INFORMATION ACCESS VIA VOICE

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee.

____
Yapin Zhong

Certificate of Approval:

____
Kai H. Chang
Professor
Computer Science and Software Engineering

____
Juan E. Gilbert, Chair
Assistant Professor
Computer Science and Software Engineering

____
Dean Hendrix
Associate Professor
Computer Science and Software Engineering

____
Stephen L. McFarland
Acting Dean
Graduate School

INFORMATION ACCESS VIA VOICE

Yapin Zhong

A Dissertation

Submitted to

The Graduate Faculty of

Auburn University

In Partial Fulfillment of the

Requirements for the

Degree of

Doctor Philosophy

Auburn, Alabama

December 19, 2003

INFORMATION ACCESS VIA VOICE

Yapin Zhong

____

Signature of Author

____

Date

Copy sent to:

<u>Name</u> <u>Date</u>

DISSERTATION ABSTRACT

INFORMATION ACCESS VIA VOICE

Yapin Zhong

Doctor of Philosophy, December 19, 2003

170 Typed Pages

Directed by Dr. Juan E. Gilbert

This dissertation concentrates on the problem of designing and developing a spoken query retrieval (SQR) system to access large document databases via voice. The main challenge is to identify and address issues related to the adaptation and scalability of integrating automatic speech recognition (ASR) systems and information retrieval (IR) systems. Additionally, the mechanics of designing an effective and efficient speech user interface (SUI) pose yet another significant challenge, especially since the aim is to facilitate voice queries of large document databases. The resulting system should enable users to access large document databases effectively and efficiently. Furthermore, its language model should be capable of adapting to updates of the document databases. In this research, a framework allowing information access to large document databases via voice is presented and several approaches designed to cope with the issues of adaptability, scalability, effectiveness and efficiency are described in detail. Through

experiments performed on the TREC-9 document dataset, the performances of the new approaches were evaluated and their potential was demonstrated.

ACKNOWLEDGMENTS

The author would like to express his deep gratitude to his advisor, Dr. Juan E. Gilbert, for his patient guidance, valuable advice, and continued encouragement throughout his studies. Sincere thanks are also due to his two graduate committee members, Dr. Kai H. Chang and Dr. Dean Hendrix, for their reviewing and advising efforts. In addition, the author would like to thank Celeste German for her reviewing and valuable comments. Finally, deepest thanks to author's wife, Weihong Hu, for her help while conducting the experiment and constant support.

Information Access via Voice

Yapin Zhong

Dept. of Computer Science & Software Engineering
Auburn University

Outline

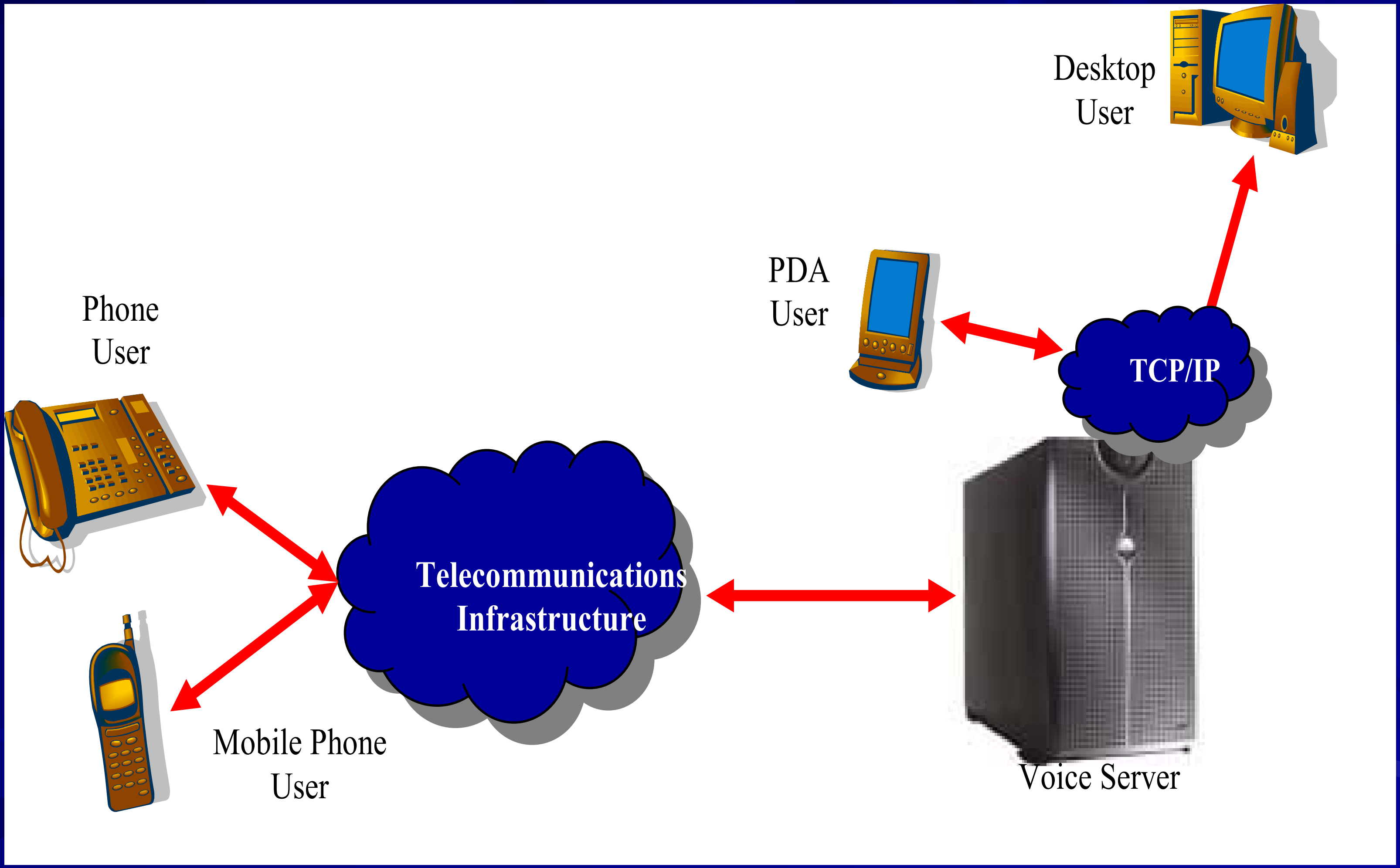
- Motivation
- Background
- Research Challenges
- A Framework of Spoken Query Retrieval
- Experiments and Research Findings
- Conclusions

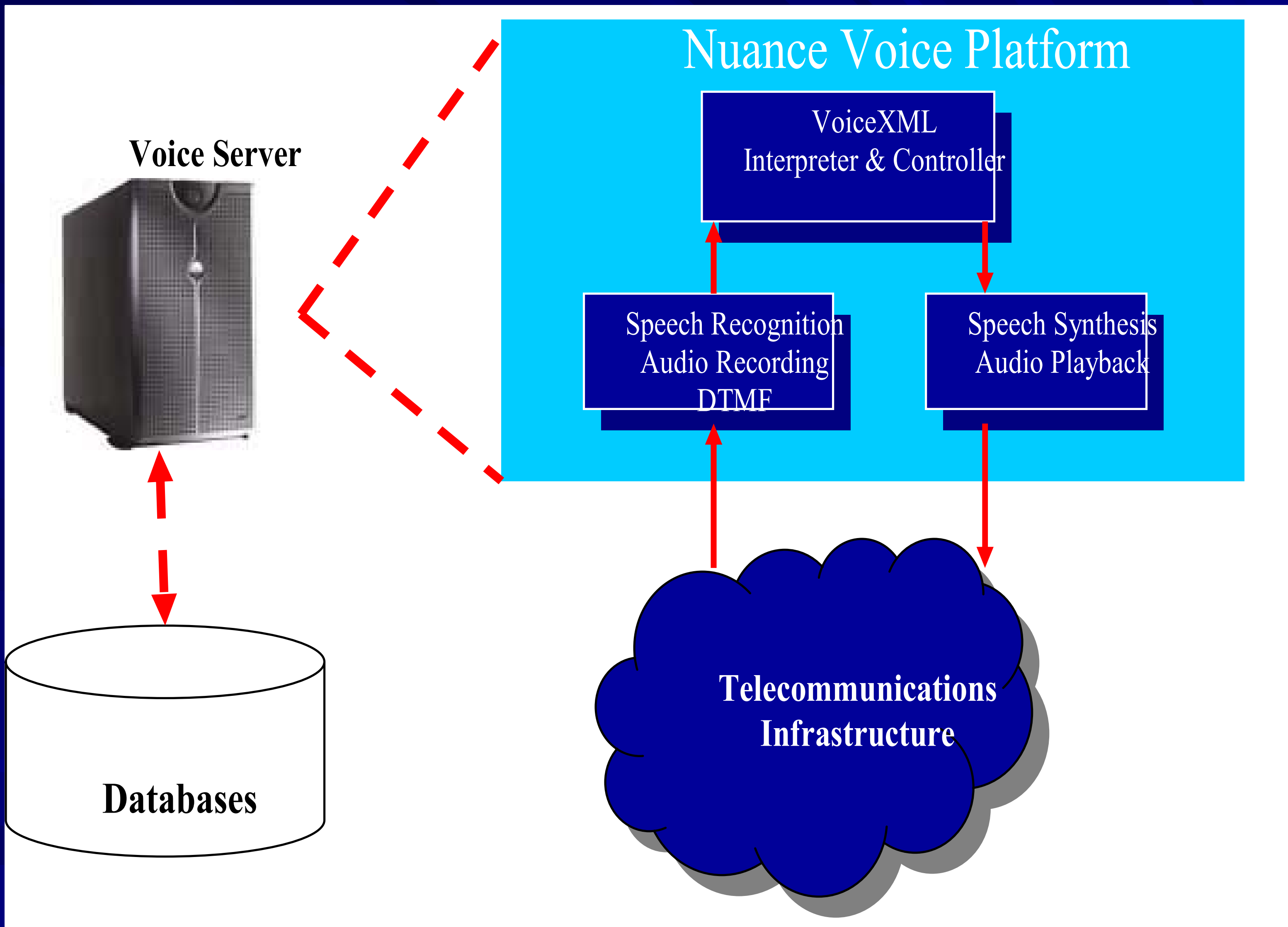
Motivation

- A very large part of the world population does not have access to either computers or the Internet
- Very tiny visual interfaces make users feel quite uncomfortable
- Blind or partially-sighted users are not able to access information visually

Background

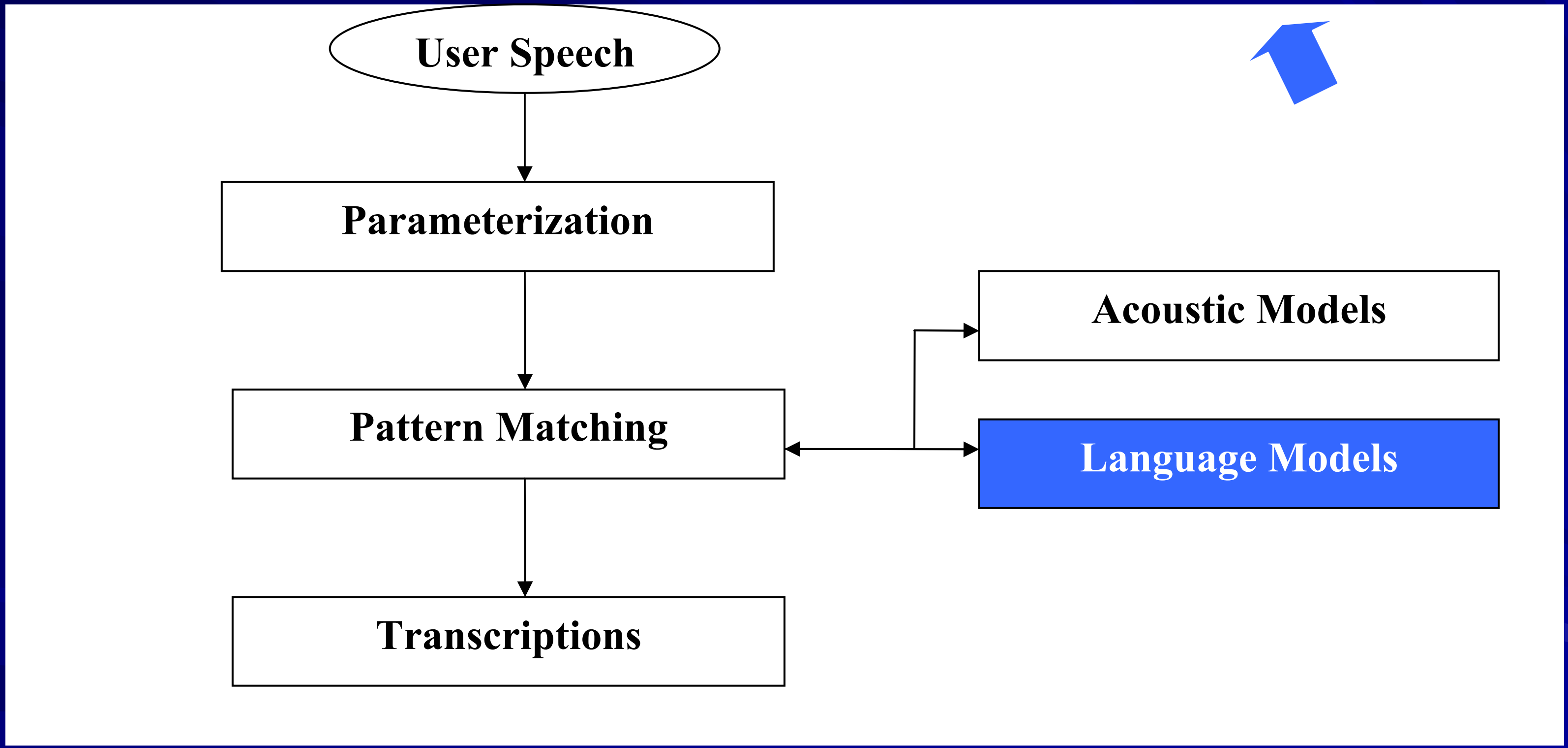
- Two categories: spoken document retrieval (SDR) and spoken query retrieval (SQR)
- In SDR, written queries are used to search speech archives for relevant speech information
- SQR uses spoken queries to retrieve relevant textual information


Desktop
User
PDA
User
TCP/IP
Phone
User
Telecommunications
Infrastructure
Mobile Phone
User
Voice Server


Nuance Voice Platform
VoiceXML
Interpreter & Controller
Speech Recognition
Audio Recording
DTMF
Speech Synthesis
Audio Playback
Voice Server
Databases
Telecommunications
Infrastructure

Automatic Speech Recognition

$$\arg\max_{W} P(W \mid X) = \arg\max_{W} P(X \mid W) \bullet P(W)$$



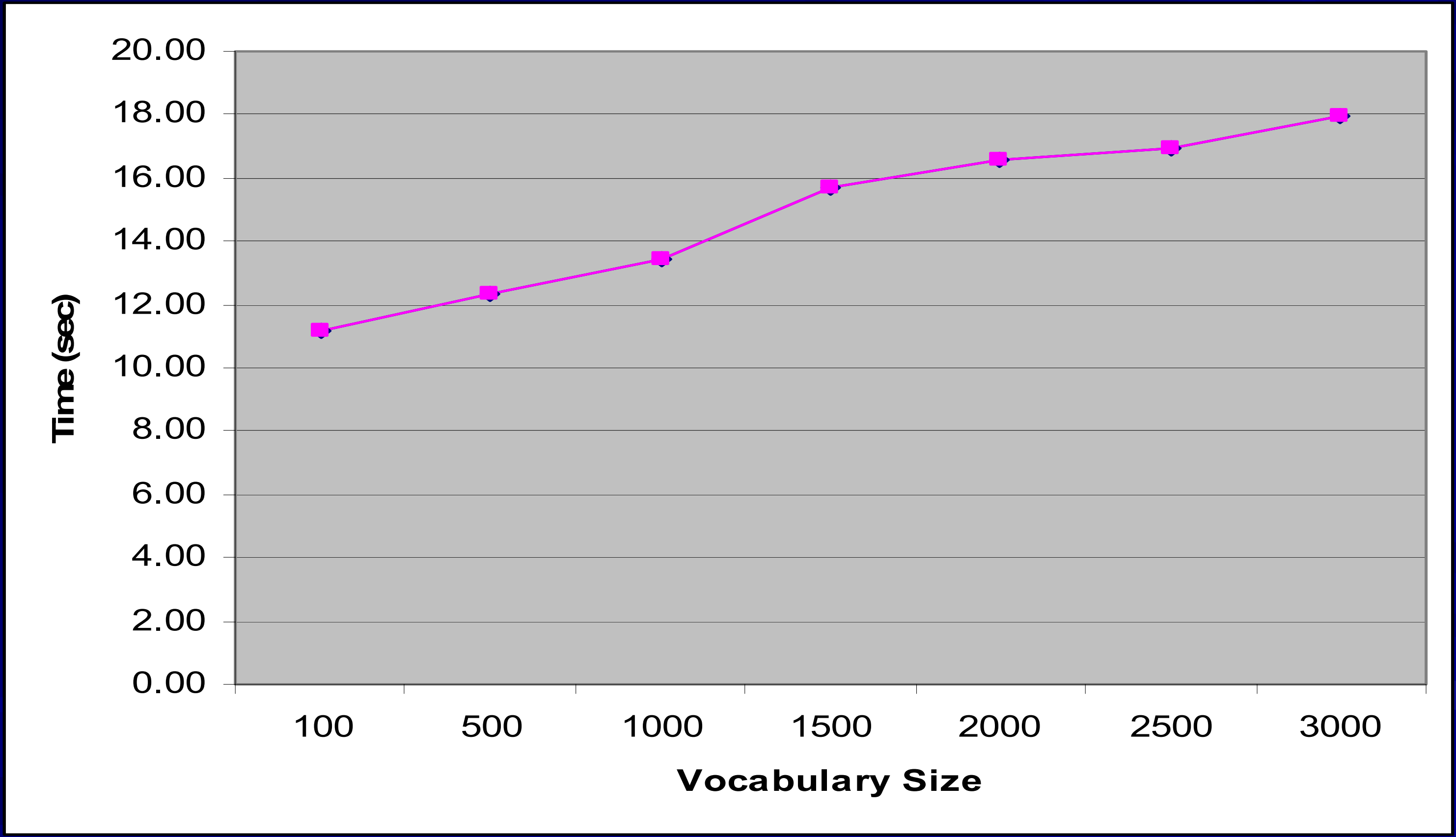
Three major properties in SQR

- Spoken queries are usually very short
- Spoken queries usually need a very large vocabulary
- Query processing is required to be in close to real time

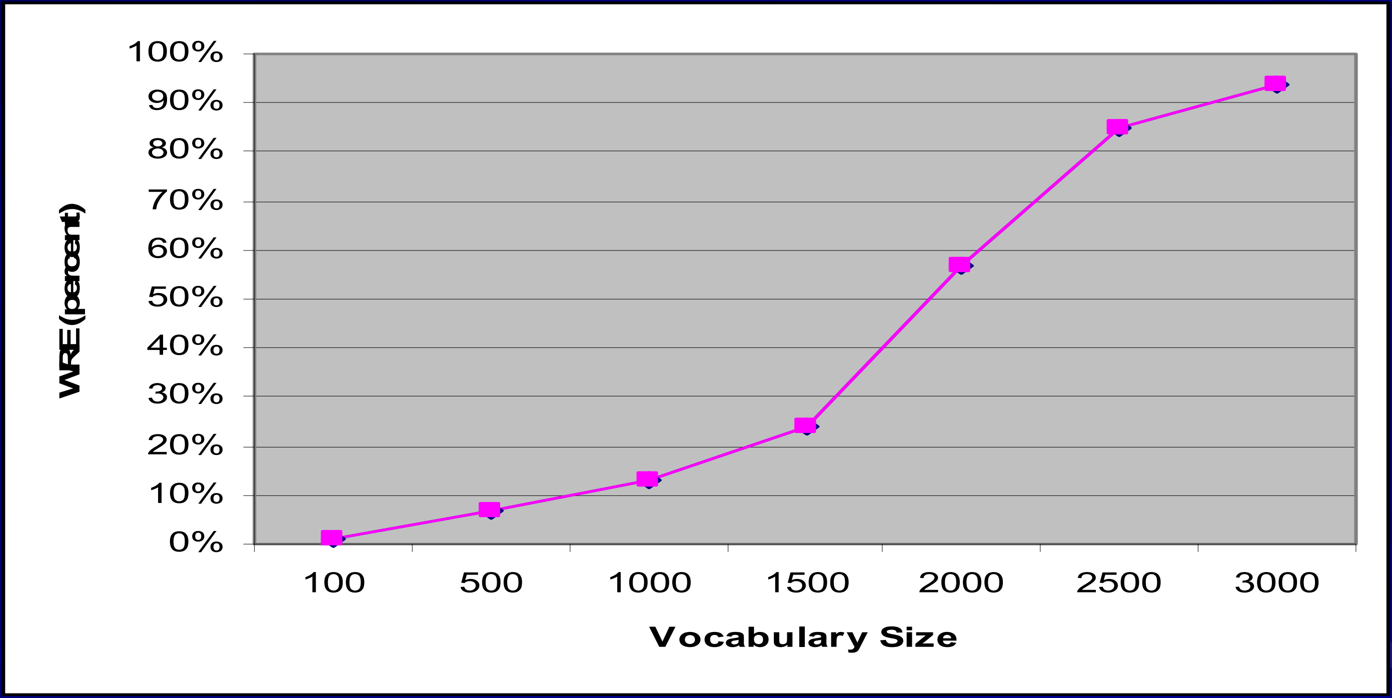
Research Challenges

- A lack of adaptability and scalability in integrating language models between ASR and document Information Retrieval systems
- Difficult to meet the critical aspect of “real time” user expectations
- A general lack of effectiveness and efficiency in designing speech user interfaces

The Language Size Vs. The time Consumed



The Language Size Vs. the Word Recognition Error (WRE) Rate



Query Coverage Compared to the Language Size

Vocabulary Size (k)	25	100	200	300	400	500
Query Coverage [FM02] (percent)	62.2	79.2	83.9	85.9	87.1	87.9

Problems in SQR Interfaces

- Speech is transient but graphics are persistent
- Speech is invisible
- Speech is asymmetric

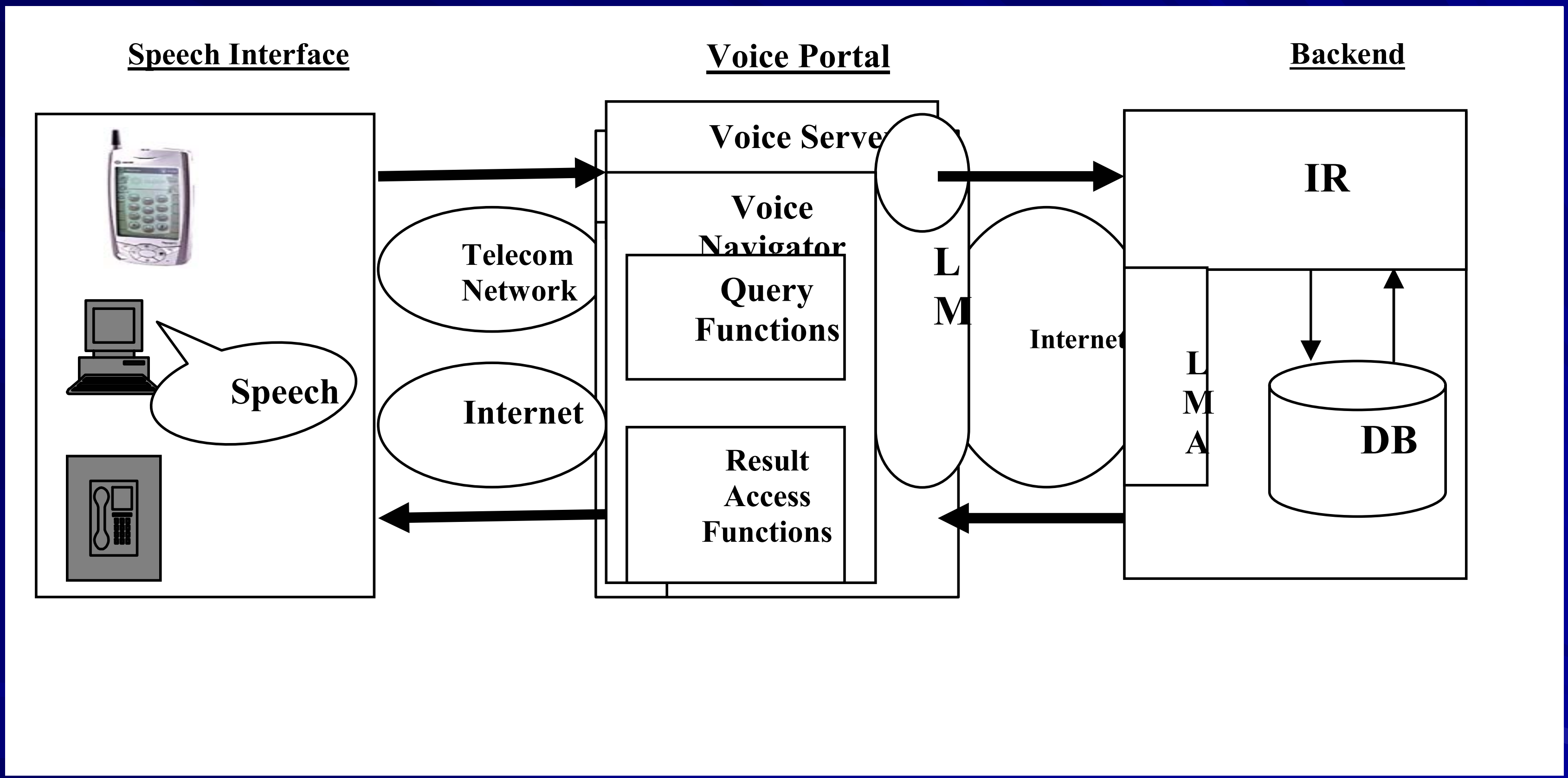
A Framework of SQR

- Design Principle
- System Architecture
- Context-Aware Language Model
- Bisecting K-Medioids Method
- Voice Navigator
- Information Verbalization

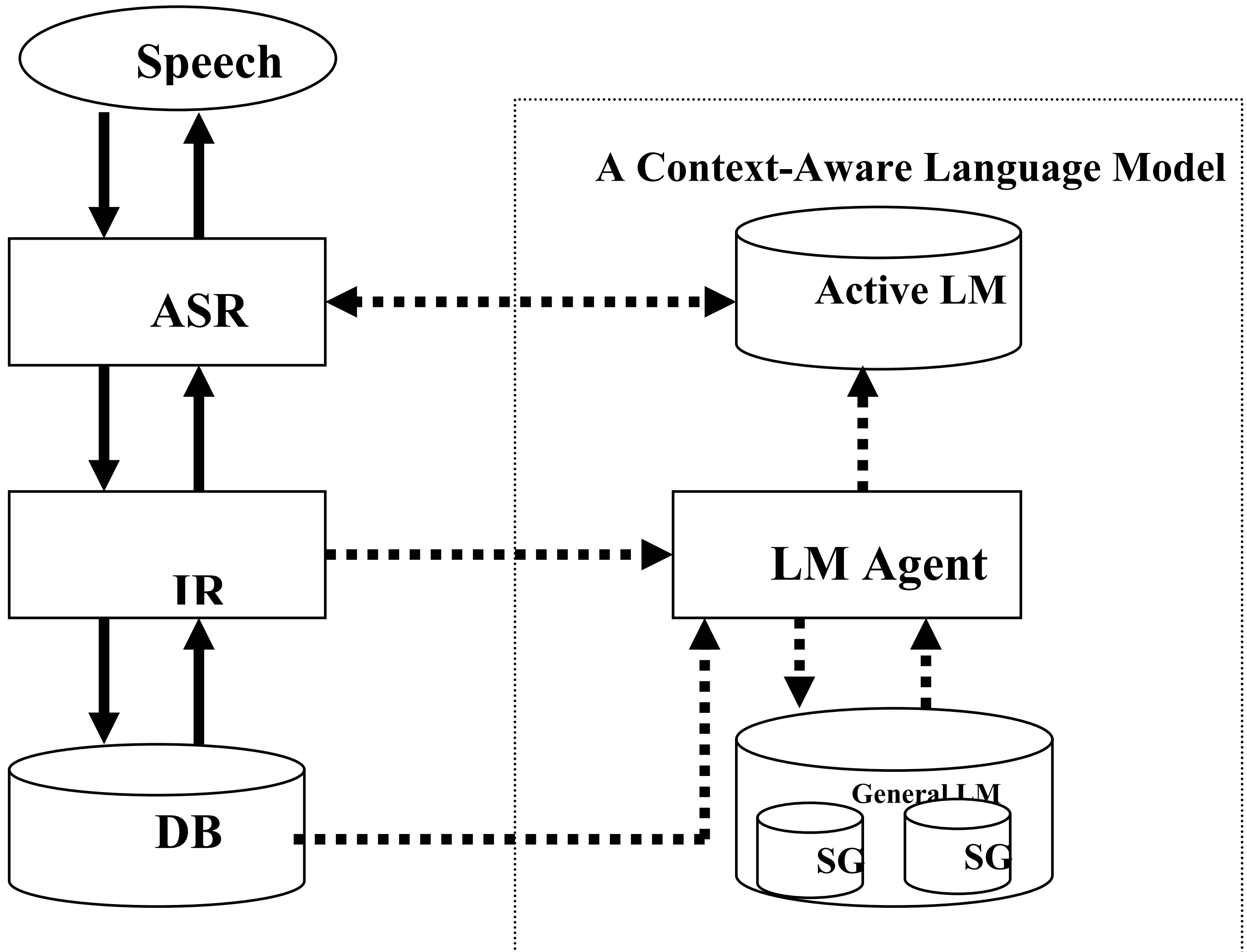
Design Principle

- Integrating ASR systems with existing IR systems, but not "simply combined by the way of input/output protocol" [FII02]
- Language models that will enable both adaptability and scalability so as to satisfy the document retrieval requirements for large databases
- Effective and efficient SQR user interfaces

System Architecture

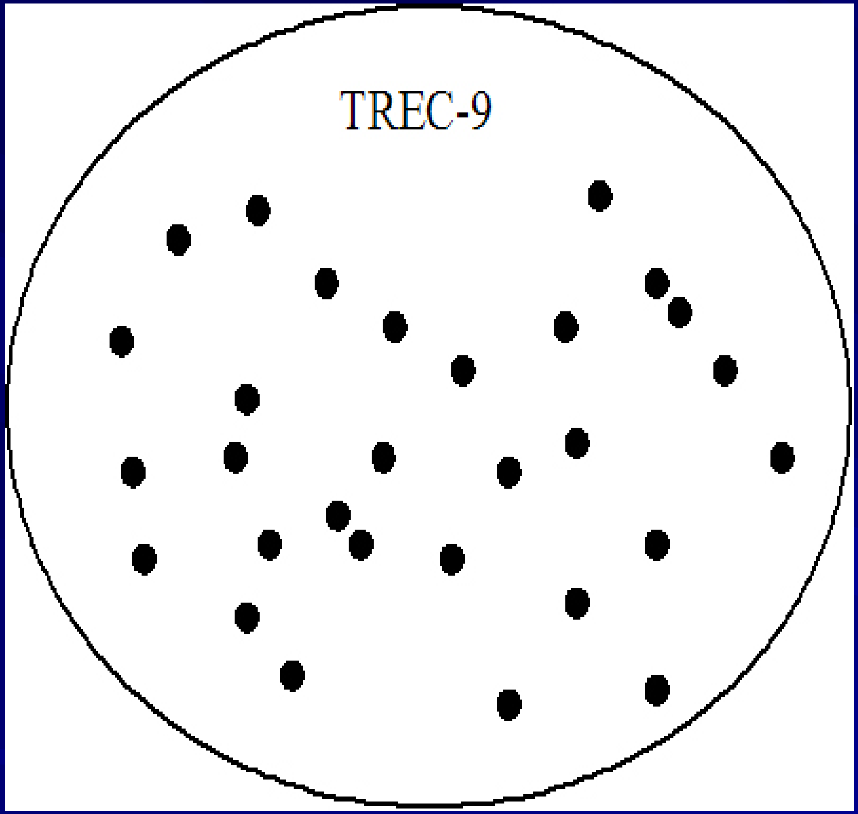


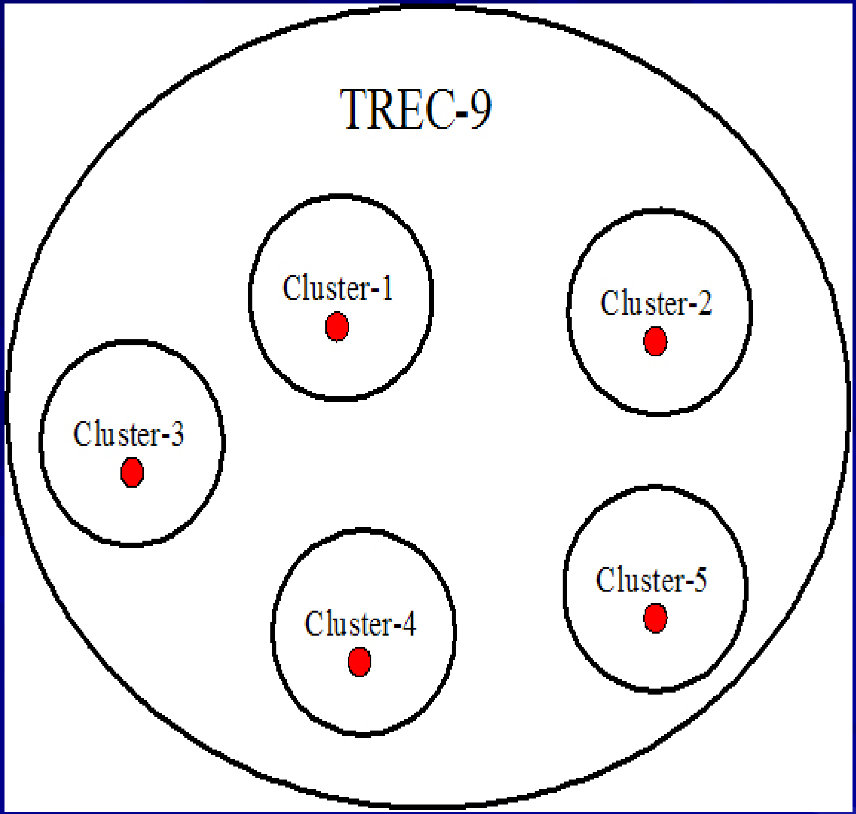
Context-Aware Language Model (CALM)



Procedure to construct CALM

- Preprocess and Index the collected documents
- Represent each document with a vector
- Cluster the collected documents into certain groups
- Represent the center of each group with an important document
- Construct the CALM with a set of important keywords from the centered document


TREC-9


TREC-9
Cluster-1
Cluster-2
Cluster-3
Cluster-4
Cluster-5

Document Clustering Analysis

- Assign a set of documents to the different groups based on their similarity
- closely associated documents tend to be relevant to the same requests
- document clustering should result in more effective, as well as more efficient, retrieval
- Hierarchical and partitioning clustering

Bisecting K-Medioids (BKMdd)

- A medioid representative
- An objective function to control the iterative optimization:

$$J = \sum_{i=1}^{2} \sum_{j=1}^{n} d_{ij}^{2}$$

- A two-phase clustering method

BKMdd

- The bisecting phase:

Set iter = 0;
Repeat
 Compute by using (2.2.1);
 Assign $V^{old} = V$;
 Compute the new medioid set by using $J = \sum_{i=1}^{2} \sum_{j=1}^{n} d_{ij}^{2}$;
Until (= or iter = *MAX_ITER*)

- The K-Medioids phase:

Set K;
Repeat
 Pick a cluster to split by using $\partial(M) = \frac{1}{M} \sum_{j=1}^{M} d_{ij}^{2}$;
 Find two sub-clusters by using Bisecting phase;
Until K

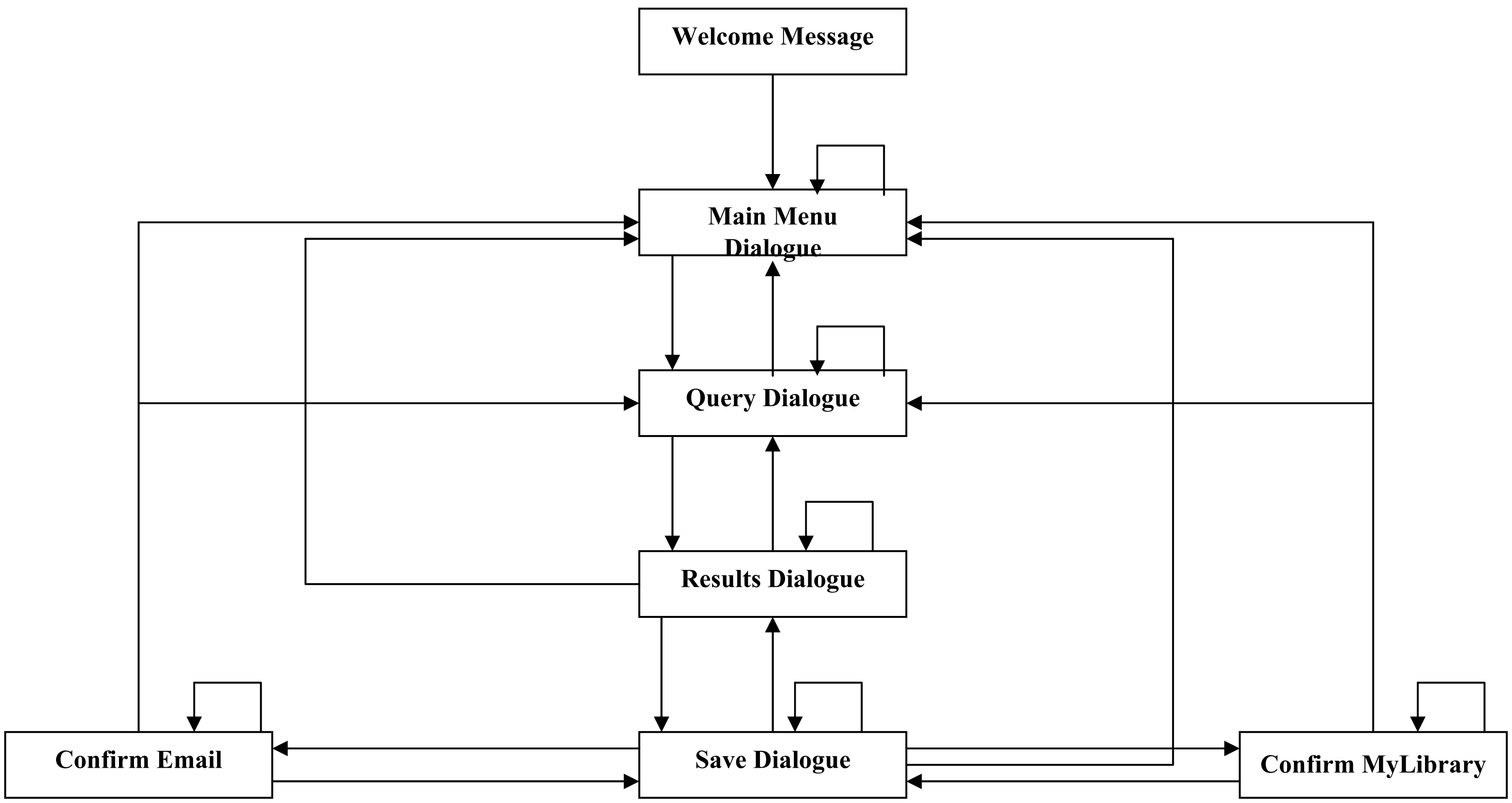
Voice Navigator (VN)

Category	Functions
System	Help
	Main menu
	Try again
	Exit (Goodbye)
Query	By source
	By field
Results	Literal Response
	Cooperative Response
Browsing	Read by title
	Read by abstract
	Previous
	Next
	Repeat
	Stop
	Save

VN Dialogues

- Main Menu Dialogue
- Query Dialogue
- Results Dialogue
- Save Dialogue

Diagram of the VN



Information Verbalization

- The use of computer supported, auditory interactions to amplify understanding of abstract and/or large data
- Literal Response (LR)
- Cooperative Response (CR)
- Mixed Intelligent Response (MIR)
- Cluster-based Intelligent Response (CIR)

MIR Strategy

- MIR

  Combine LR and CR strategies to present all documents in the ranked results set one by one if there is no response from the user. If the system receives any responses from the user, MIR will stop the current presentation immediately, then process the action quickly

CIR Strategy

- The results are clustered before present action
- The results are ranked within each cluster
- The top ranked documents are selected from each cluster
- The number of the documents to be presented is usually five but never beyond nine

Experiments and Research Findings

- Experimental protocol
- Data collection methods
- Participants and procedure
- Evaluation Metrics

Experimental Protocol

- Materials
- Large document databases:

Data Set:	OHSUMED (1988-1991)
Number of documents	348,566
Size of collection in Mb	381.5
Average document length	195
Average document length (unique terms)	111

Data Collection Methods

- Dialogue recordings
- System logs
- User surveys

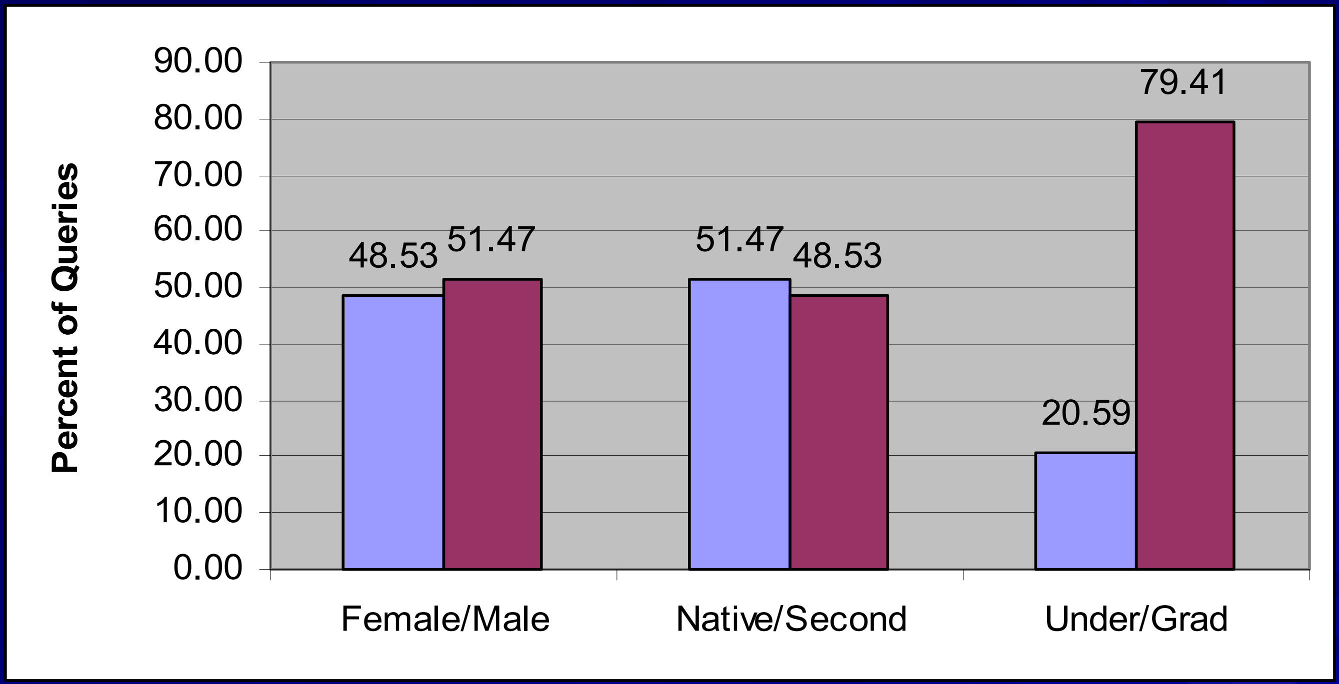
Participants and Procedure

- 39 college level students
- Read the instructions
- Access the system
- Search documents
- Fill out the survey

Evaluation Metrics

- Spoken query metrics
- Task success metrics
- Interface efficiency and quality metrics
- User satisfaction

Spoken Query Metrics

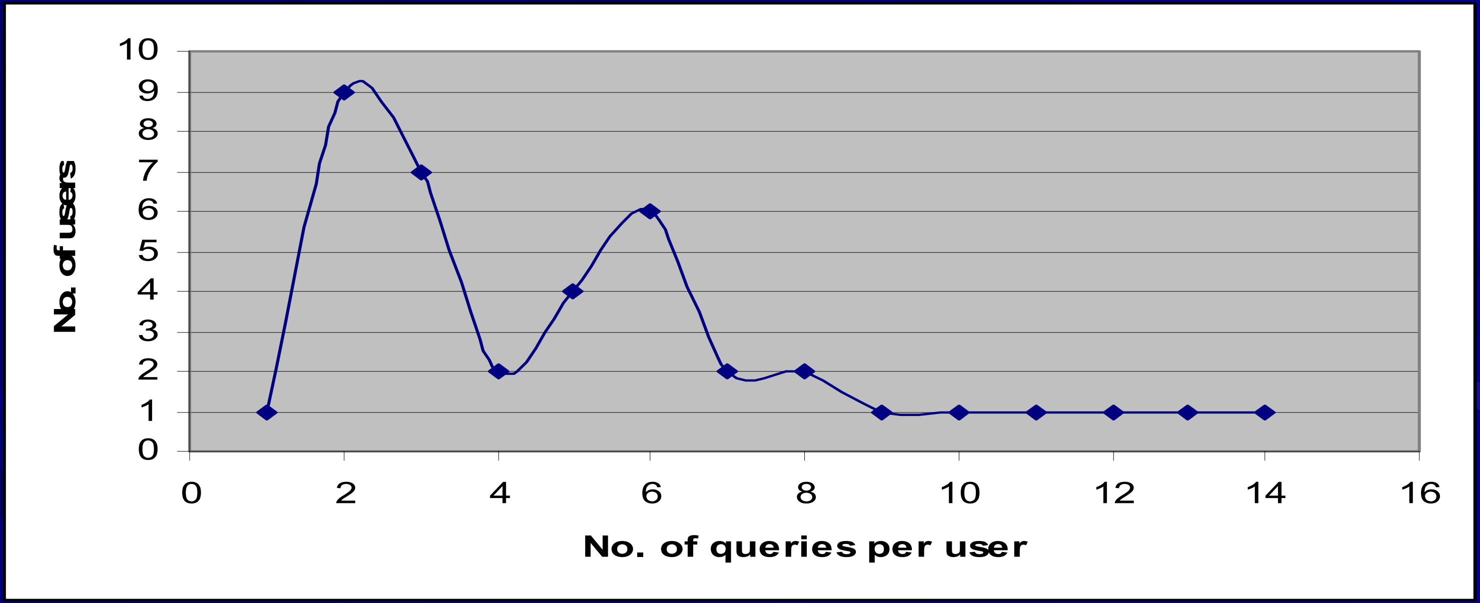


Numbers of Users, Queries, and Terms

Total Number of Participants	39
Total Number of Spoken Queries	203
Average Number of Spoken Queries per User	5.21
Number of Unique Queries	138
Total Number of Spoken Terms	542
Total Number of Uniquely Spoken Terms	99
Mean Number of Terms	2.66

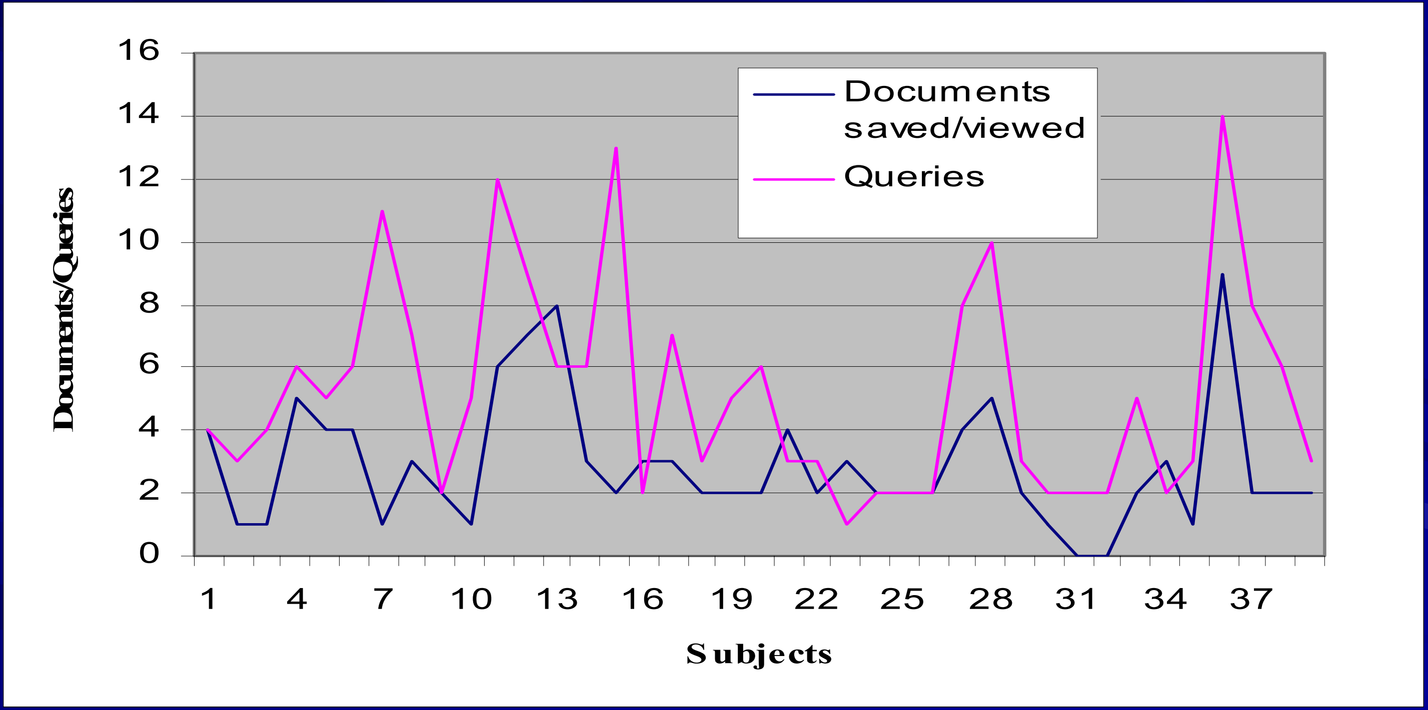
Users by Number of Queries

- SQR: 70% of users more than a single query
- Excite: 67% of users had one and only query
- Spoken query modification was a strong trend



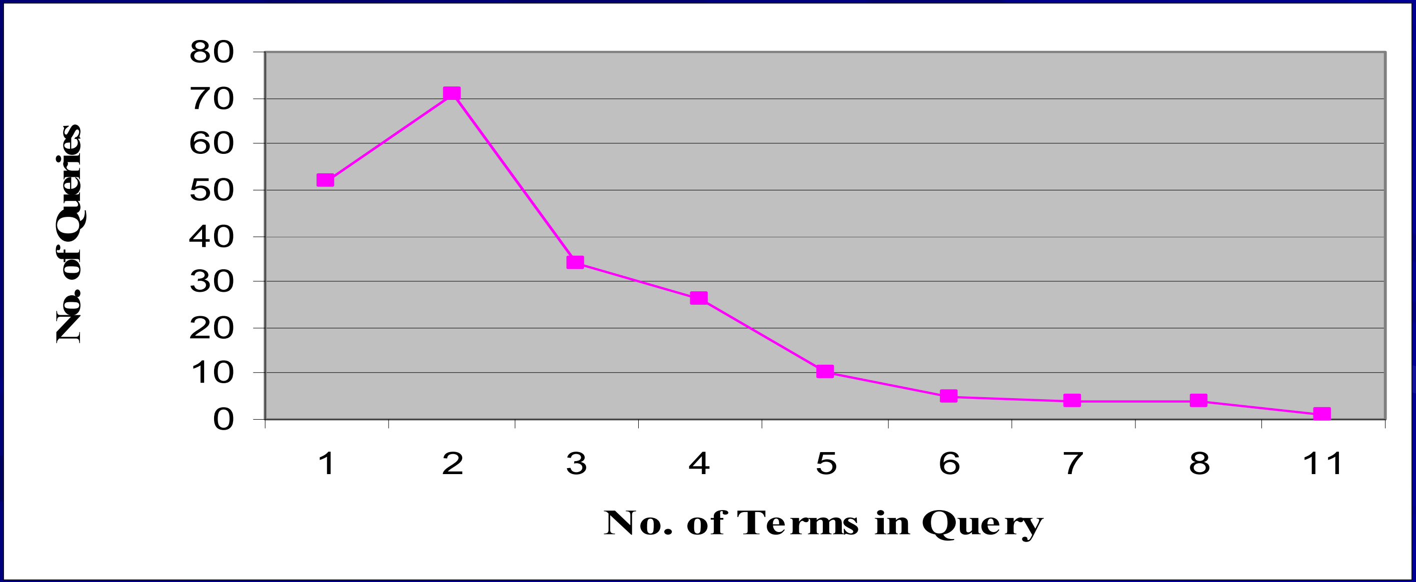
Spoken Queries & Documents Reviewed

- Significant association between the spoken queries and the document reviewed



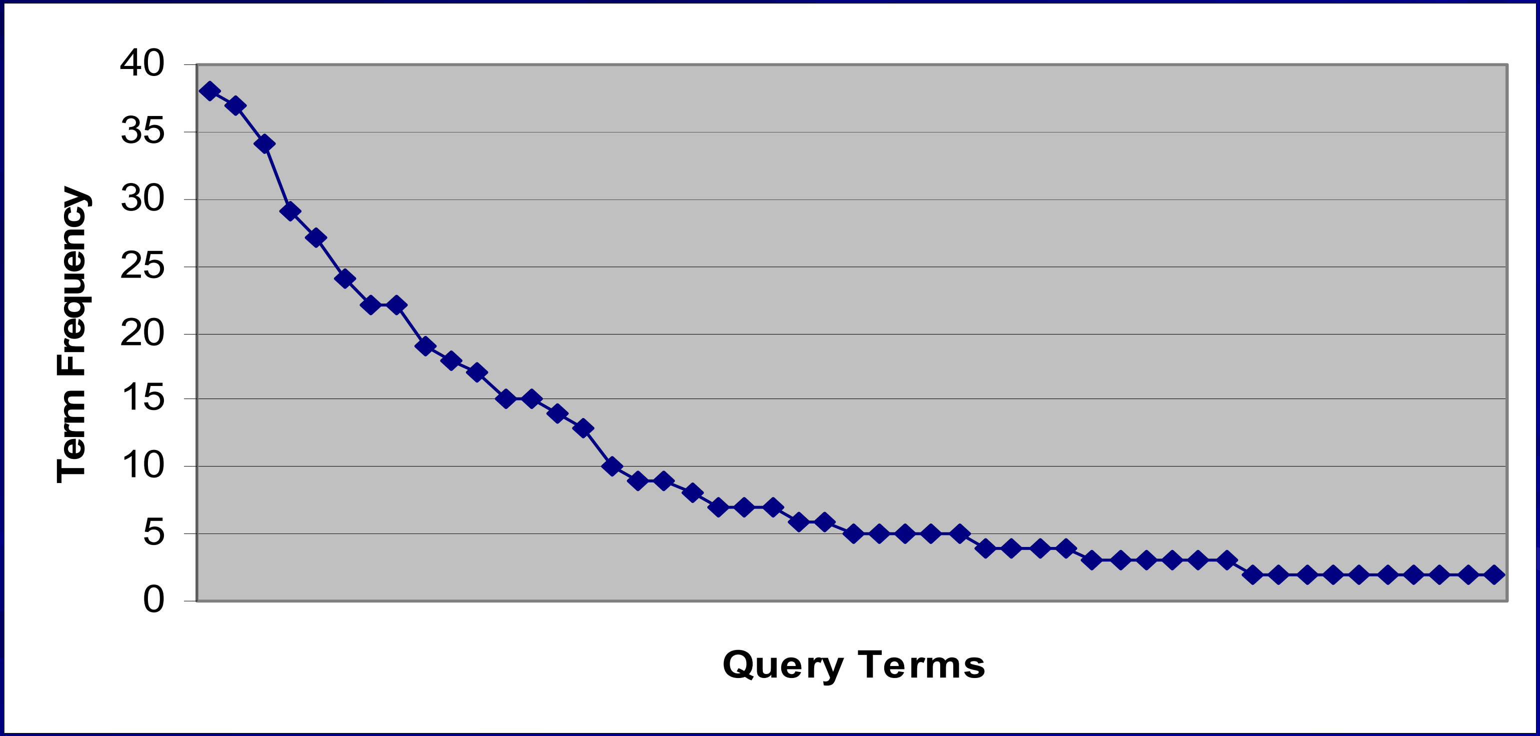
Queries by Number of Terms

- SQR: More than 60% of the queries less than or equal to 2 terms. the mean of terms was 2.66
- Excite: the mean of Web search terms was 2.21



Term Frequency Distribution

- 24 terms covers 82.8% of all queries



The CALM Coverage

- The CALM consisted of 389 terms which represent at least a 93% coverage of all the most important terms found in all documents
- There were 138 unique spoken queries covering 99 unique terms
- There were 8 unique terms that were spoken by participants that did not appear in the CALM
- The experimental coverage of the CALM was 93.10%

Task Success Metrics

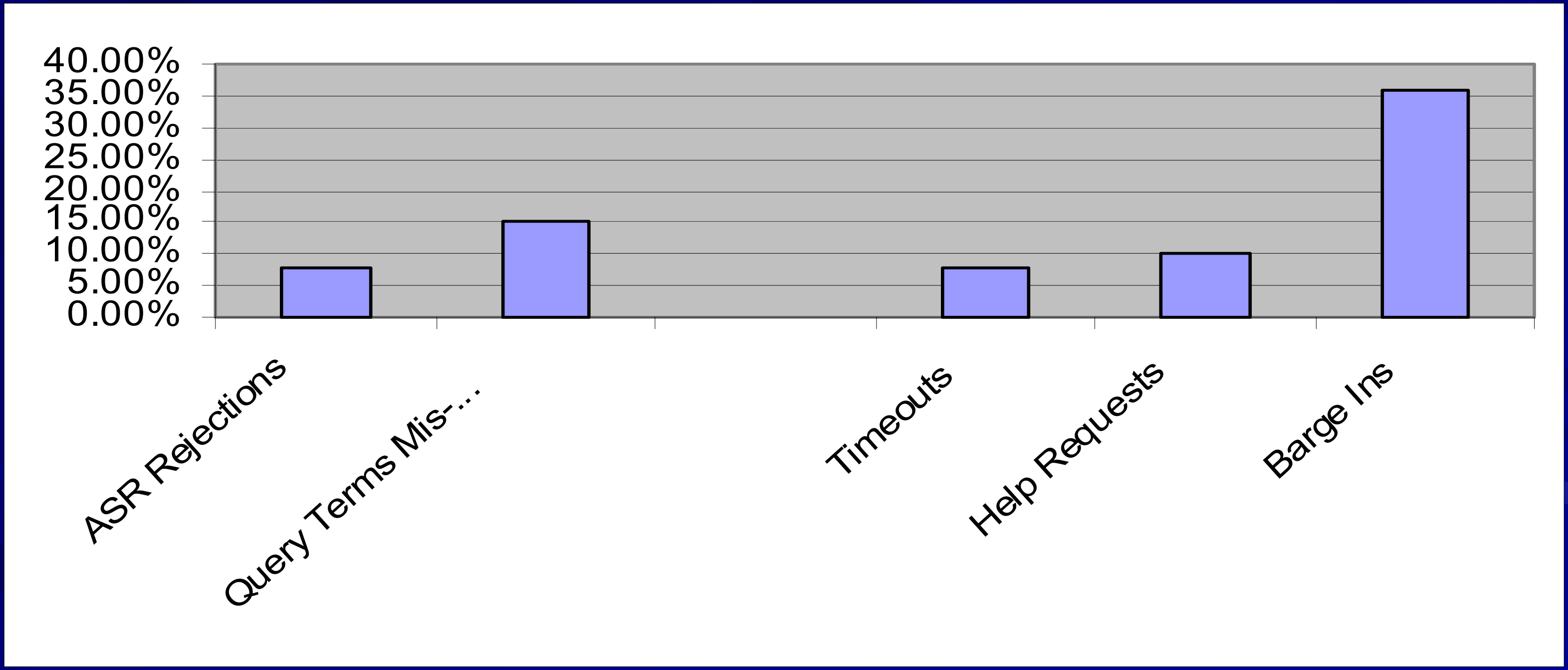
- The mean of Documents Found was 1.54
- Significant as a function of users' experience
- No significant difference between the MIR &CIR strategy
- Significantly negatively correlated to Barge Ins, Query Terms Mis-recognition, and Word Recognition Error (WRE)

Interface Efficiency Metrics

- System Turns and User Turns were positively associated with Query Term Mis-recognitions and Barge Ins
- Elapsed Time was positively associated with Query Term Mis-recognitions, Barge Ins, and Spoken Queries

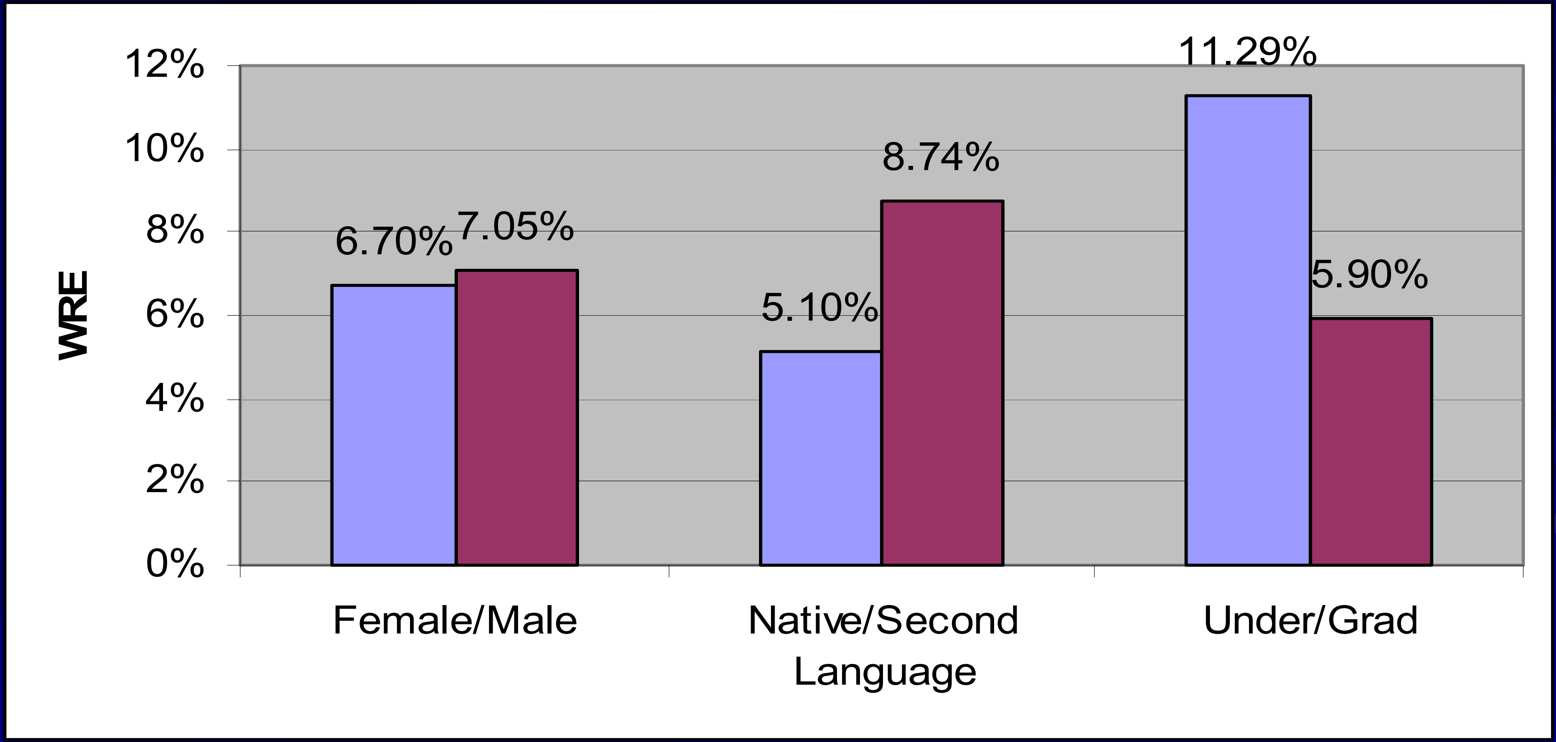
Interface Qualitative Data

- ASR rejections were positively associated with WRE, Barge Ins, and Spoken queries



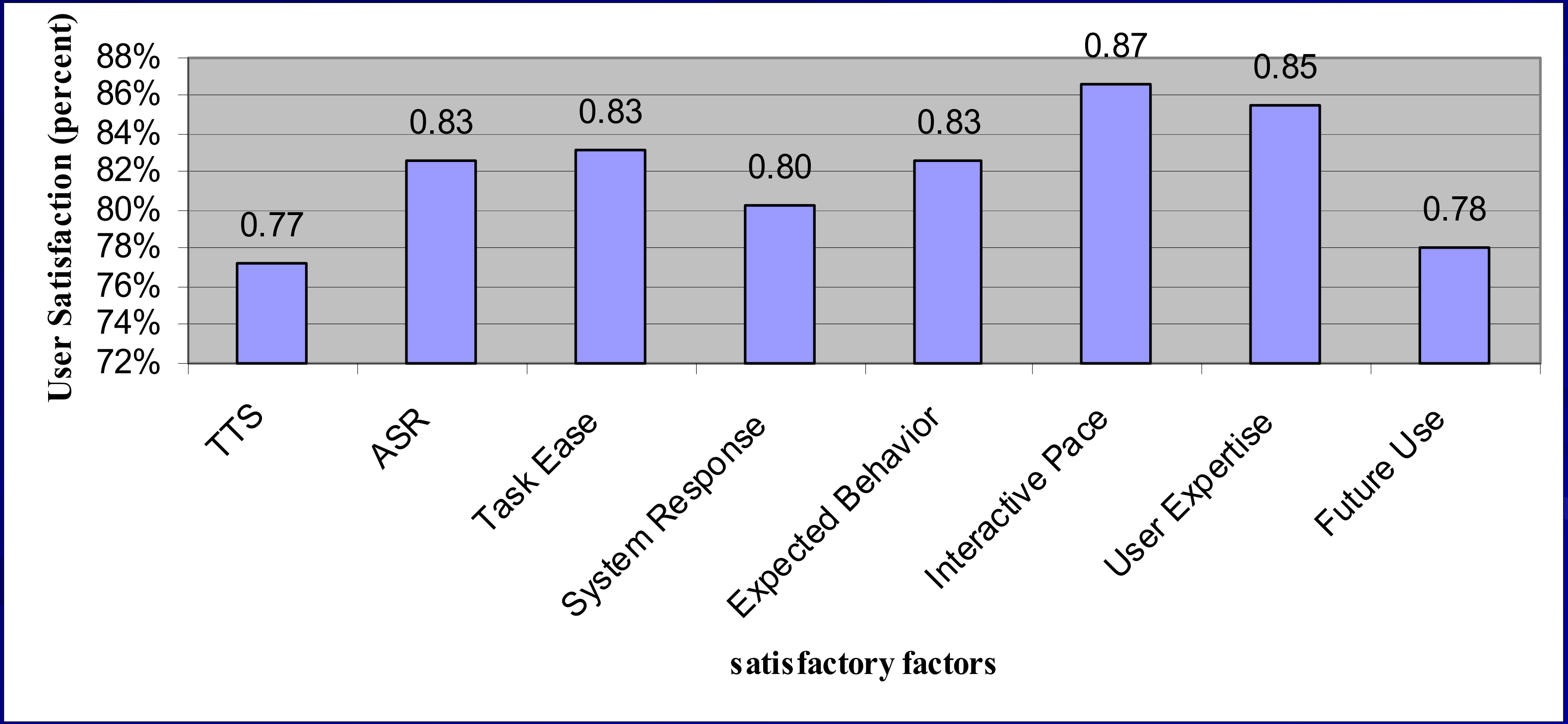
Word Recognition Error (WRE)

- The WRE rate for the entire set of spoken query utterances was 6.87%



User Satisfaction

- The average user satisfaction rating was 82% based on 31 satisfactory factors

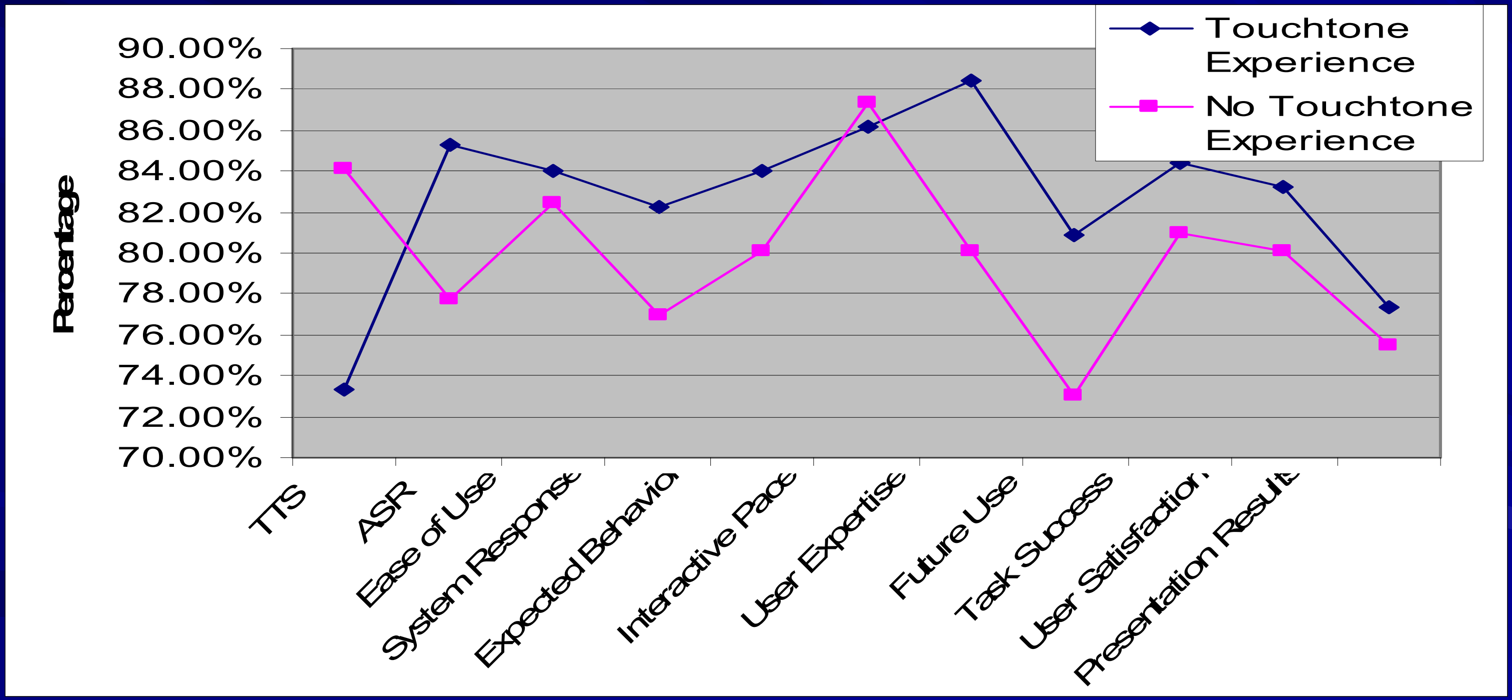


Task Success & Results Presentation Strategies

- Task Success was significantly positively related to Results Presentation Strategy
- No significant difference in the Task Success measure as a function of the Results Presentation Strategy

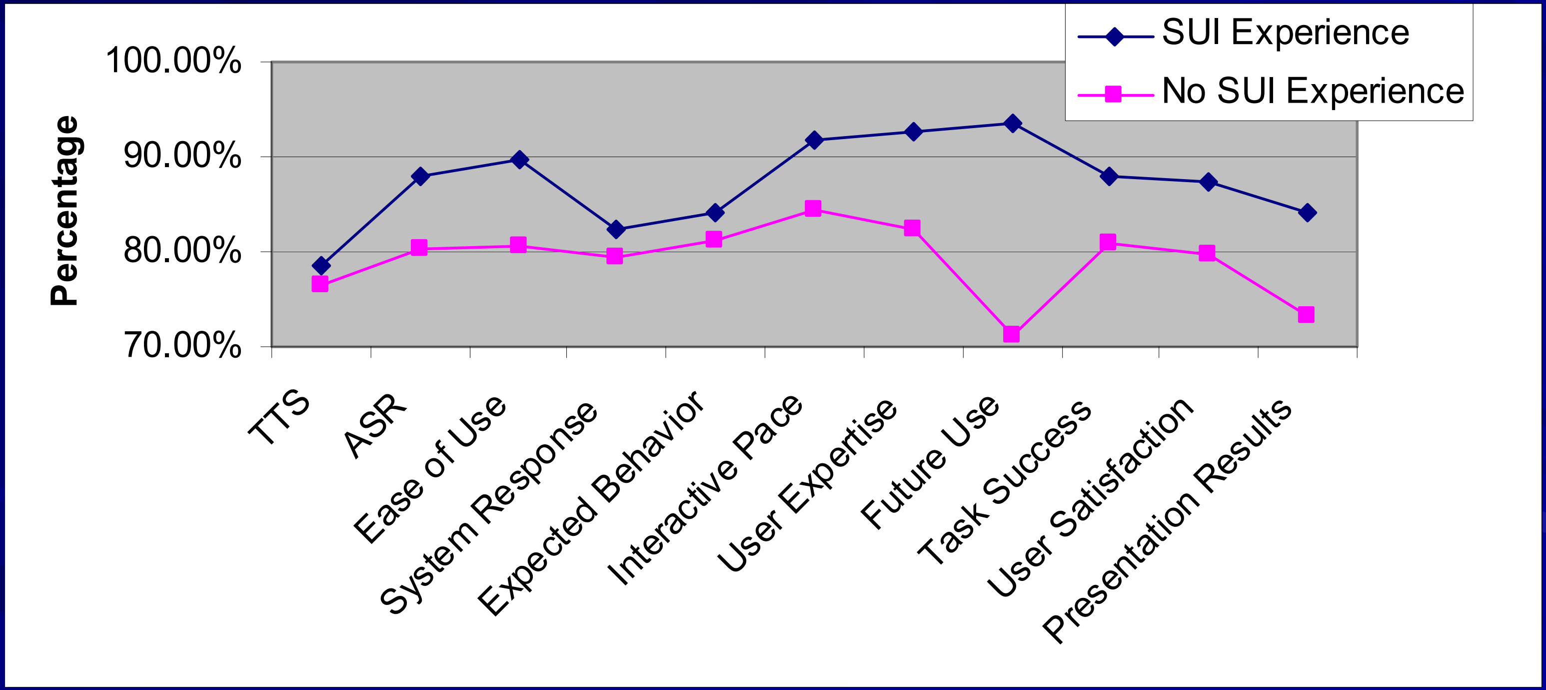
Users' Satisfaction Perceptions with Touchtone Experience

- No significant difference between subjects who had touchtone and no touchtone experience



Users' Satisfaction Ratings with SUI experience

- No significant difference as a function of subject SUI experience



Conclusion

- Achieved a high user satisfaction rating
- Achieved a high Task Success rating
- Performed well with regard to interface efficiency
- Kept its promise of an improved interface quality
- The CALM was developed with a high coverage

Contributions

- Identified and addressed the issues and constraints that the effects of a language model on the ASR performance
- Addressed the issues related to SUI that arise in performing SQR tasks
- Studied and investigated document clustering techniques and VoiceXML technologies

Contributions (Cont'd)

- Proposed an architecture of SQR systems for large document databases
- Defined a Context-Aware Language Model (CALM). A document clustering technique was employed to build such a CALM
- Defined an effective and efficient dialogue framework to facilitate SQR tasks
- Proposed two information verbalization strategies to present the retrieval results.

Future Research

- Study how to combine a visual and verbal user interface to interact with the user to enhance the user's information access
- Document summarization may improve the user's information access
- Temporal and contextual factors may affect the user's information needs

Thanks

- Questions ?
- Comments ?

References

- *[FII02] Fujii, A., Itou, K., and Ishikawa, T. A Method for Open-Vocabulary Speech-Driven Text Retrieval. Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP2002), pp.188-195, July. 2002*
- *[FM02] Franz, Alexander, Milch Brian. Searching the web by voice. Proceedings of the 19th International Conference on Computational Linguistics (COLING), pages 1213-1217, 2002.*