

## SESSION 13: NEW DIRECTIONS

*Ralph Weischedel*

BBN Systems and Technologies  
70 Fawcett Street  
Cambridge, MA 02138

The three papers of Session 13 address issues differing from those in the remainder of the workshop. Two employ a methodology to discover preferences for speech input in a multimodal interface. The third raises issues of processing human language without any assumption that the speech or text has been converted to an online sequence of ASCII characters (or other character codes).

As an introduction to the first two papers, consider Wizard of Oz experiments such as used in collecting ATIS data. In such an experiment, the subject is asked to use a system to solve one or more problems. The "system" could be a person who simulates a proposed capability, for instance to determine language and interface properties for a proposed computer capability. Alternatively, the system might be an existing capability.

Perhaps the first such experiment was performed by Ashok Malhotra (1975) to collect data that would suggest how varied (and challenging) textual queries would be in an interactive query application. Malhotra simulated the whole system, a very labor-intensive task.

The first paper ("Mode Preference in a Simple Data-Retrieval Task") employs fully implemented components to measure user preference for spoken input, versus filling a form, versus employing a scroll bar to look up telephone numbers in an online telephone book. The paper immediately got my attention with the following statement in the introduction, "For activities in a workstation environment, formal comparisons of speech with other input modes have failed to demonstrate a clear advantage for speech on conventional aggregate measures of performance, such as time-to-completion . . . ". The author's experiments demonstrate a flaw in the analysis of previous results and go on to measure a marked preference for speech input, even when speech may not give the best time-to-completion results.

The second paper, "A Simulation-Based Research Strategy for Designing Complex NL Systems," involves a person behind the scenes (the wizard) simulating the system, though much is automated. The resulting environment being simulated for the user is quite rich, allowing both speech and handwriting input. Careful preparation of the experimental environment enabled automated support so that response to the user is streamlined, thereby allowing the user to move at his/her own pace. To illustrate the kind of studies the methodology supports, the authors show some results suggesting that syntactic ambiguity is less when filling out a form (rather than when producing unconstrained input) and is also less in handwriting than in speech.

The third paper, "Speech and Text-Image Processing in Documents," assumes minimal signal processing. For instance, they describe editing and indexing of audio forms rather than the text file resulting from continuous speech recognition. Similarly "text-image" processing, is the editing of the bitmap representation resulting from scanning a document in, rather than editing a sequence of bytes in some character code such as ASCII. One of the tools described is therefore aptly named "Image Emacs". A third effort described in this paper is document image decoding, a framework for processing scanned-in documents.

### REFERENCES

Malhotra, A. "Design Criteria for a Knowledge-Based English Language System for Management: An Experimental Analysis", Massachusetts Institute of Technology, Cambridge, Ma., MAC TR, No. 146, February, 1975.