Examples	*Here are some examples of what you can say. I NEED TO RESET MY WEB PASSWORD or I'D LIKE TO TALK TO SOMEONE ABOUT SAVINGS PLANS.*
Directed	*To help me direct your call, please choose from the following options: WEB PASSWORD RESET, COURSE ENROLLMENT, DIRECT DEPOSIT, or BENEFITS. If you didn't hear what you want, say MORE OPTIONS.*

Table 3:Prompt types and wordings for Study B

Experiment	Prompt construction
1	Hello + Preceding examples + Open
2	Hello + Open + Delayed help
3	Hello + Preceding examples + Mid
4	Hello + Mid + Delayed help
5	Hello + Main Menu

Table 4: Experiments of Study B

3.3. Utterance classification

The caller's first response to each interaction was classified into three main categories, invalid utterances, routable utterances (which translates to a successful utterance) and failure. For details see Table 5.

4. Discussion

4.1. Discussion of Study A

Table 6 shows the results of Study A. Experiment 5 has the highest percentage of correctly routed utterances and the lowest percentage of unroutable utterances. Experiment 5 was a directed-dialog strategy, combined with the earcon, named persona, and recently changed notice (see Table 3). These results indicate that for an application with a large caller population, very few repeat callers and a large, disparate set of tasks, directed-dialog is better than an open strategy.

Category	Sub-category label	Sub-category description
Invalid (excluded from analysis)	I	Invalid utterance – background noise, voices, end-pointer error, etc.
Routable (success)	R	Routable: utterance contained enough information
Failures	S	Silence – end-pointer didn't trigger
	C	Obvious caller confusion – e.g., "Hello?"
	A	Explicit agent request
	U	Cooperative caller but unroutable for reasons not listed above

Table 5: Call classification rules

Exp	Num	R	S	C	A	U
1	250	47%	45%	3%	0%	4%
2	60	31%	39%	25%	3%	5%
3	196	54%	38%	4%	0%	0%
4	53	36%	14%	36%	0%	15%
5	131	67%	29%	2%	1%	1%
6	44	46%	22%	18%	0%	13%

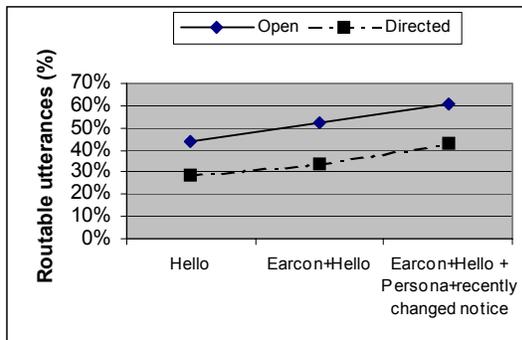*Table 6: Result summary for Study A*

Figure 1: Routability results for open and directed strategies for Study A

From Figure 1 it can also be seen that the directed strategy yields higher routability rates than the open strategy for every prompting combination. In other words, the directed strategy appears more suitable to this task (as indicated by callers' very first reaction) than the open strategy independent of any additional prompting.

This observation is confirmed by the fact that the confusion rates, *C*, for all three open strategy experiments are higher than the confusion rates for the directed strategy, again this independent of any additional information in prompts preceding the main prompt.

Furthermore, we hypothesized that the category U (unroutable utterances) that contained only a product name but no indication of task that caller wants to perform could have been caused by the language of the prompt – "what can I help you *with*?" as opposed to "How can I help you?" If

those utterances with a product name in isolation are counted as members of the category *R* (routable), this diminishes the gap in routability between the Open and Directed strategies, but does not alter the ranking.

4.2. Discussion of Study B

In this study four different open prompt combinations and one directed strategy were tested. The results are shown in Table 7. The routability rates do not vary to a large degree, but Experiments 3 and 4 clearly cause less routability and more silence than the other 3 experiments. Both Experiment 1 and Experiment 5 have the highest percentages of routable utterances. Also, Experiment 1 has a higher percentage of silence whereas Experiment 5 has more confusions and un-routable utterances.

Exp	Num	R	U	S	A	C
1	168	72%	0%	25%	2%	0%
2	175	73%	2%	22%	2%	2%
3	161	67%	2%	26%	1%	3%
4	174	67%	2%	28%	0%	3%
5	180	71%	2%	24%	0%	4%

Table 7: Result Summary for Study B

To get a better understanding of the differences between these two experiments, we next looked at the routability of responses to a second prompt (occurring after a recognition failure). In cases where the first utterance did not get successfully recognized, the system in Study B would re-prompt the caller with a more specific prompt, see Table 4.

Figure 2 compares the percentage of routable utterances at the initial and the secondary prompt. Interestingly, using a more specific second prompt after an open prompt yields a higher percentage of routable utterances. This is the opposite of the directed strategy where the routability of the second utterances decreases. Probably the reason for this pattern is that in the combination of open prompt followed by directed prompt two strategies are combined two different approaches to elicit an answer from the caller whereas in the latter case, the same strategy gets repeated twice, no new information is added.

This finding shows that, at least in the domain of Study B, a dialog design that starts with an open prompt and follows with a directed prompt in the case of a recognition failure has the best of both worlds. It can enable a spoken dialog system to both catch the utterances of expert users and those who need additional guidance. Also, the increase in routable utterances from the first to the second utterances shows that callers are taking their cues from the prompts and are co-operating with the system to complete their task.
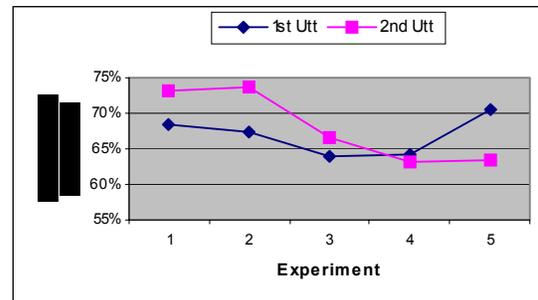
4.3. Comparison of both Studies

Figure 3 depicts a comparison of both studies for the average results of the directed and open prompting strategies. The open strategy in Study A has a significantly larger amount of unroutable and confused utterances (U and C) compared to the open strategy in Study B. This indicates that for a domain in which the caller population is internal and familiar with the task, the open prompting strategy might be a good dialog design choice, whereas in the case of an external caller population that does not have a clear expectation of task details, a directed approach might yield higher caller satisfaction than an open approach due to significantly less confusions and failures.
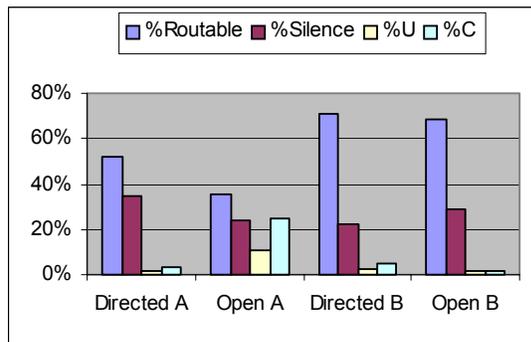

Figure 3: Open vs. Directed for both Studies

Next, both studies experimented with combining the main prompt with additional informational prompts. Study A has shown that no matter what prompting strategy is used, preceding the main prompt with an earcon, a welcome prompt or a recently changed notice significantly decreases the caller confusions. Additionally, the results of Study B have shown that the actual wording of the open prompt has a direct effect on the routability rate, in this case the "mid" prompt was less successful than the open prompt.

Finally, Study B also showed that using preceding examples rather than delayed-help improves routability. This again agrees with the previous findings, see [1] and [2], that callers feel more comfortable the more they know about their choices.

5. Conclusions

The results of the two studies described in this paper demonstrate the effect of different prompt designs for call routing tasks on the task completion rate. Open dialog strategies were contrasted with directed dialog strategies in two different domains. In the domain where the callers are more familiar with the task details and the system an open strategy is more successful than a directed strategy. However, in the case of less familiarity with the task, callers preferred more guidance in the prompting.

The fact that callers prefer additional guidance was also apparent from the observation that additional information preceding the main prompt increases the routability of utterances.

Viewing these results in combination with findings from previous studies, see [1,2,5,6], shows that more often than not a directed approach is preferred over an open dialog strategy, even though the open strategy is often thought to be more natural thus more likely to be accepted by callers. However, the answer whether a open or a directed design is more appropriate depends on the domain and the caller populations.

6. References

[1] D. J. Litman et al., "Evaluating Response Strategies in a Web-Based Spoken Dialogue Agent." In *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] B. Carpenter and J. Chu-Carroll, "Natural Language Call Routing: A Robust Self-Organizing Approach." *Proceedings of ICSLP 1998*, Sydney, 1998.

[6] F. McInnes et al., "Effects of prompt style on user responses to an automated banking service using word spotting." *BT Technical Journal* (British Telecom: United Kingdom). No 1, Vol 7, January 1999. pp 160-171.