Research-Based User Interface Recommendations for Natural Language Call Routing

Technical Report BCR-UX-2007-0001
November 7, 2007

James R. Lewis

IBM Software Group
Boca Raton, Florida

Abstract

This report provides research-based speech user interface recommendations for natural language call routing systems.

Keywords

Voice systems
Speech systems
Call router
Interactive Voice Response (IVR) systems
Speech user interface
Natural Language Call Routing
NLU

Contents

EXECUTIVE SUMMARY **1**

INTRODUCTION **3**

ANNOTATED BIBLIOGRAPHY **5**

Gorin, A. L., Riccardi, G., and Wright, J. H. (1997). How may I help you? *Speech Communication*, *23*, 113-127. 5
Jelinek, F. (1997). *Statistical methods for speech recognition*. Cambridge, MA: MIT Press. 5
Chu-Carroll, J., and Carpenter, B. (1999). Vector-based natural language call routing. *Computational Linguistics*, *25*, 361-388. 5
McInnes, F. R., Nairn, I. A., Attwater, D. J., and Jack, M. A. (1999). Effects of prompt style on user responses to an automated banking service using word-spotting. *BT Technology Journal*, *17*, 160-171. 6
Boyce, S. J. (1999). Spoken natural language dialogue systems: User interface issues for the future. In Daryle Gardner-Bonneau (Ed.), *Human Factors and Voice Interactive Systems* (pp. 37-61). Norwell, MA: Kluwer. 6
Enterprise Integration Group. (2000). *Speech Recognition 1999 R&D Program: User interface design recommendations final report*. San Ramon, CA: Author. 7
Balentine, B., and Morgan, D. P. (2001). Natural language interfaces. In *How to Build a Speech Recognition Application* (pp. 197-244). San Ramon, CA: Enterprise Integration Group. 7
Suhm, B., Bers, J., McCarthy, D., Freeman, B., Getty, D., Godfrey, K., and Peterson, P. (2002). A comparative study of speech in the call center: Natural language call routing vs. touch-tone menus. In *Proceedings of CHI 2002* (pp. 283-290). Minneapolis, MN: Association for Computing Machinery. 8
Sheeder, T., and Balogh, J. (2003). Say it like you mean it: Priming for structure in caller responses to a spoken dialog system. *International Journal of Speech Technology*, *6*, 103-111. 8
Williams, J. D., and Witt, S. M. (2004). A comparison of dialog strategies for call routing. *International Journal of Speech Technology*, *7*, 9-24. 9
Blanchard, H. E., and Stewart, O. T. (2004). Conversational re-prompting in natural language dialog. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 708-711). Santa Monica, CA: Human Factors and Ergonomics Society. 10
Knott, B. A., Bushey, R. R., and Martin, J. M. (2004). Natural language prompts for an automated call router: Examples increase the clarity of user responses. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 736-739). Santa Monica, CA: Human Factors and Ergonomics Society. 10
Cohen, M. H., Giangola, J. P., and Balogh, J. (2004). Voice user interface design. Boston, MA: Addison-Wesley. 11
Balentine, B. (2007). It's better to be a good machine than a bad person. Annapolis, MD: ICMI Press. 12

RECOMMENDATIONS **13**

Recommendation 1: Consider using natural language call routing as the front end for IVRs that would otherwise have excessively deep menu structures. 13
Recommendation 2: Support open-ended prompts with example prompts. 13
Recommendation 3: For applications that will have many repeat callers, place example prompts after the open-ended prompt. For applications that will have few repeat callers, place them before the open-ended prompt. 13
Recommendation 4: It's OK for applications to refer to themselves as "I", but they should not be excessively self-referential. 13
Recommendation 5: Example prompts should be phrases that the application can recognize with high accuracy and should be requests that callers frequently make. 13
Recommendation 6: Consider using an open-ended prompt that asks for a brief input. 13

Executive Summary

An important source of information for the development of speech user interface design recommendations is the published literature. Data from an annotated bibliography (included in this report) of information from key books and scientific papers supports the following recommendations:

1. **Consider using natural language call routing as the front end for IVRs that would otherwise have excessively deep menu structures.** Although natural language (NL) call routing might not be the only viable design solution for an IVR, there is published data that shows that NL call routers can be very effective. Even when they don't work for all callers, they can substantially speed up task completion for many callers.
2. **Support open-ended prompts with example prompts.** Example prompts help to teach callers how to provide valid responses to the prompt.
3. **For applications that will have many repeat callers, place example prompts after the open-ended prompt. For applications that will have few repeat callers, place them before the open-ended prompt.** When callers will learn over time how to provide valid responses to the open-ended prompt, interactions will be faster when the examples follow the open-ended prompt (typically played only if there is a nomatch or noinput after 2-3 seconds). Providing examples before the open-ended prompt increases the likelihood of a caller providing a valid response, but slows down the interaction.
4. **It's OK for applications to refer to themselves as "I", but they should not be excessively self-referential.** It's important to keep the application focused on the caller rather than the system.
5. **Example prompts should be phrases that the application can recognize with high accuracy and should be requests that callers frequently make.** There's no point in offering examples that have low relevance to callers and are hard for the application to recognize.
6. **Consider using an open-ended prompt that asks for a brief input.** The evidence supporting this recommendation is relatively weak and does not provide specific prompt wording. Some researchers have reported that prompts that request a brief input elicit more precise responses than a less constraining prompt such as "How may I help you?" Some examples of prompts consistent with this recommendation are "Please briefly tell me why you're calling." and "Please tell me, briefly, the reason for your call."

Introduction

This annotated bibliography lists a set of important scientific papers published since 1997 on the topic of NLU call routing. Each entry in the bibliography includes the full citation, the authors' abstract, and my comments on the work.

1997 was an important year in the history of automated speech recognition. Before that year, most computers were not powerful enough to handle unconstrained continuous speech recognition. All commercial dictation systems required users to use discrete speech – saying each word one word at a time with a small pause between each word. After 1997, discrete speech was a thing of the past. Every commercial speech recognition dictation product allowed users to speak naturally and continuously. There were two landmark publications in 1997 – a book on statistical methods for speech recognition that described practical methods for continuous speech recognition (Jelinek, 1997) and a paper on using statistical techniques to automatically route telephone calls (Gorin, Riccardi, and Wright, 1997).

In 1999, Chu-Carroll and Carpenter reported considerable success in building an NLU call router that routed callers to one of 23 destinations. Suhm et al. (2002) studied an NLU call router designed to route calls to one of five destinations (order specialist, repair specialist, billing agent, payment specialist, and default agent) and found that callers overwhelmingly preferred NLU call routing to the DTMF (touch-tone) directed dialog alternative.

The first studies comparing different prompting strategies and other aspects of leading practice for call routing came out in 1999, and this continues to be an area of active research (Boyce, 1999; McInnes et al., 1999; Enterprise Integration Group, 2000; Balentine and Morgan, 2001; Sheeder and Balogh, 2003; Blanchard and Stewart, 2004; Cohen, Giangola, and Balogh, 2004; Knott, Bushey, and Martin, 2004; Williams and Witt, 2004).

Annotated Bibliography

Gorin, A. L., Riccardi, G., and Wright, J. H. (1997). How may I help you? *Speech Communication*, *23*, 113-127.

Abstract
We are interested in providing automated services via natural spoken dialog systems. By natural, we mean that the machine understands and acts upon what people actually say, in contrast to what one would like them to say. There are many issues that arise when such systems are targeted for large populations of non-expert users. In this paper, we focus on the task of automatically routing telephone calls based on a user's fluently spoken response to the open-ended prompt of How may I help you?. We first describe a database generated from 10,000 spoken transactions between customers and human agents. We then describe methods for automatically acquiring language models for both recognition and understanding from such data. Experimental results evaluating call-classification from speech are reported for that database. These methods have been embedded within a spoken dialog system, with subsequent processing for information retrieval and form-filling.

Comments
This was one of the first publications to describe NLU call routing based on an SLM.

Jelinek, F. (1997). *Statistical methods for speech recognition*. Cambridge, MA: MIT Press.

Abstract
This book reflects decades of important research on the mathematical foundations of speech recognition. If focuses on underlying statistical techniques such as hidden Markov models, decision trees, the expectation-maximization algorithm, information theoretic goodness criteria, maximum entropy probability estimation, parameter and data clustering, and smoothing of probability distributions. The author's goal is to present these principles clearly in the simplest setting, to show the advantages of self-organization from real data, and to enable the reader to apply the techniques.

Comments
Since its publication, this has been an influential book with regard to the breadth and depth of its coverage of the mathematical and statistical techniques used in modern automated speech recognition. The author, Frederick Jelinek, led the IBM Continuous Speech Recognition Group for over 20 years.

Chu-Carroll, J., and Carpenter, B. (1999). Vector-based natural language call routing. *Computational Linguistics*, *25*, 361-388.

Abstract
This paper describes a domain-independent, automatically trained natural language call router for directing incoming calls in a call center. Our call router directs customer calls based on their response to an open-ended How may I direct your call? prompt. Routing behavior is trained from a corpus of transcribed and hand-routed calls and then carried out using vector-based information retrieval techniques. Terms consist of n-gram sequences of morphologically reduced content words, while documents representing routing destinations consist of weighted term frequencies derived from calls to that destination in the training corpus. Based on the statistical discriminating power of the n-gram terms extracted from the caller's request, the caller is 1) routed to the appropriate destination, 2) transferred to a human operator, or 3) asked a disambiguation question. In the last case, the system dynamically generates queries tailored to the caller's request and the destinations with which it is consistent, based on our extension of the vector model. Evaluation of the call router performance over a financial services call center using both accurate transcriptions of calls and fairly noisy speech recognizer output demonstrated robustness in the face of speech recognition errors. More specifically, using accurate transcriptions of speech input, our system correctly routed 93.8% of the calls after redirecting 10.2% of all calls

to a human operator. Using speech recognizer output with a 23% error rate reduced the number of correctly routed calls by 4%.

Comments
This is another highly-cited paper from the late 1990s describing NLU call routing. The authors reported considerable success using their techniques to route callers to one of 23 destinations.

McInnes, F. R., Nairn, I. A., Attwater, D. J., and Jack, M. A. (1999). Effects of prompt style on user responses to an automated banking service using word-spotting. *BT Technology Journal*, *17*, 160-171.

Abstract
Two experiments were performed to measure the effects of differing styles of prompt wording in a simulated telephone banking service incorporating a speech recogniser with word-spotting capabilities. It was found that users gave longer input utterances in response to an 'open' style of prompt ('How can I help you?') than in response to a 'closed' prompt which was more specific as to what input was expected, and that when a help facility was offered most users said 'help' straight away. However, no significant difference was found in attitudes to the different versions of the service. Also no attitude difference was found with varying recognition accuracy. This may have been partly due to inadequate vocabulary coverage obscuring the effect of accuracy within the vocabulary, but the results suggest also that small numbers of recognition errors may have little impact on user attitude provided that the intended result of the call is achieved — this is a topic where further research could be valuable.

Comments
This was one of the first reports of an experiment comparing user satisfaction with open versus directed initial prompts (but note that their definition of "directed" is different from current practice). The three prompts were:

- (open) "Main Menu – How can I help you?"
- (mid) "Main Menu – Which service do you require?"
- (closed) "Main Menu – Please say 'help' or the name of the service you require."

The open and mid prompts were more successful at initially eliciting a service name than the closed prompt (to which many callers responded with "help"). The applicability of these findings to current practice is limited, however, because the prompts did not use the current popular and effective strategy of example prompting.

Boyce, S. J. (1999). Spoken natural language dialogue systems: User interface issues for the future. In Daryle Gardner-Bonneau (Ed.), *Human Factors and Voice Interactive Systems* (pp. 37-61). Norwell, MA: Kluwer.

Abstract
Technology advances in automatic speech recognition (ASR) and natural language understanding (NLU) in recent years have brought us closer to achieving the goal of communicating with machines via unconstrained, natural speech. Until recently, most applications of speech recognition technology required that the user know a restricted set of command words. In contrast, the research system that is described in this chapter can understand and act upon fluently spoken language. This markedly changes the nature of the dialogue between the human and the computer. The objective of this research is to evaluate user interface design alternatives for these new natural spoken dialogues between humans and machines. In this chapter the focus is on a particular experimental vehicle, that of automatically routing telephone calls based on a user's fluently spoken answer to the question "How may I help you?". This chapter summarizes results from several studies conducted to determine how best to design the user interface for a spoken natural dialogue system.

Comments

The experiments described in this chapter support a number of now common practices among speech user interface designers. Among these is the finding that it's OK to script for the automated voice using personal pronouns such as "I" – at least in the initial prompt.

Enterprise Integration Group. (2000). *Speech Recognition 1999 R&D Program: User interface design recommendations final report.* San Ramon, CA: Author.

Abstract
During 1999, Enterprise Integration Group (EIG) conducted a subscription-supported project on speech recognition user interfaces for telephone-based customer self-service. The work was carried out for ten charter subscribers. This document is the final report.

Comments
This report describes the results of a comprehensive set of usability studies designed to investigate leading practices for speech user interface design. Much of the work reported here first has appeared publicly in the second edition of the book *How to Build a Speech Recognition Application* (see below).

Some of the major findings that have influenced leading practices in the design of applications that use open-ended prompts were:

- Prompt by example – provide an example of a well-formed natural language input.
- Do not present instructions before the user clearly needs them.
- Do not fail to present instructions when needed.
- "An application can successfully employ both natural language and directed dialogue design types. Users can move between the two approaches, in a single application, without either confusion or failure to accomplish tasks. Since each approach has its unique value, designers can safely build an application using both. In particular, when the directed dialogue is kept simple, users are seldom aware of it so the application appears to users to be seamless." (p. 22)

Balentine, B., and Morgan, D. P. (2001). Natural language interfaces. In *How to Build a Speech Recognition Application* (pp. 197-244). San Ramon, CA: Enterprise Integration Group.

Abstract
This chapter is devoted to the unique attributes of natural language (NL) technology and the design of user interfaces that exploit them. An NL application "resembles a command-line interface in that it hides the application's functionality." The implication is that NL technology allows systems to accept more complex and varied user input than can be conveyed through simple prompting or casual explanation. The sophistication of these interfaces lead to the conclusion that prompting schemes and dialogue structure may – indeed often must – be designed according to an entirely different set of assumptions than those for directed dialogues.

Comments
This is the chapter on NL speech user interfaces from the second edition of one of the most influential books published on speech user interface design. Their use of the term "natural language" (NL) is different from most other usage. For Balentine and Morgan, NL refers to any user interface that uses a nondirective prompt, regardless of the technology used for recognition (hand-crafted finite state grammars or statistical language models). Most of the chapter is concerned with NL user interfaces for form-filling applications – applications in which the most desirable user behavior is to provide an utterance that completes several fields of a form such as a car reservation form. They provide some discussion of NL call routing using statistical language models (SLMs) toward the end of the chapter (Section 8.5.4.1. Use SLM for Routing and Other "One-Shot" Interactions). With regard to mixing open-ended (SLM) and closed-ended (directed) user interactions, they suggest that moving from an open to a more closed interaction is consistent with the user experience of talking to a human agent, but it is less common to move the other direction – from closed- to open-ended prompting.

They suggest that the type of prompting that they have tested for general NL user interfaces (general, open-ended prompt supported by example prompting for nomatch and noinput events) is also appropriate for NLU call routing.

Suhm, B., Bers, J., McCarthy, D., Freeman, B., Getty, D., Godfrey, K., and Peterson, P. (2002). A comparative study of speech in the call center: Natural language call routing vs. touch-tone menus. In *Proceedings of CHI 2002* (pp. 283-290). Minneapolis, MN: Association for Computing Machinery.

Abstract
This paper presents a field study that compares natural language call routing with standard touch-tone menus. Call routing is the task of getting callers to the right place in the call center, which could be the appropriate live agent or automated service. Natural language call routing lets callers describe the reason for their call in their own words, instead of presenting them with a list of menu options to select from using the telephone touch-tone keypad. The field study was conducted in a call center of a large telecommunication service provider. Results show that with natural language call routing, more callers respond to the main routing prompt, more callers are routed to a specific destination (instead of defaulting to a general operator who may have to transfer them), and more callers are routed to the correct agent. Our survey data show that callers overwhelmingly prefer natural language call routing over standard touch-tone menus. Furthermore, natural language call routing can also deliver significant cost savings to call centers.

Comments
The annual CHI (Computer-Human Interaction) conference, sponsored by the Association for Computing Machinery (ACM), is one of the premier conferences covering advances in human-computer interaction. This paper described a field study in which an NLU call router and a directed menu (touchtone) call router were directly compared using a deployed system and calls from real customers. The researchers concluded that NLU call routing was very successful in the application they investigated. This is one of the main studies supporting NLU call routing as a leading practice, using the approach of providing an open-ended initial prompt followed by example prompts before falling back to some other approach (for this study, a default agent – the more common practice is to fall back to a directed menu). Note that this approach has also been supported in studies conducted in 2000-2001 by the Enterprise Integration Group along with other industry participants (including IBM – Bruce Balentine, personal communication).

The initial prompt used in this study was "Please tell me, briefly, the reason for your call today" because unpublished, informal research at BBN had indicated that this prompt elicited more precise caller responses than "How may I help you?" It is important to note that there are still unresolved questions about the best way to construct open-ended prompts for NLU call routers.

The system studied by these researchers had a fairly small number of terminals – just five (order specialist, repair specialist, billing agent, payment specialist, default agent).

Overall, the results indicated that the NLU call router was very successful, both in absolute terms and relative to the directed (touchtone) menu. Calls routed through the NLU call router were handled much faster overall, and callers clearly preferred the NLU call router.

Sheeder, T., and Balogh, J. (2003). Say it like you mean it: Priming for structure in caller responses to a spoken dialog system. *International Journal of Speech Technology*, *6*, 103-111.

Abstract
In this paper we report results of a study undertaken to evaluate the initial prompts of open prompt style call-routing applications. Specifically, we examined how placement and phrasing of examples in the initial query affected caller responses and routing success. We looked at the comparative effectiveness of placing examples

before and after the initial query and of phrasing these examples such that they promoted either a succinct structure in the form of a keyword or phrase, or a more complex but natural structure in the form of a question or statement. Findings indicate that examples encouraging a more natural structure, when presented prior to the initial query, result in significantly improved routing performance. We discuss this result in the context of using initial prompts to prime for desired structure in caller responses.

Comments

This paper describes an experiment that manipulated two variables – the positioning of examples (*preceding* the open-ended prompt or *following* the prompt after a 2.5 sec noinput timeout) and the type of example (*keyword* list or more *natural* expression). There were 18 participants for each of the four possible conditions (*preceding keyword*, *preceding natural*, *following keyword*, *following natural*) for a total of 72 participants. Participants completed three telephone account management tasks (activate service, bill reprint, and change plan) with their assigned condition. The researchers concluded that the *preceding natural* condition had the best outcomes in this experiment. They also point out that it is possible that a following strategy might be the better approach for applications in which most callers are expected to be frequent, experienced users of the application. For applications in which most callers will use the application infrequently, there appears to be an advantage in providing examples before the open-ended prompt.

Williams, J. D., and Witt, S. M. (2004). A comparison of dialog strategies for call routing. *International Journal of Speech Technology*, *7*, 9-24.

Abstract

Advances in commercially-available ASR technology have enabled the deployment of "How-may-I-help-you?" interactions to automate call routing. While often preferred to menu-based or directed dialog strategies, there is little quantitative research into the relationships among prompt style, task completion, user preference/satisfaction, and domain. This work applies several dialog strategies to two domains, drawing on both real callers and usability subjects. We find that longer greetings produce high levels of first utterance routability. Further, we show that a menu-based dialog strategy produces a uniformly higher level of routability at the first utterance in two domains, whereas an open-dialog approach varies significantly with domain. In a domain where users lack an expectation of task structure, users are most successful with a directed strategy, for which preference scores are highest, even though it does not result in the shortest dialogs. Callers rarely provide more than one piece of information in their responses to all types of dialog strategies. Finally, a structured dialog repair prompt is most helpful to callers who were greeted with an open prompt, and least helpful to callers who were greeted with a structured prompt.

Comments

The authors were apparently not aware of the paper by Suhm et al. (2002 – see above) when they stated in their introduction, "To our knowledge, there are no studies which have made direct comparisons between a menu strategy and an open strategy." Suhm is not in their list of references, but to be completely fair, the Suhm study was open versus touchtone, not open versus directed dialog.

Based on data from six initial prompting styles (directed vs. open prompts, preceded by hello, hello plus an earcon, and hello plus an earcon plus a named persona plus a recently changed notice), the researchers concluded that the style that produced the most routable utterances was a directed prompt preceded by hello plus an earcon plus a named persona plus a recently changed notice. It is important, however, to take this as a provisional recommendation since the difference in percentage of routable calls between that condition and a simpler condition (hello plus an earcon) was not statistically significant ($X^2(1)=2.4$, $p=.12$), and the generally accepted leading practice is to avoid playing unnecessary information in the introduction (e.g., avoid playing recently changed messages or naming the application's persona – for example, see Knott and Bushey, 2004 below). Because the open prompt did not include examples in this experiment, it is unsurprising that the directed dialog produced a higher percentage of routable utterances.

At least for the application designs investigated in this first experiment, the directed designs were more successful than the open designs, possibly because callers did not have a strong mental model of the task. When the application design was closer to the designs investigated by Sheeder and Balogh (2003 – see above), the routability percentages were higher, but it didn't seem to matter if the examples appeared before or after their open question ("What would you like to do?"). As Williams and Witt acknowledge, this could be due differences in the applications studied or the types of participants (usability subjects in Sheeder and Balogh; real callers in Williams and Witt).

Blanchard, H. E., and Stewart, O. T. (2004). Conversational re-prompting in natural language dialog. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 708-711). Santa Monica, CA: Human Factors and Ergonomics Society.

Abstract
Natural language understanding technology has reached a level of sophistication where it can be profitably employed in interactive voice response systems in telephony. This paper describes a call routing application, where callers state a request in unconstrained, natural speech. The system then routes the call to the correct destination system or attendant. If there is a problem in understanding, then the caller must be re-prompted. This paper looks at two cases of re-prompting in the second turn of dialog based on the caller's response in the first turn: (a) when a caller's initial request is to speak to a real person instead of stating the reason for their call, and (b) when callers are too vague in their initial response. A strategy of conversational re-prompting is introduced with fits into the greater naturalness of the dialog, and, we show, increases the performance of the system in terms of successful fulfillment of user requests.

Comments
The focus of this research is on how to reprompt an open-ended prompt under two situations (a caller asks for a human; a caller otherwise fails to respond to the initial open-ended prompt). They created reprompts for these two situations and found that the following reprompts were effective for the application they were studying:

> After receiving a request for an agent: "Okay. In order to direct your call please tell me if you need to refill an existing medication, find out the status of your prescription order, or anything else you want to speak to the customer service representative about."
>
> After receiving a vague request about an order: "I'm sorry, do you need to refill an existing medication or get status on an order you've already sent in? Please tell me how I may help you.
>
> After receiving any other vague request: "Okay. What's your question?"

Note that while these specific prompts were found to be effective in getting callers to provide more specific responses, the researchers did not explicitly test their style of priming the response (a fairly natural-sounding but short list of actions) against the more commonly practiced style of example prompting ("For example, you might say 'refill an existing medication', 'find out the status of a prescription order'," etc.).

Knott, B. A., Bushey, R. R., and Martin, J. M. (2004). Natural language prompts for an automated call router: Examples increase the clarity of user responses. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 736-739). Santa Monica, CA: Human Factors and Ergonomics Society.

Abstract
Speech-enabled interfaces are proving to be an effective option for service-center call routing applications. However, the effectiveness of a call routing application is dependent upon the speech recognizer correctly interpreting the caller's utterance. One approach to increase the clarity and routability of callers' initial requests is to provide examples within an open-ended prompt to provide context and guide caller's speaking behavior.

Two experiments were conducted to investigate the effect of including examples in the opening prompt for a natural language call routing application. In both experiments, callers were asked to say the purpose of their call with an open-ended prompt. Half of the callers were presented a prompt that contained examples, and half of the callers were presented with a prompt that did not contain examples. The results showed an advantage for the use of examples in terms of the percentage of routable caller utterances. The design implications for natural language prompts are discussed.

Comments

These experiments were similar to that of Sheeder and Balogh (2003 – see above), but used a deployed system to capture initial caller responses to two open-ended prompts – one that started with examples and one that never presented any examples. In Experiment 1 the participants were 4,058 callers from California and Nevada; in Experiment 2 there were 4,479 callers from Illinois, Indiana, Michigan, Ohio, and Wisconsin. For the prompts used in Experiment 1, the application did not refer to itself by a name, but in Experiment 2 it did ("Hi, I'm Bill"). The researchers concluded that providing examples up front improved the routability of the caller's initial utterance – a finding consistent with that of Sheeder and Balogh. They also found, however, that providing examples up front increased the likelihood that the caller would request an agent rather than making an attempt to provide a routable input (from 4.5 to 9.4% in Experiment 1; from 5.4 to 7.9% in Experiment 2). They also found that when the application did not refer to itself by name, there was no significant difference in the wordiness of caller responses as a function of presenting examples up front, but when the system did refer to itself by name, caller responses were more verbose in the no examples condition. As with the results of Sheeder and Balogh, the findings regarding the placement of examples are very likely limited to applications in which callers are infrequent users of the application rather than frequent users. In addition to the primary findings of the experiments, Knott and Bushey also point out that because callers had a strong tendency to mimic the examples presented in the prompts, it is important to select examples that the system can recognize with high accuracy and are requests that callers frequently make.

Cohen, M. H., Giangola, J. P., and Balogh, J. (2004). Voice user interface design. Boston, MA: Addison-Wesley.

Abstract

The biggest challenge now is the design of the user interface. There are far too few practitioners who have the knowledge and skills to create all the systems needed and to advance our understanding as new technology enables new capabilities. Current practitioners come from a wide variety of backgrounds: speech technology, user interface design, cognitive psychology, linguistics, and software development. All of these fields have contributed to our current understanding of voice user interface design. In fact, the field has benefited substantially from the diversity of influences. However, the need to pull together information from diverse fields has made it difficult to codify and teach the rationale for design.

In this book we aim to offer in one place much of the background information needed for practitioners to design specific applications and to contribute to the advance of the field. We try to take a principled approach to deriving best practices, with the hope that designers will then have a basis for approaching new design situations and new technologies.

Comments

Along with the book by Balentine and Morgan (2001 – see above), this is required reading for anyone interested in the design of voice user interfaces.

Balentine, B. (2007). It's better to be a good machine than a bad person. Annapolis, MD: ICMI Press.

Abstract
Design philosophy is not an easy subject. Inherently multidisciplinary, it quickly gets too technical for the humanists, too touch-feely for the pragmatists, too theoretical for the engineers, too subjective for the behaviorists, and too abstract for the businessmen. So this play – like the subject itself – is also multidisciplinary.

Comments
This book is a collection of essays by Bruce Balentine, a leading voice user interface designer. It touches only briefly on natural language call routing, but Essay 118 (Express Lanes and Problem Lines) describes a rationale for presenting help in user-system dialogs only when there is evidence that help is necessary (in other words, start everyone in the express lane, moving only exceptional cases to the problem line). An implicit assumption in the argument is that users will be repeat callers ("The callers have no special incentive or knowledge of the system – at least not at first. But we can condition their behaviors quickly by establishing and enforcing the behavioral imperatives that keep them in the express lane and out of the problem line.") This supports the playing of example prompts shortly after open-ended prompts rather than providing examples before open-ended prompts, at least for systems that will have a significant percentage of repeat callers.

Recommendations

Recommendation 1: Consider using natural language call routing as the front end for IVRs that would otherwise have excessively deep menu structures.

Although natural language (NL) call routing might not be the only viable design solution for an IVR, there is published data that shows that NL call routers can be very effective (Chu-Carroll & Carpenter, 1999; Gorin, Riccardi, & Wright, 1997) and under certain circumstances callers strongly prefer NL call routing to directed dialogs (Suhm et al., 2002). Even when they don't work for all callers, NL call routers can substantially speed up task completion for many callers (Suhm et al., 2002; EIG, 2000).

Recommendation 2: Support open-ended prompts with example prompts.

Example prompts help to teach callers how to provide valid responses to the prompt (Balentine & Morgan, 2001; EIG, 2000; Knott, Bushey, & Martin, 2004; Sheeder & Balogh, 2003; Suhm et al, 2002; Williams & Witt, 2004).

Recommendation 3: For applications that will have many repeat callers, place example prompts after the open-ended prompt. For applications that will have few repeat callers, place them before the open-ended prompt.

When callers will learn over time how to provide valid responses to the open-ended prompt, interactions will be faster when the examples follow the open-ended prompt (typically played only if there is a nomatch or noinput after 2-3 seconds); providing examples before the open-ended prompt increases the likelihood of a caller providing a valid response, but slows down the interaction (Balentine, 2007; Sheeder & Balogh, 2003).

Recommendation 4: It's OK for applications to refer to themselves as "I", but they should not be excessively self-referential.

It's important to keep the application focused on the caller rather than the system (Boyce, 1999; Balentine, 2007).

Recommendation 5: Example prompts should be phrases that the application can recognize with high accuracy and should be requests that callers frequently make.

There's no point in offering examples that have low relevance to callers and are hard for the application to recognize (Knott, Bushey, & Martin, 2004).

Recommendation 6: Consider using an open-ended prompt that asks for a brief input.

The evidence supporting this recommendation is relatively weak and does not provide specific prompt wording. Suhm et al. (2002) reported that prompts that request a brief input elicit more precise responses than a less constraining prompt such as "How may I help you?" Some examples of prompts consistent with this recommendation are "Please briefly tell me why you're calling." and "Please tell me, briefly, the reason for your call."