Audient: An Acoustic Search Engine

Ted Leath

Supervisor: Prof. Paul Mc Kevitt

Research plan. Faculty of Engineering, University of Ulster, Magee, Londonderry

Abstract

Most current Spoken Document Retrieval (SDR) systems involve the production of intermediate text for the purposes of indexing, searching and retrieval. The work described in this research plan proposes *Audient*, an acoustic search engine for both audio and video files with an SDR system core that uses phonogrammic streams derived from phonemic streams for internal data representation, avoiding the time penalties, overheads and errors introduced through the production of intermediate text.

Audient has a wide range of potential indexing, search, retrieval and monitoring applications and also provides tools for philosophical and cognitive investigation. Core modules are to be developed using The Hidden Markov Model Toolkit, Festival Speech Synthesis System, Apache HTTP Server and VoiceXML.

Keywords: spoken document retrieval, information retrieval, audio mining, word spotting, acoustic search engine, phonemic stream, phonogrammic stream

1. Introduction

1.1 Background

Within the broad area of information discovery, there is an ever increasing requirement for an effective means of indexing, searching and retrieving audio information. Since most video material also contains an audio portion, any developments within the audio area also have implications for video indexing, searching and retrieval. Most current Spoken Document Retrieval (SDR) systems involve the production of intermediate text. This intermediate text is either derived from an audio stream via Automatic Speech Recognition (ASR) (Jones et al., 1997), manually transcribed (Takeshita et al. 1997), or partially derived from associated artefacts like metadata, closed captioning (in the case of video with an audio stream) or by other textual annotations (Hauptmann and Witbrock, 1997, Mani et al. 1997, Maybury, 1997).

It may be possible to effectively index, search and retrieve audio material avoiding the time penalties and errors introduced through the ASR phase.

Indexing problems inherent in some current architectures include:

1. Closed vocabulary prohibiting recognition of Out of Vocabulary (OOV) words (particularly proper nouns) and new words. Current Large Vocabulary Continuous Speech Recognition (LVCSR) systems typically have a vocabulary of around 5,000 to 60,000 words (Jurafsky and Martin 2000) while the Oxford English Dictionary currently has in excess of 290,000 entries.

2. Mispronunciation within the audio stream

3. Unintelligible and truncated speech

4. Less success with longer search terms

5. Queries dependent on spelling rather than phonetics

6. Lack of granularity for user determined search parameters

It is envisaged that a speech-centric model, using an abstraction of the phonemic stream rather than text for internal data representation, would address some of these shortcomings. Research suggests that processing of speech is handled differently by humans than non-speech acoustic information (Liberman, 1982). Others are examining the retrieval of non-speech acoustic information like music (e.g. Harvey, 2003, Leman, 2002, Cater and O'Kennedy, 2000) and sound effects. It is not within the scope of this work to address non-speech audio.

1.2 Potential Areas of Application and Further Investigation

1.2.1 Potential Applications

Audient has a wide range of potential indexing, search, retrieval and monitoring applications including Internet audio files, broadcast media, services for the blind, library services, surveillance and intelligence gathering, voice mail, audio mining and trend analysis (topic detection and tracking).

1.2.2 Potential Areas for Philosophical and Cognitive Investigation

Since Audient is modelled in part on what is understood of human speech perception, it has the potential for facilitating research into artificial self-learning systems, philosophical investigations of speech-centric versus text-centric methods, research models for cognitive science and consciousness theories and examination of behaviourist versus cognitive semantic recognition of speech.

Audient may also allow exploration of philosophical views on the differences between verbal, non-verbal and written communication (Palmer, 1997, Powell and Howell, 1996).

1.3 Objectives

The primary aims of Audient are to:

1. Create a unique alternative to existing word-based LVCSR systems.

2. Develop a speech-centric model which uses a standards-based phonogrammic stream for internal data representation.

3. Allow both text and audio queries.

4. Test against audio corpora used in the evaluation of other Information Retrieval (IR) systems.

2. Literature Review

2.1 TREC

The Text REtrieval Conference (TREC) began in 1992, and is jointly sponsored by the National Institute of Standards and Technology (NIST) and the Defense Advanced Research Projects Agency (DARPA) (TREC 2002). Its purpose is to support research within the information retrieval community. TREC conferences run annually and consist of a set of "tracks" - areas of focus in which particular retrieval tasks are defined. One of these tracks which ran from TREC-6 (1997) through TREC-9 (2000) was the Spoken Document Retrieval (SDR) track. Figure 2.1 below represents a typical TREC SDR process (Garfolo et al., 2000).

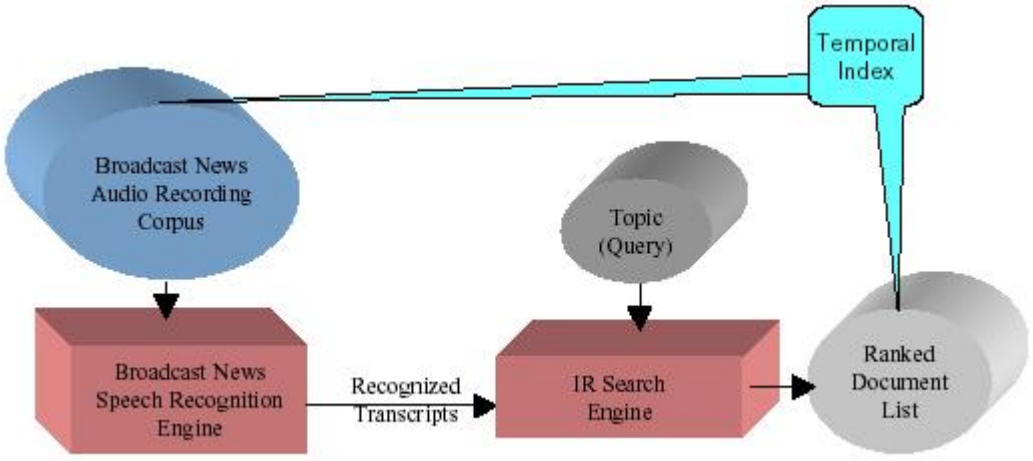

Figure 2.1 A typical TREC SDR process

Several notable research efforts in SDR have been participants in the TREC SDR track, including:

- The Informedia projects at Carnegie Mellon University (Hauptmann and Witbrock, 1997, Informedia, 2003)
- The Video Mail Retrieval and Multimedia Document Retrieval projects at Cambridge University (Jones et al., 1997, Video Mail, 1997, Tuerk et al., 2000)
- The SCAN system at AT&T Research (Choi et al., 1999)
- The THISL project at Sheffield University (Abberley et al., 1999, THISL, 2000)

2.2 SpeechBot and NPR Online – Public Internet Search Sites

SpeechBot (HP SpeechBot, 2003) claims to be the first Internet search site indexing streaming spoken audio on the Web (Quinn, 2000). SpeechBot is currently applied to broadcast radio shows, routinely indexing about 20 shows. Some shows have been indexed as far back as July 1996. There are currently over 6,500 hours of shows indexed. SpeechBot uses ASR to automate the transcription and indexing of audio streams that do not have accompanying manuscripts or other associated descriptive artefacts. Users can interactively search an index of transcribed shows. The transcript that is output by the speech recognition software is rarely an exact match of what was originally spoken.

National Public Radio (NPR) in the USA has manual transcripts for radio shows stretching as far back as 1990. NPR Online's archive search (NPR Archives, 2003) allows retrieval of both textual and audio information through an index derived from these transcripts.

2.3 Digital Libraries Initiative Phase II – The National Gallery of the Spoken Word

The US National Science foundation funded a project beginning in September 1999 called "The National Gallery of the Spoken Word" (NGSW, 2003). The objective of NGSW is to make historically significant voice recordings freely available and easily accessible via the Internet. The estimated total budget is $3,599,989 USD (NSF, 2002) and the project is due for completion at the end of August, 2004. The University of Colorado at Boulder is the key collaborator in the engineering of the NGSW project data storage and retrieval (Hansen et al. 2001). A flow diagram for the search engine is given below in Figure 2.2. The search engine is to be transcript-free, and will be searched by users submitting text search sequences (Hansen et al. 2000).

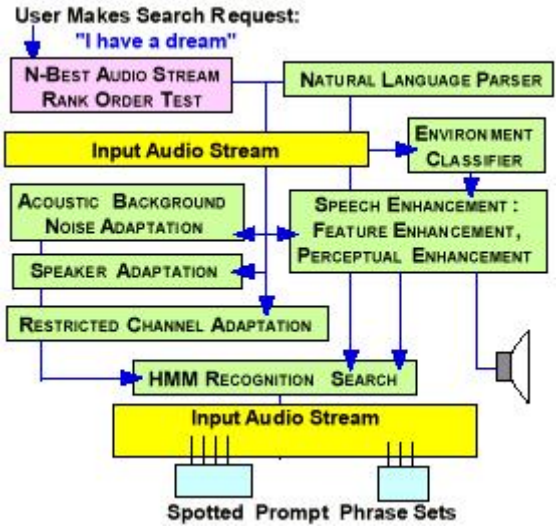

Figure 2.2 Flow diagram of the audio-stream search engine under development for the NGSW

2.4 Commercial Audio Mining Products

Several companies have released commercial audio mining software, and industry observers expect the number of products to increase during the next few years. Currently, accuracy levels are relatively low and some products expensive with high-end software packages costing in excess of $100,000 US dollars for full scale deployment (Leavitt, 2002).

2.4.1 BBN Rough 'n' Ready

The BBN *Rough'n'Ready* system (Rough'n'Ready, 2003) produces rough transcriptions of audio files using large-vocabulary speech recognition, topic spotting and relationship extraction. Transcriptions include the following features:

- Segmented continuous audio input into stories, passages, or sections based on topic
- Speaker identification

- Text transcript
- Information denoting the speakers designation within the organisation
- Indexing by speaker, topic, or concept

This then allows for text-based approaches to information extraction and retrieval.

2.4.2 Fast-Talk

Most commercial audio mining solutions (as in current research systems) use intermediate text for the purposes of indexing, searching and retrieval. An exception is the Fast-Talk system from Fast-Talk Inc. (Clements et al., 2001a) which is referred to as a "phonetic search engine". It does not employ intermediate text, but rather uses an approach called "high-speed phonetic searching" (Clements et al., 2001b). In Fast-Talk a "search track" is created in the pre-processing phase. This is comprised of a highly compressed, proprietary representation of the phonetic content of the original digitised speech.

2.4.3 ScanSoft Dragon MediaIndexer

The Dragon MediaIndexer creates an XML speech index of spoken content using ASR while simultaneously creating a streamable, encoded version of the content in real time (MediaIndexer, 2003). A related product is the ScanSoft Audio Mining Development System which includes an SDK and other tools for developers (AudioMining, 2003).

2.5 Comparison of SDR Systems

While the previously mentioned spoken document retrieval systems are functionally and contextually disparate, Figure 2.3 below attempts to compare them with some of the features of Audient. The following features are compared:

- **Does not use LVCSR** – indicates that audio input is not evaluated using LVCSR and producing an automated speech recognition transcript. Use of LVCSR implies word-based data representation.
- **Both audio and text queries** – evaluates whether only textual queries or both audio and text mode queries are allowed. In this context, WYSIWYG graphics interfaces are also considered as textual queries since direct manipulation within these environments ultimately produces text.
- **Not word-based** – indicates that the system does not use a lexical approach to abstraction, indexing and retrieval with words as minimal units of evaluation. Among those few systems that are not word-based, most use sub-word units for evaluation.
- **Free text searches** – denotes whether free text queries may be performed, or whether queries are restricted to keywords only.
- **No transcript required** – indicates that the system does not require a transcript of the audio information in advance of operation.
- **Open standard phonogrammic data** – shows whether audio processing results in a phonogrammic standards-based data format.

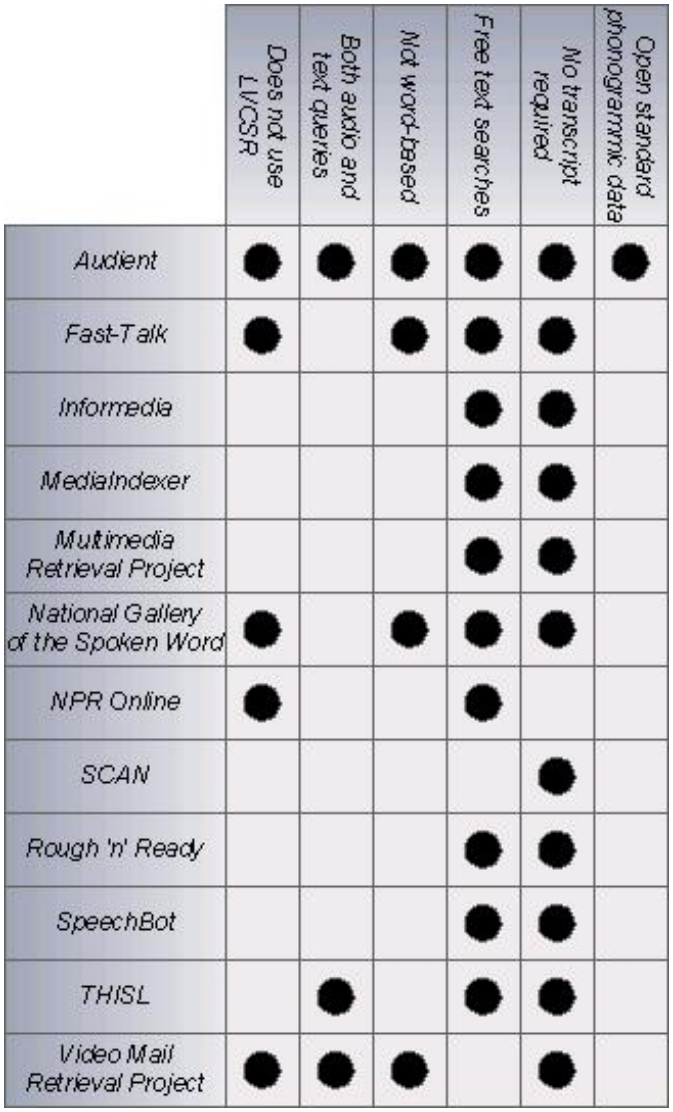

Figure 2.3 SDR system comparison chart

Of those SDR systems compared, only 3 other than Audient are not word-based. The proposed phrase recognition component of the National Gallery of the Spoken Word has yet to be fully developed. It would seem that the phone lattice approach utilised in the Video Mail Retrieval Project was abandoned in favour of a word-based approach in the later Multimedia Retrieval Project. Fast-Talk is the system most architecturally similar to Audient of those SDR systems compared. There are several differences in approach between Fast-Talk and Audient, but one of the primary differences is the proposed open standard VoiceXML phonogrammic streams to be used as internal data representation for Audient. Another difference is Audient's allowance for multi-modal queries.

While not implemented in a full-featured SDR system, relevant research also exists comparing the effectiveness of different sub-word units in SDR – phone n-grams (phone sequences from 1 to 5 phones long), broad phonetic class sequences, phone multigrams (non-overlapping, variable length phonetic sequences) and syllables (Ng, 2000).

3. Project Proposal

The project involves the design, construction and integration of core modules toward the development of a search engine for digitised audio streams (initially optimised for spoken English). Each of these core modules is outlined below along with data flow diagrams. Figure 3.1 below shows the proposed system architecture for Audient's core modules. Figure 3.2 is a top level context diagram for the core modules of Audient and Figure 3.3 is a more detailed level 1 data flow diagram showing the interaction of the core modules.

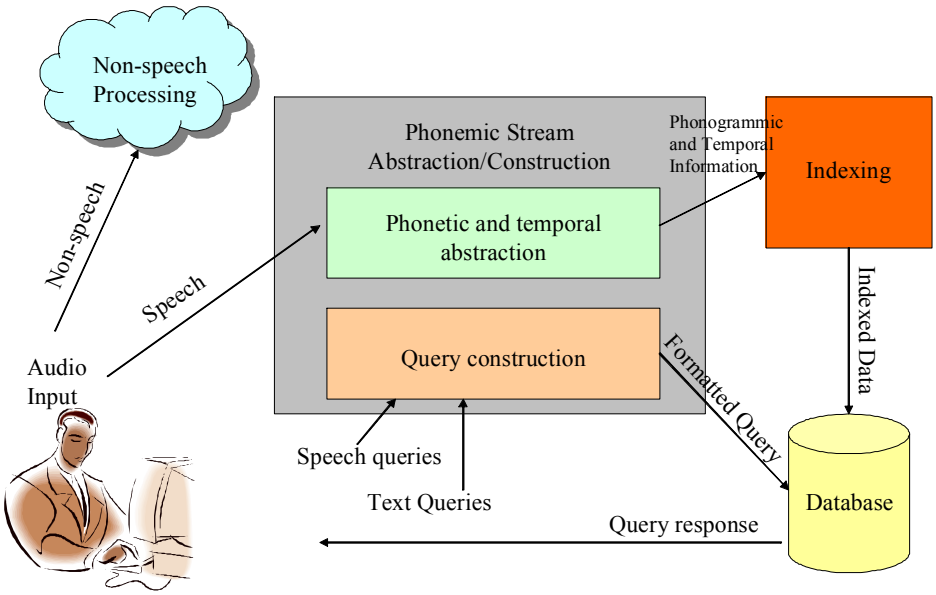

Figure 3.1 System architecture of Audient core modules

- **Phonemic Recognition and Abstraction Module**: Module for the conversion, abstraction and storage of digitised audio streams into abstracted phonogrammic streams with associated temporal information (the possibility of capturing prosodic information will also be investigated). Phonogrammic streams will be orthographical representations of phonemic streams.

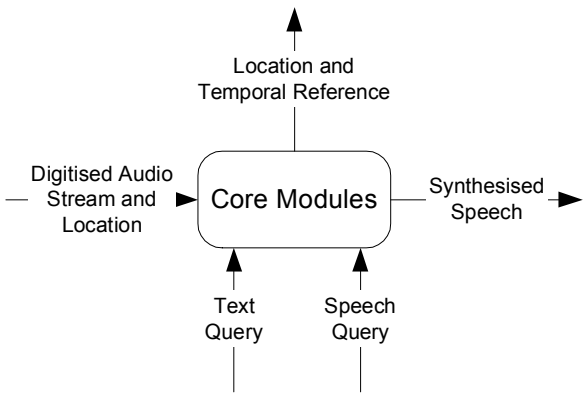

Figure 3.2 Context diagram of core modules

It is desirable to construct phonogrammic streams with the minimal amount of semantic and syntactic interpretation, modelling a behavioural "first pass" type of recognition (unconscious perception and intelligent action). However, semantic and syntactic evaluation may be necessary to achieve acceptable levels of accuracy. This could be thought of as modelling what Dennett refers to as the "Multiple Drafts" model of consciousness (Dennett, 1991) in which speech comprehension seems to occur in a continuous temporal stream, but is actually being revised imperceptibly. A sub-task of this module is the definition of the internal data structures required for abstraction, storage and effective indexing.

- **Stream to Speech Module**: Module to produce synthesised speech from phonogrammic streams. For the purposes of this research, this module is required in the first instance for the development of the Phonemic Recognition and Abstraction Module. It is planned that full advantage be taken of the human aptitude for the evaluation of speech in fine-tuning this module. Later, the Stream to Speech Module should provide the output for query results.

  For the purposes of future research, these first 2 modules should allow for a kind of computer "parrot" – the computer having the ability to "hear" spoken audio information, and repeat the information in a synthesised voice.

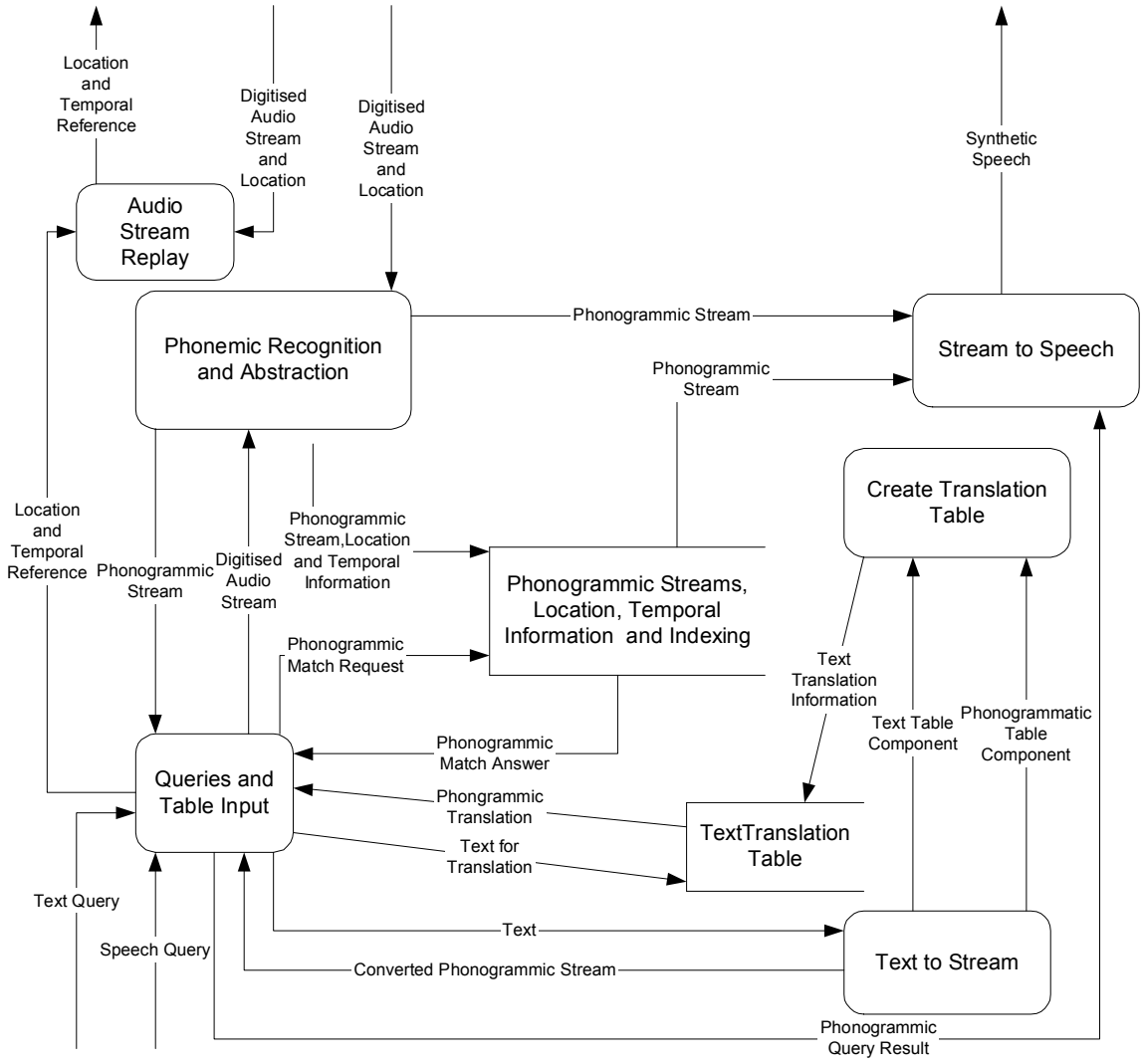

Figure 3.3 Level 1 Data Flow Diagram of Core Modules

- **Text to Stream Module**: Module for producing phonogrammic streams from plain text incorporating Text to Speech (TTS) conversion tools. This will be used for text queries and provide input for another module for the automated production of a table pairing text search terms and keywords with their phonogrammic translation.

- **Queries and Table Input Module**: Module to service queries originating in either textual or spoken form by reducing them to phonogrammic stream segments, accessing storage and presenting query results to the user. This module also provides text input to populate the Text Translation Table.

- **Audio Stream Replay Module**: Module to fetch audio files, and to replay files from specific temporal reference points.

- **Create Translation Table Module**: Module to create pairs of text with their phonogrammic equivalents for the Text Translation Table.

Creation and integration of these modules provides the core functions for Audient.

Having created and integrated the modules providing the core functions, modules will be tested. The first phase of testing will examine the efficacy of the conversion and abstraction functions. It is proposed that this be examined as follows:

- Audio streams with existing transcripts should be abstracted and converted to phonogrammic streams.

- These streams should then be output via TTS to a human listener who will transcribe the synthesised speech word by word.

- The resulting transcription should be compared to the original transcription for accuracy.

The next phase of testing will involve Information Retrieval (IR) functions being tested against audio corpora used in the evaluation of other IR systems. Iterative testing results will be compared throughout, and where possible, modules will be improved and optimised. Finally, search engine crawler elements are then to be integrated with the core functions, and features and interface further refined.

3.1 Prospective Project Tools and Technologies

3.1.1 The Hidden Markov Model Toolkit (HTK)

The Hidden Markov Model Toolkit (HTK) (HTK, 2003) was originally developed at the Speech Vision and Robotics Group of the Cambridge University Engineering Department and contains a set of library modules and tools available in C source form used primarily for speech recognition research. It is anticipated that HTK be used at least for phone level transcription (Young, 1994). Editing and re-estimation tools exist within HTK, and while these may prove useful and/or necessary, it is desirable that Audient's phonemic transcription and abstraction be as context-free as is possible.

3.1.2 Linux and C++

While it should be possible to build all of the main HTK tools on any machine supporting ANSI C and either X-Windows or MS-Windows, it is currently planned that the Linux operating system be used, along with either the Intel C++ or GNU C++ compilers. Both of

these compilers are ANSI compliant. The Linux operating system is a very rich environment for systems integration. Being open source, it is also very accessible, and has extensive X-Windows management and development facilities.

3.1.3 Festival

Festival is a speech synthesis system developed at The Centre for Speech Technology Research, University of Edinburgh (Festival, 2003). Festival offers a full text to speech capability. It is written in C++. This tool is to be used most fully in the Stream to Speech core module of Audient, and in the Text to Stream module where text input is converted to phonemic representation.

3.1.4 VoiceXML and the SGML Family

Significant compression should be achieved from the abstraction of phonemic and temporal information from the spectral features of the initial audio stream. Phonemic information is to be translated into a phonogrammic stream, preferably in an existing non-proprietary, standards-based form. VoiceXML contains elements from the Speech Synthesis Markup Language (SSML) which allows for the encoding of phonemic, prosodic and other information relating to speech synthesis (VoiceXML, 2003) which may be suitable for these purposes. The use of VoiceXML should allow the leveraging of currently available software, particularly with regard to browsing and speech synthesis elements of the project.

3.1.5 The Apache Web Server

The Apache HTTP Server Project is a collaborative software development effort which has created an efficient and extensible HTTP server whose source code is freely available (Apache Web Server, 2003). The project is jointly managed by a group of volunteers located around the world, using the Internet and the Web to communicate, plan, and develop the server and its related documentation. The Apache HTTP Server is currently the most widely used HTTP server in the world.

After the development of modules for the core functions of Audient, it will be necessary to allow users to interface with the modules. The Apache HTTP Server will provide the engine for this interface.

4. Project Schedule

The work proposed requires several tasks to be undertaken in order to achieve the objectives. Table 4.1 in Appendix A outlines the main tasks and schedule of the project.

5. Conclusion

In conclusion, the objectives of Audient are to:

- 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

Potential applications include:

- 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)

References

Abberley D., D. Kirby, S. Renals and T. Robinson (1999) "The THISL Broadcast News Retrieval System" In Proc. of ECSA (European Speech Communication Association) ETRW (ESCA Tutorial and Research Workshop) on Accessing Information in Spoken Audio, 14 – 19, Cambridge, U.K.

Apache Web Server (2003) "Welcome! - The Apache HTTP Server Project" http://httpd.apache.org/

AudioMining (2003) ScanSoft – "ScanSoft - AudioMining Development System" http://www.scansoft.com/audiomining/developers/

Cater, A. and N. O'Kennedy (2000) "You hum it, and I'll play it" In Proc. of the 11th Conference on Artificial Intelligence and Cognitive Science, Galway, Ireland.

Choi, J., D. Hindle, F. Pereira, A. Singhal and S. Whittaker (1999) "Spoken Content-Based Audio Navigation (SCAN)" In Proc. Of the ICPhS-99 (International Congress of Phonetics Sciences).

Clements, M., P. S. Cardillo, M. S. Miller (2001a) "Phonetic Searching vs. LVCSR: How to Find What You Really Want in Audio Archives" In 20th Annual AVIOS (Applied Voice Input/Output Society) Conference, San Jose, California, USA.

Clements, M., P. S. Cardillo, M. S. Miller (2001b) "Phonetic Searching of Digital Audio" In 2001 Broadcast Engineering Conference Proceedings, Las Vegas, Nevada, USA.

Dennett, D. C. (1991) *Consciousness Explained*, London, England: Penguin Books Ltd.

Festival (2003) "The Festival Speech Synthesis System" http://www.cstr.ed.ac.uk/projects/festival/

Garfolo, J. S., G. P. Auzanne and E. M. Voorhees (2000) "The TREC Spoken Document Retrieval Track: A Success Story" In Proc. of the Eighth Text REtrieval Conference (TREC-8), E. M. Voorhees and D. K. Harman (Eds.), 107 – 130, Gaithersburg, Maryland, USA.

Hansen, J., B. Zhou, M. Akbacak, R. Sarikaya and B. Pellom (2000) "Audio Stream Phrase Recognition for a National Gallery of the Spoken Word: "One Small Step"" In Proc. of International Conference on Spoken Language Processing (ICSLP-2000), 1089 – 1092, Beijing, China.

Hansen, J., J.R. Deller and M. Seadle (2001) "Engineering Challenges in the Creation of a National Gallery of the Spoken Word: Transcript-Free Search of Audio Archives" In Proc. of the IEEE and ACM JCDL-2001: Joint Conference on Digital Libraries, 235 – 236, Roanoke, Virginia, USA.

Harvey, F. (2003) "Name That Tune" In Scientific American, June 2003, 84 – 86.

Hauptmann, A. G. and M. J. Witbrock (1997) "Informedia: News-on-Demand Multimedia Information Acquisition and Retrieval" In *Intelligent Multimedia Information Retrieval*, M. Maybury (ed.), 215 – 239, Menlo Park, California, USA: AAAI Press/MIT Press.

HP SpeechBot (2003)
http://speechbot.research.compaq.com/

HTK (2003) "HTK Web-Site HTK FAQ"
http://htk.eng.cam.ac.uk/docs/faq.shtml

Informedia (2003) "Informedia Home Page"
http://www.informedia.cs.cmu.edu

Jones, G., J. Foote, K. Spärck Jones and S. Young (1997) "The Video Mail Retrieval Project: Experiences in Retrieving Spoken Documents" In *Intelligent Multimedia Information Retrieval*, M. Maybury (ed.), 191 – 214, Menlo Park, California, USA: AAAI Press/MIT Press.

Jurafsky, D. and J. H. Martin (2000) *Speech and Language Processing*, Upper Saddle River, New Jersey, USA: Prentice-Hall.

Leavitt, N. (2002) "Let's Hear It for Audio Mining" In IEEE Computer, Vol. 35, 23 – 24.

Leman, M. (2002) "Musical Audio Mining" In *Dealing with the Data Flood: Mining data, text and multimedia*, J. Meij (ed.), Rotterdam, Netherlands: STT Netherlands Study Centre for Technology Trends.

Liberman, A. M. (1982) "On Finding That Speech Is Special" In American Psychologist, Vol. 37, No. 2, 148 – 167.

Mani, I., D. House, M. Maybury and M. Green (1997) "Towards Content-Based Browsing of Broadcast News Video" In *Intelligent Multimedia Information Retrieval*, M. Maybury (ed.), 241 – 258, Menlo Park, California, USA: AAAI Press/MIT Press.

Maybury, M.T. (ed.) (1997) *Intelligent Multimedia Information Retrieval*, Menlo Park, California, USA: AAAI Press/MIT Press.

MediaIndexer (2003) "ScanSoft - Dragon MediaIndexer"
http://www.scansoft.com/mediaindexer/

Ng, K. (2000) “Subword-based Approaches for Spoken Document Retrieval” Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, February 2000.

NGSW (2003) The National Gallery of the Spoken Word. Project Information Site
http://www.ngsw.org/

NPR Archives (2003)
http://www.npr.org/archives/index?loc=hometext.html

NSF (2002) Award#9817485 - DLI-2: A National Gallery of the Spoken Word
https://www.fastlane.nsf.gov/servlet/showaward?award=9817485

Palmer, D. D. (1997) *Wittgenstein for Beginners*, New York, USA: Writers and Readers Publishing.

Powell, J. and V. Howell (1996) *Derrida for Beginners*, New York, USA: Writers and Readers Publishing.

Quinn, E. (2000) “SpeechBot: The First Internet Site for Content-Based Indexing of Streaming Spoken Audio” Technical Whitepaper, Compaq Computer Corporation, Cambridge, Massachusetts, USA.

Rough ‘n’ Ready (2003) What We Do: Speech & Language Processing: Rough'n'Ready[tm]
http://www.bbn.com/speech/roughnready.html

Takeshita, A., I. Takafumi and T. Kazuo (1997) “Topic-based Multimedia Structuring” In *Intelligent Multimedia Information Retrieval*, Maybury (ed.), 259 – 277, Menlo Park, California, USA: AAAI Press/MIT Press.

THISL (2000) “THISL: Thematic Indexing of Spoken Language“
http://www.dcs.shef.ac.uk/research/groups/spandh/projects/thisl

TREC (2002) “Text REtrieval Conference (TREC) (2002) Home Page”
http://trec.nist.gov

Tuerk, A., S.E. Johnson, P. Jourlin, K Spärck Jones and P.C. Woodland (2000) “The Cambridge University Multimedia Document Retrieval Demo System” In Proc. of ACM SIGDIR Conference, p. 394, Athens, Georgia, USA.

Video Mail (1997) “Video Mail Retrieval Using Voice”
http://svr-www.eng.cam.ac.uk/research/Projects/vmr/vmr.html

VoiceXML (2003) “Voice Extensible Markup Language (VoiceXML) Version 2.0”
http://www.w3.org/TR/voicexml20/

Young, S. J. (1994) “The HTK Hidden Markov Toolkit: Design and Philosophy" CUED/F-INFENG/TR.152, Cambridge University, Cambridge, England.

Appendix A: 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

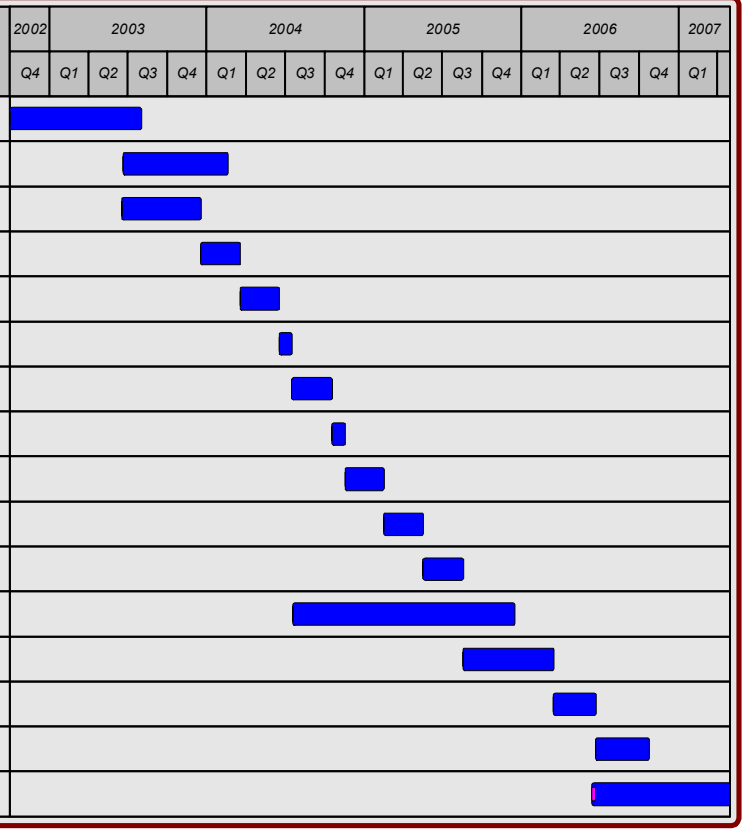

Table 4.1 Project Schedule