ACM-CHI, 2001 (Seattle, WA)

Just (All) the Facts, Ma'am

Dawn Dutton
AT&T Labs-Research
D103, Bldg. 103, 180 Park Avenue
Florham Park, NJ 07932 USA
+1 973 236 6522
dldutton@att.com

Selina Chu
Information and Computer Science
444 Computer Science Bldg, UC Irvine
Irvine, CA 92697 USA
+1 949 509 9762
selina@ics.uci.edu

James Hubbell
Human Factors International
Suite 11, Bldg. 1, 1 Bethany Road
Hazlet, NJ 07730 USA
+1 732 335 8888
jhubbell@humanfactors.com

Marilyn Walker
AT&T Labs-Research
E103, Bldg. 103, 180 Park Avenue
Florham Park, NJ 07932 USA
+1 973 360 8959
walker@research.att.com

Shrikanth Narayanan
Speech and Image Proc. Institute
EEB 430, Dept. of EE-Sys., USC
Los Angeles, CA 90089 USA
+1 213 740 6432
shri@sipi.usc.edu

ABSTRACT

AT&T Communicator is a speech-enabled telephony-based application that allows the end-user to select and reserve airline itineraries. We report the results of an experiment exploring how the amount and structure of information presented in complex lists influences the user experience and the ability of subjects to complete an itinerary selection task. Presenting all the relevant information needed for a decision at once was the factor that most positively influenced successful task completion and the user experience.

Keywords

dialog, user interface, spoken language, lists, selection

INTRODUCTION

AT&T Communicator is a state of the art spoken dialogue system that allows the end-user to select and reserve various travel related resources, in particular, airfare, hotel, and rental cars [2]. One of its most challenging dialog modules is that for presenting information that allows the user to select a desired air travel itinerary. While selecting between multiple possible itineraries in a visual domain is a relatively simple task because most of the selection criteria are listed in a single scrollable page, the same task is likely to have a much higher cognitive load in an audio-only domain [1, 3]. The selection criteria for each candidate flight must be presented to the user serially, leading to higher cognitive demands that could result in errors in itinerary selection. This paper describes an experiment in which we vary the amount of information about an itinerary and the number of itinerary options that AT&T Communicator presents. We predict that the amount of information presented affects the ability of users to successfully select the optimal itinerary within a set.

METHODS AND PROCEDURES

Subjects, Apparatus and Procedure

A Wizard of Oz (WOZ) methodology was used. Relevant aspects of the AT&T Communicator user interface were prototyped using the Unisys Natural Language Speech Assistant software which runs on a PC using the Windows NT operating system. Sixty-four subjects recruited at a local shopping mall over a five day period called into the Communicator prototype using an analog telephone and interacted with the system by voice.

Experimental Design

All itineraries presented to the subjects were round-trip.

Independent Variables

This was a factorial experiment with two factors, one factor between subjects and the other within subject.

Selection Itinerary Content. There were be two levels of this between subjects factor. *Terse:* presented itineraries included: airline, number of stops, and departure time. *Verbose:* presented itineraries included: airline, flight number, number of stops, departure time, and arrival time.

Number of Legs Before Question. Each level is actually a combination of two separate, but related, factors. *Combined vs. Separate:* whether outbound and inbound legs are presented separately or in combination. *Number of legs:* the number of legs that are presented before asking the subject (S̲) to make a decision. Four levels of this factor were chosen. In all cases (1) the total number of flights 'found' was 5, and, (2) the question was, "Would you like to hold [that flight/any of those flights]?". *Separate 1:* the outbound and return legs of the trip are presented separately and after each flight the S̲ is asked the question. *Separate 3:* the outbound and return legs of the trip are presented separately and after the third flight the S̲ is asked the question. *Separate 5:* the outbound and return legs of the trip are presented separately and after the last flight the S̲ is asked the question. *Combined:* the outbound and return legs of the trip were presented at the same time and after each set of flights the S̲ is asked the question.

Tasks

Each subject was asked to complete four tasks in the course of this experiment. In each task the subject was given a set of criteria that the subject had to meet in selecting both an outbound and a return leg. The tasks used in this experiment exercise representative selection criteria that are typically used by individuals purchasing airline tickets.

Dependent Measures Collected

At the end of each dialogue, the wizard labeled the dialogue for task completion. Each subject also filled out a survey after each call using a Likert scale to probe user attitudes toward the amount of information presented, user satisfaction, ease of use and the speed of the interaction.

RESULTS AND CONCLUSIONS

Terse or Verbose?

A two-way Analysis of Variance (ANOVA) was run for each of 5 dependent measures: successful task completion, amount of information presented about each flight, satisfaction, ease of use, and speed of interaction. For each dependent measure, no significant interactions were found. (Throughout the experiment, the alpha level used to determine significance of an effect was $p<.05$.) A significant main effect for Terse/Verbose was found for the subjective measure of the amount of information presented about each flight ($p=.001$), No other significant main effects were found for any of the dependent measures. The optimum value for the dependent measure amount of information is '2' (Just the right amount of information about each flight). The average value for the Verbose condition (across the 4 levels of # of Legs) was 2.06, while the equivalent average for the Terse condition was 2.24. That is, the subjects felt that the amount of information presented in the Verbose condition was more appropriate than the amount of information presented in the Terse condition.

Number of Legs?

The above analyses indicate that the amount of information presented in the Verbose condition better met the expectations of subjects. The next question then was, within the verbose condition, which level of the number of legs before the question factor showed the best performance. A one-way ANOVA was run for the verbose condition for each of the five dependent measures: successful task completion, amount of information about each flight, satisfaction, ease of use, and speed of interaction. A significant main effect was found for successful task completion ($p=.005$). No significant effects were found for the other four dependent measures. The significant main effect was probed using the Tukey test. Only one pairwise comparison was significant ($p<.05$). Tasks attempted in the Separate 5 condition were significantly more likely to be completed successfully than tasks attempted in the Separate 3 condition. The successful task completion rates were: Separate 5=90%, Combined=83%, Separate 1=60%, and Separate 1=57%. Anecdotal evidence sheds some light on which condition (Separate 5 or Combined) is preferred by subjects. In the Verbose condition, the last 17 subjects run in the experiment were asked a few questions that provide evidence concerning their subjective impressions of the four levels of the number of legs before question factor. Twelve subjects noticed a difference between the 4 conditions. The Combined version was selected by 5 of 12 of the subjects as the version they felt to be the 'worst.'

DISCUSSION

Presenting all the relevant information about a given leg *at once* seemed to be the single overarching factor that most positively influenced successful task completion and the user experience. Subjects wanted to hear *all of the relevant information* about a flight needed to make the best choice.

Within the Separate conditions (Separate 1, Separate 3 and Separate 5), the task completion rate was highest for the Separate 5 condition. That is, when *all of the flights* were presented at once, without any intervening system questions. The Separate 5 and Combined conditions had similar task completion rates and were not significantly different. However, the Combined condition was the only condition considered 'worst' by subjects. Thus, the condition that maximized both successful task completion and user experience was the Verbose Separate 5 condition. A major concern in the design of this experiment was that the audio presentation of lists of complex information, in this case lists of multiple airline flights each containing multiple pieces of information, would result in cognitive overload. These findings argue that our concern about the increased cognitive load in an audio-only domain was unfounded. Many subjects apparently dealt with the increased cognitive load by taking notes, with flight times, etc., while completing the experimental tasks.

Eighty percent of the subjects stated that the most important criterion when personally selecting a flight was price. A number of subjects stated that they were willing to trade off other important criteria, e.g. airline, number of stops, in order to get a better price. Such a negotiation between user and machine might, in future experimentation, lend itself well to exploring machine-user dialog in a natural language telephony-based system.

REFERENCES

1. Blanchard, H.E. & Lewis, S.H. (1999), The Voice messaging user interface, *in* D. Gardner-Bonneau (ed.), *Human factors and voice interactive systems*, Kluwer Academic Publishers, pp. 257-284.
2. Levin, E., Narayanan, S., Pieraccini, R., Biatov, K., Bocchieri, E., Di Fabbrizio, G., Eckert, W., Lee, S., Pokrovsky, A., Rahim, M., Ruscitti, P., and Walker, M. (2000), *The AT&T-DARPA Communicator mixed-initiative spoken dialog system*, Proc. of the International Conference of Spoken Langurage Processing, (Beijing, China), pp. 122-125.
3. Schneiderman, B. (1992). *Designing the user interface* (2nd ed.). Reading MA: Addison Wesley.