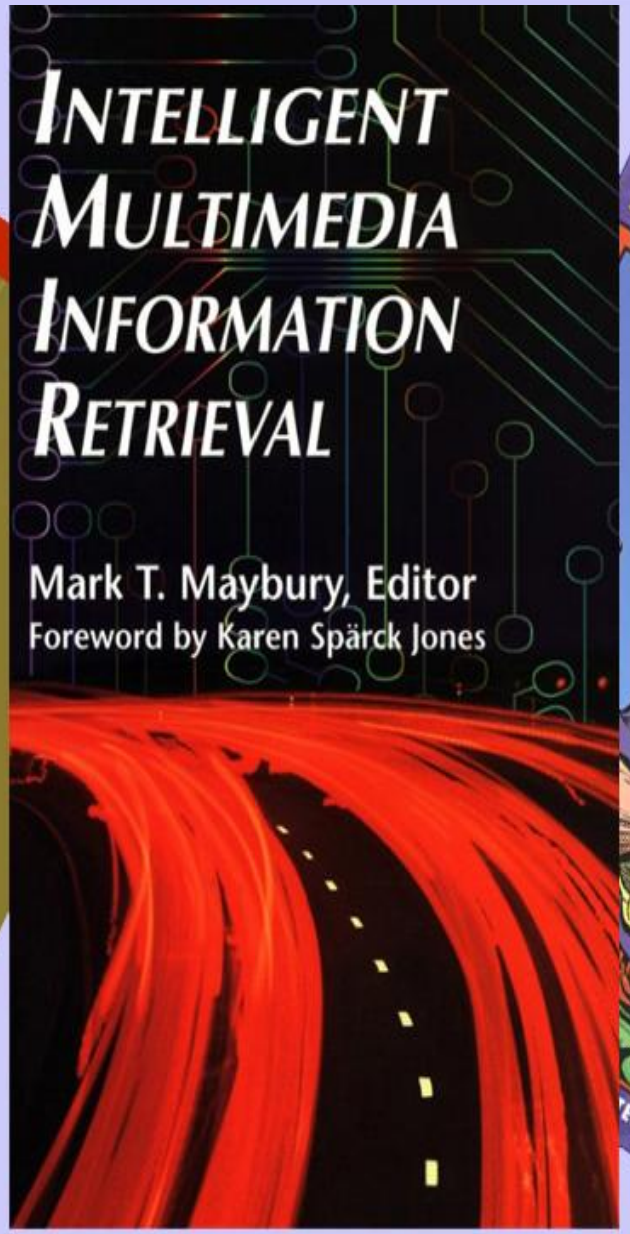
Audient:
An Acoustic Search Engine

By Ted Leath
Supervisor: Prof. Paul Mc Kevitt

School of Computing and Intelligent Systems
Faculty of Engineering
University of Ulster, Magee

Food for Thought



Existing SDR Systems

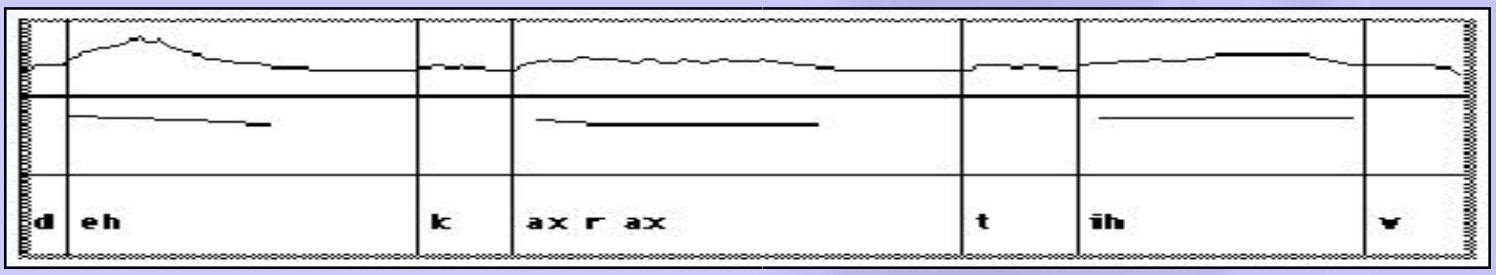
- Involve the production of intermediate text for the purposes of indexing, searching and retrieval
- Require a high level of semantic processing for word recognition
- Have a limited vocabulary
- Have a high word recognition error rate

Things can be done differently!

Non-word Representations of Speech

- Could be features of the audio signal

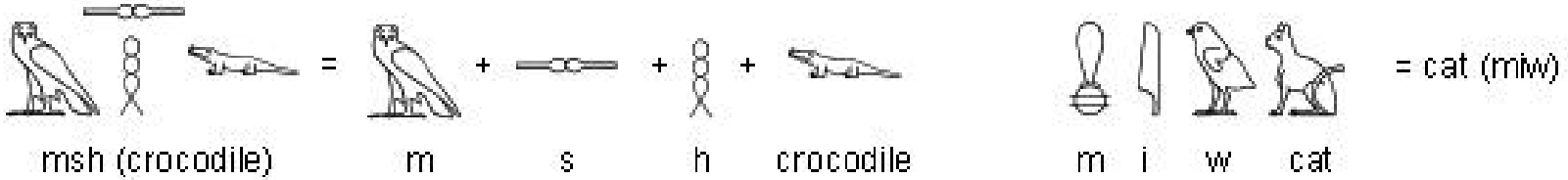
- Could be phonemes



Phonemic and Phonogrammic Streams

Phonogrammic streams are orthographical representations of phonemic streams. This abstraction is ancient, and partially inherent in the English alphabet.

The glyphs have both semantic and phonetic values. For example, the glyph for crocodile is a picture of a crocodile and also represents the sound "msh". When writing the word for crocodile, the Ancient Egyptians combined a picture of a crocodile with the glyphs which spell out "msh". Similarly the hieroglyphs for cat, *miw*, combine the glyphs for m, i and w with a picture of a cat.



Egyptian hieroglyphs with semantic and phonetic value. Ref. http://www.omniglot.com/writing/egyptian.htm

Project Goals

- Create a unique alternative to existing word-based LVCSR speech retrieval systems along with potential tools for future cognitive and philosophical investigation
- Develop a speech-centric model which uses standards-based phonogrammic streams as primary internal data representation
- Allow both text and nonlexical phonemic audio queries of varying length
- Test against audio corpora used in the evaluation of other Information Retrieval (IR) systems

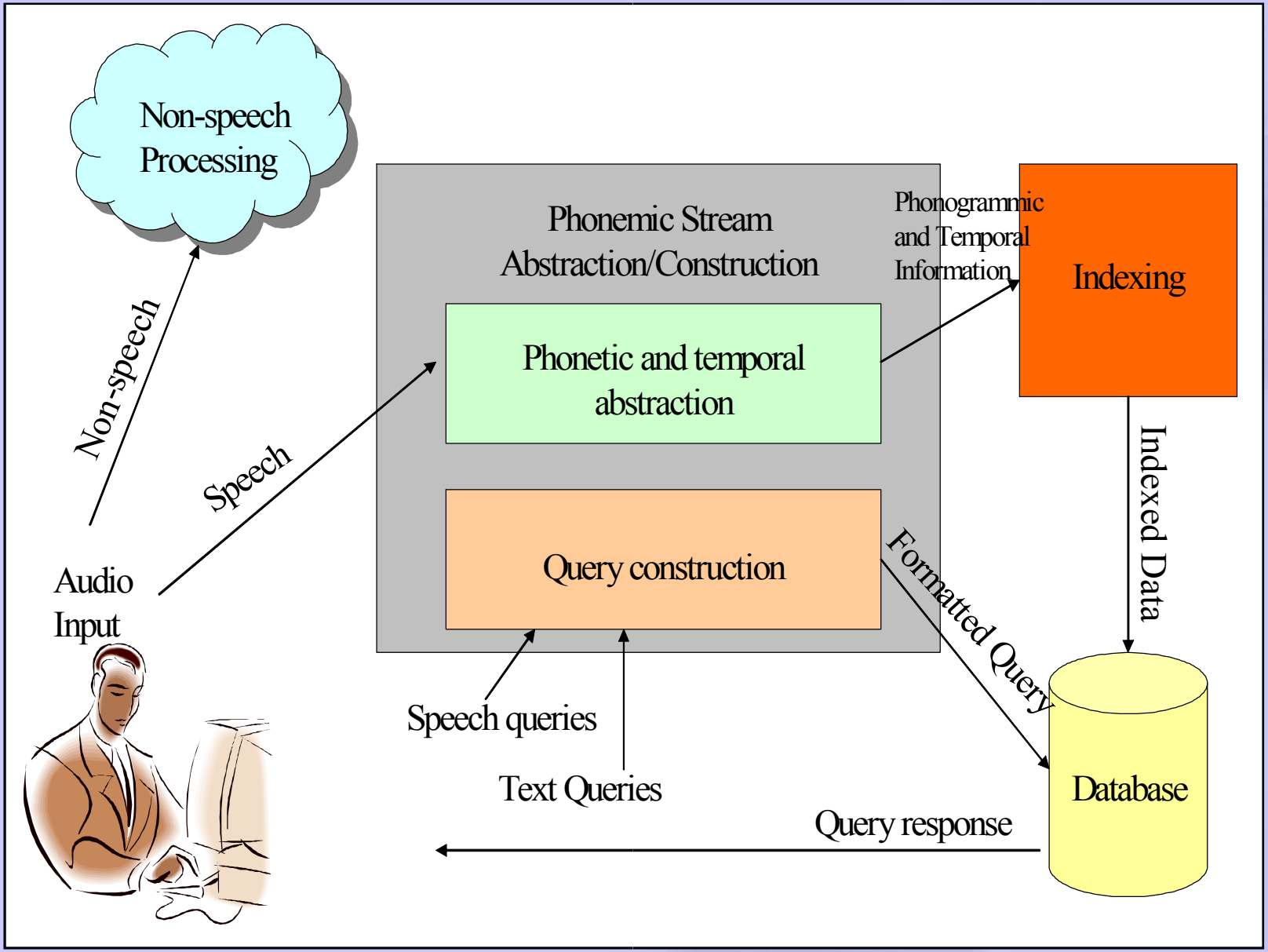
Previous Research/Systems

- TREC
  - The Informedia projects at Carnegie Mellon University
  - The Video Mail Retrieval and Multimedia Document Retrieval projects at Cambridge University
  - The SCAN system at AT&T Research
  - The THISL project at Sheffield University
- SpeechBot and NPR Online – Public Internet Search Sites
- The National Gallery of the Spoken Word
- BBN Rough ‘n’ Ready
- Fast-Talk

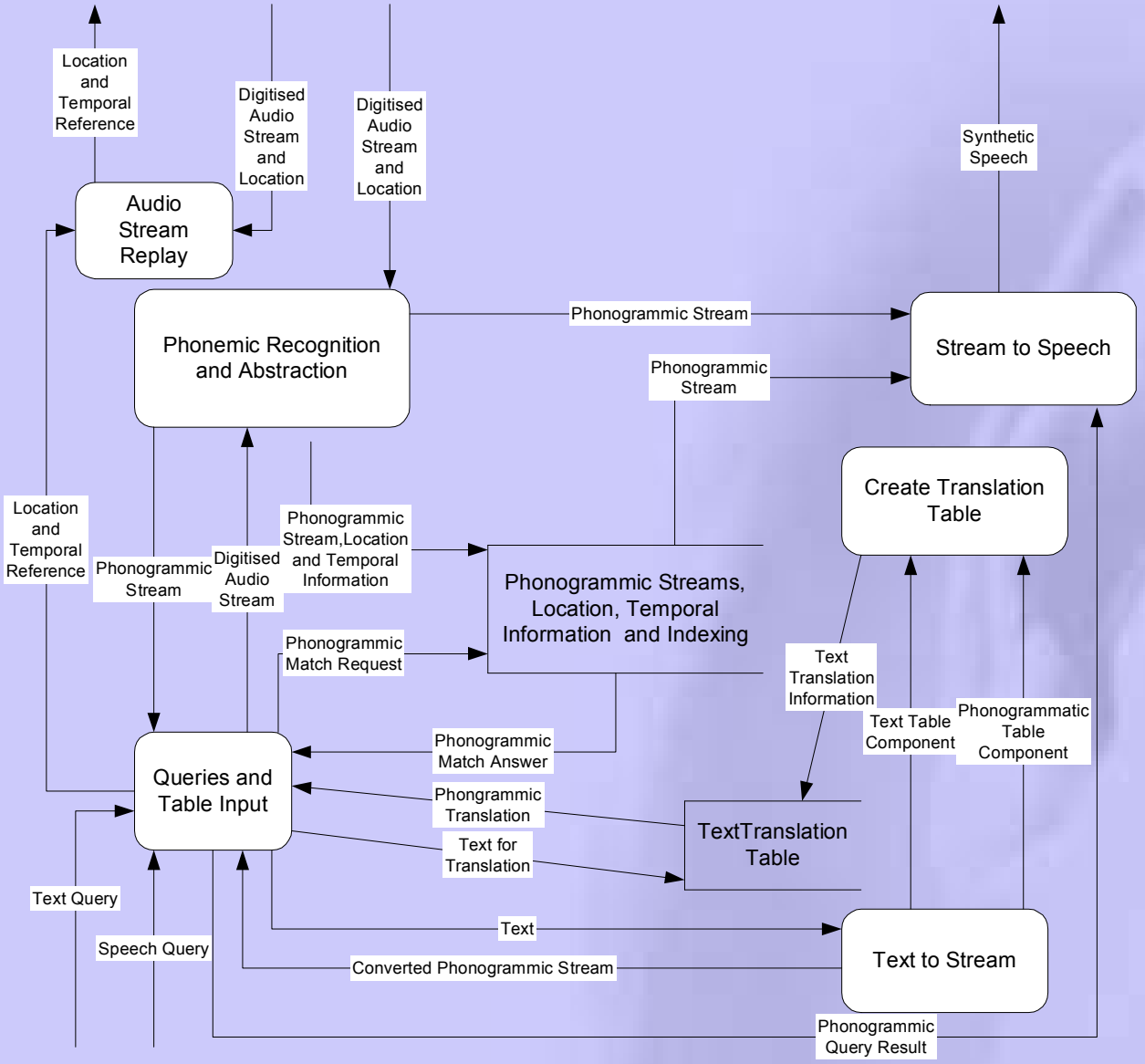
SDR System Comparison Chart

	Does not use LVCSR	Both audio and text queries	Not word-based	Free text searches	No transcript required	Open standard phonoparamic data
Audient	●	●	●	●	●	●
Fast-Talk	●		●	●	●	
Informedia				●	●	
MediaIndexer				●	●	
Multimedia Retrieval Project				●	●	
National Gallery of the Spoken Word	●		●	●	●	
NPR Online	●			●		
SCAN					●	
Rough 'n' Ready				●	●	
SpeechBot				●	●	
THISL		●		●	●	
Video Mail Retrieval Project	●	●	●		●	

Audient System Architecture



Core Modules

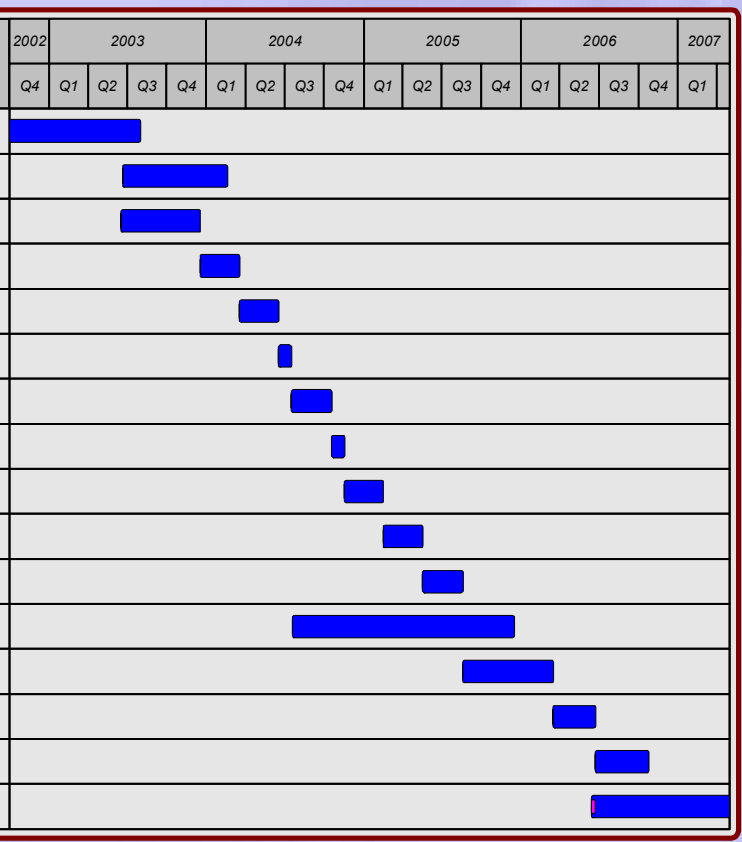


Proposed Tools

- The Hidden Markov Model Toolkit (HTK)
- Linux and C++
- Festival
- VoiceXML and the SGML Family
- The Apache Web Server

Project Schedule

ID	Task Name	Start	End	Duration
1	Literature Survey	01/08/2002	01/08/2003	262d
2	Write up literature review	20/06/2003	19/02/2004	175d
3	Selection, installation and integration of tools	17/06/2003	18/12/2003	133d
4	Construct Phonemic Recognition and Abstraction Module	18/12/2003	18/03/2004	66d
5	Construct Stream to Speech module	18/03/2004	17/06/2004	66d
6	Test and refine modules	17/06/2004	16/07/2004	22d
7	Construct Text to Stream module	16/07/2004	18/10/2004	67d
8	Test and refine modules	18/10/2004	17/11/2004	23d
9	Construct Queries and Table Input module	17/11/2004	15/02/2005	65d
10	Construct Create Translation Table module	15/02/2005	18/05/2005	67d
11	Construct Audio Stream Replay module	18/05/2005	18/08/2005	67d
12	Integrate and test core modules	19/07/2004	16/12/2005	370d
13	Test core modules against other IR systems using corpora and optimise	18/08/2005	17/03/2006	152d
14	Populate index and demonstrate	17/03/2006	22/06/2006	70d
15	Incorporate search engine elements	22/06/2006	25/10/2006	90d
16	Finish thesis	14/06/2006	29/05/2007	250d



Conclusion

- Create a unique alternative to existing word-based LVCSR speech retrieval systems along with potential tools for future cognitive and philosophical investigation
- Develop a speech-centric model which uses standards-based phonogrammic streams as primary internal data representation
- Allow both text and nonlexical phonemic audio queries of varying length
- Test against audio corpora used in the evaluation of other Information Retrieval (IR) systems

Applications

- Searching, indexing and retrieval of Internet audio and video files
- Searching, indexing and retrieval of broadcast media
- Services for the blind
- Library services
- Surveillance and intelligence gathering
- Voice mail
- Audio mining
- Trend analysis (topic detection and tracking)