Ideal II
Initial Requirements for a Next-Generation Usability Evaluation Support Tool
Client: Virginia Tech HCI Labs

Points of contact:
Dr. Hartson (VP, Marketing),
Ms. Mridu Darshani (Project Manager, Usability Process)

Version 1.2, 13 January 1997

**Overall purpose of system**
Support gathering and analysis of usability data within the process of formative usability evaluation.

**Terminology**
The term *user* refers to the person using the software application, the usability of which is being evaluated. In typical usability evaluation this person is also called the subject or participant. To distinguish, the term *evaluator* will be used to refer to the person using Ideal II to perform usability evaluation.

**Goals**
- Extract and support most essential features from Ideal I that directly support formative usability evaluation.
- Provide broad support for both quantitative and qualitative data gathering.
- Use video and voice technology to take this product well beyond anything currently available.

**Existing comparable (present-generation) units**
We have each of the following available for inspection and analysis by the Ideal II development team:
- Ideal I
- TRC, Inc., portable usability lab (to arrive by end of 1997)
- Norm Wilcox Associates, Lab-in-A-Bag

**Data gathering requirements**
- Video and audio data
  - Both user and evaluator audio are synchronized to the video and are considered part of the video stream.
- Quantitative data
  - Most common: time on task, counts (e.g., errors, mouse clicks)
  - Time on task: Need a way to start and stop the timing
    - Real-time
      - Evaluator has explicit control; clicks on something to start and stop timer
    - Probably the most convenient and cost effective for evaluator
  - Ex Post Facto
• Evaluator starts and stops timer while reviewing video, associating start and stop times (for timed intervals) with points in the video sequence.
• Can get accuracy to a single frame
• Easiest to implement from system designer’s view
• **Instrumentation of application (automated data gathering)**
  • An advanced, experimental approach to explore
  • Requires software communication with application to start and stop timer
  • Same high accuracy as ex post facto
  • Less hassle and work for evaluator, a *large user consideration*
  • Difficult to standardize over range of application development platforms; therefore, this probably has to be customized for each application, a *large developer consideration*.

• **Counts**
  • Cannot be automated
  • Evaluator must explicitly control, usually by clicking on a button
  • Single design should allow in real-time or ex post facto

• **Qualitative data**
  • Identifying critical incidents
    • Each time and evaluator identifies a critical incident, a critical incident record is created, containing:
      • Textual (or voice) comment, mostly for labeling, by evaluator
      • Link to the Usability Problem Classifier (UPC) starting point for classifying a usability problem associated with the critical incident
      • Link to the Usability Problem Database (UPDB) to review the classification of a classified usability problem associated with the critical incident
      • Link (tag or pointer) to location of associated critical incident in video stream.
    • New critical incident record can be generated:
      • in real-time during data capture and
      • retrospectively, during video review.

• Critical incident record data to be displayed and accessed in same format as used by the Ideal database, i.e., organized by project, version, usability specification, usability attribute, task, session, date, user id (see below).

• Critical incidents associated with a single session (e.g. one user, one task, one date) to be displayed in table form, one row per critical incident. Clicking on any row causes video for that critical incident to be played.

• Playback and review of critical incidents
  • Clicking on a row in the critical incident table causes playback of video associated with that critical incident
Evaluators can add typed or video comments during playback (how many different channels of audio can be used?)

Verbal protocol gathering not treated separately; each verbal protocol episode is simply considered a critical incident.

Data analysis requirements

- Ideal II is to support usability data analysis of for purposes of project management decision making, including:
  - rating usability problems with respect to importance-to-fix
  - cataloging candidate solutions
  - cost estimates for implementing candidate solutions.
  - calculating cost-importance ratios and priority rankings for usability problems
  - deciding which usability problems to fix
  - deciding when to stop iterating (if usability specifications are also supported)

Data storage, retrieval, and dissemination requirements

- Need for underlying, embedded database management system to store and retrieve critical incident records
- Critical incident records stored and retrieved by project, version, usability specification, usability attribute, task, session, date, user id (not name), other attributes (TBD—need to consult with industrial partners for these)
- Critical incidents associated with a single session (e.g. one user, one task, one date) to be displayed in table form, one row per critical incident. Clicking on any row causes video for that critical incident to be played.
- Access to all data via similar Ideal II stations remotely located. This is important for development teams distributed over time and space.
- All critical incident tables and data gathering results (and usability specification table, if they exist) will be printable in an acceptable format.

Development process requirements

- The usability of the Ideal II user interaction itself is to be developed using a prototype or similar method, with iterative usability evaluation.
- The Usability Methods Research Lab (UMRL) is highly suited for this kind of meta-development.

Desirable, but not mandatory, features

- Physical portability for field work
- Support for establishing usability specifications
  - Strong market demand
  - Necessary to fully support use of quantitative data
  - Would distinguish us from competitors
  - But, electing not to do this is a possibility for limiting scope in Version 1
• If usability specifications are supported, it is to be through the same tool (not a separate tool as is done in Ideal I), using additional functionality.

**Early engineering decisions**

- Hardware/software development platforms
- Video streams
  - Two streams, recorded in parallel
    - Camera view of user plus scan converted or direct digital screen activity
    - Twice the storage requirement for data
    - Somewhat more complex, requiring synchronization
    - Better for making video clips for executive summaries and training, education
    - No data loss
  - One stream: mixed in real-time
    - More cost effective
    - Less storage
    - Less complex
    - Possible data loss
  - I lean toward two-stream, because it is an experimental system, meaning flexibility is important.
- Explore where best to apply speech recognition technology. We expect it to apply wherever typed or verbal/audio inputs are mentioned above from the evaluator. It is not expected that it will apply to audio inputs from the user.
  - Verbal input from the evaluator is raw data. Will it add efficiency to have this available as text for editing?

**Roles of contact people**

- Dr. Rex Hartson, VP, Marketing:
  - Deliver seminar on the usability evaluation process
  - Represent view and interests of evaluators (users of Ideal)
  - Issues of general marketability beyond Virginia Tech usability labs
  - Liaison with possible partners in industry
- Ms. Mridu Darshani, Project Manager, Usability Process:
  - Represent technical needs to fit into context of larger usability development process
- Anyone else want to be on this list?

**Possible partners in industry**

- IBM's User-Centered design & Evaluation Group, Toronto, Canada, has developed UCDCam, a software-based digital "video recorder" that constantly monitors screen activity during user task performance. We have already worked with this group in our work on remote usability evaluation. Could be used for screen capture part of our qualitative data. We have permission to
make limited use of their experimental systems. (All involved might have to sign a non-disclosure agreement.)

- Ted Szostak is President of Triangle Research Collaborative, Inc. of Research Triangle Park, N. C., a manufacturer of portable usability labs, once of which we are slated to receive on 29 Dec 97. They have developed, built, and delivered numerous present-generation usability evaluation support tools and can provide real-world market-savvy inputs. (All involved might have to sign a non-disclosure agreement.)

- The Social Security Administration (SSA), a leader in usability evaluation and instrumentation within government. We have worked with them in the past and they are highly suitable as a real-world source of requirements.