Structured Speech Input for Clinical Data Collection

Michael A. Grasso, PhD, MS-I
grasso@gwu.edu
http://home.gwu.edu/~grasso/

George Washington University School of Medicine
7th Annual Health Sciences Annual Research Day
April 19, 2002

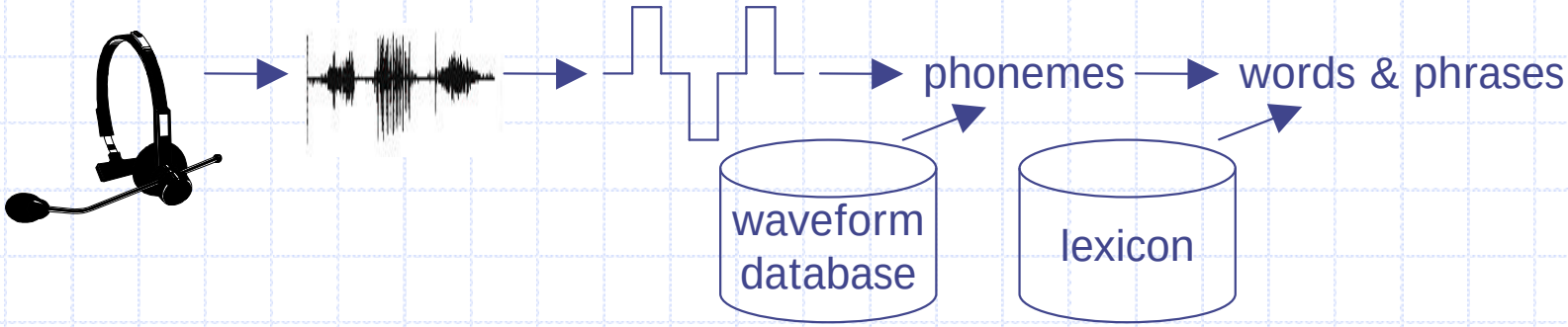
Supported by NIH grant 2R44RR07978

Introduction

- A common application of speech recognition software is to dictate clinical narratives and reports.
- However, free-form dictation has several problems.
  - Inefficient operation.
  - High error rate.
  - Low user acceptance.
  - Can not analyze narratives by computer automation.
- We address these limitations with the following.
  - Structured speech input to increase efficiency and reduce speech recognition errors.
  - Nomenclature types and validation rules to support a wide range of clinical studies.
  - Strongly typed observations in an event-oriented data model for distribution and analysis according to study protocol.

How Speech Recognition Works

- Speak into the microphone.
- Convert signal to digital waveform.
- Convert digital information to phonemes.
  - A phoneme is the smallest unit of speech.
  - Compare to database of waveforms for all known phonemes.
- Assemble phonemes into words and phrases.
  - Use lexicon and stochastic model.



Recent Speech Recognition Trends

- Large vocabulary.
  - Systems can recognize up to 40,000 words or more.
- Speaker independence.
  - Little or no training is required to teach the system to recognize your voice.
- Continuous recognition.
  - You can speak in a natural rhythm with no pauses between words or phrases.
- Major speech recognition products.
  - IBM ViaVoice (www.ibm.com/software/speech/).
  - Dragon NaturallySpeaking (www.lhsl.com/naturallyspeaking/).

Clinical Dictation with Speech Recognition

- Dictate medical narratives directly into a computer.
  - Used in Radiology, Pathology, and Endoscopy.
- Advantages.
  - Clinicians speak in a natural manner.
  - Narratives do not need to be transcribed by clerical personnel.
  - Decrease turnaround time.
- Disadvantages.
  - Computer can only recognize words, but not their meanings.
  - Clinical narratives can not be interpreted or analyzed.
  - Free-form text dictation has a high error rate.

Research Objectives

- Use structured speech input to increase efficiency and minimize speech recognition errors.
- Constrain speech input through a set of nomenclature types and validation rules that can be customized for a wide range of studies.
- Store speech input as a series of coded and strongly typed observations to facilitate data analysis and interpretation.

Materials

- Software prototype.
  - Pentium-based computer running Windows 98.
  - Dragon NaturallySpeaking and the Microsoft Speech API.
  - Microsoft Visual Basic and Microsoft Access.
  - Roughly 1,500 lines of code.
- Event-oriented data model.
  - Supports a list of observations for each subject on a study.
- Validation rules.
  - Ensure the completeness and accuracy.
  - Follow a simple "if-then" format based on Boolean criteria.
  - Used to identify missing data, verify normal ranges, and identify trends in longitudinal data.

Methods

- Evaluation.
  - Prototype supported microscopic pathology data collection.
  - Compared structured vs. unstructured speech input.
  - Twenty pathologists in a within-groups design counterbalanced on structure and task order.
  - Measured speed, accuracy, and acceptance.
- Extrapolation.
  - Added configurable nomenclature types and validation rules.
  - Added an event-oriented data model.
  - Five subjects attempted to apply the system to a wide range of clinical studies.

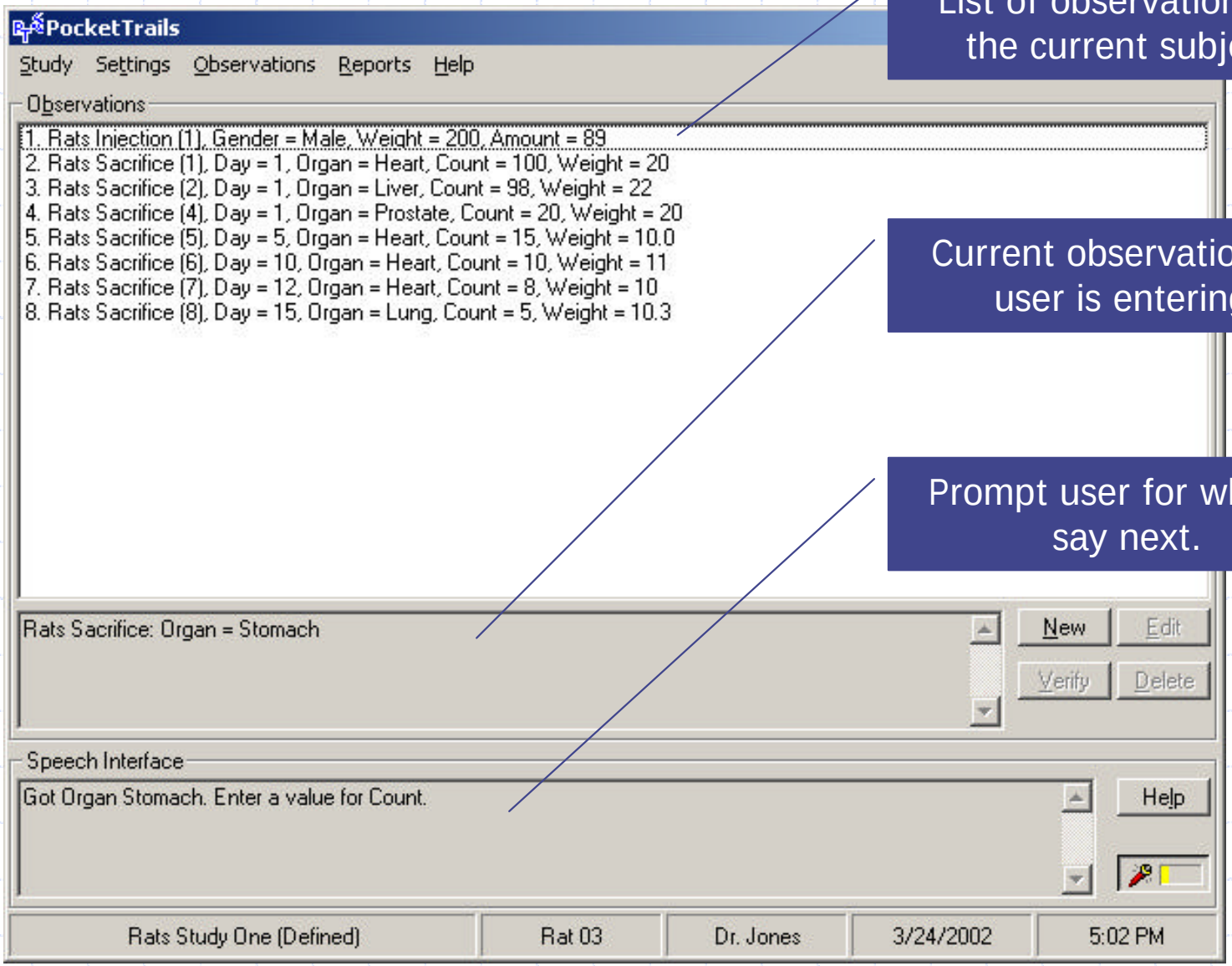
Results

- Evaluation of microscopic pathology prototype.
  - A 22% decrease in task completion time ($p < .001$).
  - A 36% reduction in speech recognition errors ($p < .01$).
    - Reduction due to the application of user interface principles, not improved speech recognition algorithms.
  - A 2.4% improvement in user acceptance ($p < .05$).
- Extrapolation of prototype to various applications.
  - Collect animal drug toxicology data.
  - Collect aquatic pathology data.
  - Compute the dose deposition of tumor antiferritin in tumor versus normal tissue.
  - Collect survey data for a Pfiesteria cohort study.
  - Collect patient vitals.

Conclusions

- Successfully developed a prototype speech-driven system for hands-free clinical data collection.
- The prototype address the limitations of free-form clinical dictation using the following.
  - Structured speech input increase efficiency, reduce speech recognition errors, and increase user acceptance.
  - Nomenclature types and validation rules to support a wide range of clinical studies.
  - Strongly typed observations in an event-oriented data model for distribution and analysis according to study protocol.

Data Collection Prototype

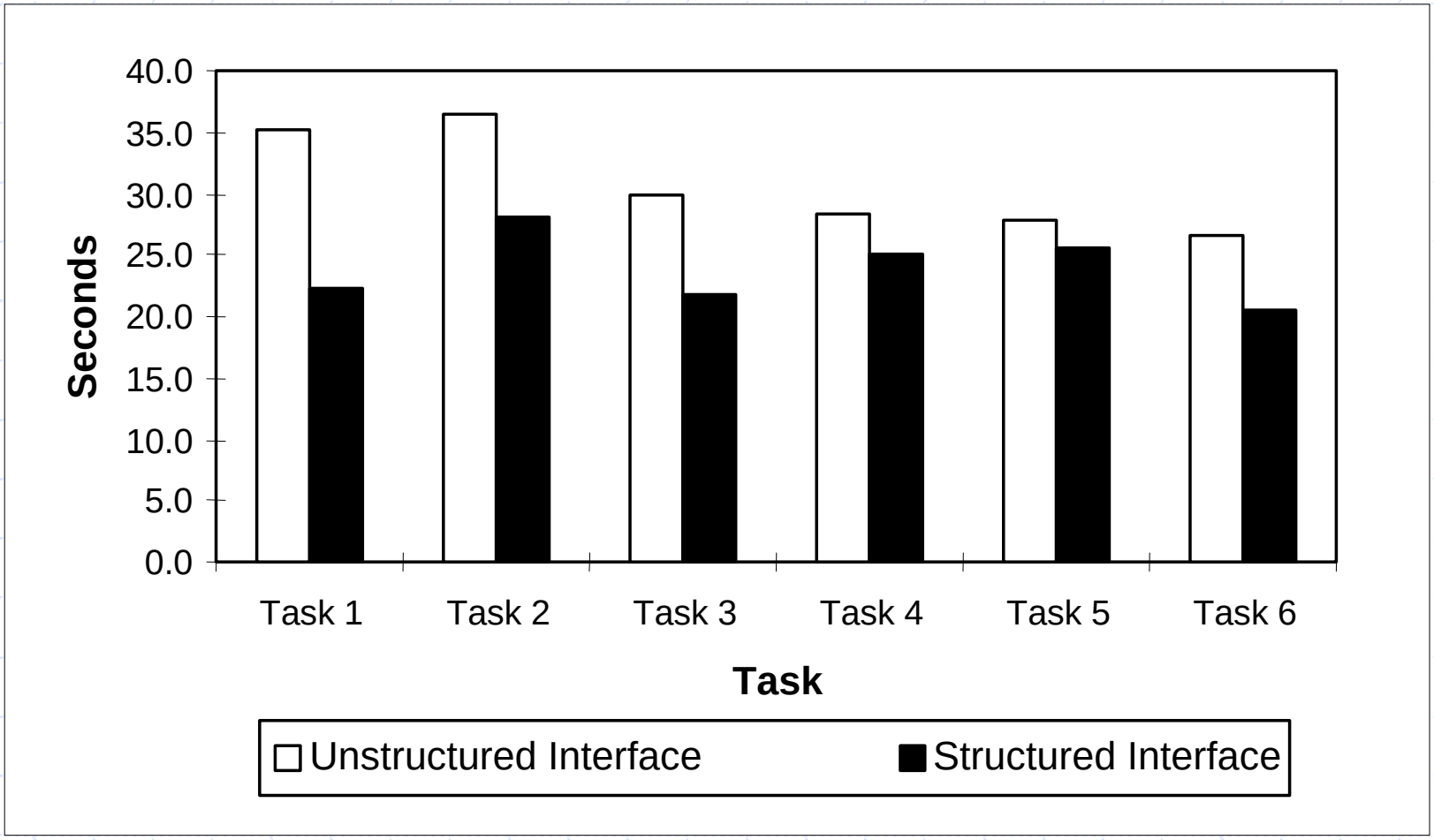


List of observations for the current subject.

Current observation the user is entering.

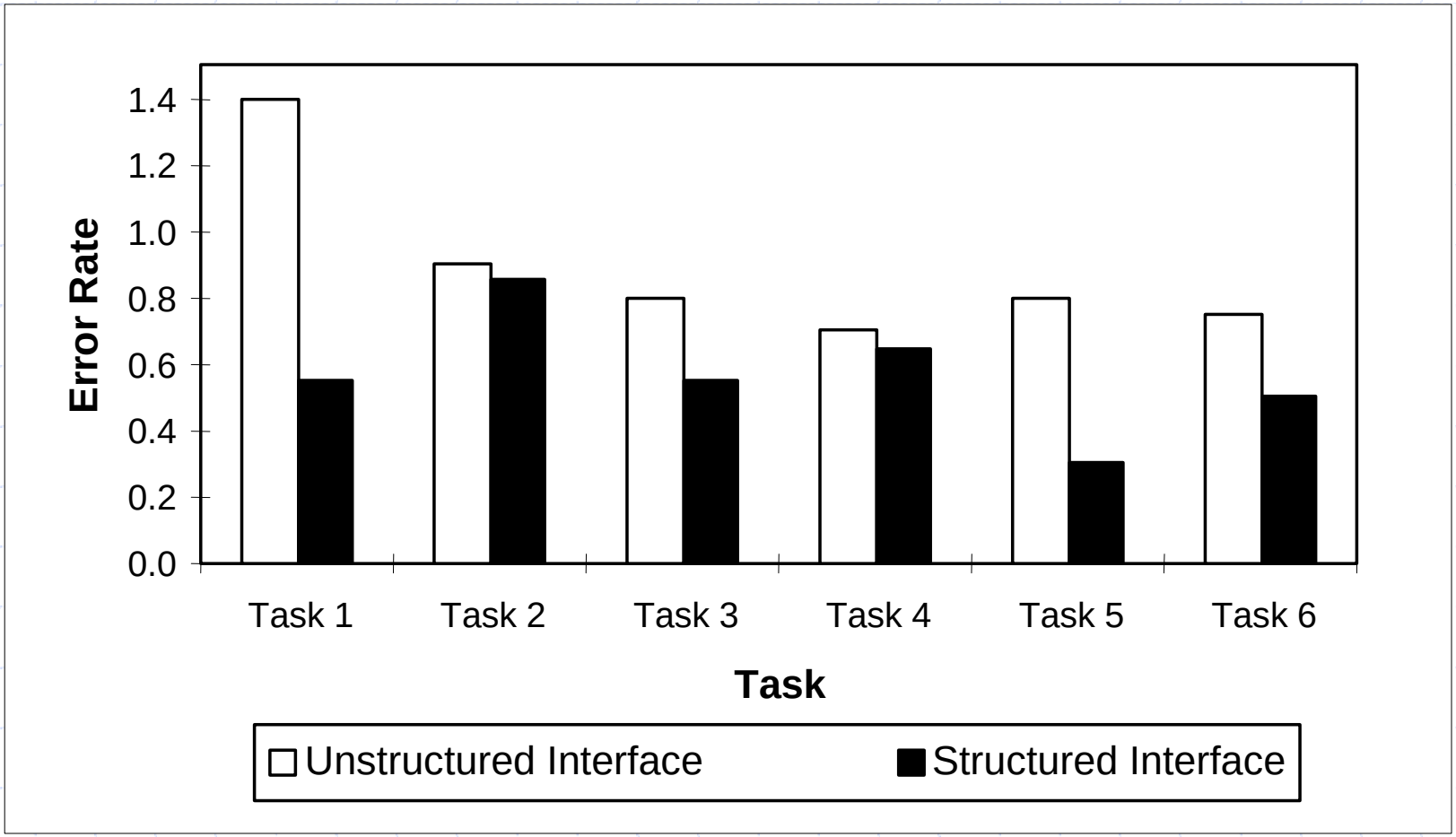
Prompt user for what to say next.

Mean Task Completion Times



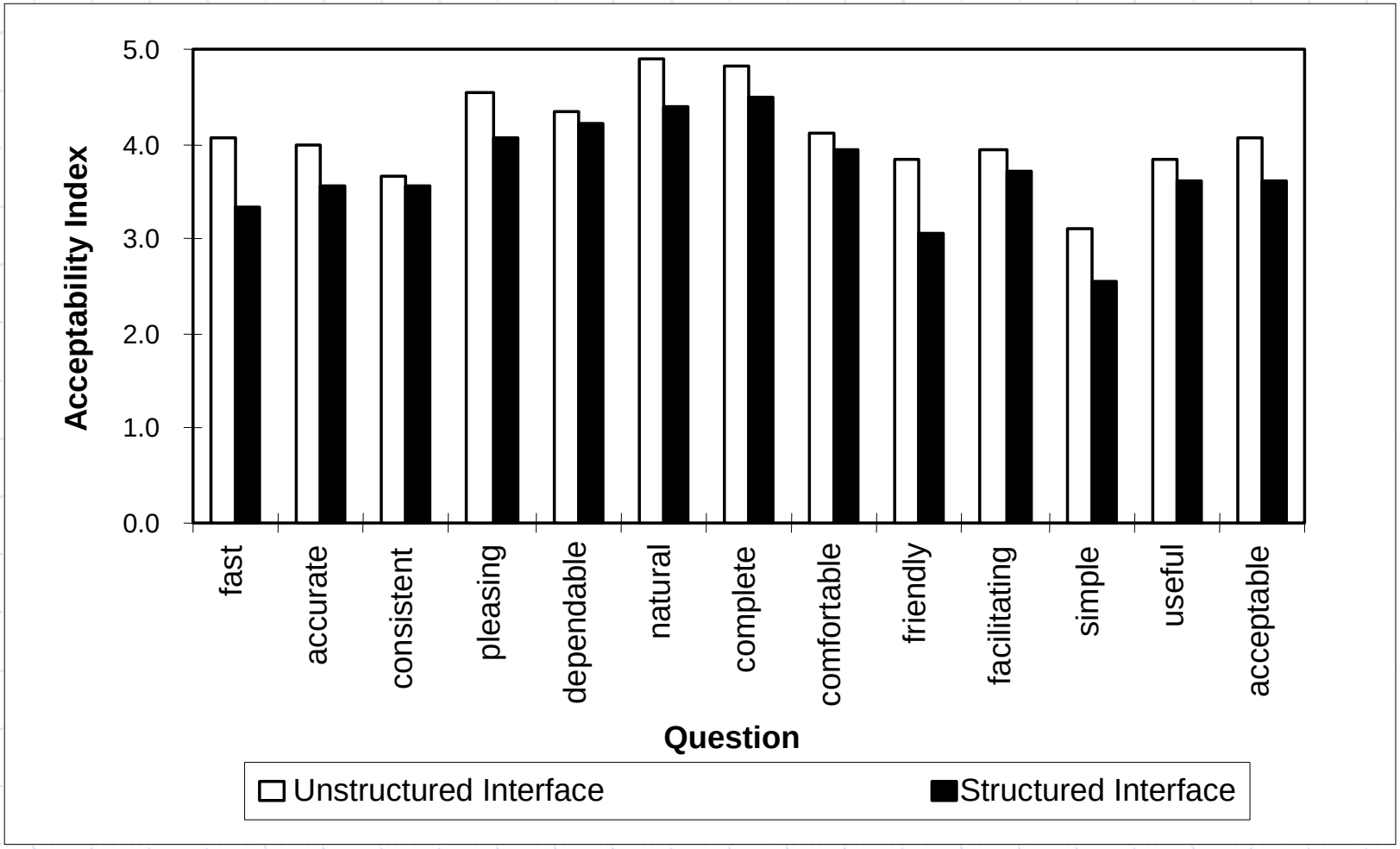
A 22% decrease in task completion time ($p < .001$).

Mean Speech Error Rates



A 36% reduction in speech recognition errors ($p < .01$).

Acceptability Index by Question



A 2.4% improvement in user acceptance ($p < .05$).

References

1. Grasso MA. Structured Speech Input for Clinical Data Collection, CBMS 2002 - 15th IEEE Symposium on Computer Based Medical Systems.

2. Grasso MA, Ebert DS, Finin TW. The Integrality of Speech in Multimodal Interfaces. ACM Transactions on Computer-Human Interaction. 1998;5(4):303-325.

3. Grasso MA, Ebert DS, Finin TW. Acceptance of a Speech Interface for Biomedical Data Collection. AMIA 1997 Annual Fall Symposium, 739-743.

4. Grasso MA. Automated Speech Recognition in Medical Applications. M.D. Computing. 1995;12(1):16-23.