Experimenter Instructions for End-to-End Evaluation

MADCOW

October 18, 1993

The experimenter instructions provide a step-by-step description of what needs to be done for running an end-to-end evaluation. As the physical arrangements at every site are different, the procedure outlined here (and in the Subject Instructions) may need to be adapted to local circumstances. All sites should prepare a document that details how their procedure deviates from that outlined in these documents. This document should be submitted to NIST prior to the start of data collection.

General characteristics of the evaluation process are described in the document **A Proposal for Experimental End-to-End Evaluation**. The experimenter should refer to it for details that are not covered in the current document.

1 Before the evaluation

You should have received the following materials from NIST:

- two sets of 4 scenarios each.
- a set of scenario sequences to run for each subject.

The subject instructions will contain stubs for site-specific information. These should be filled out with the necessary information. Your site's customized instructions are to be submitted to NIST (in addition to the document detailing procedural differences).

2 Identifying subjects

You will need to schedule the appropriate number of subjects that fit the constraints described in section 3.3 of the **End-to-End** document. The prudent experimenter understands that some subjects may not show up or that equipment may malfunction; the scheduling procedure should be sufficiently flexible to accommodate such eventualities.

Subjects should be qualified prior to a session (say by telephone). Please refer to section 3.3 of the End-to-End Evaluation document for the criteria you should apply. At the time of initial contact you should describe the nature of the task that will be performed.

3 The collection room

The data should be collected under normal office conditions, subject to common sense constraints (e.g., choosing a location that minimizes extraneous interruptions). The seating arrangement, monitor position, etc. are left entirely to the site's discretion. The site should, however, should prepare a complete description of the physical circumstances of data collection, to be submitted to NIST.

No materials pertaining to the ATIS task, other than those specified by this document should be present in the collection room. These materials are:

- Subject instructions and crib sheet.
- A notepad with pencil.
- Scenarios, with only the current scenario visible to the subject at any one time.

4 General procedure

It is essential that the counterbalancing scheme outlined in the End-to-End Evaluation document be adhered to. As specified in that document, each site will run eight subjects, 4 male and 4 female. Two of each will be assigned to the first Latin square, the other two to the other Latin square. Each subject will be given the four scenarios in their row of their square, in the order prescribed.

The following procedure will be followed during the collection session:

- When the subject arrives, explain again the nature of the activity (in general terms what will be done, the time involved, etc.)
- Present subject with a consent form for approval and signature. If the subject refuses to sign the consent form, release them from the experiment at this time.
- Go through the subject instruction script (as detailed in **Subject Instructions for End-to-End Evaluation**) and answer questions about the procedure. Note that the instructions will contain stubs for site-specific information, such as a description of the input protocol, display layout, specific system features, etc.

- Present the practice scenario (and announce it as such). At this time make sure that the subject is operating the system properly. It is appropriate to intervene with comments on the mechanics of interaction.

- Present a (real) scenario to the subject. The specific scenario to use will be specified by the ordering information received from NIST. The scenario should be easily accessible throughout the session, say placed on a typing stand next to the monitor. The experimenter should be absent from the experimental room during the time that the subject is reading or working on a scenario.

  When the subject indicates that a scenario is complete or is being abandoned, remove the scenario description from view. At this time, the user is to mark down the solution to the scenario problem, making use of the template provided by NIST. If the subject so wishes, a short break may be taken at this time. (The method used to notify the experimenter is left to the site's discretion.) The break should not last longer than 5 minutes. The subject may leave the room during the break.

  This procedure will be followed for each of the scenarios in a set (i.e., "go to step a").

- When all scenarios have been completed, present the subject with the Subject Questionnaire. When it's completed, present the subject with an Information Form.

- Debrief the subject. Debriefing involves giving a short verbal explanation of the experiment, then soliciting questions. The following is a minimalist debriefing statement; sites may wish to expand upon it.

> The purpose of the experiment you have just participated in is to gather data on how people interact by voice with a computer. The system you used automatically interpreted your spoken commands and displayed information as best it could. As you may have noticed the technology is not perfect and various errors occurred. By having you and others use the system, we hope to improve it. During the session a variety of information was recorded and will be analysed later to help us better understand what went on. Everything you spoke to the system has been recorded, as well as the system's responses to you and various timing information. Your data will be combined with that of others during analysis and your contribution will remain anonymous. We would ask you not to discuss this experiment with others as this may compromise their effectiveness if they choose to also participate.
>
> Thank you very much for helping us with our work. Do you have any questions about what went on or about what we're doing?

- Present the motivator (if any).

At the site's discretion, a complete record of the session may be kept, either on audio or video tape. The site recording a session may later use this information to interpret user behavior.

5 Things to watch for during the session

It is important that the experimenter at no time communicate any information about appropriate task behavior to the subject: do not give any hints on how the task can be completed and do not volunteer any information about the task domain (if necessary, refer the subject to the written subject instructions).

At no time prior to the debriefing should you comment on the experiment, its purpose or design (past the details explicitly presented to the subject). You may, however respond to question dealing with the equipment (e.g., "is this the button I should click?") or about general procedure (e.g., "can I start now?").

You may come across the occasional subject who is very talkative. If you find yourself in this situation, remind the subject that you have been asked to keep communication down to a minimum, but that you will be glad to discuss any aspect of the experiment once the session is over. The rule of thumb is: don't give any hints that may affect how the subject solves the problem described in a scenario.

You may interrupt and correct the subject if one of the following conditions arise:

- The microphone becomes dislodged without the subject's awareness. (The subject should have been made aware of microphone placement issues during the instruction process.)
- The subject fails to correctly input speech (e.g., by forgetting to click on the appropriate button).

6 When the system crashes

- Inform the subject immediately that the system is malfunctioning and that they should suspend working on the scenario. Depending on the severity of the problem, you may ask the subject to leave for a few minutes.
- Restore the system.
- Ask the subject to continue the task from the point at which they left off. For logging purposes, the restart should be considered a continuation of the current scenario, with

events and files numbered accordingly. If due to the nature of the crash this is not possible, then logging should be restarted. However, the pre-crash log (even if partial) will be retained and used in scoring the session.

In certain extreme cicumstances (a hardware malfunction, for example) it may not be possible to proceed with a session. In such a case it is permissible to cancel the session and replace the subject. All crashes should be noted, together with explanatory notes and this information included with the submission to NIST.

7 Submission Checklist

The following items need to be submitted prior to the start of data collection (nominally 1 October 1992).

1. Site-customized subject instructions
2. Site-customized experimental procedure
3. Description of experimental environment

The following materials should be submitted at the conclusion of the experiment, nominally 31 October 1992:

1. For each subject, all collected data (for each scenario, whether successfully collected or not). This includes the following files:

   - `.log` file
   - `.wav` file, for each utterance.
   - `.sro` file, for each utterance. Transcribed at the site.
   - `.squ` file a machine readable version of the subject questionnaire answers, as per the format defined by the committee.

   In addition, a `sys-out.hyp` file should be submitted for each site. The definitions for most of these files can be found in the directory madcow/doc/, at the NIST archive site (`jaguar.ncsl.nist.gov`).

2. Subject information forms; the standard demographics as currently obtained during data collection.
3. Session notes; a log of unusual occurrences and other information potentially useful for interpretation (e.g., noting subjects who have already used the Atis system, explaining a crash, etc.).