

Authors

◆ **Dawn Dutton**

AT&T Labs – Research
dldutton@att.com

◆ **Selina Chu**

UC Irvine
selina@ics.uci.com

◆ **Marilyn Walker**

AT&T Labs – Research
walker@research.att.com

◆ **James Hubbell**

jameshubbell@hotmail.com

◆ **Shrikanth Narayan**

USC – Speech & Image Proc. Ir
shri@sipi.usc.com

AT&T Communicator

- ◆ Funded by DARPA
 - Multi-year, multi-site project whose goal is to advance spoken dialog research.
- ◆ State-of-the-art spoken dialog system that:
 - allows user to select and reserve travel related resources.
 - is telephony-based (auditory-only interface)
 - allows user to speak in relatively unconstrained, natural way.
 - For the most part, this is not a “Please say ...” interface.
 - Has a mixed-initiative dialog design
 - The scope of allowed user input can include information that has not directly asked for by the system. For example, when the system asks *where are you flying from?* the user can respond *“I’m flying from London to Seattle.”*
 - Most users would interact with the system infrequently.

Selecting from a Complex List

- ◆ In terms of usability, the itinerary selection module is one of the most challenging.
 - **Visual domain:** relatively easy because all info is available for user's perusal.
 - **Audio-only domain:** Serial presentation of information exerts a potentially strong negative influence on successful task completion and satisfaction.
- ◆ Why is this a complex list?
 - Communicator presents the user with a numbered list of options (potential itineraries). Each option itself contains multiple sub-items (e.g. departure time, airline, etc.).
 - Selecting the single correct item from such a list is a potentially daunting cognitive task.

Initiative and Flow of Control

- ◆ Mixed initiative dialog strategy.
 - First prompt is open-ended, “*How may I help your travel plans?*”
 - All subsequent prompts request specific info from the user, e.g. “*When would you depart?*”
 - Users can provide multiple informational elements in each utterance.
 - Subsequent steps can be skipped if subject provided multiple pieces of information.

Sample Dialog

(edited for length)

S: Hello, this is AT&T Communicator. How may I help with your travel plans?

U: Yeah, I'd like to fly from Newark to Seattle.

S: OK, a flight from Newark, New Jersey to Seattle, California. And, will this be a roundtrip?

U: Yes, and I need to arrive in Seattle by 4 PM on April 22nd.

S: OK, a roundtrip, arriving by 4 PM on Sunday April 22nd.

S: I found five outbound United Airlines flights.

Option 1 is flight number 63; has one stop; it departs at 10: AM and arrives at 2:42 PM.

Option 2 is flight number 75; has one stop; it departs at 12: PM, and arrives at 4:37 PM.

Option 3 is flight number 83; a non-stop; it departs at 2 PM and arrives at 5 PM.

Would you like to hold any of those flights?

U: I guess I'll take option 1.

S: Got it.

Goal of Experiment

- ◆ For each task, there is one itinerary that best matches a set of constraints given to the user.
- ◆ Explore how the amount of information presented affects:
 - the ability of users to successfully select the optimal itinerary within a set, and,
 - user satisfaction.

Methods and Procedures

- ◆ **Subjects** (N=64)
 - Run in local mall, ½ hour duration, paid \$20.
- ◆ **Wizard of Oz** (WOZ)
 - No real ASR/NLU.
 - Experimenter clicks buttons, acts as recognizer.
- ◆ **Apparatus and Materials**
 - AT&T Communicator UI prototyped on PC.
 - S called into prototype.
 - S completed 5 surveys, one after each task and a final survey.
 - S had access to pad and pen, although pad and pen were not men by experimenter.

Independent Variables

Selection Itinerary Content

- ◆ 2 levels (between subjects)
 - **Terse:** airline, number of stops, departure time
 - Potential benefit: Minimize cognitive load, user can engage in dialog with system to obtain additional information only when they need it, e.g. *“When does arrive?”*
 - **Verbose:** airline, number of stops, departure time + flight number, arrival time
 - Potential benefit: All the relevant information for selecting any flight in this experiment is available in every itinerary presented. The user does not need to ask the system for any additional information.

Independent Variables

Number of Options before Q

- ◆ 4 levels (within) -- Composed of 2 separate or related factors:
 - ◆ **Combined vs. Separate:** Outbound and return legs presented separately, or in combination.
 - ◆ **Separate:** System first presents the outbound options, e.g. From Newark to Seattle. Once the user selects one of the outbound options the system moves onto the return options, e.g. From Seattle back to San Francisco.
 - ◆ **Combined:** System presents one outbound leg and one return leg together. The system presents these *combinations* of outbound and return legs one-by-one.
 - ◆ **Number of options:** # of options presented before subject is asked a question.

Independent Variables

Selection Itinerary Content (Between)			
Terse		Verbose	
Outbound / Return	# of Options Before Question	Outbound / Return	# of Options Before Question
(Within)		(Within)	
Separate	1	Separate	1
Separate	3	Separate	3
Separate	5	Separate	5
Combined	1	Combined	1

Tasks

- ◆ Completed 4 tasks, order counterbalanced.
 - **Departure only.** The task criteria for both the outbound and return legs require the subject to choose flights based on departure time only.
 - **Arrival Only.** The task criteria for both the outbound and return legs require the subject to choose flights based on arrival time only.
 - **Departure & Arrival.** The task criteria require the subject to choose the outbound flight based on departure time and the return flight based on arrival time.
 - **Specific Flight.** The task requires the subject to book a particular flight for both the outbound and return legs.

Dependent Measures

- ◆ **Successful task completion.**
 - Binary, determined on the spot by Experimenter.
- ◆ **Satisfaction:** “Overall, how satisfied...” [5 pt.]
 - 1=Very Satisfied
 - 5=Very Dissatisfied
- ◆ **Ease of use:** “Overall, how easy to use...” [5 pt.]
 - 1=Very Easy
 - 5=Very Difficult
- ◆ **Speed of Interaction:** “Overall, how quickly.. [5 pt.]
 - 1=Much too fast
 - 3=Just the right speed
 - 5=Much too slow
- ◆ **Right amount of info:** “For EACH flight...” [3 pt.]
 - 1=Too much ...
 - 2=Just the right amount ...
 - 3=Too little ...

Results

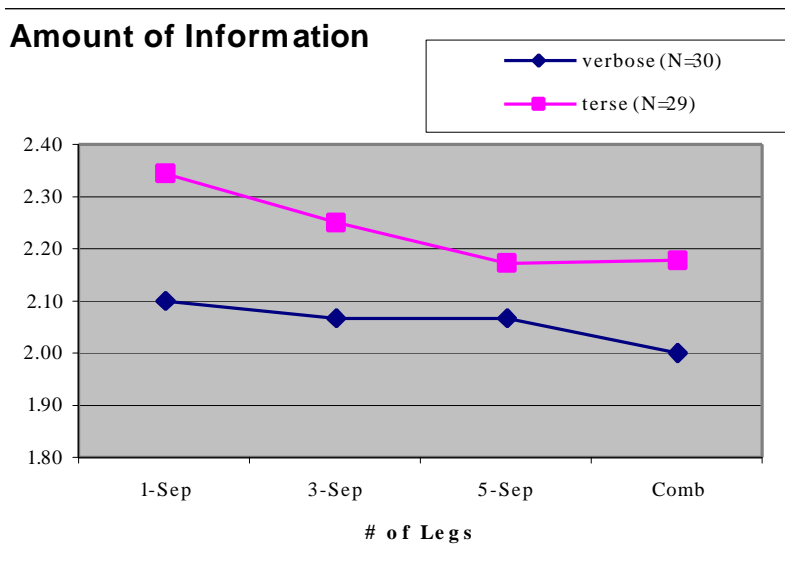
◆ **Statistical Overview**

- ANOVAs run on 4 dependent measures.
 - Successful task completion, satisfaction, ease of use, and speed of interaction.
- Tukey test used to probe main effects.
- $p < .05$ for all analyses.

◆ **It Should be Noted.**

- Most subjects were unfamiliar with spoken dialog systems.
- It is possible that for subjective measures, effects of experimental manipulations were overwhelmed by novelty interaction style.

Terse vs. Verbose?



Verbose and Terse subject ratings to the Amount of Information question.

3 pt. Scale:

1=Too little information about each flight

2=Just the Right amount of information about each flight

3= Too much information about each flight

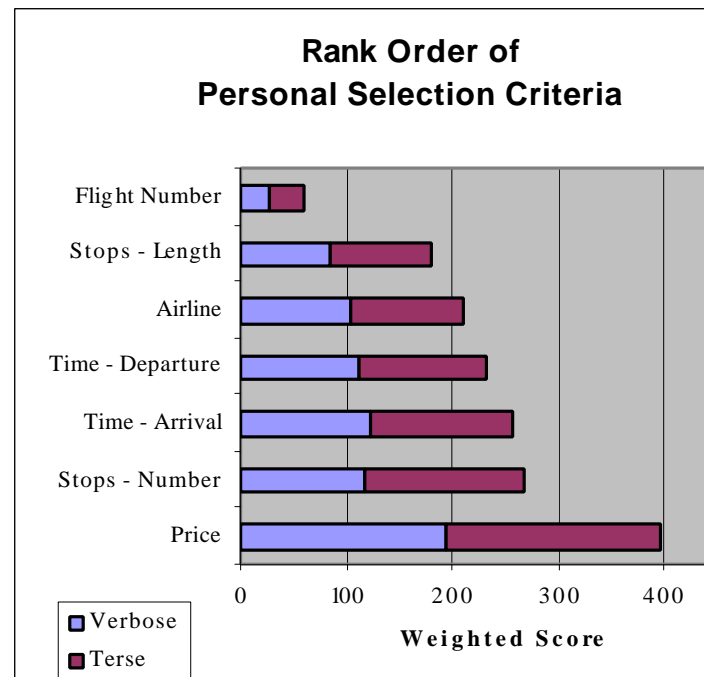
◆ **Verbose prompts w preferred**

- Only significant main effect for Selection Itinerary Content was Amount of Information
- Verbose selection itineraries included: **airline, flight #, number stops, departure time arrival time.**

Subject's Own Criteria

◆ Choosing between multiple flights

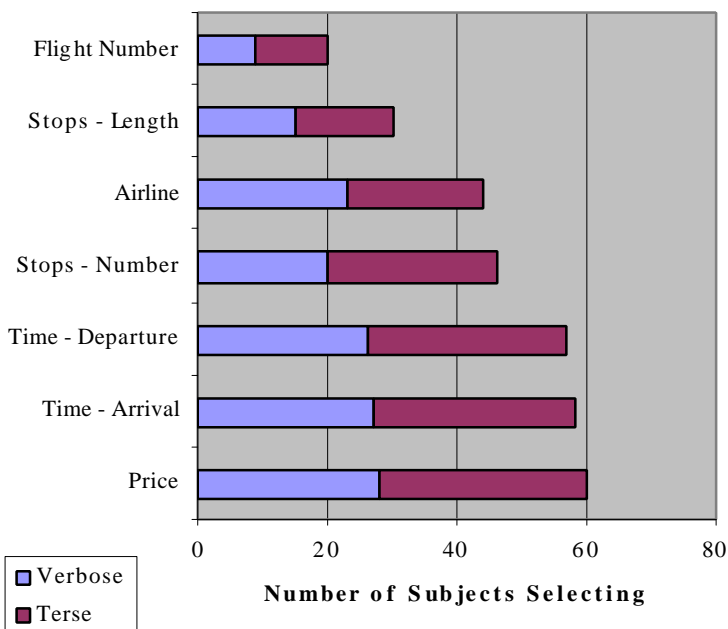
- Subjects rank ordered 7 criteria.
- Weighted score was produced.
 - Rank of 1 = 7 points.
 - Rank of 7 = 1 point.
 - etc.
 - Weighted scores summed for all subjects.



Summed *Weighted Scores* of subjects ordering of their personal selection

What Criteria Subjects Want

What Information Should AT&T Communicator Present?



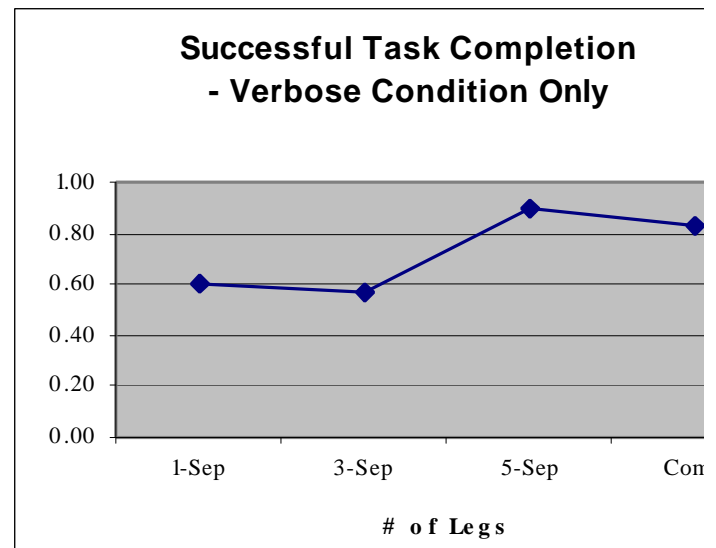
Number of subjects indicating that each selection criteria should, by default, be presented by AT&T Communicator.

- ◆ **Choosing between multiple flights**
 - Simple sum of criteria selected by all subjects.
- ◆ Should include:
 - **Price, arrival time, departure time, number of stops, and airline.**
- ◆ Do not need to include:
 - **Flight number, length of stops.**
 - But, length of a stop might be useful when it is inordinately long or short.

Number of Options before Q'

◆ *Within Verbose condition.*

- Not possible to unambiguously choose one condition over all others.
- Significant main effect for task completion.
- Significant pairwise: Sep3 vs. Sep5.
- Why?
 - A lot of variability in the data.
 - Tasks too easy?



Successful task completion (in percentage) across the four levels of the # of Legs Before Question condition (Verbose

Mid-stream Correction

- ◆ Experimenter informally probed customer perceptions/reactions.
 - Unanticipated observations:
 - Many subjects had very strong feelings, some positive, some negative concerning conversational natural language interface. It seemed possible that these subjects based subjective responses on interface style and experimental manipulations.
 - Many subjects in the Separate 3 condition were unaware that never listed all available flights, although heard prompt like “*I found five outbound Northwest flights.*”
 - Last 17 Verbose subjects were asked 3 questions:
 - Notice difference between versions of system? 12/17 = yes.
 - Which best? No consistent preference.
 - Which worst? 5/12 = Combined, 7/12 = None specified.
 - ◆ Some subjects disliked it when one leg matched their constraints and the other did not. These subjects wanted to choose the legs separately.

Take Home Lesson

- ◆ When choosing between multiple flights, subjects wanted to *hear all relevant information* about each flight.
- ◆ Task completion improved when *all flights were presented at once*, without any intervening questions from the system.
- ◆ Goal of the users:
 - The goal of the user is to book the best possible flight. Their goal is *not* to have a conversation with the system; to figure out the rules for interacting with the system. “Just (all) the facts, ma’am.”

Prompt & Collect vs. Selecting from List

◆ Prompt & Collect

- System plays a 'Prompt' and 'Collects' a spoken response from the user.
- The user knows the information, and the system must obtain it from them.
- **Terse is better!** The standard UI practice is to minimize the length of prompts.

◆ Selecting from a List

- System presents a list of options, the user chooses item from the list.
- The system knows the possible options, the user must understand them and choose the best.
- **Verbose is better!** Give user all the information that is relevant to their decision
 - But good UI practice would dictate irrelevant info should not be presented.

Adaptation in Selection Criteria Content

- ◆ Users like to hear all relevant information at once a selection task.
- ◆ While is agreement concerning important selectic criteria, are idiosyncratic differences, e.g. some users have strong preferences about layover location.
- ◆ **Implication:** If a user asks a question concerning a selection criteria that is not by default listed by the system, e.g. “*Where does that flight stop?*”, the information presented about all subsequent flight should include the location of any stops.