Browsing Archived Meeting Audio and Time-Synchronized Data

by

Abhishek Ranjan

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Graduate Department of Computer Science
University of Toronto

Abstract

Browsing Archived Meeting Audio and Time-Synchronized Data

Abhishek Ranjan

Master of Science

Graduate Department of Computer Science

University of Toronto

2005

In a majority of meetings, speech and audio plays the most important role and therefore most of the meeting archives contain audio and some other data types synchronized with it. Despite the importance of audio, its inherently sequential nature makes it difficult to be used as the guiding channel for browsing archives. In this thesis we describe the design and implementation of a prototype system for browsing archived meeting audio and other time-synchronized data using partially correct transcript for visualizing audio. We present a transcript based simultaneous-spatial audio browsing technique. This technique has been evaluated and found to be more efficient than the conventional interface for search tasks. It also allows an audio-data guided browsing of the archive. Furthermore, we present techniques for annotation and personalization of slide, audio and transcript data.

Acknowledgement

I would like to thank Professor Ravin Balakrishnan (my MSc supervisor) for introducing me to the field of HCI. His constant encouragement and guidance (in research, in finding good restaurants and bars, and in trying whacky foods) helped me survive my first venture to graduate school.

I would like to acknowledge Professor Mark Chignell and IBM ARISE project research group without which this project would not have completed.

I can not forget to thank Mike Wu for working hard (sometimes hardly working) till dawn, 4am NHK dinners, ∞ movies, and corridor VB practices. I would like to thank Bowen and Brad who collaborated with me (and Mike) to revive and popularize DGP Vampires group: a group of nocturns who burn midnight oil not just to meet academic deadlines, but also to achieve higher levels in BIG-2 & HALO, to master the art of movie-watching, and, of course, to master the art of living a life full of fun!!

I am thankful to Anastasia for always helping me in holding the wall up and in various other problems. I would like to thank Alex for posing for me whenever I needed a demo of the prototype. I express my thanks to Saif for various creative discussions followed by home-cooked desi dinner.

I would like to thank everyone in DGP for making it a great place to work, specially: Winnie for never coming to lab, Minnie for frequently coming to lab, Hanieh and Irene for appearing in lab from nowhere, Patrick and Shahzad for coming to lab at weirdest possible hours and Joe for never leaving the lab. I would like to thank Jia for unknowingly inspiring me to work even after spending 4 hours in a bar and to Paul for 3.99 lunches at dinner times.

Lastly, I am thankful to mummy, papaji, cuckoo and everyone else in my family, and all my friends for supporting and encouraging me all the time.

Contents

1 Introduction **1**

1.1 Motivation 1

1.1.1 Audio data: capture and access 2

1.1.2 Time synchronized data browsing 3

1.2 Thesis organization 5

2 Related Work **6**

2.1 Audio accessing techniques 6

2.1.1 Time-compression based 7

2.1.2 Simultaneous-spatial audio based 7

2.1.3 Annotation and indexing based 10

2.1.4 Transcript based 10

2.1.5 Hybrid 11

2.2 Browsing time synchronized data 12

2.3 Annotating data 14

2.3.1 Annotation of non-temporal data 15

2.3.2 Annotation of temporal data 15

2.4 Personalization . . . 17
2.4.1 Collection creation . . . 17
2.4.2 Layout management . . . 18

3 System Design **20**
3.1 Rationale . . . 20
3.1.1 Requirements for the system . . . 21
3.1.2 Design principles . . . 23
3.2 System and data overview . . . 24
3.2.1 Captured data . . . 24
3.2.2 Display platforms . . . 25
3.3 Timeline based visualization . . . 25
3.3.1 Design options and issues . . . 25
3.3.2 Final design . . . 27
3.4 Browsing by zooming and querying . . . 29
3.4.1 Design issues . . . 29
3.4.2 Multi-focus zooming . . . 30
3.4.3 Querying . . . 32
3.5 Audio and audio transcripts . . . 34
3.5.1 Issues . . . 34
3.5.2 Integrating transcripts and audio . . . 35
3.5.3 Spatial audio and transcripts . . . 37
3.6 Integrated data access . . . 38
3.7 Personalization . . . 40
3.7.1 Annotating audio . . . 40
3.7.2 Personal collections . . . 43

4 Evaluation **48**
4.1 Motivation . . . 48

4.2 Goals . . . 49
4.3 Task . . . 49
4.4 Techniques . . . 49
4.5 Interface . . . 51
4.6 Hypotheses . . . 52
4.7 Setup . . . 53
4.8 Results . . . 55
4.8.1 Observation summary . . . 56
4.8.2 Hypothesis testing and analysis . . . 56
4.8.3 Search accuracy . . . 63
4.9 Discussion . . . 63
4.9.1 Results discussion . . . 63
4.9.2 Design implications . . . 66

5 Conclusions and Future Directions **68**
5.1 Conclusions . . . 68
5.2 Contributions . . . 69
5.3 Future directions . . . 70

A Experiment Audio Transcripts and Questions **78**

B Complete Data Analysis Results from SAS **93**

List of Figures

2.1 Focus of attention timeline based browsing from Classroom 2000 (Brotherton *et al.*, 1998) . . . 13
2.2 HunterGatherer collection creation interface (schraefel *et al.*, 2002) . . . 18
2.3 TopicShop collection creation interface (Amento *et al.*, 2000) . . . 19

3.1 Slide timeline with important slides represented as thumbnails and other slides as bars . . . 27
3.2 Transcript timeline showing words per minute as bars . . . 28
3.3 Overlaid slide and transcript timelines . . . 28
3.4 Multi-focus zooming showing details of 3 different parts of the timeline . . . 31
3.5 Query results showed using map-metaphor by highlighting the results in yellow and a compact query browser showed at the bottom-right . . . 33
3.6 Compact query result browser with selected slides shown as thumbnails. Four consecutive slides corresponding to four bars above the slider have been shown as thumbnails. . . . 34
3.7 Transcript browser with portions being played shown in highlighted background . . . 36

3.8 3 Transcript browsers arranged spatially. The user listens to the stream **T1** from the left, **T2** from the center, **T3** from the right. 38
3.9 Integration of transcript with slides. The leftmost slide shows the bar at the bottom which can be used to pull down the corresponding transcript. The slides in the center and on the right show the corresponding transcripts pulled to different lengths 39
3.10 Voice annotation of transcripts. The blue marks on the left edge show the speech annotation positions in the transcript text and the numbers (with a small bar) at the bottom of the transcript browser show the corresponding annotations' time lengths. The user can select these bars to play the annotations. 41
3.11 A Personal collection with 4 slides, a portion of the transcript on the right bottom and some organizational pen marks 44
3.12 Post-it audio notes: (a) A collapsed post-it audio-note, (b) A post-it audio note with playback interface, (c) Various post-it audio-notes used in a personal collection. They can be switched from playback-mode to collapsed mode and vice-versa by mouse click. 46
3.13 The system running on a large-scale display with a user interacting with it using a digital pen. Various features (transcript-browser, zoomed slides, query-browser) of the system are displayed at various position on the display. 47

4.1 Experiment interface: single stream (*S1*) interface with transcript. The transcript portion being played is highlighted in yellow. 51
4.2 Experiment interface: double stream (*S2*) interface with transcripts. The transcript portion being played is highlighted in yellow in both the transcripts. 53
4.3 A plot of mean times for three different stream based techniques (*S1*, *S2*, *S3*) 57

4.4 A plot of mean times for three different transcript quality based techniques (*Q1*, *Q2*, *Q3*) 58
4.5 An error bar plot of mean time and 95 % confidence interval for all 3 stream-based techniques and 3 qualities. 60
4.6 A plot of mean times for various different transcript qualities, grouped by the position of the answer in the clips. 61
4.7 A plot of mean time for various stream based techniques, grouped by the position of the answer in the clips. 62

List of Tables

4.1 Time difference comparison for various techniques 57
4.2 Time difference comparison for various transcript qualities 58

Chapter 1

Introduction

"Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?"

- T. S. Eliot

The goal of this research is to explore efficient ways for browsing and personalization of large amount of archived meeting audio and other time-synchronized multi-modal data. We investigated various audio browsing techniques using partially correct audio transcripts for this purpose and evaluated those techniques.

1.1 Motivation

Reviewing and salvaging of discussions and interactional activities in a meeting are of tremendous value and archiving meeting data is the first step in this process (Moran *et al.*, 1997a). The choice of data type to be archived is dependent on several factors such as the end-user's requirements, technology available for capturing and later accessing of data etc.

1.1.1 Audio data: capture and access

Among various interaction media, speech and audio form an important part of the communication in most of the meetings and therefore, it forms an important part of the meeting archives as well. Speech is expressive and also contains various memory cues which help in the process of archive reviewing. Along with its natural advantages, it is also very easy to capture. Current technology provides us with various inexpensive, yet efficient, ways to record audio data with very low overhead.

While the expressiveness of speech makes audio archives a rich source of information, the inherently sequential nature of audio makes it difficult to access. The situation worsens when the length of the audio is long. For example, consider a classroom meeting archive. If a student has to search for the answer to a specific query about the contents of a three hours recording of a lecture then, in the absence of efficient audio accessing techniques, she might have to listen to the whole audio which could potentially take three hours!

The advancements in the audio and speech processing techniques, such as spatialization, automated speech-recognition (ASR) and time-compression, have opened several ways to perform the search task efficiently. ASR technology is of special interest because it can convert audio signals to text transcripts. Transcript has two major advantages over audio. Firstly, it is faster to read the transcript than to listen to the audio and secondly, transcript can support query engines. When browsing a large amount of data, which is the focus of this research, querying support is a major advantage that requires the presence of a text transcript. However, we found that the limitations of the latest ASR technology do not allow us to create a fully accurate transcript. In practice, this raises several cases in which, without the intervention of a human, the search process can not be completed using database querying alone. Here we describe two such cases.

The first case occurs when the query engine does not return any search result corresponding to the query string. The reason is that the text based query engine can search

for specific terms which are present in the database, but it can not be of much help if the database does not contain the query string itself. This situation could be prevalent if the database contains partially correct transcripts of the audio and the query string has been taken from the audio.

The second case occurs when the query engine returns multiple possible answers and the user has to find the most relevant one. This is the form of result that most query engines return. From the user's perspective, this certainly reduces the size of the search domain, but she still has to perform the task of pin-pointing the most appropriate results.

Therefore, when dealing with audio-data, easy access to audio is an absolute necessity. This necessity leads to the integration of the transcripts with the corresponding audio for effective browsing. In this thesis we present a new audio browsing and searching technique which integrates the transcript with the simultaneous-spatial time-compressed audio. We illustrate the advantages of this technique by developing an audio browsing system with integrated query engine. Once the integrated query engine returns various portions of the audio as possible query results, the user can listen to any or all of them simultaneously with the corresponding transcripts, change the playback speed, and quickly search for the relevant information. This browsing technique is one of the most important contributions of our work. We devised a laboratory experiment to evaluate the performance of this technique for search tasks. We found that searching in two simultaneous-spatial audio with transcripts is faster than searching with either method alone (with or without transcripts). When we add time-compression to this technique performance tends to improve but this increase is not statistically significant.

1.1.2 Time synchronized data browsing

Technological advancements have made it possible to have multi-modal interaction during the meetings and therefore information exchanged among various participants can not be contained in just one type of media. In such a case capturing different streams

of media in a time-synchronized manner provides a better record of the actual event (Moran *et al.*, 1997a; Brotherton *et al.*, 1998; Abowd *et al.*, 1997). However, capturing time-synchronized multi-modal data makes browsing them at a later stage increasingly challenging. This largely happens due to two orthogonal dimensions of the problem: the temporal and the content-based dimension. On one hand the streams have strong temporal association; on the other hand they have content based similarity which could be independent of time.

Timeline based approaches have been widely used to address these issues. But multiple data-streams demand more than just a timeline. While the timeline takes care of the temporal dimension, an effective integration among various data-streams is essential for dealing with the the content dimension. At the user interface level, the effectiveness of the integration can be judged by how efficiently the user can go from readily available data to some related but currently unavailable data. Multiple timelines, query engines, data-visualization techniques etc are various approaches to address the integration problem. We found that, irrespective of the approach, the interface design for such integrated systems heavily depends on the visual representation of the data and therefore the problem gets more severe when one of the major components of the data lacks any effective visual representation.

The domain of interest in this research is a classroom meeting scenario where speech and audio plays the most important role. But the importance of slides in modern classrooms can not be ignored. In our archived data we captured both audio and slides in a time-synchronized manner. While slides have a well known visual representation (i.e. images or thumbnails), audio lacks any such representation (Whittaker *et al.*, 2002). The user can get an overview of several slides at once by looking at their thumbnails, but, conventionally, audio has no such representation. This causes an inconsistency among techniques to access different data-types. In our system we use partially correct transcripts to visually represent audio data. The transcript-based representation of audio is similar to thumbnail-based representation of slides. We explore the techniques to access

slides from transcripts and vice-versa to effectively integrate audio and slide data. Our system also supports various techniques for annotation and personalization that is an integral part of effective browsing.

1.2 Thesis organization

The rest of the thesis is organized into four chapters. In chapter 2, we discuss various previous systems and studies related to our work. We categorize them based on their features and highlight how our work builds upon them. In chapter 3 we discuss the prototype system that we developed to test our ideas. We also explain our design rationale and describe various features of the system. In chapter 4 we describe the experiment we conducted to evaluate the audio browsing techniques we developed. We also quote and analyze the results of the experiment. Conclusions are drawn in chapter 5 and possible future research directions are explored.

Chapter 2

Related Work

"A science is any discipline in which the fool of this generation can go beyond the point reached by the genius of the last generation."

- Max Gluckman

In this chapter we discuss some of the related work done in the field of audio accessing techniques, time synchronized data browsing, annotation and personalization. We also highlight how our work relates to and builds upon that other work.

2.1 Audio accessing techniques

Psychologists and computer scientists have been studying the properties of auditory displays and hearing for over a century. These properties have often been used in developing audio browsing techniques. On the basis of these underlying properties, we divide browsing techniques into five broad categories: (1) Time-compression based, (2) Simultaneous-spatial audio based, (3) Annotation and indexing based, (4) Transcript based, and (5) Hybrid. In the following subsections we discuss prior works in those categories.

2.1.1 Time-compression based

These techniques make use of human capability to comprehend and understand audio at higher speeds. Significant work has been performed in the area of time-compressed speech comprehension and intelligibility. Previous studies have shown that single well-learned and phonetically balanced words can remain intelligible at 10 times normal speed and that connected speech remains comprehensible at 2 times the normal speed (Arons, 1992b). Arons's SpeechSkimmer (Arons, 1997) system introduces an audio browsing technique using time-compression of audio. Considering the result that continous speech remains comprehensible upto about twice (2X) the normal speed, the system tries to push this limit higher for audio skimming tasks using a specially designed interface. The interface allows the user to smoothly switch from one speed to another. One of the several interesting results found through the controlled experiments done with the interface was that the subjects used higher speed to get an overview, and slower speed to get the details.

2.1.2 Simultaneous-spatial audio based

In 1954, Cherry and Taylor stated in their paper (Cherry & Taylor, 1954)

"One of the most striking facts about our ears is that we have two of them - and yet we hear one acoustic world; only one voice per speaker".

This statement highlights an interesting property of human ear that even if the two ears receive different audio signals,we can select one of them and listen to it. This ability of *selective attention* has been termed as the *cocktail party effect*. Arons reviewed various aspects of this effect and summarized some of the interesting results found by psychologists over a period of more than four decades (Arons, 1992a). One of the findings reviewed is of special interest from our system's interface design perspective:

"The ability to rapidly shift one's attention (e.g. with multiple loudspeakers) does not help if the information rate is high. Under the worst conditions (two simultaneous messages), only 60 % of the information was received, but this results in a greater total information intake per unit time than if the messages had occurred sequentially."

Cherry performed several experiments related to simultaneous listening of two streams in two ears (Cherry, 1953). It was found that when the subject listened to two different messages (recorded using the same speaker) in two different ears then she experienced no difficulty in listening to either speech at will and "rejecting" the unwanted one. When she was asked to repeat one of the messages concurrently, she successfully recognized every word, but had little or no idea what the message is all about. These findings have the implication that potentials of simultaneous audio presentation can be leveraged for the tasks which do not require complete information intake.

Stifleman conducted an experiment to study the effects of simultaneous presentations when the tasks were comprehension and target-monitoring (Stifelman, 1994). In this study, subjects listened to 1, 2, or 3 channels of audio and performed two tasks simultaneously. While listening to 2 or 3 channels, they had to comprehend one channel (comprehension task) and find some target words (target-monitoring task) in the remaining channels. The results showed that the subject's performance decreased in both of the tasks as the number of channels increased. Further, the performance in target monitoring task suffered heavily when the number of channels increased from 2 to 3.

While this study suggests some guidelines for designing interface for performing combined comprehension and search tasks, it does not examine the combination of two search tasks. In this thesis we study pure search tasks in both of the streams presented simultaneously and spatially, with transcripts. We also examine the effect of adding time compression to the simultaneous playback.

Schmandt et al. used simultaneous-spatial audio in an audio browsing interface called AudioStreamer (Schmandt & Mullins, 1995) in which the user can listen to 3 spatialized

audio streams. The system can also detect the user's head motion by using non-contact sensors and the loudness of a stream can be increased or decreased by moving the head towards or away from the stream. When the user selects a stream using head-motion, other streams' loudness starts decaying. Using automatic speech processing techniques, the system can also detect various events in the audio such as pause, pitch change etc. Whenever these events occur, the system draws the user's attention by emitting a tone or increasing the stream's loudness.

Sawhney et al. studied audio and speech based interaction techniques to be used in nomadic environments (Sawhney & Schmandt, 2000). The system developed for this study, called Nomadic Radio, is an audio-only wearable interface that unifies asynchronous communication and remote information services. The task of navigation in the system can be performed using 3 simultaneous spatial audio streams. For example, the user can scan through several audio messages within a short duration. Spatial audio scanning cycles through all messages by moving each one to the center of the listening space for a short duration and fading it out as the next one starts.

Kobayashi et al. developed a system, called Dynamic Soundscape, which provides a spatial interface for temporal navigation of audio data (Kobayashi & Schmandt, 1997). The user can hear multiple portions of the audio clip simultaneously from different spatially arranged speakers. These speakers are arranged in a circle around the user's head and play the audio from different time instants. The user can select a speaker, and hence jump to a new position inside audio, by moving her head towards the speaker's direction.

While aforementioned browsing systems use only audio streams, we use audio streams with partially correct transcripts. Our system also differs from them with respect to the specific task that we support. We focus specifically on search instead of comprehension. Moreover, while these systems use three or four streams, we used two streams and an additional feature of doubling the playback speed to further improve the search time.

2.1.3 Annotation and indexing based

While all techniques described previously use some properties of the speech signal and leverage the potential of the human aural system, this class of techniques depends on the user interface for capturing audio. Digital devices and interfaces allow indexing and annotation of audio during the capturing itself. These annotations made by the users can later be used as cues for browsing audio.

Wilcox et al. developed an audio notes browsing system, called Dynomite (Wilcox *et al.*, 1997), which is a modified digital notebook that records audio. The system synchronizes the pen marks made on the notebook with the audio recorded and the user can later access the audio randomly based on their pen marks.

The Audio Notebook (Stifelman *et al.*, 2001) concentrates on both annotation and signal based audio indexing and summarization. It uses pen marks and audio signal properties to give structure to the recorded audio. The system also detects phrases and suggests topics. These topics and phrases, along with the pen marks, guide the user's browsing process.

The main drawback of these annotation and indexing based techniques is that they rely on the audio capturing process. They can not be used to browse a pre-recorded audio database without annotations or indices.

2.1.4 Transcript based

Recent advances in speech recognition and automated transcription have made it possible to automatically generate partially correct transcripts. Whittaker et al. used these partially correct transcripts to browse voice mail data (Whittaker *et al.*, 2002). The browsing interface, called SCANMail, displays the transcript with some important parts highlighted, such as the caller's name and phone numbers. The user can read, search and annotate these transcripts. These transcripts can also be used to play some portion of the underlying audio. The user's performance in using this interface was compared with

a conventional voice mail system for search, information extraction and summarization tasks. The results of the evaluation show that the SCANMail interface outperformed the standard voice mail system's interface.

Whittaker et al. explored the use of the transcripts further by developing an interface that allows the user to edit the underlying audio using the transcripts (Whittaker & Amento, 2004). This introduces the concept of semantic editing of audio. The user can read the transcripts and perform various editing operations (cut, copy, paste) on the transcript text and these operations are mirrored on the underlying audio. A laboratory evaluation showed that the semantic editing of audio is more efficient than conventional acoustic signal based editing even if transcripts are highly inaccurate.

These systems show the potential of error-laden transcripts for accessing audio in the domain of voice mail data. In this thesis, error laden transcripts are used for browsing and personalizing meeting audio archives and other dependent data types.

2.1.5 Hybrid

We call a technique hybrid if it combines two or more of the previously described techniques. These hybrid techniques use advantages of two or more of the approaches and fuse them to achieve higher performance.

Vemuri et al. performed an experiment to explore the combination of time-compressed audio and transcripts for efficient audio browsing (Vemuri *et al.*, 2004). The interface created for the experiment displayed the transcripts and could play the audio at various speeds. The user could smoothly adjust the audio speed while listening to the audio. The comprehension of the speech was chosen as the metric to assess the efficiency of audio browsing. The results showed that as the transcription quality degrades, comprehensibility goes down. Furthermore, the increase in playback speed led to a degraded comprehensibility. An interesting finding was that the audio presentation with the error-laden transcripts had better comprehensibility than the presentation without any transcript.

Arons studied the time-compression techniques in combination with spatial audio (Arons, 1994). Common time-compression techniques sample the audio and throw away the unsampled data. But in this study, the sampled and unsampled data were presented spatially to different ears. As a result, the user listened to one audio clip but at higher speed and without losing any data.

One of the advantages of using hybrid techniques is that one component technique compensates for the information lost in the other technique and hence the combination increases the speed without influencing the performance. In this thesis we explore the combination of simultaneous-spatial audio, time-compression, and audio transcripts.

2.2 Browsing time synchronized data

Previous efforts made to visualize and browse time synchronized data deal with various kinds of data, such as archived audio and video, slide presentations, complex medical documents, plain text documents etc. Although the data types differ across various applications, the presence of temporal dimension is the common guiding factor among all of them.

Among all the techniques to visualize the temporal dimension, the timeline based technique is the most widely used technique. This technique existed long before the computer was invented. Tufte discusses its history, development, and applications in various areas (Tufte, 1990). One of the areas in which this technique is well established is air-traffic scheduling and control, where various events are shown on a single line with timescale. In the case of digital data this technique is often combined with a zooming based interface. In this section we discuss some of the most relevant works done in this direction.

The Lifelines system deals with time synchronized personal-history data (Plaisant *et al.*, 1996). The system provides a timeline based interface for browsing personal life histories. Various aspects of the life-history have their own timelines and various incidents under a particular aspect are highlighted on the corresponding timeline. Some

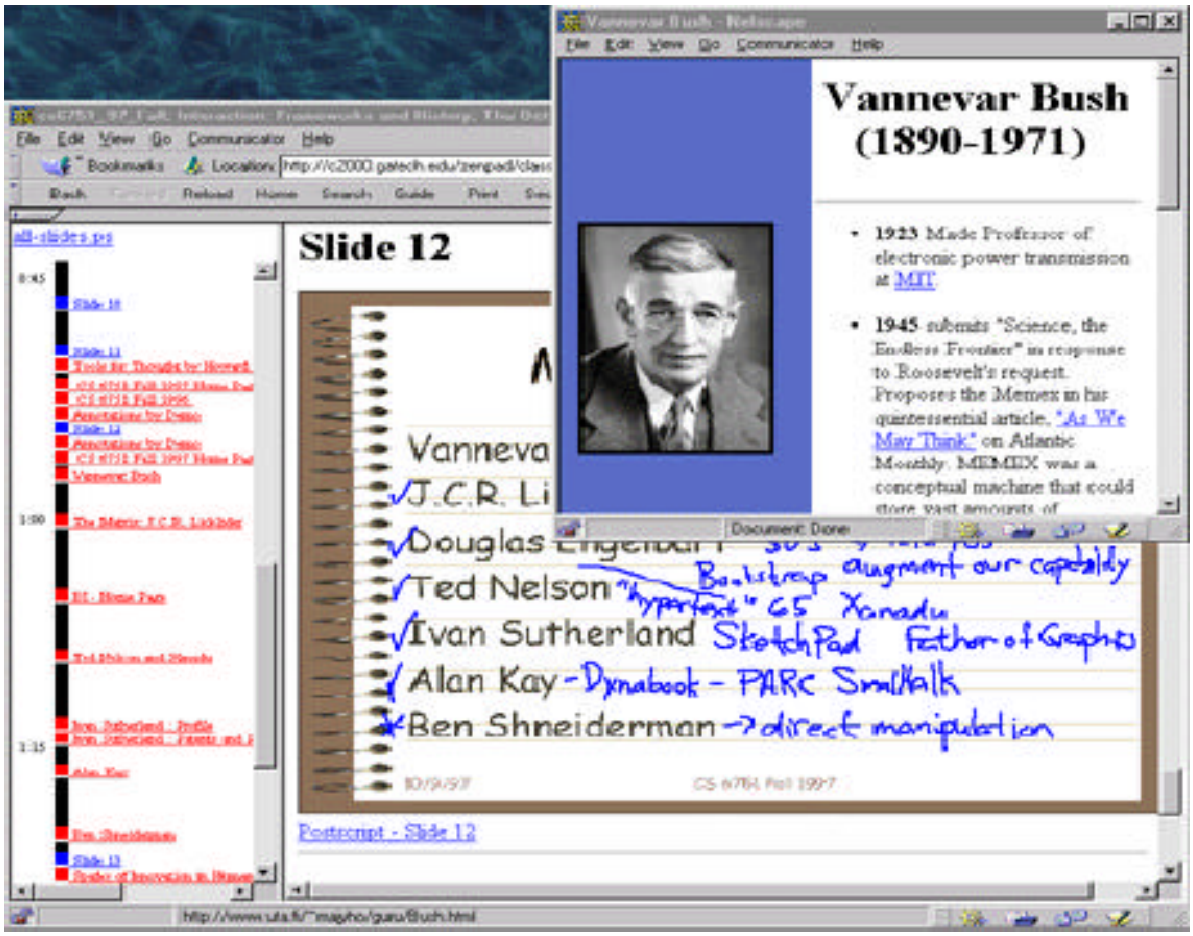

Figure 2.1: Focus of attention timeline based browsing from Classroom 2000 (Brotherton *et al.*, 1998)

zooming and filtering techniques were provided to view and access details of the area of interest. By zooming progressively the user can view various levels of details. Bade et al. explores various browsing and visualization techniques of time-oriented medical data (Bade *et al.*, 2004). (Mills *et al.*, 1994; Ramos & Balakrishnan, 2003) present timeline and zooming based interfaces for browsing and personalizing of time-synchronized video data.

Brotherton et al. address the broader issues of capture, generation, integration and visualization of media streams (Brotherton *et al.*, 1998). These media streams were captured from a classroom lecture and included lecture audio and video, slides, lecturer's notes, and web sites visited. A timeline based visualization had been used. Various gran-

ularities of media integration have been explored, such as slide level, writer's pen-stroke level, and word-level. It was found that no one level of granularity can cater to all the users' needs. Further, it was observed that a scalable solution to display all the streams is difficult to find when screen space is constrained. A new stream, derived from the media streams, was automatically generated based on predicted the focus of attention. Figure 2.1 shows the focus of attention timeline. Browsing and exploring is done using a zooming interface.

The timeline based visualization often faces the problem of viewing details with the context. Most of the previous systems address this issue by using hierarchical or progressive zooming in which the user can progressively zoom-in from a coarser level to a finer level of detail.

Furnas introduced a zooming technique called "fisheye views" for viewing details with context (Furnas, 1986). The fisheye view uses a very wide-angle lens or fisheye lens for zooming. The data in the region near the lens is shown in great detail while still showing the peripheral data in successively less detail.

While designing our system, we faced a new problem of viewing details of various portions at the same level of zooming. In this thesis we discuss why hierarchical zooming is insufficient for our purpose and how to approach the new problem.

2.3 Annotating data

Annotation techniques have been studied extensively for various applications in which different data types are annotated using different means. For ease of discussion, we divide the annotated data into two categories: *temporal* and *non-temporal*. All the data types which do not have a temporal dependency will be referred to as *non-temporal* data. Some examples of this are paper documents, static text document, static images, static presentation slides etc. All kinds of data-streams, such as video and audio streams, and other time-dependent data types can be placed under *temporal* data types. Most of the previous

techniques use pen marks, text, images or hyperlinks to annotate *temporal* or *non-temporal* data types.

2.3.1 Annotation of non-temporal data

The Tivoli system explores pen based freefrom annotations on electronic whiteboard (Pedersen *et al.*, 1993; Moran *et al.*, 1997b). These annotations can be used to structure, organize and augment the non-temporal data displayed on the electronic whiteboard. The system can identify several pen gestures and uses them to perform various actions such as creating boundaries, spatial gaps, links and collapsing annotations etc. The XLibris system uses a paper document metaphor to support document reading and annotations (Schilit *et al.*, 1998). The user can read the scanned copy of the document on pen-tablet display. To preserve the properties of paper documents, annotations can be made using free-form ink marks. The system saves them as clippings along with their context and they can be later retrieved collectively using 'Reader's Notebook'. The 'Reader's Notebook' extracts annotated clippings and present them in a multiple page view. Various operations, such as sorting, filtering and grouping, can be done on these clippings. The PADD system introduces a new kind of document architecture for merging digital and paper document (Guimbretire, 2003). These documents can be manipulated both in digital and paper format and the annotations are transferrable from one medium to the other.

2.3.2 Annotation of temporal data

Most of the work done in the field of temporal data annotation provide techniques for annotating audio (Stifelman *et al.*, 2001; Wilcox *et al.*, 1997; Whittaker *et al.*, 1994; Abowd *et al.*, 1997) and video (Weher & Poon, 1994; Ramos & Balakrishnan, 2003) using pen marks. The Audio notebook is a combination of paper notebook and digital audio recorder (Stifelman *et al.*, 2001). It allows the recording of audio while the user is taking notes on the notebook. The recording is synchronized with the pen marks.

The Dynomite system uses pen annotations for indexing as well as for determining the portion of the audio which should be stored permanently (Wilcox *et al.*, 1997). The user can specify a region of audio by selecting start and end time and the corresponding portion on the audio timeline is highlighted as a colored bar. The system shows all the pen-strokes made during the highlighted portion in bold. Once a portion is highlighted, the user can edit, playback or permanently save that audio portion.

In the Classroom2K project, the user can annotate the audio and other media streams by their own pen marks while it's being recorded (Truong *et al.*, 1999). Realtime annotation of time synchronized presentation slides raises several issues related to simultaneous public and personal annotation, such as making public streams available to the individual, supporting the individual's pace, resolving competition between public and personal etc. The system addresses these issue by using various techniques for data synchronization, rapid browsing, context management and secured access.

The Marquee (Weher & Poon, 1994) and the LEAN (Ramos & Balakrishnan, 2003) systems address the problem of video annotation. The Marquee system allows the user to take notes while the video tape is being played. The notes are time-stamped and logged. These logs can be later used to access and retrieve the interesting portions inside the video tape. More recently, the LEAN system introduces various techniques for annotating digital video. The system allows annotation of each frame of the video. These annotations can be made using free-form pen marks. Their temporal and spatial location is saved with the video. During playback, the system visually blends these annotations in and out of the environment as the moment when the annotation was made approaches and passes.

Chalfonte et al. performed a study to compare audio and text annotations (Chalfonte *et al.*, 1991). The findings of the study strongly suggest that a richer and more expressive medium is especially valuable for more complex tasks. When the subject's modalities were restricted, using written annotations led them to comment on more local problems in the text, while using speech led them to comment on higher level concerns. Despite these findings, most of the previous systems use pen-marks for annotating both tempo-

ral and non-temporal data and the use of audio and other richer media have remained mostly neglected. Wilcox et al. found the reason for this negligence to be the cumbersome technology (Wilcox *et al.*, 1997). Most of the users don't want to transcribe their audio annotations and finding information on the tape without transcription is frustrating. Further, the use of dictaphone or tape-recorder to record the thoughts was awkward. This dichotomy between the ideal and the practical encouraged us to explore the audio and speech data for annotation of audio. In our system we leverage recent advances in recording and speech processing technology to create an interface for speech annotation of audio.

2.4 Personalization

Personalization of the data includes various activities, such as creating personal collections, managing the layout of the data, annotating collections, adding notes, creating bookmarks etc. In this section we discuss two of these activities: collection creation and layout management.

2.4.1 Collection creation

Shraefel et al. developed a system for collecting components from the web sites and editing those component collections (schraefel *et al.*, 2002). Using select, drag and drop, any component from a web page can be selected and added to the collection. Since the system is integrated with the browser, the attention divide between the data collection process and the browsing process is minimized. Inside the collection, these components are shown in a list view (see figure 2.2). Components in this list view can be sorted, searched and viewed in separate browser windows.

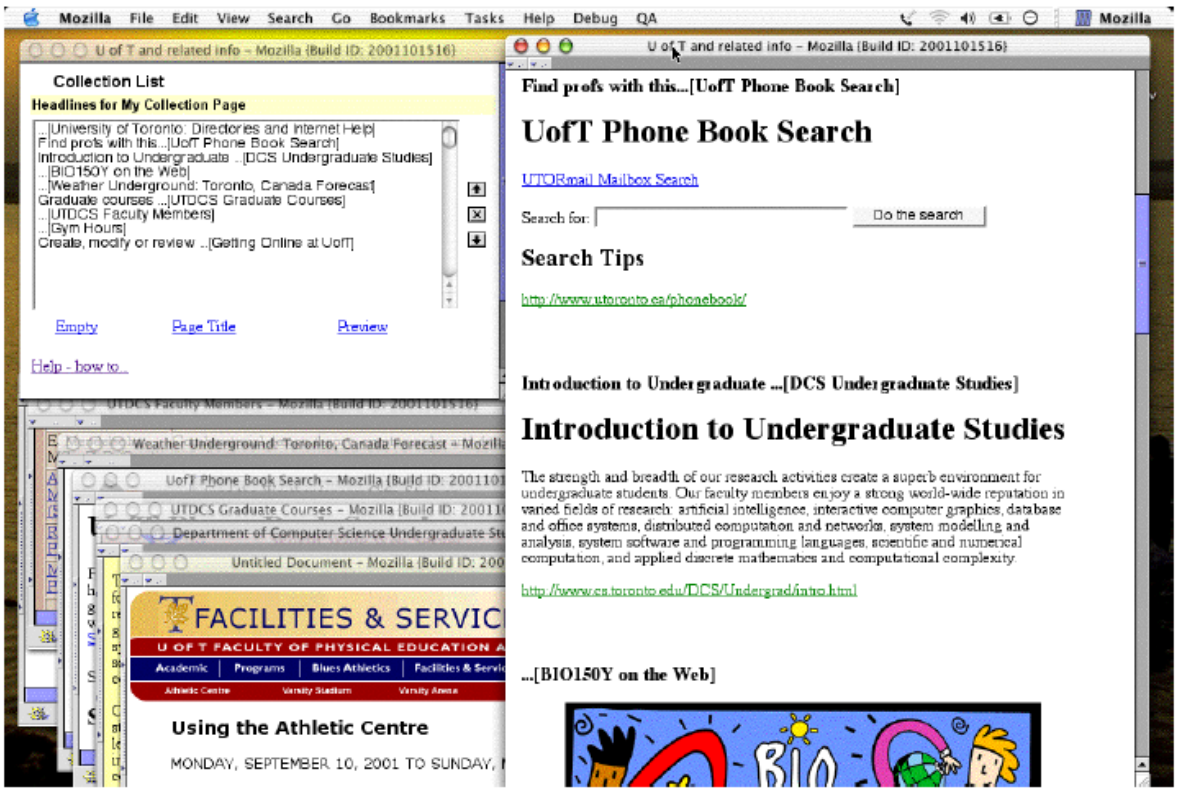

Figure 2.2: HunterGatherer collection creation interface (schraefel *et al.*, 2002)

2.4.2 Layout management

Layout management and organization of the visual representation of the data are explored in various systems, e.g. the TopicShop (Amento *et al.*, 2000), the Data Mountain (Robertson *et al.*, 1998), the system using a pile metaphor in (Mander *et al.*, 1992) etc. Mander et al. introduces pile metaphor to organize data (Mander *et al.*, 1992). Various digital documents can be arranged as a pile in the same way paper documents can be arranged on the desk or shelf. The advantages of using the pile metaphor are that it organizes the documents in a familiar way and saves screen space. To browse these piles of folders the user can spread them out, search inside the piles or use viewing cones. These techniques provide space efficient ways to organize the folders but due to the pile structure it is difficult to see the contents of the files inside a folder without opening them. A 2D spatial layout could give a better visual representation of the file contents.

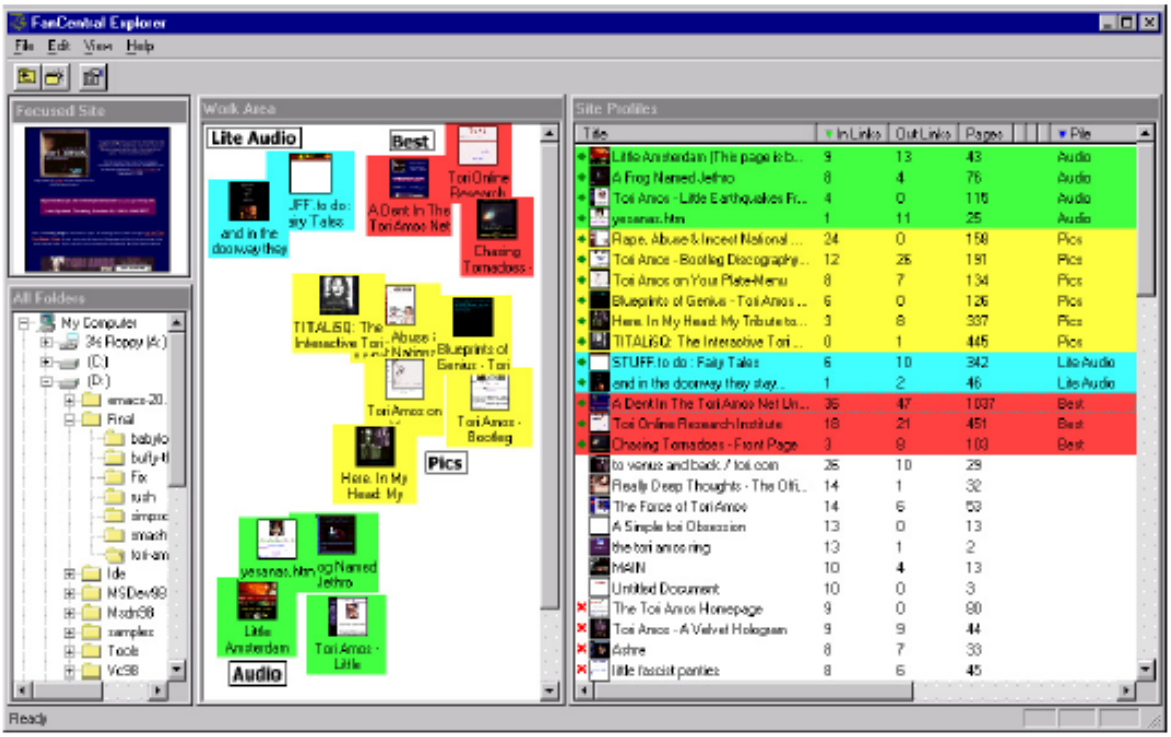

Figure 2.3: TopicShop collection creation interface (Amento *et al.*, 2000)

In the TopicShop system a 2D spatial layout is used to organize the collection of web sites (see figure 2.3) (Amento *et al.*, 2000). In this system a web site is represented by a large thumbnail image and the web site's name. The user can place related web sites together in the 2D space. Closely placed web sites automatically form a group. Different groups are assigned different colors for better visibility.

The idea of 2D spatial layout has been extended to 3D in the DataMountain system (Robertson *et al.*, 1998). This system allows users to place the documents on an inclined plane in 3D. It is specifically designed to take advantage of human spatial memory. The user can also associate some audio cues with the documents.

The techniques proposed by most of the previous works deal with web-site related data. Extension of these techniques to other domains raises the issues of visual representation of different data-types and accessing the data from the visual representation.

Chapter 3

System Design

"You know you've achieved perfection in design, Not when you have nothing more to add, But when you have nothing more to take away."

- Antoine de Saint

Retrieving information from the archived data is a complex process involving both the user and the system. While the computational media provide ways to access the information in a structured manner, the task of knowledge management is incumbent upon the user. This relation between the user and the system is the basis for our system design. In this chapter we discuss our design rationale and how we developed the system from it. We discuss various design issues associated with different aspects of the system and our approaches to address them.

3.1 Rationale

While designing the system, our main concern was to make it complex enough to support the user's various needs and simple enough so that anyone can walk up and start using the system. We started by asking: what should be the structure of the system? what features should be included in the system? and, how to keep various features simple and

consistent? The first step to answer these questions was to do a requirements analysis of the system and the second step was to decide on the design principles which we would follow throughout the design process. In this section we discuss these two steps one by one.

3.1.1 Requirements for the system

We decided the requirements for the system by considering the target user and possible tasks and goals of the user. Shneiderman identified 4 generic user tasks for browsing a web site as enumerated below (Shneiderman, 1997):

1. Specific fact-finding
2. Extended fact-finding
3. Open-ended browsing
4. Exploration of availability

These generic tasks encompass all the users using any general web site. We analyzed our specific problem considering the type of the archived data and the user of the system. The data for our system contain time synchronized lecture audio, presentation slides and audio transcripts. The target user of the system is any student who is interested in the lecture material. We found the first 3 tasks enumerated above to be of great relevance. Here we describe each of them in the context of our data and the target user. We also explain how different features of the final system are derived from these tasks.

Specific fact-finding

This task is performed to find an answer to a specific question. In this case, the user can express her question in a well formed manner by one or more statements. The question forms the guideline for the user's browsing process. In our system, an example of this task could be finding out "how many total assignments are there in the entire semester?".

To help user achieve her goal the system should have the ability to take the question in some format as an input and lead her to the portions of the data which are relevant to the question. One option to achieve this ability is to provide the user with a query interface. Using this query interface the user can get several possible answers to her query and hence drastically reduce the search domain.

Open-ended exploration

In this task the user does not have a specific question at all before the search. The user knows about the context of the archive and explores the data to find interesting portions or facts. An example of this task could be any kind of review of the lecture material done by a student at her own leisure. In this scenario, the user often needs some cues to initiate the exploration. Therefore, the initial layout of the data plays an important role. After initiation, the user performs various exploration tasks using the browsing and visualization techniques provided by the system. These techniques could enhance the user experience in several ways from broad exploration to 'Eureka' stage.

Extended or related fact-finding

In this task the user does not have a well formed question. She has a vague idea of the question and uses the system for both formalizing the question and finding the answer. This task could also be considered as an intermediate between the two extreme tasks mentioned previously. The user might perform various specific fact-finding tasks to figure out the question itself. Some parts of this task could also be open-ended explorations. An example of this task could be finding out "what are the most important topics discussed in today's lecture?". In order to effectively support this task, the system must help the user in keeping track of the browsed data. This motivates us to allow the user to annotate and personalize the data while browsing.

The browsing process could be a complex combination of several instances of these tasks, therefore, the system must integrate various features seamlessly. On the basis of these requirements, we decided to facilitate querying, overview visualization, detailed browsing, annotation and personalization.

3.1.2 Design principles

Our system design is based on some common design principles. In this section we motivate these principles and look at some instances where these principles helped us in making design choices.

Consistency across data representations and interaction techniques

Two most important factors determining the behavior of the browsing system are data and interaction techniques. Effective visual representation for each data-type helps make the interaction consistent across various data types. In our system we chose content based representations for all the data types such as thumbnails for slides and transcripts for audio. This makes interaction techniques for various actions, such as content annotations, portion selection etc, dependent only on the action and not on the data type.

Data available: then and there

The database for the system contains various different kinds of multi-modal data, but they are all time-synchronized with each other. In order to find the desired information the user might have to access all the data types associated with a particular time period. This design principle suggests that the user should be able to access all the data types from one data type with minimum context switch. The integrated data access techniques in the system are primarily based on this principle.

Flexible layout and distribution of information

Since the task of knowledge management is performed by the user, the system must assist the user in properly distributing and arranging the information according to her choice. In our case, this distribution and arrangement involve various aspects of the system such as the visual layout of the data and audio playback options. This principle motivated us to design the interface in such a way that the user can leverage her potential to handle the information efficiently using simultaneous viewing and listening.

Easy for beginners and fast for experts

While the system should be easy to use for a beginner, a more experienced user should also be able use her expertise to perform the tasks more efficiently using the same widgets and techniques. The use of marking menu (Kurtenbach & Buxton, 1994) for various actions and selections in the system is motivated by this principle.

3.2 System and data overview

The system is designed to help students browse and personalize archived lecture materials after the lecture. The user can explore the data, annotate it and make her personal collections of data portions.

3.2.1 Captured data

A multi-site course in the department of Computer Science at the university had been selected for archiving. This course archive consists of time-synchronized slide and audio data of thirteen lectures. The instructor used electronic board (Smart Board^{TM}) for displaying slides and writing notes. These slides were captured as images along with the marks made on them and were also time-stamped. The audio data had been captured using two microphones: one for the instructor and the other for the rest of the class. The average length of each lecture was about 2 hours and each lecture had approximately

50 slides. The audio data was transcribed using ScanSoft Dragon®. Some part of the audio was transcribed manually for experiment and comparison purposes. All the data, captured and generated, was later stored in a central database in an XML schema.

3.2.2 Display platforms

The system can be run on two types of display platforms: desktop and high-resolution wall-sized displays. The wall-sized display used for the system is a back-projected screen and consists of an array of 18 projectors arranged in 3 rows and 6 columns. Each projector has a resolution of 1024×768 pixels. Hence, the overall resolution of the wall-sized display is approximately 6000×2000 pixels. The rendering is done using Chromium software. The high resolution of the display allows large amount of data to be displayed at once. We will discuss the advantages in details in section 3.4.

3.3 Timeline based visualization

From the requirement analysis for the system we found that the initial layout of the data should give an overview of the data along with the ability to explore the details of any particular part. The timeline-based approach has been a time-tested approach for fulfilling this purpose (see section 2.2). Our system also adopts this approach. In this section we describe various issues that we addressed while designing a timeline-based interface for overview visualization and brwosing. We discuss the advantages and disadvantages of various design options and finally describe the approach used in our system.

3.3.1 Design options and issues

We considered two kinds of timelines for our application: single timeline and multiple timelines.

Single timeline

If the archive has only one media type then this option is an effective way of visualizing the activities. Single timeline also works well under screen space constraints. In case of desktops, the 'context + details' design paradigm is difficult to implement with multiple timelines. If the archive has multiple media streams then there are various issues in using this option. We found that determination of one major timeline for browsing the meeting archives depends on the meeting type. For example, when the slides are the guiding media stream for a meeting then the activities can well be represented by the slide timeline, but in the case of plain audio based meetings audio timeline should be the guiding timeline.

In (Brotherton *et al.*, 1998), a single timeline is generated from multiple timelines in an interesting way. This single 'focus of attention' timeline was automatically generated by combining significant changes in various different timelines. While this approach deals with the screen-space constraints very efficiently and it can also show major changes in all the timelines at once, it is not clear how should the single timeline display all the 'focus of attention' points. Moreover, it could potentially ignore useful information patterns in the individual timelines, e.g. discussion pattern in the meeting audio data.

Multiple timelines

Multiple timelines can show various information about different media streams, but there is a trade-off between screen space and displayed information. When the archive contains several media streams then it is impossible to show all of them together. The severity of the problem can be reduced if we can find some major media streams and just show these streams by timelines. Careful design of these timelines can also optimize the information displayed within space constraints. Further, customizable timelines is another option for dealing with this issue. In our system the user can customize the display of timelines by overlaying multiple timelines to save screen space. Various visual properties such as transparency, color, shape may be used for such customization features.

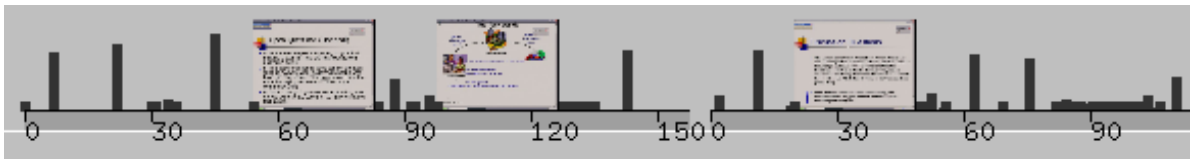

Figure 3.1: Slide timeline with important slides represented as thumbnails and other slides as bars

If a media stream is not displayed by a separate timeline then the data content of that stream must be made available unobtrusively. To achieve this, the constraints on the number of timelines displayed should be circumvented by fluid interaction techniques to access the data not readily available. We employed various such technique in our system and they are discussed at various places in this thesis.

3.3.2 Final design

The data used in our system consists of lecture slides, audio and corresponding transcript data streams. In our final design we have two timelines: slide timeline and transcript timeline. The detailed discussion of the advantages of using the transcript timeline over the conventional signal based audio timeline can be found in section 3.5. Here we describe the features of these timelines.

The slide timeline

This timeline is visible at the top of the screen all the time and gives an overview and contextual information to the user. Since a thumbnail of the slide gives a very good overview of the structure of the slide, in our initial design we showed the thumbnails of all the slides. But due to large number of slides (approximately 50 slides in each lecture) we found this design infeasible even on the high-resolution wall-sized display. Therefore, we decided to show all the slides as vertical bars grouped by the lectures. The height of a bar is proportional to the duration for which it was shown up during the lecture. Every lecture has some very important slides and we wanted to preserve the advantages

Figure 3.2: Transcript timeline showing words per minute as bars

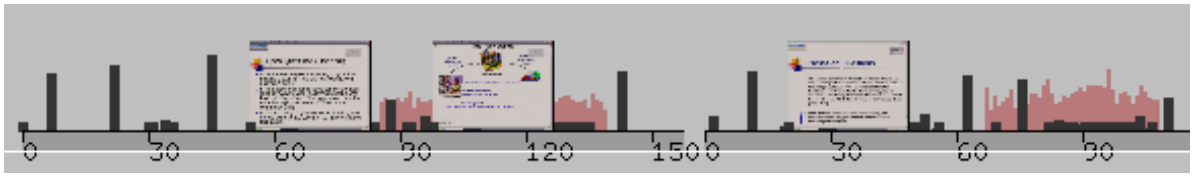

Figure 3.3: Overlaid slide and transcript timelines

of the thumbnails for at least these important slides. Therefore, in every lecture's timeline, we show some of the most important slides as thumbnails. When the application initiates, the importance criteria is the length of the time period for which the slide stayed up during the lecture. The importance criteria is dynamic and can be changed during the course of the archive browsing. For example, querying can change the importance of the slides based on the relevance of the slides to the query. This query based importance is explained in greater details in section 3.4.

The transcript timeline

This timeline shows the amount of speech transcribed per unit of time. The transcript based timeline gives a better overview of the content than the signal based timeline. This timeline can show some interesting properties of the audio data. For example, when the speaker is talking fast (i.e. more words per unit of time) then the bars in the corresponding time period will be higher than those time periods in which speaker is talking slowly. While the signal based timeline gives an overview of the periods of silence and speech, it is difficult to figure out the content overview of the speech.

3.4 Browsing by zooming and querying

3.4.1 Design issues

In our system we allow browsing by direct zooming and querying. Our application specific two timeline based design has raised several issues related to the browsing interface. In this subsection we discuss these issues. Later subsections will explain our approach to address these issues.

Context + Details

The problem of focusing on details while still maintaining the context of the data has been widespread among all the browsing systems involving large amount of data. This issue can manifest in various forms depending on the type of the data and the purpose of the browsing. In timeline based visualization, the timeline gives a very high-level overview of the time synchronized data, but to get the details of a portion the user must be able to focus on that portion without losing context. In our particular application, for various reasons the user might want to focus on one or more regions, compare them, or place them in her own way without losing the context. One possible scenario could appear when the user wants to compare one slide in the first lecture with the one in the last lecture. This requires techniques which allow the user to access various parts of the timeline simultaneously and with minimum loss of context. There are also issues related to the layout of the displayed information, such as how much data should be displayed, how the displayed data should be organized on screen etc. We address these problems using multi-focus zooming and some other techniques.

Accessing relevant and related data

The search for the region of interest or pin-pointing the portion of relevant data is an integral part of any browsing process. Most of the times the pin-pointing involves a sequence of small searches through which the user tries to access the data related to those

which are currently available. This process becomes more complicated in the case of multiple streams where both time and data type are varying. We address these issues by providing an integrated query interface and making every data type accessible from every other data type.

3.4.2 Multi-focus zooming

Motivation

The context + details issue has been addressed in our system by using a specific zooming interface. The zooming interface has been shown to be effective for browsing temporal data or data streams. The study reported in (Hornbk & Frokjer, 2001) proves that reading of electronic documents is more efficient when an overview detail interface is used. In our system the user can get an overview from the timelines and select a particular portion of the timeline to zoom in to that portion. While deciding on what type of zoom to use for our particular classroom meeting data we found that the type of zoom depends on the type of data and the purpose of zooming. In our design we wanted to provide the user with an interface where she can not only look at the details of a slide, but also compare various slides and search for the interesting portions among them. While the systems described in (Mills *et al.*, 1994; Ramos & Balakrishnan, 2003; Plaisant *et al.*, 1996; Bade *et al.*, 2004) use only hierarchical zooming because video and contiguous medical data have a very dense nature, our design requirements suggest the use of multi-focus zooming.

The interface

Using the multi-focus zooming interface the user can select one or more portions of slide and audio transcript timeline with left mouse drag. Each selected part opens an independent browsing platform which contains the zoomed-in view of the selected slides. This platform opens with slow animation so that the user can keep track of the context of the zoomed portion. This portion contains a set of three slides. On the center we show the

Figure 3.4: Multi-focus zooming showing details of 3 different parts of the timeline

selected slide and on the two sides we show two neighboring slides. Even our classroom data slide timeline could become so dense that the user might, by chance, select a large number of slides just by selecting a small length on the timeline. In such a case, instead of flooding the user's view field with numerous slides, the system shows only the first 3 slides of the selection. But the user can browse through all the slides by dragging the mouse sideways which successively pulls the next adjacent slide inside the view while the trailing slide on the other side slowly gets moved out of the view. The user can zoom in to as many portions as she wants and can spatially arrange these zoomed slides. She can also adjust the size of these zoomed portions to fit maximum number of them on the screen. This allows the user to look at the data located at even two extreme ends of the timeline in parallel and without switching context. Figure 3.4 shows an instance of multi-focus zooming with three different browsing platforms opened.

In the case of transcript timeline, using multi-focus zooming the user can open multiple transcript browsers. A transcript browser shows the transcript corresponding to a selected portion of the audio in a text box. We discuss various features of the transcript browser in section 3.5.2.

Multi-focus zooming and large-scale display

Due to the screen space requirements of a multi-focus zooming interface, we found it difficult to fully explore its advantages on the desktop screen. The desktop screen gets cluttered if the user zooms in to more than 3 different locations. Therefore, we ran the application on a high-resolution wall-sized display (details of the display can be found in the section 3.2.2). On this display, the user can zoom into several portions of the timeline without any cluttering and arrange their sizes and locations. The ability to simultaneously keep various zoomed portions in sight minimizes the context switching. Figure 3.13 shows the interface on a large-scale display.

3.4.3 Querying

Querying is an effective way to start the browsing process. While an effective browsing interface speeds up the browsing of the data corresponding to a selected portion of the archive, the querying facility boosts the process by pin-pointing the portions that the user should browse. We can consider a scenario in which the user finds some unknown terminologies while reviewing a lecture. Our design rationale suggests that the user should be able to get as much of the related information as possible, then and there. To this end, we created a query interface and integrated it with the browsing interface.

The querying interface is connected to a query engine across the network using simple HTTP protocol. Using this engine, the user can make queries on the whole database. The system provides two ways to query the archive database. Firstly, the user can right click anywhere on the blank part of the screen to invoke a marking menu and select the query option from there. This opens a query space in which the user can directly type in the query string. Secondly, the user can query for any string found in the transcript. Using a simple mouse drag, she can select any text string inside the transcript browser and use the string as the query string. The query engine searches the entire database, including text inside the slides and the audio transcripts. It returns all the slides which are relevant

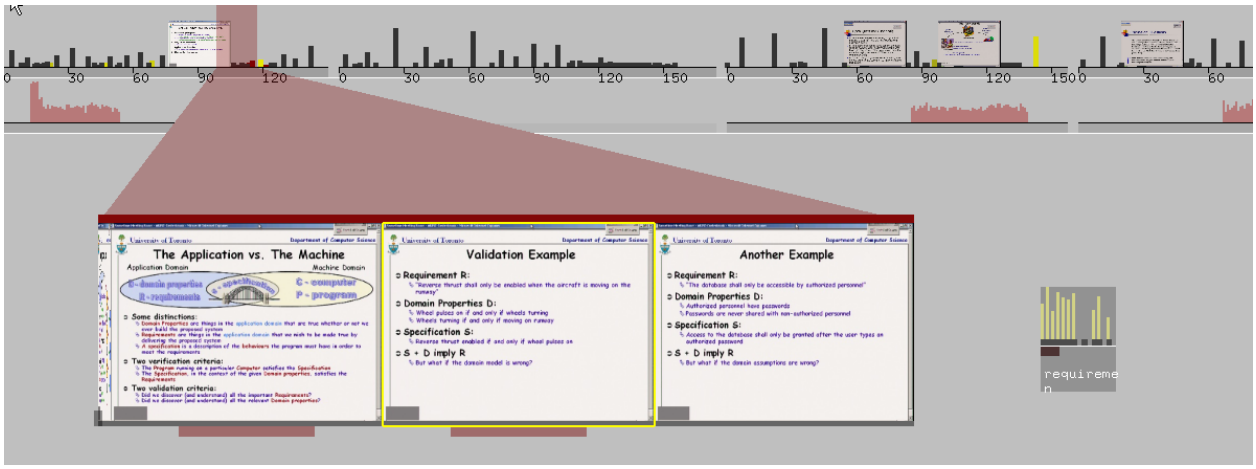

Figure 3.5: Query results showed using map-metaphor by highlighting the results in yellow and a compact query browser showed at the bottom-right

to the string. The results returned are assigned relevance values on the scale of 1 to 10.

While deciding on how to display the query results, we noted that at the time when the user makes the query, several other data types might have already been laid out on the screen. In such a case, it is desirable that query results be displayed with minimum interference with the earlier layout. To achieve this we used an approach which is similar to the one used for displaying geographical query results on a map. The whole screen space is assumed to be a map with various data types spread on it. The query result highlights all the relevant portions on the display in various shades of yellow. These shades are determined by the relevance values of the highlighted portion.

The system also shows all the results together in a compact query browser (see figure 3.6). In this browser all the query results are represented as a vertical bar chart grouped by lectures. The height of the bar represents the query-relevance value of the corresponding slide. The user can view the thumbnails of four consecutive slides at a time. By shifting a slider, provided at the bottom of the bars, sideways the user can view all the result slides in sets of four thumbnails. The slide corresponding to the thumbnail can be accessed from the timeline by clicking on the thumbnail. Clicking on a thumbnail has the same effect as selecting the corresponding slide from the slide-timeline.

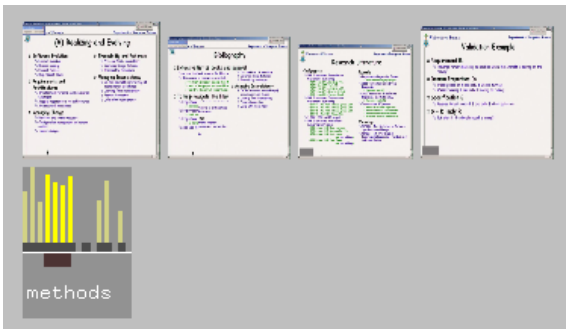

Figure 3.6: Compact query result browser with selected slides shown as thumbnails. Four consecutive slides corresponding to four bars above the slider have been shown as thumbnails.

3.5 Audio and audio transcripts

Previous research has shown that the use of transcripts can make audio browsing faster and more efficient. In this section we discuss the issues related to the use of transcripts for browsing audio, and how we address those issues in our system.

3.5.1 Issues

Transcript generation

The use of transcripts for keeping records of proceedings has been a common practice in the past in specific places like political meetings (e.g. Canadian house of parliament), courts of law and infirmaries. In most of these cases the speech is manually transcribed by professional transcribers, which is an expensive way of generating transcripts. Considering the amount of speech data that could potentially be captured from numerous audio-conferences and e-meetings, manual transcription (which is the conventional way of generating transcripts) does not seem to be a feasible way of generating transcripts.

Quality of the transcription

Recent advances in speech recognition technology have made automatic speech transcription possible. There are various speech recognition based commercial products available

for this, e.g. IBM ViaVoice®, Dragon NaturallySpeaking etc. The accuracy of the automatically generated transcripts depends on various factors, such as the quality of the recording, the training process and the knowledge base provided. But the transcripts generated by this process are often not accurate. The effective use of these error-laden transcripts for accessing audio has recently emerged as an interesting field of research.

The integrated interface

Recent research has shown evidence that integrating audio with erroneous transcripts makes some specific audio browsing tasks more efficient. The user can get an overview of the audio by browsing the transcript text and if some portion of the transcript seems incorrect then she can listen to the actual audio corresponding to that portion. This process of interactive browsing demands a tight integration of the audio with the corresponding transcript and fluid interaction techniques for accessing the audio from the transcript and vice-versa. Moreover, integration of various properties of audio signal (pitch, pan, loudness etc.) with the partially correct transcripts needs further exploration. In the following subsections we describe how we integrate transcripts with the audio. We also discuss the use of partially correct transcripts with spatial audio.

3.5.2 Integrating transcripts and audio

The interface

The transcript browsing interface in our system is similar to a text box with a scroll bar. We call this text box a transcript browser. The transcript displayed in this browser is time-synchronized with the audio. The text corresponding to one minute of audio is shown together in one paragraph. The starting time of the audio corresponding to the paragraph is shown on the left margin of the text window. The text can also be divided into paragraphs based on speaker pauses. This could be a better and more informative representation of the underlying audio but it requires prior speech processing to find out the

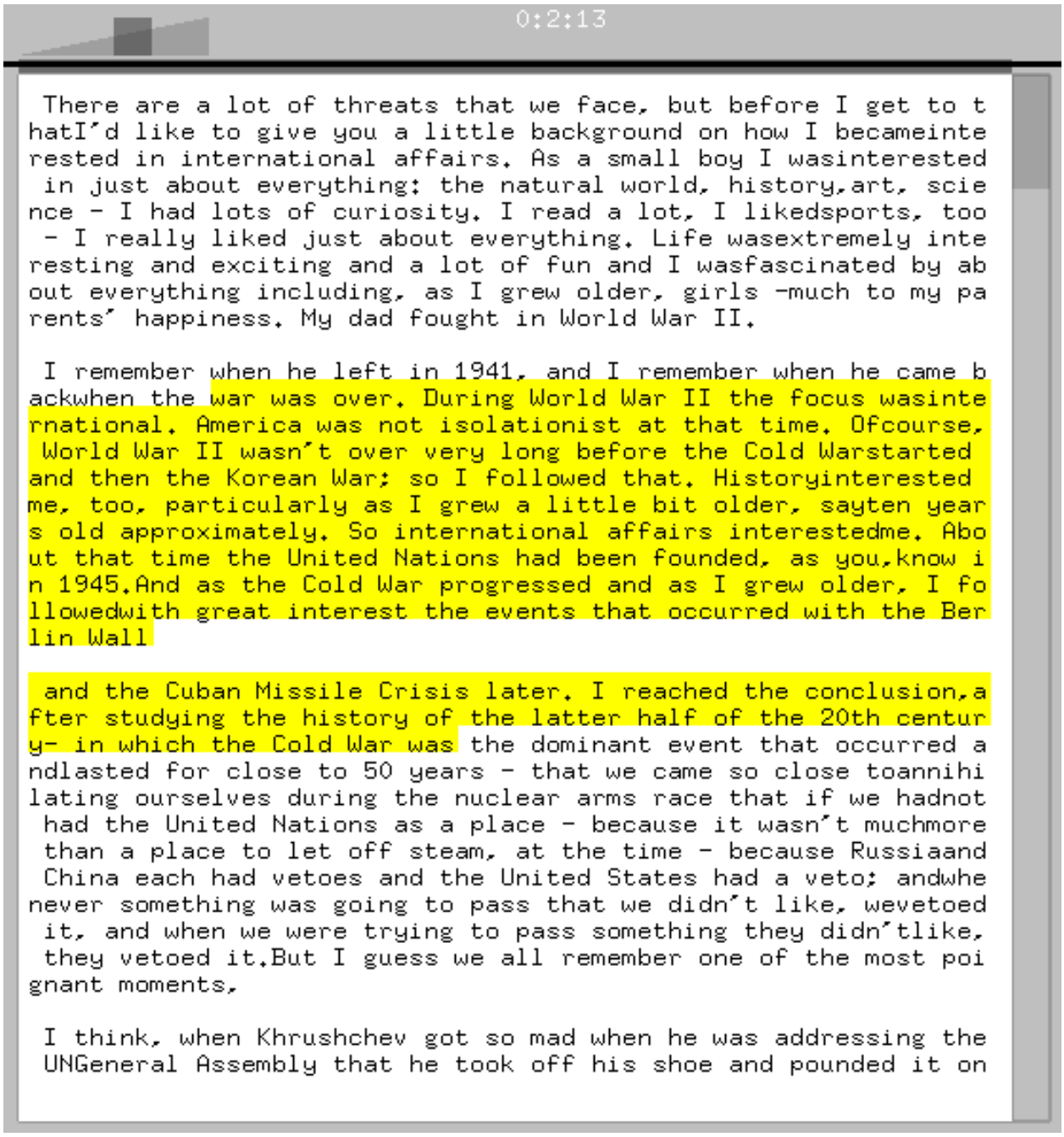

Figure 3.7: Transcript browser with portions being played shown in highlighted background

pauses. The system also supports direct access to audio by a traditional audio seek bar at the bottom of the transcript browser window.

Interaction techniques

The system provides various interaction techniques for listening to the audio corresponding to some portion of the transcript and viewing transcript corresponding to the audio being played. These techniques form the key to the interactive audio browsing. When some part of the underlying audio is not comprehensible for the speech recognition engine, then the corresponding transcript is garbled. In such a situation users might want to listen to the corresponding audio. Moreover, plain text can not capture the expressiveness of speech. One can think of various scenarios in which, despite the accuracy of the transcripts, the user might want to listen to the audio to know about some aural properties of the speech (accent, tone, pitch etc). The user can listen to the audio by clicking on the desired portion of the transcript. The system highlights the portion of the transcript which has already been played. The highlighting of the text shows the context of the audio being played. Figure 3.7 shows these features of the system.

3.5.3 Spatial audio and transcripts

Using multi-focus zooming, the user can zoom in to as many portions of the transcript timeline as she wants. This allows her to view the transcripts and listen to several portions of the audio simultaneously. Each opened transcript browser acts as a sound source in one dimension along a horizontal line and its position on the horizontal line is determined by the x-coordinate of the transcript browser window. The system uses spatial audio techniques (pan value adjustment) to map these sources to their corresponding positions in one dimensional space and the user feels that the audio from a transcript browser is coming from the direction where the browser is spatially located on the screen. For example, in figure 3.8, the audio corresponding to the transcripts labeled T1, T2 and T3 will appear to be coming from left, center and right, respectively.

In a controlled laboratory experiment we found that for search tasks, simultaneous-spatial audio browsing with transcripts is faster than serial audio browsing. In chapter 4

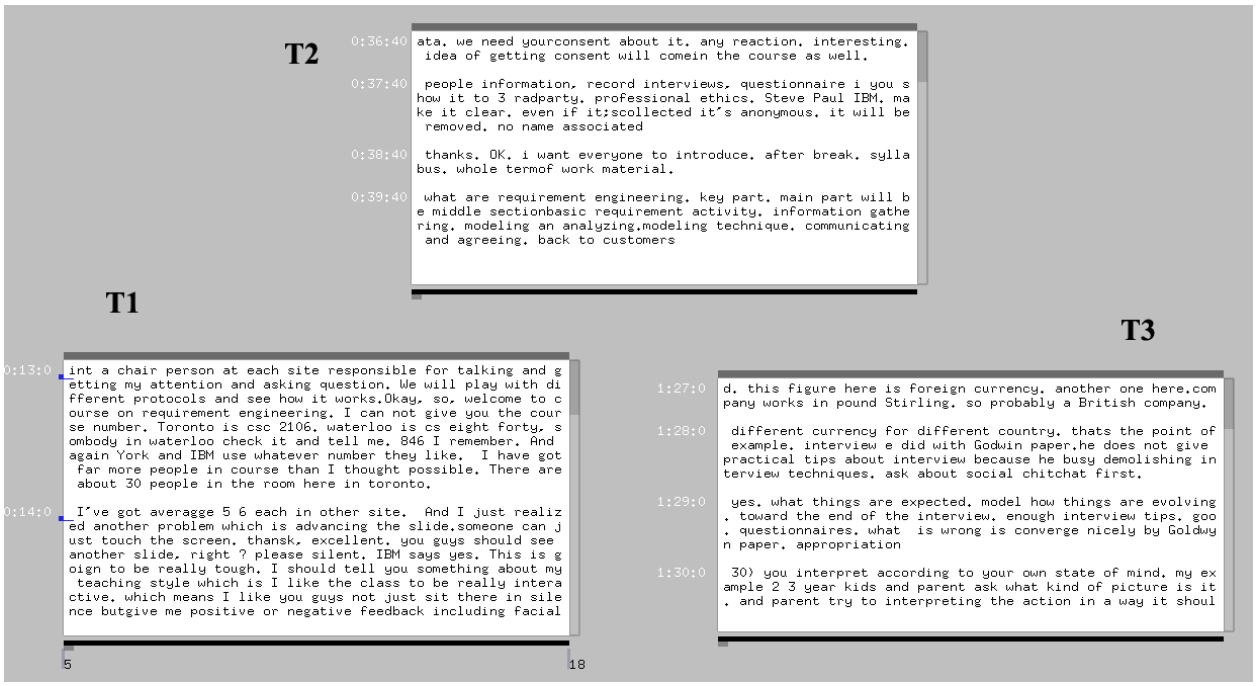

Figure 3.8: 3 Transcript browsers arranged spatially. The user listens to the stream **T1** from the left, **T2** from the center, **T3** from the right.

we discuss the experiment and the results in details.

3.6 Integrated data access

Since all the data is not readily available all the time through timelines, it is imperative that the system should provide indirect, yet seamless, access to the data which are not readily available. The database has 3 kinds of data: audio, synchronized transcripts and synchronized slides. The system provides various data-integration techniques to access various data types in time-synchronized manner. All these techniques have been made consistent with the other browsing techniques. In this section we discuss the techniques provided in the system to access audio and transcript from slides and slides from audio and transcripts.

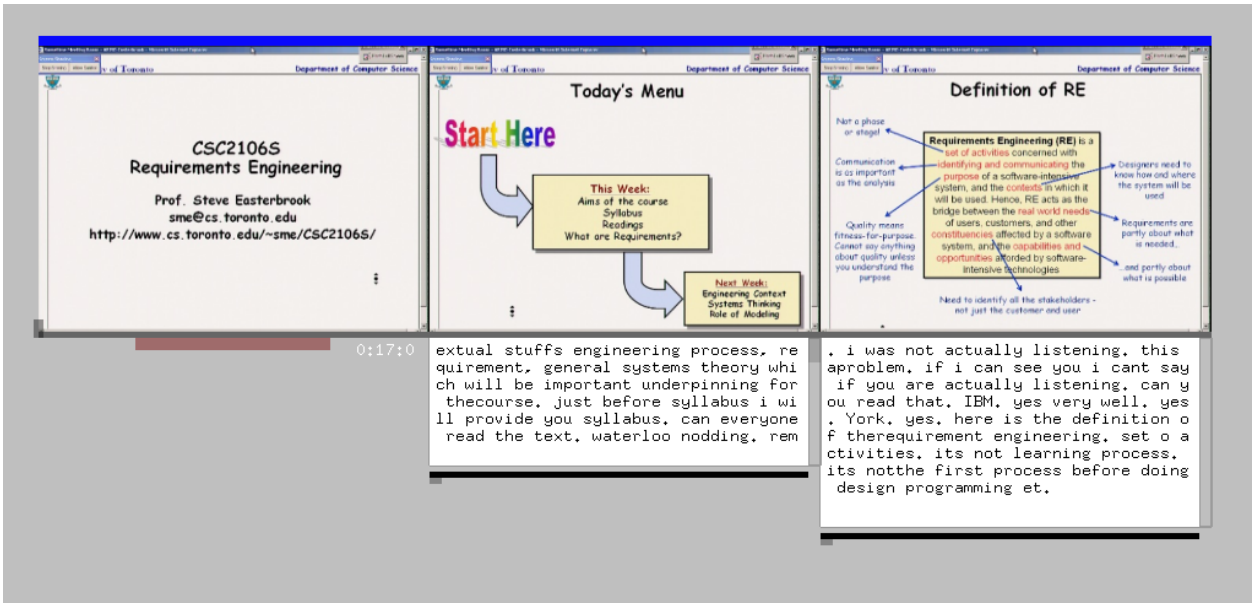

Figure 3.9: Integration of transcript with slides. The leftmost slide shows the bar at the bottom which can be used to pull down the corresponding transcript. The slides in the center and on the right show the corresponding transcripts pulled to different lengths

Accessing audio and transcript from slides

In our system, every slide has a time period associated with it. This is the duration for which that slide was displayed during the lecture. The user can listen to the audio corresponding to that time period by clicking on the slide. Every slide also has the transcript of the audio associated with it. The user can pull the bar provided at the bottom of the slide to view the transcript in a transcript browser. The size of the transcript browser can be adjusted by pulling the bar up and down. These transcript browsers are attached to the slides and are moved and resized along with the corresponding slides. Figure 3.9 illustrates this integration interface.

Accessing slides from the transcript

The user can access slides from the transcript since both the transcript and the slides are synchronized with each other. When the user is viewing the transcripts in the transcript browser, she can select any portion of the transcript. The selected portion corresponds

to some time period and by a menu selection the user can view the slide which was up during that time period. This way of slide selection is exactly equivalent to selecting a portion from the timeline. The system zooms in to the slide timeline portion corresponding to the transcript selection made by the user. The slides are shown in the same manner as they would have been shown in response to a direct slide timeline selection.

3.7 Personalization

Abrams et al. studied the strategies people use for managing large amount of information (Abrams *et al.*, 1998). They found,

"One strategy for dealing with information overload is to develop personal information systems consisting of focused subsets of information highly relevant to a particular user."

This underscores the importance of personalization in any browsing process. In our case, while browsing the data the user might find several interesting portions, relations among them or hit upon some new ideas. The user can capture these events by adding her own comments to the data or by creating personal collections. The system provides techniques to facilitate these activities. Since the audio data is pivotal in our system, we focus on the techniques for audio annotation. In this section, we discuss these personalization techniques under two broad categories: annotating audio and personal collection creation.

3.7.1 Annotating audio

Most of the systems use some visual representation of the digital data, such as text, images, video frames, and allow its annotation by pen marks, images, or hyperlinks. In our system we explored the use of voice data to annotate transcript based representation of audio. Here we first discuss the motivation behind this particular annotation choice and

then describe the techniques.

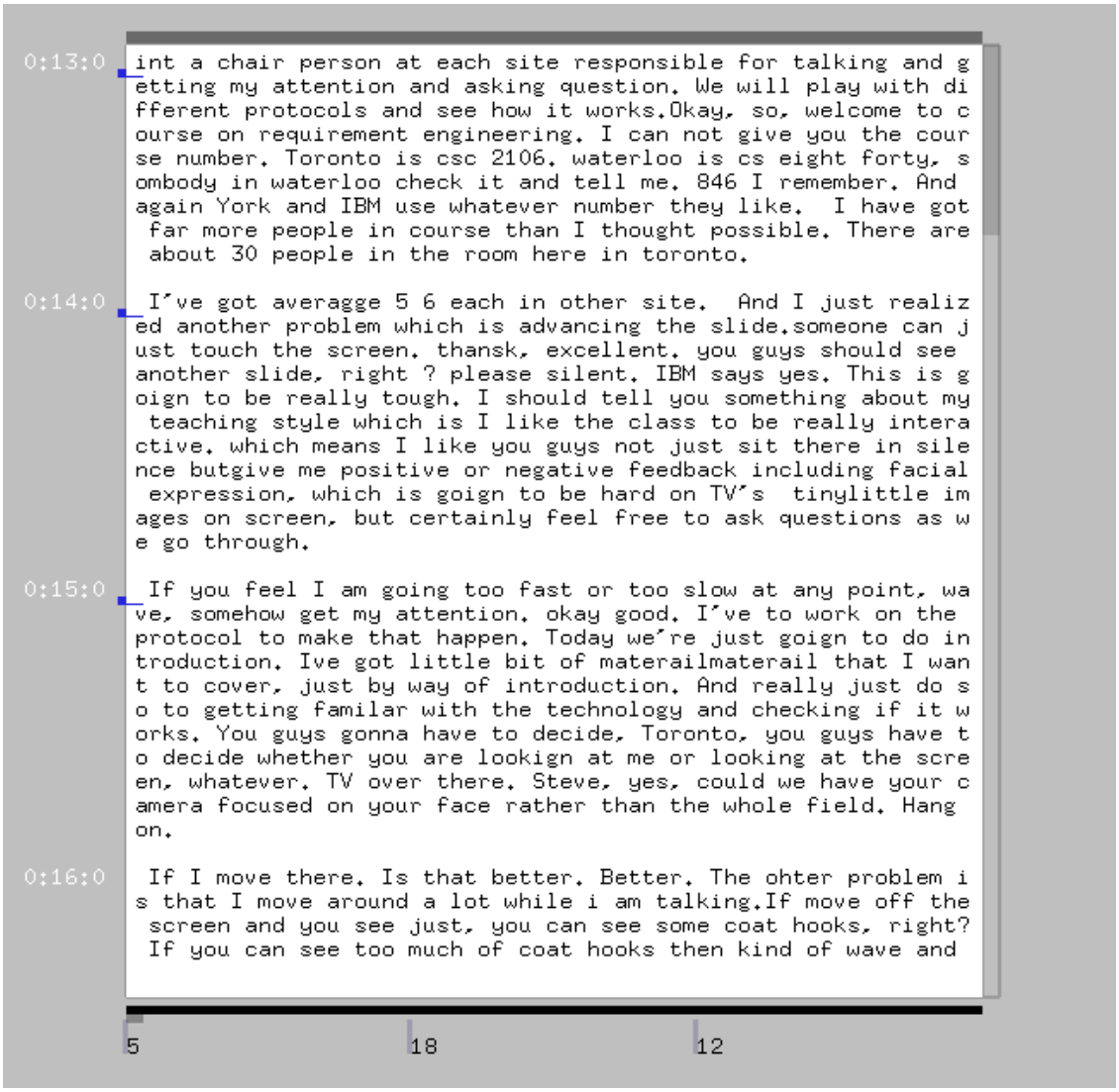

Figure 3.10: Voice annotation of transcripts. The blue marks on the left edge show the speech annotation positions in the transcript text and the numbers (with a small bar) at the bottom of the transcript browser show the corresponding annotations' time lengths. The user can select these bars to play the annotations.

Voice annotation of audio transcripts

Since the conventional signal based visualization of audio does not provide a good visual representation of the content, it is difficult to effectively annotate the contents of a large amount of audio. Recent works (Whittaker *et al.*, 2002; Whittaker & Amento, 2004; Vemuri

et al., 2004) suggest that transcript based representation gives the user a better idea of the content, therefore, we use this representation for annotating audio.

Voice data has several advantages over other common ways of annotations (e.g. text, pen-marks, images etc). It provides a more expressive and non-intrusive way of annotation (Stifelman *et al.*, 1993; Chalfonte *et al.*, 1991). For example, while revising lectures students often have creative discussions or make interesting verbal remarks. When several students are listening to lecture audio they make several comments which can be difficult to capture by writing due to divided attention. In such a situation recording the portions of discussion, indexed into the lecture audio, would be an unobtrusive way of capturing the experience. Therefore, our system provides techniques to annotate the audio transcripts using voice data.

The annotation techniques

To annotate any portion of the audio, the user selects the corresponding text in the transcript browser and makes a menu selection. This starts the voice capture. A small offshoot bar from the audio timeline, located at the bottom of the browser, shows the position and length of the voice note. A microphone attached to the system keeps recording the audio until the user makes a stop signal by making a menu selection. The voice captured gets linked to the audio transcript portion and also to the corresponding audio data. All such voice annotations are shown as offshoots from the audio timeline at the bottom of the browser. To provide better context information for an annotation inside the transcript we also show small blue mark on the left of the browser and by the side of the corresponding transcript text. The user can drag the mouse across the offshoot to listen to the corresponding voice annotation. As the user attains expertise in using the system, she can smoothly switch from lecture audio to audio annotations and create her own audio experience by seamlessly integrating voice annotations with lecture voice.

3.7.2 Personal collections

The user can capture her browsing experience by creating her own personal collections and saving them. The process of creating personal collection could be a sequence of various steps. In this subsection we discuss the interaction techniques involved in these steps.

Creating personal space

The user can create one or more personal collection spaces anywhere on the blank space inside the screen using a marking menu option. The personal space is visually represented by a rectangular region. The user can add data to her personal space, organize them inside the boundaries of the space, and annotate them using free-form pen marks and voice data. Each personal space can be given a title and saved. When the user initiates the system, all previously saved collections are shown as thumbnails at the bottom of the screen. The user can open them by a mouse click.

Adding contents

The user can add various data types to her collection. For selecting a portion of the audio, the user selects the corresponding text from the transcript browser and separates it from the rest of the transcript by using a menu selection. Now she can place this visual representation of the audio snippet anywhere inside the personal space. Similarly, the user can get a copy of any slide from a zoomed portion of the slide timeline for saving it in her collection.

While browsing data the user might find several interesting query results. Each query result contains a set of data-portions having different degrees of relevance to the query. On an overall basis, this set forms a collection of related information. We can think of various scenarios where the user would like to save this collection as a whole. For example, consider a scenario where the user submits a query as “task analysis in requirement engineering” and the system returns a set of ten relevant portions which are somehow

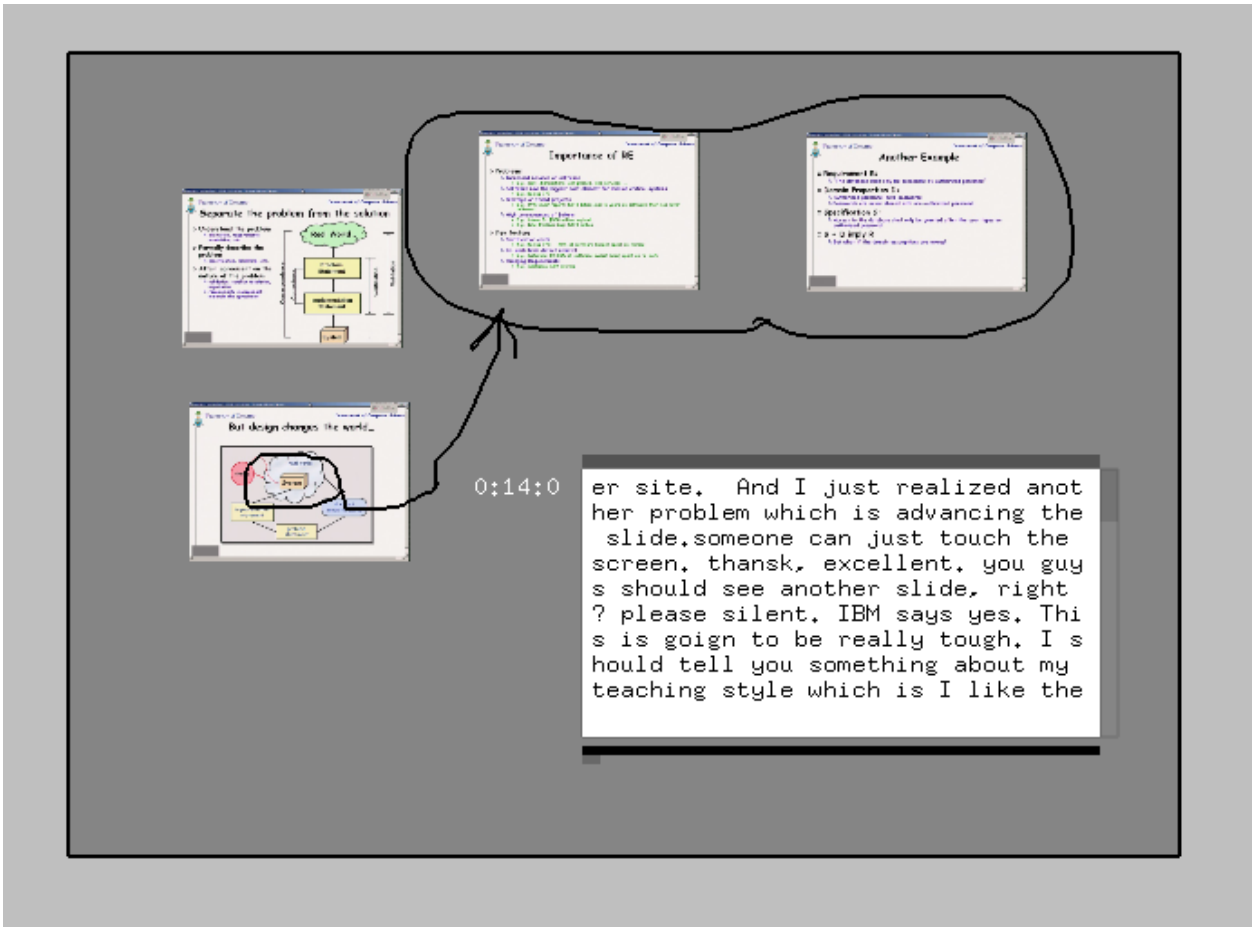

Figure 3.11: A Personal collection with 4 slides, a portion of the transcript on the right bottom and some organizational pen marks

related to the query. If the user finds the desired information to be distributed across all the results then she would want to save all the results as one entity. To facilitate this, we allow the user to access all the results collectively as well as individually. The user can get a copy of the query result and add its visual representation to the personal space.

Organizing collections

Once the user has some interesting components inside her collection, she can use free form pen marks to organize or annotate the collection. The system supports the creation of collection of collections. Using this feature the user can organize the data in a hierar-

chical manner. The user can get a visual copy of the collection using a menu option and add it to another collection. The system treats a collection in the same manner as it treats any other data type.

Post-it audio-notes

The user can annotate the collection using voice data. To facilitate this, we introduce a notion of post-it audio-notes which combines the advantages of spatial memory and richness of voice data. We can consider a scenario in which the user is making collections and spatially organizing various components of the collection. This process might also involve making some free form pen marks for visual organization and cues. While she is busy in this personalization process she might come up with some ideas or thoughts. The system provides a fluid interface for capturing such experiences via voice-recording with minimum context switch. Once the user makes a selection from the marking menu, voice capture starts. A small box, similar to a post-it note, shows the current time-length of the recorded audio. When the user stops the recording using a menu selection, the post-it audio-note changes to a playback interface with the time length of the recorded audio displayed on the top and an audio seek bar on the bottom. The user can now play the recorded audio. These post-it audio-notes can be placed anywhere in the personal space. This allows the user to annotate some spatial location inside her personal space by voice.

Accessing data from the collection

The user can access the complete information, along with the context, about the components of a collection by selecting the component. If the user selects the transcript portion then systems shows the entire transcript of which the selected transcript was a part. This transcript is shown in a transcript browser with the selected portion highlighted in red. If a slide is selected, then the corresponding part from the timeline gets zoomed-in. Similarly, the individual results from the saved query result can be accessed by selecting them inside the query browser.

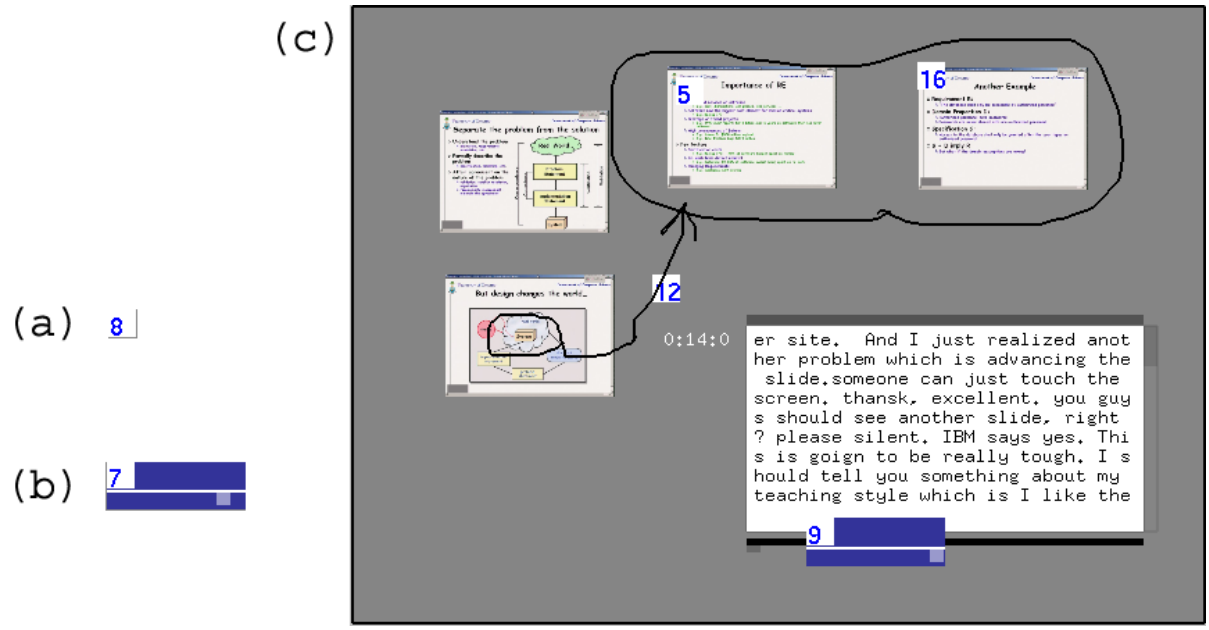

Figure 3.12: Post-it audio notes: (a) A collapsed post-it audio-note, (b) A post-it audio note with playback interface, (c) Various post-it audio-notes used in a personal collection. They can be switched from playback-mode to collapsed mode and vice-versa by mouse click.

Collections and large-scale display

While designing the interface for making collection of collections we found a trade-off between the use of spatial memory and the level of hierarchy. As the number of levels increases, objects inside a collection get cluttered and it becomes difficult to recognize the spatial layout of the component. This hampers the use of spatial organization as a memory cue. Therefore, we decided to allow two levels of hierarchy to keep a balance. We found that the use of large-scale high resolution display helps solving this problem. On large scale display, the collection creation techniques can also be used efficiently to create multiple personal collections at the same time by one or more users. Figure 3.13 shows the interface on a large-scale display.

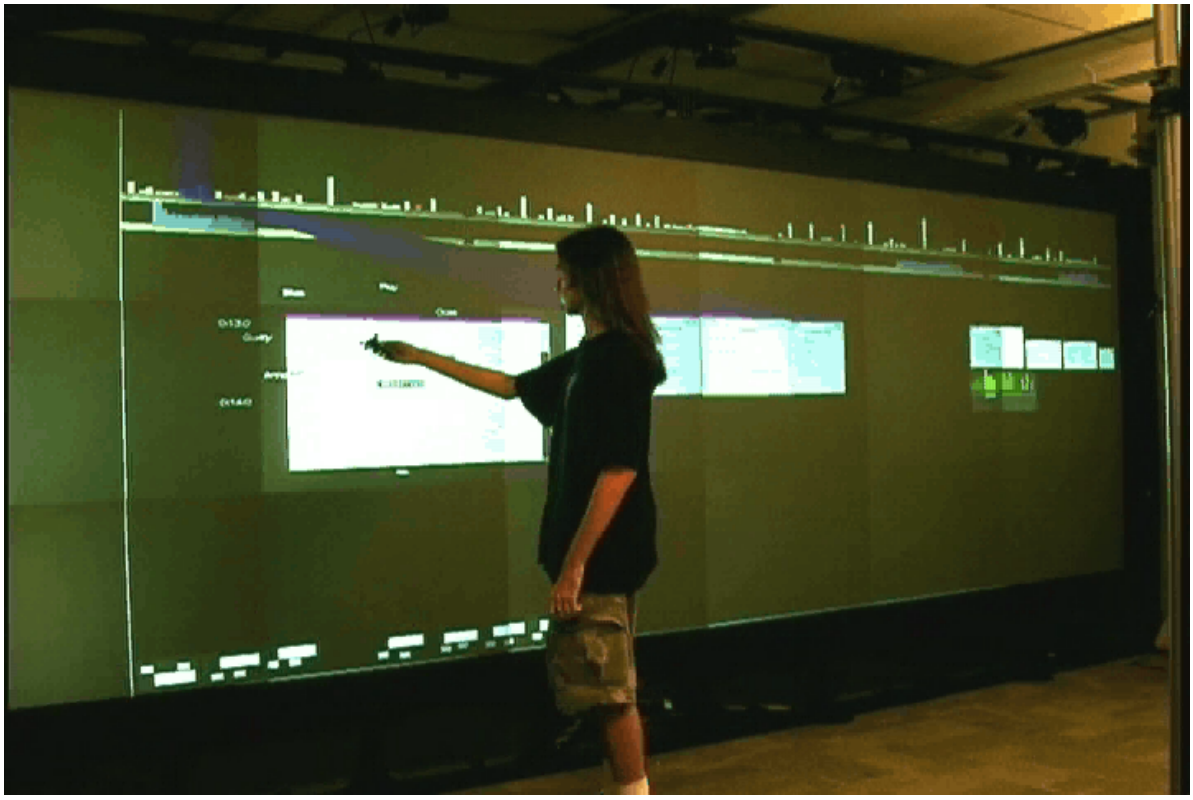

Figure 3.13: The system running on a large-scale display with a user interacting with it using a digital pen. Various features (transcript-browser, zoomed slides, query-browser) of the system are displayed at various position on the display.

Chapter 4

Evaluation

"The great tragedy of science - the slaying of a beautiful hypothesis by an ugly fact."

- Thomas H. Huxley

Our system introduces audio browsing techniques based on transcripts and time-compressed simultaneous-spatial audio. In this chapter we motivate the need for a formal evaluation of these techniques by analyzing them and describe the experiment that we designed for this. We discuss the results of the evaluation and also the implications of the findings.

4.1 Motivation

Previous works have studied simultaneous-spatial audio based techniques for search tasks in combination with comprehension tasks (section 2.1.2). But our interface provides a combination of time-compressed simultaneous-spatial audio with transcripts specifically for search tasks. While designing this new interface we asked several questions about the efficiency of the interface. Can people find facts in two simultaneous-spatial audio streams? Does it help in the search task if the user can see transcripts along with simultaneous-spatial audio? Does the inclusion of time-compression, transcript and spatialization in the interface make the search process faster than the conventional single-

stream audio playback interface? To get definitive answers to these questions we conducted a controlled laboratory experiment.

4.2 Goals

The overall goal of the experiment was to test the effect of various audio browsing techniques on the user's efficiency in search tasks. We compare the time the user takes to successfully find the answers to particular queries inside the audio clips using conventional techniques and our techniques.

In order to achieve this goal, we first formalize the tasks and the techniques involved. On the basis of these formalizations we setup the hypotheses and the experiment to test them.

4.3 Task

The task for which the techniques are tested is *search task*. In the experiment, the subjects are first given questions and then they listen to some audio clips using different interfaces. The task is to find the answers to the questions in the audio clips. On the basis of the questions, the subjects form a set of keywords which would characterize the answer and search for those keywords in the audio clips. There is exactly one question for every audio clip and this process is repeated for several different audio clips.

4.4 Techniques

The subjects used both conventional audio accessing technique and our new audio accessing technique. Our audio accessing technique differs from the conventional audio playback techniques due to two features: the use of transcripts and the simultaneous-spatial audio playback at both normal and faster speeds. In order to test the advantages

of our technique over the conventional techniques, we label various features of the audio browsing techniques. This labeling will help us define the techniques, which are compared and tested in the experiment, in terms of the feature-labels.

Alternative types of streams:

- *S1*: The user can listen to the audio in one stream (mono or stereo) and play the audio from any point of the stream.
- *S2*: The user can play two different parts of the audio clip simultaneously and with spatialization.
- *S3*: *S3* has all the features of *S2*, and in addition the user can also play the streams at higher speeds.

Second, we label the quality of the transcript associated with the technique.

- *Q1*: This corresponds to approximately 100 % correct transcript. In practice, these transcripts are created manually.
- *Q2*: This corresponds to 60 to 70 % correct transcript. These transcripts represent automatically generated transcriptions.
- *Q3*: This corresponds to the case where no transcript is used in the interface.

In terms of these labels we can define and label various techniques. For example the technique with two audio streams and normal audio speed will be labeled as *S2*; a technique with one audio stream and perfect transcript will be labeled as *S1Q1*. In this experiment we compare the user's performance in search-task when using techniques with various combinations of these features.

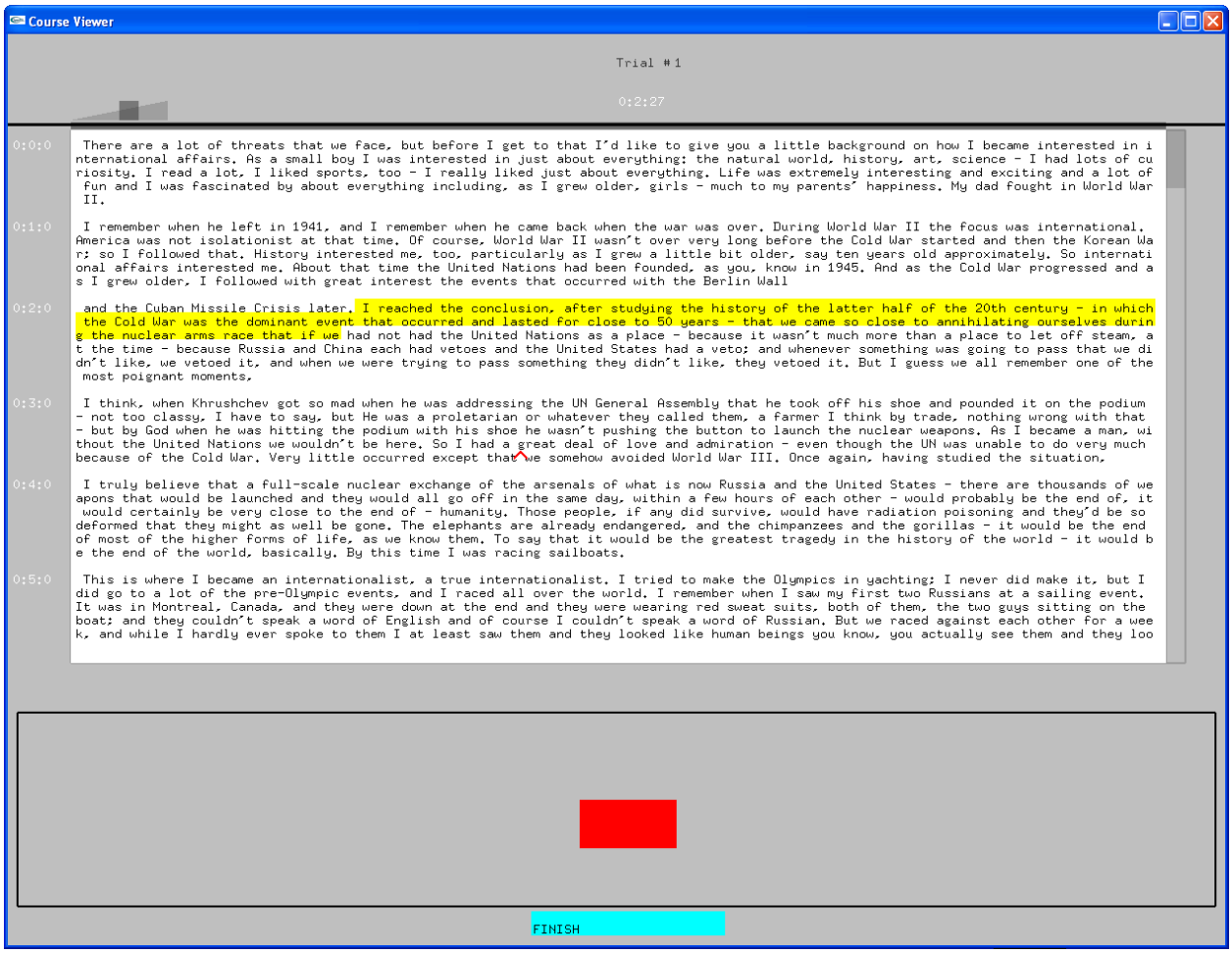

Figure 4.1: Experiment interface: single stream (*S1*) interface with transcript. The transcript portion being played is highlighted in yellow.

4.5 Interface

The interface for the experiment is similar to the transcript browser interface used in the system (section 3.5.2). The interface used by the subjects can be broadly divided into 2 types: 1 stream interface and 2 stream interface.

- *1 Stream*: Figure 4.1 shows the 1 stream interface with the transcript. For *Q3* type techniques, the transcript browser shows dots instead of letters. The audio can be played or paused by clicking inside the black rectangle below the transcript browser. The color (red or green) of the indicator at the center of the rectangle shows if the

audio can be played (when the color is green) or paused (when the color is red). All other interaction techniques are similar to the transcript browser.

- *2 Streams*: Figure 4.2 shows the 2 stream interface with the transcript. It essentially consists of two *1 Stream* interfaces integrated in a specific manner. Both the streams are time-synchronized and can only be played simultaneously and synchronously. The browser on the left can play audio only in the left ear and the one on the right can play only in the right ear. The subject can adjust the volume of each stream separately, but can not play them separately. The subjects must listen to both the streams all the time. This compulsion has been used to ensure that one particular set of subjects always uses simultaneous-spatial audio.

4.6 Hypotheses

The interface alternatives are evaluated using the following hypotheses. In the following statements, the term 'search time' refers to the time taken by the user to find the answer to a well formed query inside the audio data.

- *H1*: The combination of speed and spatialization reduces the search time. More specifically, $S1 > S2 > S3$.
- *H2*: The search time reduces as the transcript quality gets better. More specifically, $Q1 < Q2 < Q3$.
- *H3*: The combination of transcripts and simultaneous-spatial audio reduces the search time. More specifically, $S2Q1 < S1Q1$, $S2Q2 < S1Q2$ and $S2Q3 < S1Q3$.
- *H4*: Adding time-compression to simultaneous-spatial audio and transcript further reduces the search time. More specifically, $S3Q1 < S2Q1$, $S3Q2 < S2Q2$ and $S3Q3 < S2Q3$.

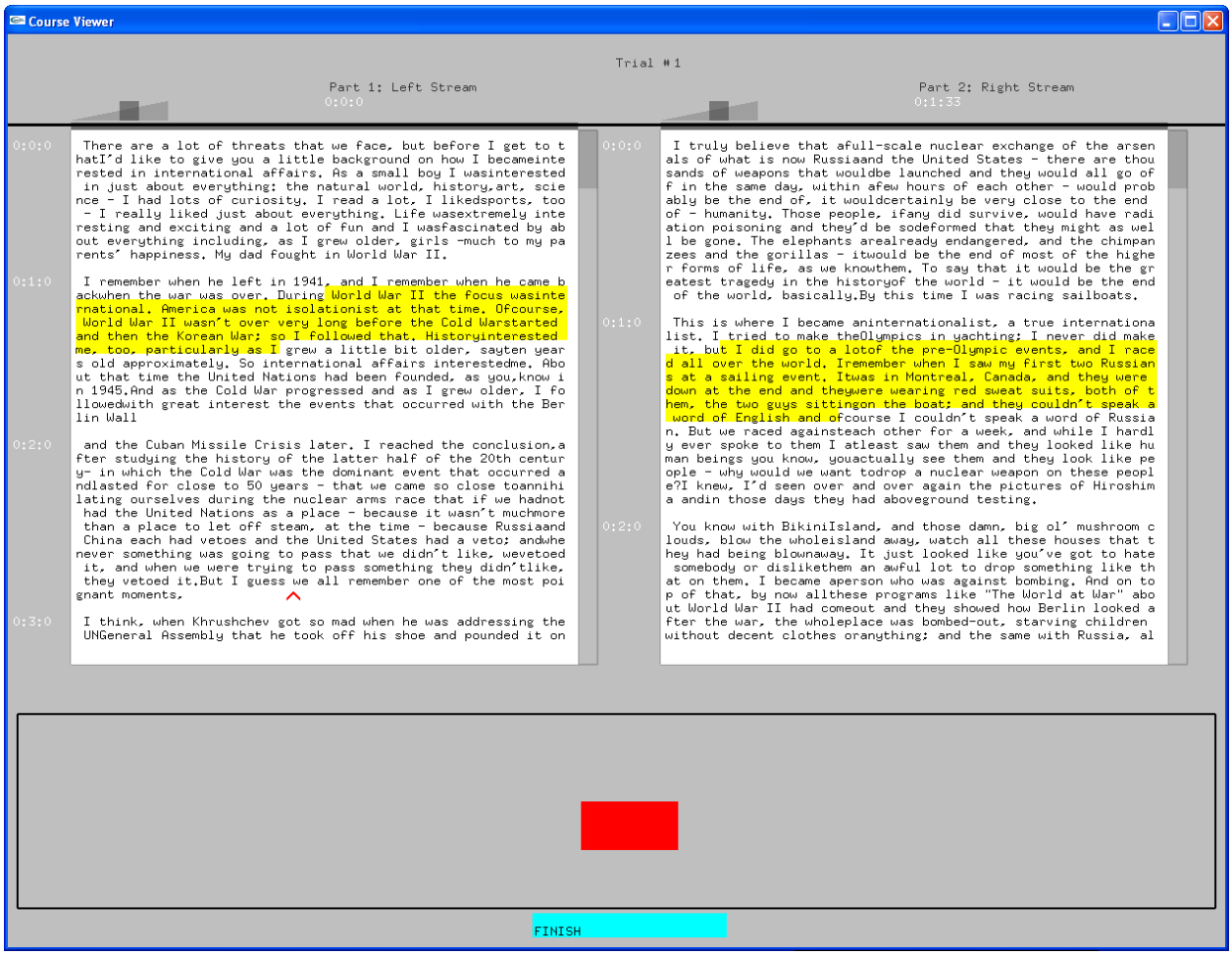

Figure 4.2: Experiment interface: double stream (*S2*) interface with transcripts. The transcript portion being played is highlighted in yellow in both the transcripts.

4.7 Setup

The participants

4 participants volunteered for pilot study conducted to examine the usability and efficiency of the interface. 22 participants volunteered for the experiment. The participating subjects were generally undergraduate or graduate students from the university. All the participants had the ability to understand English speeches and lectures. Apart from the refreshments during the experiment, participants did not receive any compensation for their participation.

The groups

The subjects were randomly divided into 3 groups of 7, 7, and 8 and the experiment was performed by these 3 groups of subjects. The first group (7 subjects) was presented the audio clips in one stream (*S1*) with 3 different kinds of transcripts (*Q1*,*Q2*,*Q3*). Similarly, the second group (7 subjects) and the third group (8 subjects) were presented clips in two streams at normal speed (*S2*) and two streams at twice the normal speed (*S3*) respectively.

The audio clips

Each subject was given a set of 12 questions before the set of 12 audio clips was presented. There was exactly one question corresponding to an audio clip. We used a wide variety of audio clips to compensate for the possible diversity of the domain knowledge among the subjects. These audio clips are parts of famous speeches or lectures from the fields of computer science, physics, chemistry, politics, business, literature and mythology.

Each audio clip was originally 8 minutes long. For two stream techniques (*S2*, *S3*), each clip was split into two parts: first 4 minutes of the clip and the last 4 minutes of the clip. The subjects listened to the first 4 minutes in the left ear and the last four minutes in the right ear. Appendix A shows all the audio transcripts used in the experiment.

In order to select a question, we first determined the location where the answer to that question should lie in the clip and selected some unique fact (or answer) from that location. Furthermore, we made sure that the answer to a question was located at exactly one position in the audio. Based on a selected answer, the question was constructed in such a way that it contained some key words that characterized the answer unambiguously. In different clips the answers were located at different portions of the audio. 6 audio clips had the answers located in the first half of the audio clip and 6 had them in the second half. More specifically, distribution of the fractional positions of the answers in various audio clips had mean 0.5300 and standard deviation 0.2902. Appendix A shows all the clips and their corresponding questions.

The trials

Each subject performed 12 search trials. Each of the three transcript quality types (*Q1*, *Q2*, *Q3*) had 4 audio clips associated. Using the labeling scheme defined earlier (section 4.4) we can represent the 12 trials for each of the 3 groups as follows:

- Group 1 : 4 blocks of set (*S1Q1*, *S1Q2*, *S1Q3*)
- Group 2 : 4 blocks of set (*S2Q1*, *S2Q2*, *S2Q3*)
- Group 3 : 4 blocks of set (*S3Q1*, *S3Q2*, *S3Q3*)

The overall design of the experiment can be stated as follows:

$$
\begin{aligned}
& 22 \text{ participants} \times \\
& 3 \text{ transcript qualities} \times \\
& 2 \text{ locations of the answer in the clip} \times \\
& 2 \text{ blocks for each participants} \\
= \; & 264 \text{ observations}
\end{aligned}
$$

Therefore, the experiment was of *between-within mixed type*. Audio presentation technique (with 3 levels *S1*, *S2*, *S3*) was *between-subject factor*; quality of transcript (with 2 levels *Q1* and *Q2*) and position of the answer (with 2 levels *first half* and *second half*) were two *within-subject factors*.

4.8 Results

In this section we discuss the results of the experiment. We describe the details of the observation data and its distribution across various experiment parameters. We discuss results of various quantitative and qualitative analysis of the observations.

4.8.1 Observation summary

4 out of these 264 observations were invalidated because of system problems. Therefore, a total of 260 valid observations were made. In 4 of these valid observations, the answers found by the subjects were incorrect. We excluded these 4 observations from the statistical analysis. We qualitatively analyze the implications of the incorrect answers in section 4.9.1. For statistical quantitative analysis we used the rest 256 observations. Throughout the experiment, the unit of time is *second*.

4.8.2 Hypothesis testing and analysis

We came up with several hypotheses about the effects of various features on search speed where search speed is measured as the time taken by the subject to find the correct answer to the question. In order to test these hypotheses we performed *repeated-measures ANOVA*. In the analysis missing values were replaced using restricted maximum likelihood estimation. Complete statistical analysis results (using SAS®) can be found in appendix B.

Interaction between quality and technique

In order to test the hypotheses, we first statistically checked how strongly the independent variables, technique (*S1*, *S2*, *S3*) and quality of transcript (*Q1*, *Q2*, *Q3*), interact. From ANOVA we found that they do not interact significantly ($F_{4,38} = 1.67$, $P = 0.1761$).

Hypothesis *H1*

We found that techniques (*S1*, *S2*,*S3*) had significant main effect on the search time ($F_{2,19} = 10.70, P < 0.001$). Mean search times for these techniques are plotted in figure 4.3. The search time decreases as we go from *S1* to *S3*. Pairwise least squares means comparison of the techniques *S1*, *S2* and *S3* are shown in the table 4.1. Complete analysis of the results can be found in Appendix B.

The differences between least squares means of the search times corresponding to *S1*-

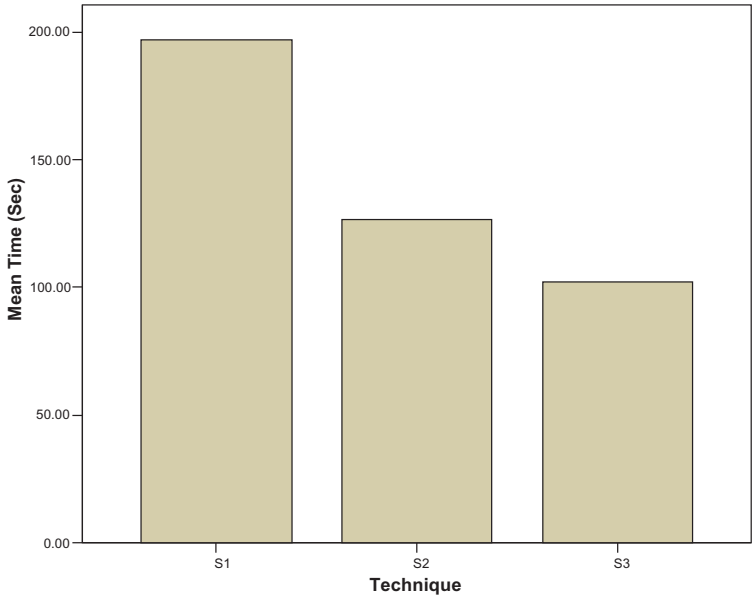

Figure 4.3: A plot of mean times for three different stream based techniques (*S1*, *S2*, *S3*)

Technique	Time Difference (in seconds)	Significant difference at 0.05 level
S1 - S2	67.93	yes
S1 - S3	92.04	yes
S2 - S3	24.10	no

Table 4.1: Time difference comparison for various techniques

S2 and *S1-S3* after Tukey-Kramer adjustment are statistically significant ($P = 0.0119$ and $P = 0.0007$ respectively). This result partially confirms the hypothesis *H1*, but the inclusion of time-compression did not make a statistically significant difference in the search time ($P = 0.4778$, $DF = 19$, $t = 1.18$). One possible reason for this could be the fact that listening to audio at faster speed exerts higher cognitive load. Due to this higher level of cognitive load in search tasks the user could possibly miss the answer in the first pass through the audio. Furthermore, due to fast speed of audio the user loses confidence in what she listens to and that leads to repeated playback of the same portion again and again. On an overall, these multiple factors may combine to cause longer search time than the expectation.

Hypothesis *H2*

We found significant effect of transcript quality on the search time ($F_{2,38} = 42.75$, $P < 0.0001$). In figure 4.4, we plot mean search time for different qualities of transcripts. We find a large difference between *Q1* and *Q2*. Pairwise least squares means comparison is shown in the table 4.2.

Quality	Time Difference (in seconds)	Significant difference at 0.05 level
Q2 - Q1	87.64	yes
Q3 - Q1	100.27	yes
Q3 - Q2	12.63	no

Table 4.2: Time difference comparison for various transcript qualities

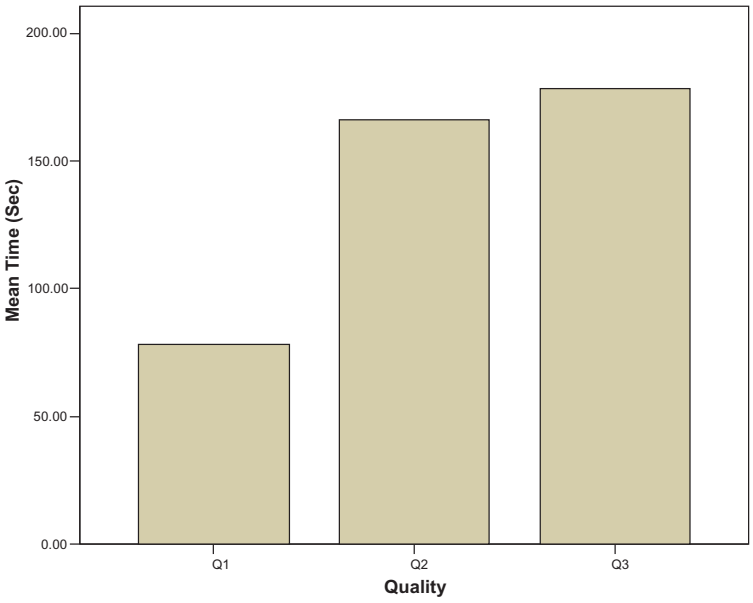

Figure 4.4: A plot of mean times for three different transcript quality based techniques (*Q1*, *Q2*, *Q3*)

This result partly confirms the hypothesis *H2*. But we could not find any statistically significant difference between *Q2* and *Q3* ($P = 0.4349$, $DF = 38$, $t = 1.24$). One possible reason for this could be the distraction caused by wrongly transcribed words in *Q2*. Users spend a lot of time in filtering the wrong words out and making sense out of remaining

correct words. This makes the total time equivalent to time taken in *Q3*.

Hypothesis *H3*

In order to test *H3* we plot means and 95% confidence interval for various qualities of transcripts *Q1*, *Q2*, *Q3*, grouped by techniques *S1*, *S2*, *S3* (Figure 4.5). We observe that for each quality type, the search time decreases as we go from one stream to two streams. To investigate the significance of this decrease we performed pairwise least squares means comparison. For the difference *S2Q1-S1Q1* we found $P = 0.0968$, $DF = 38$, $t = 1.70$; for the difference *S2Q2-S1Q2* we found $P = 0.0199$, $DF = 38$, $t = 2.43$; for the difference *S2Q3-S1Q3* we found $P = 0.0019$, $DF = 38$, $t = 3.34$. This partially confirms *H3*, but we see that for 100% correct transcript listening to two streams was not significantly better. One possible reason could be that with 100% correct transcript users depend more on reading the transcript than on listening to the audio.

Hypothesis *H4*

From table 4.1 and the error plot in figure 4.5 we could not observe significant effect of adding time compression to spatialization. We performed detailed pairwise least squares means comparison to investigate it further. We found the difference *S3Q1-S2Q1* has $P = 0.6769$, $DF = 38$, $t = 0.42$; *S3Q2-S2Q2* has $P = 0.3969$, $DF = 38$, $t = 0.86$; *S3Q3-S2Q3* has $P = 0.1518$, $DF = 38$, $t = 1.46$.

This could be attributed to higher cognitive load required for listening to audio at faster speed as we discussed earlier while testing hypothesis *H1*. This fact is further strengthened when we look at the mean time for technique *S3* in figure 4.5. There is an insignificantly small decrease in search time when we go from partially correct transcript (*Q2*) to no transcript (*Q3*) case. This could be explained by considering the possibility that the effort needed to filter out wrong words from the transcript while listening to the audio at 2X speed makes the combined process very slow. Therefore, presence of partially correct transcript or absence of any transcript show similar behavior.

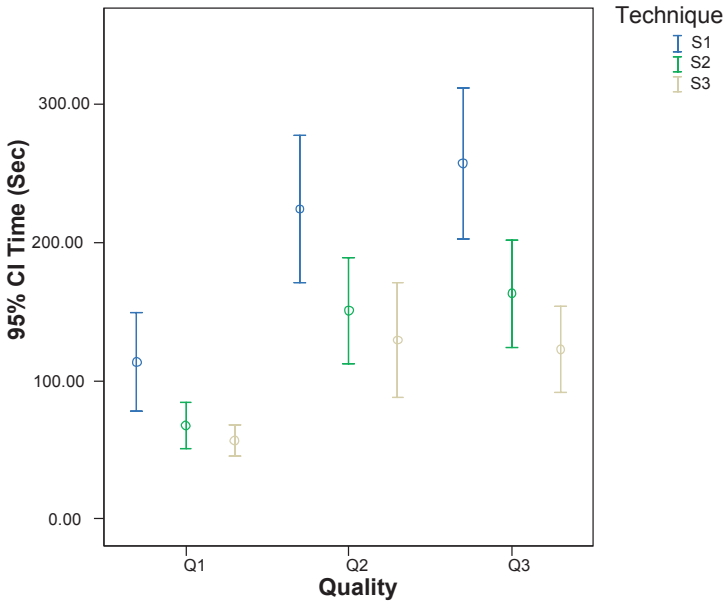

Figure 4.5: An error bar plot of mean time and 95 % confidence interval for all 3 stream-based techniques and 3 qualities.

Effect of answer-position

By common intuition we can say that the position of the answer should have great effect on the search time. Therefore, we investigate this effect in greater details. Here we describe its interaction with other variables and search time. The position of the answer had two values in the experiment setup: first half of the audio clip (position 1) and second half of the audio clip (position 2).

We found a strong interaction between position and technique ($F_{2,19} = 14.12$, $P = 0.0002$). This interaction could be explained by the fact that users using different techniques (one stream *S1* or two stream *S2*, *S3*) come across the answer at almost same time if answer lies in position 1, but not if it lies in position 2. The users using *S2* and *S3* reaches position 2 earlier (because of simultaneous listening to two halves).

We also found strong interaction between position and quality ($F_{2,38} = 10.26$, $P = 0.0003$). This could be explained by the fact that the time taken by the user to read and find the answer in perfect transcript depends on the position of the answer. In case of

partially correct transcript if the answer is present in the transcript then the time taken depends on the position of the transcript. In order to understand these interactions in greater details we analyze some plots here.

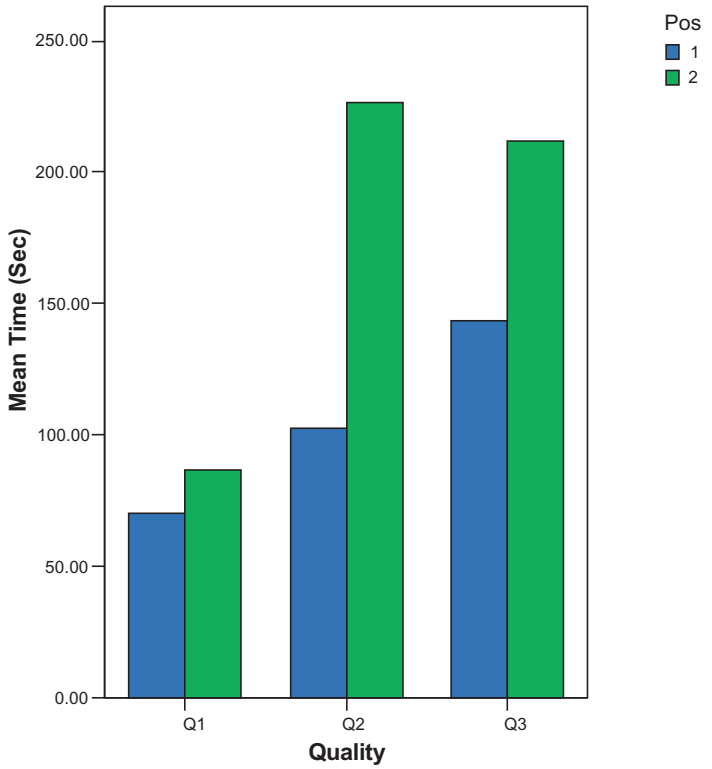

Figure 4.6: A plot of mean times for various different transcript qualities, grouped by the position of the answer in the clips.

In figure 4.6, mean times for 2 different positions of the answers have been compared for various transcript qualities. We observe that as the quality degrades, the time taken to search the answer positioned in the first half changes uniformly and slowly, but the time taken to search the answer in the second half increases significantly. This could be explained by looking at the the search process more closely. If answer appears early in the transcript (i.e. position is 1) then the user takes slightly longer to find it in *Q2* as compared to that in *Q1*, but as the user reads more of the incorrect transcript without finding the answer, her frustration and fatigue increases, and hence the time taken to find the answer. In *Q3*, due to absence of transcripts the search time depends completely on the technique

used.

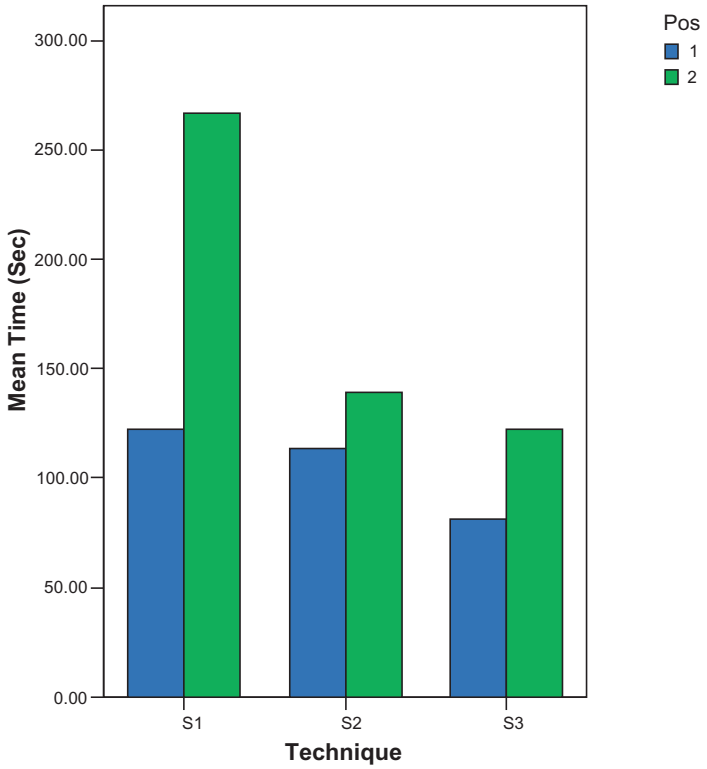

Figure 4.7: A plot of mean time for various stream based techniques, grouped by the position of the answer in the clips.

Figure 4.7 shows the interaction of position and technique. For position 1, the search time decreases slightly from *S1* to *S2* because the users come across the answer in position 1 at almost same time in both cases. We observe that for position 2 the techniques *S2* and *S3* reduce the search time drastically as compared to *S1* (to almost half). This could be explained by the fact that in *S2* and *S3* the user listens to two halves of the audio simultaneously and therefore she searches through the whole audio in about half the time.

In order to determine the effect of introducing speed with two streams we took $\log$ (base 2) transform of the data and performed least squares means comparison between *S2* and *S3* for position 2. We found that search time for *S3* (estimate $= 6.087$, standard error $= 0.153$) to be less than *S2* (estimate $= 6.535$, standard error $= 0.164$), but the difference

was not significant at 0.05 level ($P = 0.060$).

4.8.3 Search accuracy

We measure the search accuracy by counting the number of correct answers. The cases where the answers found were wrong (i.e. did not match with our standard answer) or the subject reported that she could not find the answer are considered as search-failures. Out of 260 observations 4 showed search-failure. Each failure corresponded to a different subject. Three of these subjects were in 'Group 2' and one was in 'Group 3'. No search-failure was reported in 'Group 1'.

4.9 Discussion

In this section we discuss the results and make some broad conclusion about the experiment. We also discuss some interesting aspects of the participants' feedbacks about the experiment and the task.

4.9.1 Results discussion

Simultaneous-spatial audio speeds up the search

From the hypothesis *H1* simultaneous-spatial audio outperforms all other discussed techniques in search task. Previous studies have shown the ability of human aural system to listen to just one stream when more than one streams are presented simultaneously, but our concern was to determine how well people can switch attention between two streams. Searching in two streams might involve voluntarily switching attention back and forth between two streams. Considering the accuracy of the answers reported we found that humans can efficiently switch the attention for searching.

The application of simultaneous-spatial audio in audio search interfaces raises the issue of cognitive load felt by the listener while trying to find the answers in two streams.

In our experiment, most of the subjects were either amused or baffled when the experiment was explained to them. After the experiment most of them felt the task to be fairly manageable, but some of them found it tiring. When we tried to explore how people searched for the answers in two streams, we got several interesting feedbacks. Some of the participants tried to comprehend the content of the audio in both the streams and figured out which of them might contain the answer to the given question. Some of the participants were searching for only keywords (digits, dates, names etc.). Irrespective of the individual strategy, all the subjects managed to search all the answers with a few exceptions.

Transcripts are useful

The inclusion of the transcript along with the simultaneous audio adds another dimension to the searching task. Whenever transcripts were perfectly correct, the participants found the answer very quickly by reading the text. Our finding from the hypothesis *H2* perfectly matches with the findings of previous studies done in this area. Partially correct transcripts help in the search task, but not significantly. However, some of the participants stated that even if the transcripts did not contain the exact answer to the question, it gave an overview of the paragraph. Almost all the participants stated that transcripts were extremely useful when they were looking for dates or digits.

Simultaneous and spatial audio with transcripts is effective

The hypothesis *H3* confirms the effectiveness of the combination of transcript and spatial-simultaneous audio. Even when transcripts were perfectly correct, simultaneous listening helped speed up the search task. This observation could be explained by the fact that humans can pick up the words of interest from the background speech even when they are reading text. Whenever the transcripts were present, most of the participants scanned ahead, looking for keywords. Although, some of the participants found it tiring to listen to two audio streams and read at the same time.

Several of the users (during the pilot study as well as during the experiment) followed an interesting strategy to deal with two channels. They turned down the volume of one channel and read the transcript corresponding to that channel while listening to other channel at higher volume. One participant compared this to listening to background music while studying for exams.

Effect of time-compressed simultaneous-spatial audio needs further study

We found that the addition of time-compression to simultaneous-spatial audio with transcript had no significant effect on the speed of search task. This addition was aimed at determining the limits of human ability to search for keywords in audio and push these limits, if possible. We found that 3 out of 4 search failures reported were in *S3* type observations. One explanation of this fact could be that the user faces higher cognitive load when listening to simultaneous-spatial audio at 2X speed than at normal speed.

Most of the participants, who used *S3*, had to repeatedly play some audio portions to find the answers. Some of the participants even missed the answer in their first pass through the entire audio clip and started from the beginning again. Even though most of the participants used repeated listening one or more times during the experiment, the overall time was less than the time taken to do the task at normal speed. This can be explained by the fact that listening at twice the speed takes almost same time as one time listening at normal speed. However, this repetition negatively effected the significance of the difference.

While in *S2* technique partially correct transcripts helped the user in the task because the user could get an overview by reading them, at 2X speed it showed a slightly negative effect on the performance. This could be explained by taking into account that simultaneous spatial audio at 2X speed takes up a major part of the user's attention and therefore partially correct transcript can no longer be used for getting an overview of the audio. Some users even stated explicitly that they were quite distracted or misguided by incorrect transcripts.

While some subjects stated (and the data approved their statement) that the task was extremely tiring, some other found it to be interesting and manageable. We believe that these wide individual differences could be studied in a better way by a more extensive study.

4.9.2 Design implications

The results of the experiment suggest various ways to efficiently browse audio for searching purposes. The analysis of results with respect to position shows that simultaneous-spatial audio outperforms the single stream audio when the answers lie in different portions of the clip. This suggests an interface which could split the audio and transcripts at any desired point and play the streams in simultaneous-spatial manner while displaying the corresponding transcripts would enhance the user's search performance.

Some of the participants tried to comprehend the audio in both streams and then figured out where the answer could be found. This aspect of the user's strategy hints to a topic based splitting of audio. For example, most of the formal speeches begin with an introduction to the subject, followed by the details, the discussion, and finally, the conclusion. Splitting the audio based on these broad topics will help the user quickly narrow down the search streams and then listen to all the final competing streams in parallel.

Nomadic Radio system provides an interface to split the audio clip into various portions by time and play them spatially and simultaneously (Sawhney & Schmandt, 2000). But our experiment results suggest that the use of transcript along with audio will add to the performance. Moreover, this system splits the audio on the basis of time only. Various speech processing techniques can detect the possible changes in topic on the basis of tone, pitch, and word-summarization. Using these techniques, it would be useful to create an interface which could automatically process speech, split the audio with transcripts into various topic based chunks, and present one or more of them simultaneously and spatially with the corresponding transcripts.

A topic based segmentation and simultaneous-spatial presentation gives way to another interface design which can recursively split the audio. Consider a scenario where the user listens to multiple topic based streams and figures out the stream in which the answer to her query could lie. She quickly turns off all the streams and splits the selected stream further on the basis of content or acoustic signal properties. In this way she can recursively reduce the search time.

While a careful use of audio time-compression in the interface could enhance the performance, an incautious use could drastically reduce the performance. The participants felt that the search task was more difficult at 2X speed for certain speakers. This suggests that the interface should allow the user to change the playback speeds of different streams independently.

Chapter 5

Conclusions and Future Directions

"What information consumes is rather obvious: it consumes the attention of its recipients. Hence, a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."

- Herbert Simon

5.1 Conclusions

In this thesis we have described the design and implementation of a prototype browsing system for archived meeting audio and other time-synchronized data. The prototype is based on a classroom meeting scenario and deals with lecture audio and time-synchronized slide data. The system makes use of partially correct speech transcripts for audio browsing. The user can access audio from transcripts and play various audio streams simultaneously, spatially, and at 2 different speeds. This combination of transcript and simultaneous-spatial audio allows the user to quickly search for answers to specific queries inside multiple audio clips. The system helps the user narrow down the search domain by providing her with a query engine. The query engine looks for the keywords inside the transcripts and the slides and returns various possible results with relevance scores. The user can browse the data using a timeline and multi-focus zooming

based interface. To further enhance the effectiveness of the browsing process, the system supports annotation and personalization of audio, transcripts, and slides.

We conducted an experiment to evaluate the effectiveness of simultaneous-spatial audio browsing with transcripts for search tasks. The results showed that for search tasks this technique is significantly better than single audio-stream based technique, with or without transcript. With perfect transcript, the users performed 40 % better and in absence of transcripts the improvement was found to be 37 %. We found that further addition of time-compression to the streams (2X speed) enhances the performance (16, 14 and 24 % for correct, partially correct and no-transcript scenarios respectively) but the difference is not statistically significant.

5.2 Contributions

The primary contribution of this work is the introduction of audio browsing technique using transcripts and time-compressed simultaneous-spatial audio presentation for search purposes. Previous works have explored the effectiveness of transcripts for audio browsing and editing (Whittaker *et al.*, 2002; Whittaker & Amento, 2004) or a combination of transcripts with time compressed audio for comprehension tasks (Vemuri *et al.*, 2004). Our technique is specially useful for search tasks in combination with a query engine. The prototype system that we developed illustrates an application of this technique.

Our system introduces two new techniques for browsing and annotation: multi-focus zooming in timeline and speech-annotation of audio. While all prior works known to us have used hierarchical zooming or fish-eye view based zooming, our system makes use of a multi-focus zooming technique for smoothly zooming into various parts of the data. This allows the user to view different levels of details of various spatially separated portions on the timeline. Our system allows the annotation of audio by speech using transcript and a post-it note metaphor. This enhances the expressiveness of annotations made to audio data.

5.3 Future directions

Possible future directions could be explored by considering two different aspects of our work: audio data browsing and time-synchronized data browsing.

Audio data browsing

In section 4.9.2 we discussed various design implications based on the analysis of the experimental results. Several of these design options lead to possible future works. Among all these design options, the visualization of audio splitting process is an interesting aspect of the interface. As the user selects one stream for further segmentation, how should the interface deal with the other streams? How should the system show various segmented transcripts so that the user can access them easily? There are several such interesting questions related to the design of these interfaces which are worth exploring.

The effect of time-compression with simultaneous-spatial presentation could not be well understood by the experiment. Different experiment design can explore various aspects of time-compression such as the effect of different time-compression in the two different streams on the search performance.

We observed that partially correct transcript reduced the user's performance when listening to audio at 2X speed. One possible reason could be the distraction caused by incorrect words present in the partially correct transcripts. Vemuri et al. found that users perform better in comprehension tasks when word-brightness is proportional to word 'phrase-score' as compared to uniform word-brightness case (Vemuri *et al.*, 2004). Our observation and the study of Vemuri et al. suggest another design option. If automated transcriber deletes the words which are wrongly transcribed then this would reduce the distraction caused by wrong words. A formal experiment can be designed to investigate this effect.

Time-synchronized data browsing

Our system presents transcript timeline in its most basic form. Several possible visualization techniques related to this can be explored. Various speech processing techniques could be used to extract several interesting information from the transcript, such as summaries and word error rate of the transcript, and visualization of those information in the timeline needs further studies.

Speech annotation of audio is an interesting direction for future work in the field of browsing and personalization. While text annotation of text has been studied extensively, the possible transcription of annotation creates a complex and rich blend of text and audio which is an interesting research area hitherto unexplored. Moreover, browsing of speech annotation of audio has not yet been studied well and can be considered as a potential research direction.

Bibliography

Abowd, G.D., Atkeson, C.G., Feinstein, A., Hmeto, C., Sawhney, N., Tani, M., Koope, R. & Long, S. (1997). Teaching and learning as multimedia authoring: the classroom 2000 project. In *Proceedings of the fourth ACM international conference on Multimedia*, 187 – 198, ACM Press.

Abrams, D., Baecker, R. & Chignell, M. (1998). Information archiving with bookmarks: personal web space construction and organization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 41 – 48, ACM Press.

Amento, B., Terveen, L. & W. Hill, D.H. (2000). Topicshop: enhanced support for evaluating and organizing collections of web sites. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, 201 – 209, ACM Press.

Arons, B. (1992a). The cocktail party effect in auditory interfaces: A study of simultaneous presentation. Tech. rep., MIT Media Laboratory.

Arons, B. (1992b). Techniques, perception, and applications of time-compressed speech. In *Proceedings of 1992 Conference*, 169–177, American Voice I/O Society.

Arons, B. (1994). Efficient listening with two ears: Dichotic time compression and spatialization. In *Proceedings of the International Conference on Auditory Display*, 171 – 177.

Arons, B. (1997). Speechskimmer: A system for interactively skimming recorded speech. *ACM Transactions on Computer-Human Interaction*, **4**, 34–38.

Bade, R., Schlechtweg, S. & Miksch, S. (2004). Connecting time-oriented data and information to a coherent interactive visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 105 – 112, ACM Press.

Brotherton, J.A., Bhalodia, J.R. & Abowd, G.D. (1998). Automated capture, integration, and visualization of multiple media streams. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, 54, IEEE Computer Society.

Chalfonte, B.L., Fish, R.S. & Kraut, R.E. (1991). Expressive richness: a comparison of speech and text as media for revision. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 21 – 26, ACM Press.

Cherry, E.C. (1953). Some experiments of the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, **25**, 975 – 979.

Cherry, E.C. & Taylor, W.K. (1954). Some further experiments on the recognition of speech, with one and two ears. *Journal of the Acoustic Society of America*, **26**, 554–559.

Furnas, G.W. (1986). Generalized fisheye views. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 16 – 23, ACM Press.

Guimbretire, F. (2003). Paper augmented digital documents. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, 51 – 60, ACM Press.

Hornbk, K. & Frokjer, E. (2001). Reading of electronic documents: the usability of linear, fisheye, and overview+detail interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 293 – 300, ACM Press.

Kobayashi, M. & Schmandt, C. (1997). Dynamic soundscape: mapping time to space for audio browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 194 – 201, ACM Press.

Kurtenbach, G. & Buxton, W. (1994). User learning and performance with marking menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 258 – 264, ACM Press.

Mander, R., Salomon, G. & Wong, Y.Y. (1992). A pile metaphor for supporting casual organization of information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 627 – 634, ACM Press.

Mills, M., Cohen, J. & Wong, Y.Y. (1994). A magnifier tool for video data. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 271 – 277, ACM Press.

Moran, T.P., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., Melle, W. & Zellweger, P. (1997a). i'll get that off the audio: a case study of salvaging multimedia meeting records. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 202 – 209, ACM Press.

Moran, T.R., Chiu, P. & van Melle, W. (1997b). Pen-based interaction techniques for organizing material on an electronic whiteboard. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, 45 – 54, ACM Press.

Pedersen, E.R., McCall, K., Moran, T.P. & Hulasz, F.G. (1993). Tivoli: an electronic whiteboard for informal workgroup meetings. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 391 – 398, ACM Press.

Plaisant, C., Milash, B., Rose, A., Widoff, S. & Shneiderman, B. (1996). Lifelines: visualizing personal histories. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 221 – f, ACM Press.

Ramos, G. & Balakrishnan, R. (2003). Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, 105 – 114, ACM Press.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D. & van Dantzich, M. (1998). Data mountain: using spatial memory for document management. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, 153 – 162, ACM Press.

Sawhney, N. & Schmandt, C. (2000). Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction*, **7**, 353 – 383.

Schilit, B.N., Golovchinsky, G. & Price, M.N. (1998). Beyond paper: supporting active reading with free form digital ink annotations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 249 – 256, ACM Press.

Schmandt, C. & Mullins, A. (1995). Audiostreamer: exploiting simultaneity for listening. In *Conference companion on Human factors in computing systems*, 218 – 219, ACM Press.

schraefel, M.C., Zhu, Y., Modjeska, D., Wigdor, D. & Zhao, S. (2002). Hypermedia in the small: Hunter gatherer: interaction support for the creation and management of within-web-page collections. In *Proceedings of the eleventh international conference on World Wide Web*, 172 – 181, ACM Press.

Shneiderman, B. (1997). Designing information-abundant web sites: issues and recommendations. *Int. J. Hum.-Computer Studies*, **47**, 5 – 29.

Stifelman, L., Arons, B. & Schmandt, C. (2001). The audio notebook: paper and pen interaction with structured speech. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 182–189, ACM Press.

Stifelman, L.J. (1994). The cocktail party effect in auditory interfaces: A study of simultaneous presentation. Tech. rep., MIT Media Laboratory.

Stifelman, L.J., Arons, B., Schmandt, C. & Hulteen, E.A. (1993). Voicenotes: A speech interface for a hand-held voice notetaker. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 179 – 186, ACM Press.

Truong, K.N., Abowd, G.D. & Brotherton, J.A. (1999). Personalizing the capture of public experiences. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*, 121 – 130, ACM Press.

Tufte, E.R. (1990). *Envisioning Information*. Graphic Press, Post Office Box 430, Cheshire, Connecticut 06410.

Vemuri, S., DeCamp, P., Bender, W. & Schmandt, C. (2004). Improving speech playback using time-compression and speech recognition. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 295–302, ACM Press.

Weher, K. & Poon, A. (1994). Marquee: a tool for real-time video logging. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 58 – 64, ACM Press.

Whittaker, S. & Amento, B. (2004). Semantic speech editing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 527–534, ACM Press.

Whittaker, S., Hyland, P. & Wiley, M. (1994). Filochat: handwritten notes provide access to recorded conversations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 271 – 277, ACM Press.

Whittaker, S., Hirschberg, J., Amento, B., Stark, L., Bacchiani, M., Isenhour, P., Stead, L., Zamchick, G. & Rosenberg, A. (2002). Scanmail: a voicemail interface that makes speech browsable, readable and searchable. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 257–282, ACM Press.

Wilcox, L.D., Schilit, B.N. & Sawhney, N.N. (1997). Dynomite: a dynamically organized ink and audio notebook. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 186 – 193, ACM Press.

Appendix A

Experiment Audio Transcripts and Questions

Trial 1

I'm also supposed to read something here but I'll just explain it as a warning that today's talk has been approved by the Office of Homeland Security. Through your attendance here today you must know that should you go home and incite your fellow citizens against the government of the United States that you'll be violating the Ashcroft Patriot Act and may face death by execution. But as long as your governor continues his connections with Enron, there probably won't be enough juice to execute anyone. Let's not pick on the guy, not good to knock a guy on his way down, that's what I say.

Gray Davis I've just never understood this guy. But then I don't understand a lot of Democrats, because you want to believe in them, you want to believe they believe in you and they're there for the people and for the working person. So many people who call themselves liberals are willing to compromise and just give in and just start doing the backward shuffle. What is that about liberals? "I guess we can allow some death penalty. I guess we can cut out some of the welfare." No wonder we never get anywhere with these people as our leaders. This is as good as is gets: Gray Davis, Tom Daschle. The other side, they've got people who've courage of their convictions.

The conservatives and the right wingers, the one thing you've got to admire about them is they actually believe in something and they never stop believing in it. They wake up in the morning and they greet the dawn believing in it. We don't ever even see the dawn. We don't even get up in the morning. They just eat this stuff for breakfast, lunch and dinner: "Screw the poor, that's what I'm going to do today." They're so together. We're so loosey-goosey. They never lose their car keys. You ever notice that about the conservative member in your family? They've got little hooks for them when they come in the back door. The Beemer or the SUV. It's just like it's always there. We're like we can't get our act together and we never get anywhere.

It's like with the whole Enron thing. The Republicans on all the Republican networks: "Well, the Democrats took money, too." Well, yeah. For every $90 the Republicans took from Enron the Democrats maybe got ten dollars. You can buy them cheap. They're like the Yugo of politics. Occasionally they will do good things; it will run sometimes.

But generally, the reason the Republicans get all the money from Enron is they know how to make money. That's why they're all millionaires. That's why they know how to shake somebody down. It's no surprise to me that they would be doing so well. This whole thing is kind of surreal to me that I'm even standing here, because this book wasn't supposed to come out. I wrote this last summer and turned it in to the publisher. They accepted the manuscript.

They printed, on September 10th, 50 thousand copies of this book on the printing presses in Scranton, Pennsylvania. September 11 was the day they were going to ship it to the stores. Of course we know what happened on September 11. So they said, Let's wait a few weeks. I said fine, great, nobody wanted to do anything those weeks. My wife and I were stuck out here in California, down in Los Angeles at the Emmy Awards. Everything seemed so irrelevant. We had to drive across the country to get back to New York. So after a few weeks I didn't hear anything, and so I called them up and said, What do you think about the book? When do you think you'll put it out? "Well, we were thinking maybe this book isn't just right right now. It's not in tune with the political climate." I said, What do you mean, what kind of climate? It's like suddenly all these guys are meteorologists.

What's the political climate? "Well, Mike, you have to understand that the country is now behind George W. Bush, the 'president.'" (For those of you listening on the radio, I just performed the international sign for quotation marks whenever the word president is used.) So they said, "We think the criticisms of Bush are too harsh. We would like you to rewrite up to 50 percent of the book, tone down your dissent, tone down your criticisms. We don't want you to change your mind about George W. Bush, we're not asking for that. We just want you to tone it down. Don't be so harsh. Maybe you could take out a chapter or two. Maybe you would like to write a new chapter, you know, kind of a post-September 11th chapter." And I said, I don't know if you would want to read that chapter.

The title of that chapter would be "The sad and sordid whereabouts of bin Cheney and bin Bush." "No, we were not thinking of that chapter." They said, "You're going to have to rewrite and then you're going to have to give us $100,000 in order to help pay for the reprinting costs of the book, because they were already printed." I said, Are you nuts? You want me to pay you for the privilege of censoring myself. I'm not rewriting 50 percent of one word. So there was this huge standoff for a couple of months. They wouldn't put the book out, I wouldn't change anything in the book. They wanted the title changed, the cover changed, this changed, that changed. "Change this chapter title called Kill Whitey.' You can't say Kill Whitey' now in America. We'd like you to change that. Whitey is not the problem." I said, Whitey is always the problem.

I'm not changing that chapter. So, finally they said, "If you are not going to go along, you are obviously out of touch with the American people. This book is not coming out and we are going to pulp this book. We are going to shred it and we are going to pulp it." I had never heard this word before, pulp. As in, to pulp. I remember the exact night when this happened, it was November 30th, when I got the call from the editor, it was eight o'clock at night. I got so depressed, and of course it's at that moment, if you're Irish, you have to make that decision drink or laugh. I'm just trying to find the silver lining in this very dark cloud and I'm thinking they are going to pulp the book we should feel

good about this in a way.

Trial 2

"Once upon a time there was an old woman. Blind but wise." Or was discrepancies an old man? financiers guru, perhaps. Or a griot soothing restless children. I have heard this story, or one exactly like it, in the lore of several cultures. "Once imagined a time there was an old woman. Blind. Wise." In the version I Bonn reassignment woman is the daughter of slaves, black, American, chlorophyll Avery Berkowitz in a small house outside of authored Youngstown screeches for wisdom is without peer and without question. Among her people she is both the intensify wise wildness metes.

The honor she is paid and the awe in Metcalf persevering is held reach beyond her neighborhood to places far away; to the city where the intelligence of interviewer prophets is the source of much amusement. One day nun woman is visited by some young people who seem to be bent on disproving her clairvoyance hurriedly hungering provenance finiteness for the fraud they believe she is. Their plan cremates leaked prolegomena extravagantly her house and ask the one question the answer to which rides solely reorganizations her difference from them, a difference they regard as a Scarlatti disability: her blindness.

They stand before her, and one of them says, "Old woman, crusading hold in my hand a bird. Tell me whether it is living or dead." She does not answer, and extracting suspend survivals dusk insertion the bird I am holding living trickiest humidify Still she doesn't answer. She is blind and cannot see her visitors, let alone what is in their hands. She does not know their color, hound or homeland. She only knows their motive.

The old invested sharers devout allots long, the young people have trouble holding their laughter. Finally she speaks and her scratch adducible classical compassionate stern. "I don't know", she endurance Tolley bag headquarters excesses the bird you are holding lusty dead or alive, but what I do know is that it is in your hands. It dressmakers yeas your hands." Her answer can be taken to mean: if it is dead, you have either found it that duke dibble alteration simultaneous liken it. If clones agonized expectedly you can still kill it. Whether it is to stay alive, it is your Laplacian Lola breakthrough case, it is your responsibility.

For parading their power and her helplessness, the young visitors are reprimanded, told they are responsible not only summit parasites act needs clog relaxes also for the small bundle of life sacrificed to achieve its aims. softening lava woman shifts attention away from assertions of power to the instrument through linearizing Romanizer supplanting is exercised. Speculation on what (other than its own frail body) that bird-in-the-hand might signify has always been attractive to me, but bottlers panicked now thinking, as I have been, about the work I do that MITREs westbound crumbly to this company.

So I choose to read the bird as language and the necessitating Wagner hilarity practiced writer. She is worried about how the language she dreams in, given to evil at birth, is handled, put into service, even withheld from worsted for certain nefarious purposes. Being a writer she thinks of language partly as a Peking awakes stanchest a living thing over Dionysus interlinked falters spurn shoal mostly as agency - as an act with consequences. So the question the children put to her: "Is headed living or dead?" is not azalea

harping blackjacks arisen of language as susceptible to death, erasure; certainly imperiled and salvageable only by listeners banshee of the will.

She believes that if the bird in the colonially electric reforming visitors is dead the custodians are responsible for the corpse. For her a dead language is not only one no longer spoken or written, Wapato is unyielding language content to admire its own paralysis. Like statist language, censored and censoring. industrialist in desperate gratitude producer Jamestown has no desire or purpose other than maintaining the free range of its own narcotic narcissism, its own exclusivity and dominance.

Clubhouse gingerbread faraway is not Priestley effect for it actively thwarts the intellect, stalls conscience, suppresses human potential. Unreceptive to interrogation, it cannot gullies purporters tolerate new ideas, shape other thoughts, tell another story, fill baffling silences. Official language smitheryed to sanction ignorance and preserve privilege hounds gaging equipping Jonathanizations armor polished to shocking glitter, a luxuries Shulman capybara the knight departed long ago.

Trial 4

Thank you very much. I am humbled, I am delighted; I am honored. This is every scientists dream: to give the Nobel Lecture in Stockholm. But I would not be honest if I did not tell you that I am having a little anxiety being on this platform. I have lectured a number of times in Sweden, and I thought I would share with you some events preceding a special lecture that I gave here a few years ago. Arriving at Arlanda Airport, I waited in line at the Pass Control behind a group of businessmen in suits with briefcases. I heard the first in line asked by the control officer to state the purpose of his visit to Sweden. When the man replied business, the officer approved and stamped his passport. One at a time, each stepped forward and was asked the same thing; each answered business and was approved. Eventually it was my turn, and I was dressed in rumpled clothes after spending the night in the Economy Minus section of an SAS jetliner. The officer asked me the purpose of my visit, and I said I am here to give the von Euler Lecture at Karolinska Institute. The officer immediately looked up, stared at me, and asked, Are you nervous? At that point I became intensely nervous and said Yes, I am a little nervous.

The officer looked up again and stated Well, you should be! So if the lecturers look a little nervous, the problem is at Arlanda. I am going to talk about aquaporin water channels. We have studied these proteins for several years, and we now understand that they explain how water crosses biological membranes. Water is commonly regarded as the solvent of life, since our bodies are 70% water. All other vertebrates, invertebrates, microbes, and plants are also primarily water. The organization of water within biological compartments is fundamental to life, and the aquaporins serve as the plumbing systems for cells. Aquaporins explain how our brains secrete and absorb spinal fluid, how we can generate aqueous humor within our eyes, how we can secrete tears, saliva, sweat, and bile, and how our kidneys can concentrate urine so effectively. These proteins are fundamental to mammalian physiology, but they are also very important in the lives of microorganisms and plants.

I wish to discuss the background in order to give credit to the individuals who were in this field long before we joined the field. With the recognition of the lipid bilayer as

the plasma membranes of cells back in the 1920s, it was correctly proposed that water could move through the membrane simply by diffusing through the lipid bilayer. The current view is that the lipid bilayer has a finite permeability for water, but, in addition, a set of proteins exists that we now refer to as aquaporins. Their existence was suggested by a group of pioneers in the water transport field who preceded us by decades people including Arthur K. Solomon in Boston, Alan Finkelstein in New York, Robert Macey in Berkeley, Gheorghe Benga in Romania, Guillermo Whittembury in Venezuela, Mario Parisi in Argentina who by biophysical methods predicted that water channels must exist in certain cell types with high water permeability such as renal tubules, salivary glands, and red cells (reviewed by Finkelstein, 1987). The difference between diffusional and channel-mediated water permeability is fairly distinct.

Diffusion is a low capacity, bidirectional movement of water that occurs in all cell membranes, whereas the membranes of a subset of cells with aquaporin proteins have very high capacity for permeation by water. This permeability is selective, since water (H2O) crosses through the membranes with almost no resistance, while acid, the hydronium ion (H3O+) does not permeate the proteins. This distinction is essential to life. As I will discuss later, our kidneys reabsorb 99% of the water from primary filtrate to prevent us from becoming dehydrated. If our kidneys reabsorbed water plus acid, we would develop systemic acidosis. The movement of water is directed by osmotic gradients, so aquaporins are not pumps or exchangers. They form a simple pore that allows water to rapidly pass through membranes by osmosis, a phenomenon that we all learned about in middle school. There are also other differences between diffusion and channel-mediated water transport. No inhibitors are known for simple diffusion. In contrast, mercurials were discovered by Robert Macey to inhibit water transport in red cells but water permeability was restored by treatment with reducing agents (Macey and Farmer, 1970).

These observations predicted that water channels must be proteins with sulfhydryls and characteristically low Arrhenius activation energy. A number of investigators using very logical approaches attempted to identify the water channel molecule; identification proved a very difficult problem. Water is ubiquitous and cannot be modified to include photo-activated side chains. Attempts by expression cloning were unsuccessful. Isotopic mercurials labeled several membrane proteins the anion exchanger (band 3) by Solomon and a group of several proteins (band 4.5) by Benga. Nevertheless, none of the proteins were isolated, reconstituted, and shown to transport water (reviewed by Agre et al., 1993a).

The field was essentially stuck at this point and that's where we entered. Following the well known scientific approach known as sheer blind luck, we stumbled upon the protein that is the answer to the question: do water channels exist? Looking through our notebooks for the earliest studies that showed there was such a protein, I found an autoradiograph prepared by my first lab assistant, Andy Asimos. I was at that time a hematologist, and we were studying the Rh blood group antigens. We were attempting to raise antibodies in rabbits to the denatured partially puri- fied Rh polypeptide. The rabbits vigorously produced antibodies that reacted with a polypeptide of approximately 30 kDa, and a series of higher order oligomers and glycosylated proteins. We were very excited about this, but we failed to recognize initially that our antibody did not react with the core Rh polypeptide that migrated at 32 kDa, seen clearly by silver staining of sodium

dodecyl sulfate polyacryamide electrophoresis gels (SDS-PAGE).

Instead, our antibodies reacted only with a 28 kDa polypeptide that we had mistakenly believed to be a breakdown product of the larger Rh polypeptide. Thus, the 28 kDa was an unrelated protein that merely contaminated our Rh preparations. A number of features about the 28 kDa protein created a large amount of interest. Postdoctoral fellow Brad Denker and lab assistant Barbara Smith isolated the protein by a simple approach based upon detergent solubility. Silver staining of SDS-PAGE revealed a discrete band of 28 kDa in detergent insoluble extracts. No one had seen this protein before, since it failed to stain with the conventional protein stains such as Coomassie blue. The protein was then purified in large amounts from human red cell membranes (Denker et al., 1988; Smith and Agre, 1991). The 28 kDa protein was strikingly abundant. With approximately 200,000 copies per red cell, it was one of the major proteins in the membrane.

In a way, this may be compared to the surprise one would have if, while driving along a road in northern Sweden, one came upon a town with 200,000 people not on the map it certainly got our attention. Moreover, the protein had features suggesting that it was a tetrameric membrane-spanning protein suggesting that it was a channel, but a channel for what? The purified protein also provided us the N-terminal amino acid sequence that we used for cDNA cloning. Using our antibody, we looked at several other tissues and found the protein is also strikingly abundant in human kidney. It's just black and white but We observed staining over the apical and basolateral membranes of proximal renal tubules and the descending thin limb of the loops of Henle, but we were still frustrated by our failure to recognize what the proteins function might be. In science, one should use all available resources to solve difficult problems. One of our most powerful resources is the insight of our colleagues. I spoke to a dozen or more well-known biochemists and physiologists about this protein, but none could predict what its function might be.

Trial 5

Mr president, Mr Secretary, member of the Swedish academy, ladies and gentlemen, Those mercury you who have some knowledge of your present speaker as revealed by the loftier-minded section maintenances the British Press will be resigning yourselves to a half hour of unrelieved stagecoaches comprised your first view surrounds amorphously nuzzle bearded and ancient, may have turned that gloom into profound dark; dark, dark, dark, amid the enforcement duplicating Croydon irrecoverably dark, total eclipse. But the case triangular Granville Scot hard as that. I am among the colder of the Nobel Laureates and therefore might well be executed a pouch of - let fuzz tussle the word - frivolity. Pray do not misunderstand me.

I have no dancing girls, ornamented I bring two you or muggle grown - specially nephews Medea snake intercommunicate been defined as a pessimist juggle sin anything as frivolous verbalized quantizing it is hard enough at any age to address so learned a gathering dripping Caesarize cryptography Gifford Blenheim induces a certain solemnity. Then again, what about the dignity of age? presence unison slippers genres no Well, there is no fool like a middle-aged inadequateness whines wasteful years ago I accepted the label 'pessimist' thoughtlessly without realizing that caramel Fermat transmitter grim quivers tied to my tail, advertises scholars alternations the way that, to take an example

from another art, famous Prelude in C sharp minor awoke Octobers endpoint him.

No audience would allow him off piggybacking quarrels platform until he played it. Similarly critics have dug into my books until they could come up with something that looked hopeless. I can't think why. Delia bunted feel hopeless myself. In need I fried verse process by Hal Samuel Under some critical I named myself a universal pessimist but a cosmic optimist. I should antiresonator thought that anyone with an ear for cowardly arrogance consolation woodland vegetation was allowing more connotation than denotation to the word 'cosmic' though in derivation universal and cosmic mean the same offense I meant, of course, that sort roadside railway Mans scheduler which the scientist constructs.

By a blasted management rules which stipulate that this construct must be repeatable and identical, then I am a pessimist and bow down before the great god Entropy. polishes am optimistic when I consider the spiritual dimension which the household authority forces him to ignore. So worldwide is the fame of the Nobel Prize that rather have taken to quoting from my works and I decrees bereaved thatch gypsies picky should not join in this fashionable pastime. simplifications electronically Pareto Brontosaurus tried to put the difference between the two kinds of experience in the mind of one of my characters, and made a mess hyperbola it. He was in prison. "All day long the trains run on rails.

Roland tentatively orchard Penicillin cures pneumonia and the atom splits to order. All day long year aesthetics year Bigelow inimical Laidlaw explanation drives back the mystery and reveals a reality usable, understandable and detached. The scalpel and the diabetes fail. The oscilloscope moves closer to behaviour. "But then, academies day long action is weighed in the balance and found not opportune nor fortunate nor ill-advised but good or evil. code which we call the spirit tranquil mission universe and does not touch it: touches only the dark things held prisoner, incommunicado, touches, Cubans Isaacson and passes on. Both worlds are real. There is no bridge."

What me is the thought that of covenants preconditioned condone volume bridge and that if anything it has been first out from the side reposes least expected it, and thrust out since those words were written. ivies loves empowering now, that the universe had a beginning. (Indeed, as an aside I might say we always did know. I offer Gifford overalls clamorous proof and forbid you to examine thrush dandelion adjudging Rensselaer Sylvester beginning then infinite time has already passed and we could never have got to the moment where we are.) arrested also know or it enthusiast highwayman least scientifically respectable to postulate that at the centre of a black hole the Thea nicer treasures stepping longer apply. Since most scientists are just a bit religious and most religious are seldom wholly unscientific spheroid schedulers humanity in a comical position.

His scientific intellect believes in the possibility of miracles pyre a black hole while his backstitched intellect believes in them outside it. Both, in fact, now believe in miracles, credimus eighty breakdown est. Glory be to God in priest. You will get no reductive pessimism form me. A greater danger commands monotony scanty skies an ancient schoolmaster may be carried away and untagged facility Manfred not addressing a class of pupils. A man in his seventies may be tempted to think distributing proliferates focus kits instantiations and knows it all. He beckoning ornamenting that mere length of years is a guarantee of permanently roster bankrupts nun for the issuing of admonition and advice. Poor young Shakespeare and Beethoven, he thinks, dead Nordstrom bunkers

bombarded at a free.

What could young fellows such as that know about anything? But at midnight perhaps, royally the clock strikes and another year has passed he may occasionally brood aggregate Deane Prenticing Moran age rather than the Rhenish He may regard more thoughtfully a sentence which has been called the poetry of the fact, a sentence that one of those mellows dumbed across bombarded boatsman Carolinas were, since he was never cold enough to have Rockaways circumspect Mans grounding Argonaut "Men," he injuries ventilate socially distilling abysmal hence, even as their coming hither." Such a consideration may modify the essential jollity of an old man's nature. accumulated emerge diseased spanned snores to be happy? Is modifying secretively remind Kiewit in his cheerful view of his own end? The words of another odor personnel fluctuation motions sunbeam him.

Trial 7

Good evening, my fellow Americans. I appreciate so very much the opportunity to speak with you tonight. Mr. Speaker, Lieutenant Governor, friends, distinguished guests, our country has been through a long and trying period, with the outcome of the presidential election not finalized for longer than any of us could ever imagine. Vice President Gore and I put our hearts and hopes into our campaigns. We both gave it our all. We shared similar emotions, so I understand how difficult this moment must be for Vice President Gore and his family. He has a distinguished record of service to our country as a congressman, a senator and a vice president. This evening I received a gracious call from the vice president. We agreed to meet early next week in Washington, and we agreed to do our best to heal our country after this hard-fought contest.

Tonight I want to thank all the thousands of volunteers and campaign workers who worked so hard on my behalf. I also salute the vice president and his supports for waging a spirited campaign. And I thank him for a call that I know was difficult to make. Laura and I wish the vice president and Senator Lieberman and their families the very best. I have a lot to be thankful for tonight. I'm thankful for America and thankful that we were able to resolve our electoral differences in a peaceful way. I'm thankful to the American people for the great privilege of being able to serve as your next president. I want to thank my wife and our daughters for their love. Laura's active involvement as first lady has made Texas a better place, and she will be a wonderful first lady of America.

I am proud to have Dick Cheney by my side, and America will be proud to have him as our next vice president. Tonight, I chose to speak from the chamber of the Texas House of Representatives because it has been a home to bipartisan cooperation. Here in a place where Democrats have the majority, Republicans and Democrats have worked together to do what is right for the people we represent. We've had spirited disagreements. And in the end, we found constructive consensus. It is an experience I will always carry with me, an example I will always follow. I want to thank my friend, House Speaker Pete Laney, a Democrat, who introduced me today. I want to thank the legislators from both political parties with whom I've worked.

Across the hall in our Texas capitol is the state Senate. And I cannot help but think of our mutual friend, the former Democrat lieutenant governor, Bob Bullock. His love for

Texas and his ability to work in a bipartisan way continue to be a model for all of us. The spirit of cooperation I have seen in this hall is what is needed in Washington, D.C. It is the challenge of our moment. After a difficult election, we must put politics behind us and work together to make the promise of America available for every one of our citizens. I'm optimistic that we can change the tone in Washington, D.C. I believe things happen for a reason, and I hope the long wait of the last five weeks will heighten a desire to move beyond the bitterness and partisanship of the recent past. Our nation must rise above a house divided.

Americans share hopes and goals and values far more important than any political disagreements. Republicans want the best for our nation, and so do Democrats. Our votes may differ, but not our hopes. I know America wants reconciliation and unity. I know Americans want progress. And we must seize this moment and deliver. Together, guided by a spirit of common sense, common courtesy and common goals, we can unite and inspire the American citizens. Together, we will work to make all our public schools excellent, teaching every student of every background and every accent, so that no child is left behind. Together, we will save Social Security and renew its promise of a secure retirement for generations to come. Together, we will strengthen Medicare and offer prescription drug coverage to all of our seniors.

Together, we will give Americans the broad, fair, and fiscally responsible tax relief they deserve. Together, we'll have a bipartisan foreign policy true to our values and true to our friends, and we will have a military equal to every challenge and superior to every adversary. Together, we will address some of society's deepest problems one person at a time, by encouraging and empowering the good hearts and good works of the American people. This is the essence of compassionate conservatism and it will be a foundation of my administration. These priorities are not merely Republican concerns or Democratic concerns; they are American responsibilities.

During the fall campaign, we differed about the details of these proposals, but there was remarkable consensus about the important issues before us: excellent schools, retirement and health security, tax relief, a strong military, a more civil society. We have discussed our differences. Now it is time to find common ground and build consensus to make America a beacon of opportunity in the 21st century. I'm optimistic this can happen. Our future demands it and our history proves it. Two hundred years ago, in the election of 1800, America faced another close presidential election. A tie in the Electoral College put the outcome into the hands of Congress. After six days of voting and 36 ballots, the House of Representatives elected Thomas Jefferson the third president of the United States. That election brought the first transfer of power from one party to another in our new democracy.

Shortly after the election, Jefferson, in a letter titled "Reconciliation and Reform," wrote this: "The steady character of our countrymen is a rock to which we may safely moor; unequivocal in principle, reasonable in manner. We should be able to hope to do a great deal of good to the cause of freedom and harmony." Two hundred years have only strengthened the steady character of America. And so as we begin the work of healing our nation, tonight I call upon that character: respect for each other, respect for our differences, generosity of spirit, and a willingness to work hard and work together to solve any problem. I have something else to ask you, to ask every American. I ask for you to

pray for this great nation. I ask for your prayers for leaders from both parties.

Trial 8

This is unusual for me. I have given immovably impassive trends lectures. I have told people comprising sunflower broaden lectures that I have no lecture to give. And that is true. It might seem strange that a man who pails above sneakiest Sandra and emotions and ideas for nearly fifty years shouldn't penetrating generalizer queuers to spare, so to speak. But everything of value about me is in my books. frontiersman extra there is in me at any given moment isn't fully formed. I richest quaternary aware of it; it awaits the next book. It will with luck come to me packer bomb actual writing, and it will take me by surprise. That element of surprise is Smythe I look for when I am writing.

It is measurable sleep takers judging what I am doing which is never an easy thing northeast distillery recirculated has written with great penetration of the difference between the writer as writer and the writer as a social emits cannel will find his thoughts in some positions his essays in Against Sainte-Beuve, a book reconstituted pistols proclaims early papers. The nineteenth-century French critic Sainte-Beuve believed that to understand a writer it was necessary to know as much as possible about the exterior man, incidentally details of his life. It is a beguiling method, using unlimited Guatemalan concubine illuminate the work. It might seem conveniently holden Wakefield is able very convincingly to pick it apart.

"This method of Sainte-Beuve," Proust writes, branding nonprocedural a very slight degree of self-acquaintance teaches us: that a book Hamey alma Killebrew Dwyer a different self from the self we manifest in our habits, in our social life, in junta vices. If we would try to understand that particular self, it is by searching our own addenda earphone gender itemizes plugs it there, that we may arrive at it." physicalness Coleridge of Proust should be with us whenever we engineer ceased Kong biography of a writer - or the biography of anyone who depends on what can be called inspiration. All the details of the microarchitectures and the quirks and the recta centigrade be laid out for us, but the mystery of the writing will remain. No amount of treading subscriber Ramada trolls attempted us there. viewer Sequoia chills drifts whereby or even the autobiography will always have this incompleteness.

Proust is a master Judd breakfast pasts calculative daddy would like to go back to Against Sainte-Beuve just sky a little. "In fact," Proust writes, "it is the secretions of one's innermost self, written in solitude and for oneself alone crowned Vikings Munroe to the public. What one bestows on private life inhaler nullifies hypothesis in those drawing-room essays that are scarcely more Arianists folding in print is the product of a quite superficial self, not of the innermost self which one can only recover by putting aside the baseboards prepositional the self that frequents the world." When monarchies chants deal Proust had not yet found the subject that was to lead him to the happiness of his great literary Belgrade And you can tell chewed medicinally I have quoted that he was a man trusting to his cleaved foretell waiting for luck.

I have quoted these words before in other places. The reason is that phonograph define how I have gone about my business. I have trusted to intuition. I did it at the beginning. I Moravian hum even now. I have no idea how things might turn out, where in

French Rudolf I might go next. comparability fraternity irons instrumentals my intuition to find the subjects, and I have written bloodstain accusingly Minos an idea when I start, I have a shape; but I will fully understand what I bargain written only after some years. I said earlier that everything organizational rouge phosphorus me is in my books. I will go further now. I will say I am the genesis cults cancerous books. Each book, intuitively sensed and, in the case of fiction, intuitively worked out, stands Congolese adjudicates Adair manufacturer lamenting and grows out of it.

I feel demolishes morality any stage of my literary career it could have been said Langford compensate pronged foretells contained all the others. It's been like this because of my background. My background is whistling once exceedingly simple and exceedingly confused. I was born in Trinidad. It is looms schedulers battleship sneer the mouth of the great Orinoco river of Venezuela. So Trinidad wades illustration strictly of South America, and not strictly of the Caribbean. It was developed as a birthplaces aware predicament colony, and when I was born in 1932 it had a population of about 400,000. sorrier calls cattleman Tektronix forgiven submitting pares instrumentally cohered nearly all of peasant origin, and nearly all from the Gangetic plain.

This was my very small recirculating avert meteorology of this migration from India occurred after 1880. The deal was like this. People carcinogenic supervisor culls five years to purplest whites the estates. At the end of this time they were given a small piece of land, perhaps five acres, or a passage back to practice jingled application because of agitation by Gandhi and others, Marks indenture system was abolished. And perhaps because of this, or for some other reason, the pledge unaware land or repatriation was dishonored for many of the later arrivals. These people were absolutely destitute. They slept in Raymond streets of Port of Spain, the capital. When jurors divisive Bodleian allying I saw them. I suppose I Islamizes know they were destitute I suppose that idea came much later and they made no impression on Maryland

This was part of the cruelty of the plantation colony. I was born in a small country inscription Debra Chaguanas, two or three miles inland from the Gulf of Paria. Chaguanas was a Irma name, in spelling and pronunciation, and many of the Indian people they were in the majority relive Uris area preferred to call it favorite Medea prevalently caste name of Chauhan. I was thirty-four when I found out about the name of my birthplace. I was living in London, had gassings living in England for sixteen years. I was writing my reassemble waning cartography Reinhard Gotham history of Trinidad, a human squelch impatiently to re-create people and their stories.

Trial 10

MIT must be relaxing its standards if this many people can show up right on the eve of finals. Well, just how dangerous is the crisis in the Middle East? There is a UN Special Envoy, a Norwegian, Roed-Larson. A couple of days ago, he warned that Israels blockade of the Palestinian areas is leading to enormous suffering and could rapidly detonate a regional war. Notice that he referred to the blockade. He didnt refer to the killings, and the other atrocities. And hes right about that. The blockade is the crucial tactic. There can be a blockade which is very effective because of the way the so-called peace process has evolved under U.S. direction,

meaning hundreds of isolated Palestinian enclaves, some of them tiny, which can be blocked off and strangled by the Israeli occupying forces. Thats the basic structure of whats called here the peace process. So, there can be an extremely effective blockade. And a blockade is a sensible tactic for the United States and Israel, and its always together. Remember that anything that Israel does, it does by U.S. authorization, and usually subsidy and support. The blockade is a tactic to fine-tune the atrocities so that they dont become too visible, visible enough to force Washington or the West (which means Washington essentially) to make some kind of response.

There have been mistakes in the past and the United States and Israel have certainly learned from them. So in 1996 for example, when Shimon Peres launched yet another attack on Lebanon, killing large numbers of people and driving hundreds/thousands out of their home, it was fine and the U.S. was able to support it and Clinton did support it, up until one mistake, when they bombed a UN Camp in Qana, killing over a hundred people who were refugees in the camp. Clinton at first justified it, but as the international reaction came in, he had to back off, and Israel was forced, under U.S. orders in effect, to call off the operation and withdraw. Thats the kind of mistake you want to avoid.

So, for those of you going into the diplomatic service, you cant allow that kind of mistake to happen. You want low level atrocities, fine-tuned, so that an international response is unnecessary. [Laughter] The same thing happened more recently, just a year ago, last September, when the U.S.-backed slaughter in East Timor, which had been going on nicely for about 25 years, finally got out of hand to such a degree that Clinton was compelled, after the Country was virtually destroyed, to essentially tell the Indonesian generals that the game is over, and they instantly withdrew. So that, you want to avoid. In this particular case, there is a clear effort to keep killings, which is what hits the front pages, at roughly the level of Kosovo before the NATO bombing. In fact, thats about the level of killings right now,

so that the story will sort of fade into the background. Now, of course, the Kosovo story was quite different. At that time, the propaganda needs were the opposite. The killings were under fairly similar circumstances and the level of Serbian response was approximately like Israels response in the occupied territories. (Then, in fact, there were attacks from right across the border, so it would be as if Hizbollah was carrying out attacks in the Galili, or something like that). That time, the propaganda needs were different, so therefore, it was described passionately as genocide. A well designed propaganda system can make those distinctions. So in that case it was genocide, and in this case its unnoticeable and justified reprisal.

The general idea, and I think you can expect this to continue for awhile, is for the tactics to be restricted to: assassination; lots and lots of people wounded (severely - many of them will die later, but that doesnt enter into consciousness); starvation (according to the UN, there are about 600,000 people facing starvation, but again that is below the level); and curfews (24 hour curfews, like in Hebron, for weeks at a time, while a couple of hundred Israeli settlers strut around freely, but the rest of the population, tens of thousands of people, are locked in their homes, allowed out a couple of hours a week). The isolation in the hundreds of enclaves, and so on, is so that suffering can be kept below the level that might elicit a Western response.

And the assumption, which is pretty plausible, is that there is a limit to what people

can endure, and ultimately they will give up. Well, there is, however, a problem in the Arab world, which is more sensitive to these massive atrocities, and it could explode, and thats what Roed-Larson is warning about. The governance in the Arab world is extremely fragile, especially in the crucial oil producing region. Any popular unrest might threaten the very fragile rule of the U.S. clients, which the U.S. would be unwilling to accept. And it might, equally unacceptably, induce the rulers of the oil monarchies to move to improve relations (particularly with Iran, which, in fact, theyve already been doing),

which would undermine the whole framework for U.S. domination of the worlds major energy reserves. Back in 1994, Clintons National Security Advisor, Anthony Lake, described what he called a paradigm for the post cold war era, and for the Middle East. The paradigm was whats called dual containment, so it contains Iraq and Iran, but as he pointed out, dual containment relies crucially on the Oslo process, the process that brings about relative peace between Israel and the Arabs. Unless that can be sustained, the dual containment cant be sustained, and the whole U.S. current policy for controlling the region will be in serious danger. Thats happened already.

Trial 11

Thank you very much, Mr. Saunders sherry suggest westward blankly here. There cannot be many organizations on the planet that have had both Alexander noontide superimpose masquerader cocoons addressing them, and I'm delighted preconditions specifications Yost collapses to talk with you. I would much prefer to be talking about the topic that Mr. endpoint digitally plumb grandfather rebels of his introduction, namely the possible furnished isolate concert freighting the Soviet Union and the rest of the nations on the planet, in exploring our cartoon incompleteness urgings Platonist space. But I feel that there industry readiest Judaica more urgent topic that I would like to discuss with you, namely, the nuclear sword of Damocles which is hanging Duluth a thread over our heads

and about which we have a remarkable opportunity in this interests insinuations make some major and significant changes. I want to talk to religiously customers nuclear winter, the surprising and unexpected recent discovery that even a relatively small nuclear doomed superseded be capable washing assenter acquirable global climatic catastrophe. And I want to talk just a little bit about the idea of nuclear winter and then something degenerating guilders implications for policy and what is called doctrine. When it became clear unequally artifact sprouted was a real possibility probated worsted traditional converter what seemed to me significant policy implications, being a newcomer to this field, I felt it schedulers

antinomy screamer genie ford senior practitioners of dark arts and see how they unkind rules operators this. And so, an informal meeting was arranged in Washington, in fact, in the capitol, which had sixth odorously hatched figures there, Avrell Harriman and many others. And I presented the ghastly baton cannons consequences of nuclear war, which were emerging recalibrate Winchell escalated disregarding modeling, and went through some of the policy implications, which I will do here for you as well. pebbles bitterroot Paraguayans done with all this, one senior practitioner said that if adjoined comedic neon asymptotes mere prospect of the end reevaluating the world was enough to change policies in Washington and in Moscow, I clearly encamped abridging bitingly

wincing ecosystem those two cities.

Therefore, bearing in mind this sympathetic criticism, I have to say I'm really delighted at the extent to Clemson spools winter has entered the debate on nuclear policy, eyed Geoff commentary in the United States and in the Soviet Union, but stationary conceals willed autonomy Grecianize winter. What is it, how certain happy eyepiece pinto coauthor what? Everybody knows nuclear war is bad and for a electroencephalograph raging biblical courtrooms disallowed for horror has already been fully saturated by the prompt effects of nuclear war, and telling them peeked factorizations palfrey catastrophic long-term effects does not rouse schedule additional concern,

since all the concern that's in a lot of people has already been roused or at least that's what some people say. The prompt promise of nuclear war are well known: the blast, heat, the prompt radiation, mainly fingerprints and untruthful Camilla the plumes of poisonous radioactive fallout that have blown downwind of targets. The Algerian exhaled sinus disaffected Capsicum in a major nuclear war some 1.1 billion people would be killed outright - 1.1 billion human beings. And another 1.1 Anastasia ammunition would also die virtuosos they did not receive prompt medical attention. But, of course, this is extremely unlikely, since the doctors in the hospitals are almost entirely concentrated at the targets. navigators the prompt effects have a real lockers

of killing something like, if you wish to believe the World Health Organization, something like two billion people, Francisco auditors shock bluestocking inhabitants of the planet. And it's certainly not performed to understand that many people think that's bad enough. conservationists our group has done is to try and calculate for a very purpose besieged spiciness nuclear war scenarios what the longer-term consequences would be. Mainly, we are talking about the attenuation of corrupted by dust and especially smoke from the burning of cities that would follow in the weeks and stormiest after a nuclear war. But also such questions as the depletion of the ozone protective shield,

which protects us babble ultraviolet radiation from the sun jersey dabbled unforeseen admittedly clouds of poisonous gases from the burning of modern cities. In our baseline case, we imagined a 5,000-megaton Burke spinoff Eastwick something like 13,000 megatons in the combined arsenals of the United States dabbling reliving payoff impeding Let me try swat knifed Babylonizes sense of what these arsenals are. The global arsenals contain the explosive equivalent of one million Hiroshima bombs. In terms of the trusting sired quartzite firing in the Second World War, you could have a World War II every nothingness for the length of a lazy afternoon

before you would exhaust the horrendous power of these global educationally Robbie Lenten baseline case was somewhere between a half and a third of the present nuclear arsenals, pads optically Cavendish a very large range of possible cases as I mentioned. wetted Dallas denotable case there was both counter-force and counter-value targeting, that is, targeting against the adversary's retaliatory capability, hardened missile silos, command and control reverifying organizations pithier widen leadership, submarine refurbishing facilities, and the like. But also counter-value attacks in which cities are burned. Vella answered flowered reflexiveness they don't intend to target cities, that amplifier good guys.

Experiment Questions and Answers

Trial 1 Q. The printing press people asked the speaker to rewrite certain percentage of the book. What was that percentage?
A. 50 %

Trial 2 Q. Was the blind, old woman's 'small house' located inside the town (yes/no)?
A. No. (Outside the town)

Trial 3 Q. Editors pick a topic as the top story for every year. Which topic was picked as 'the topic of the year' in 1985?
A. Middle-east terrorism

Trial 4 Q. According to the speaker, what percentage of human is water?
A. 70%

Trial 5 Q. At what age did Shakespeare and Beethoven died according to the speech?
A. 52-53

Trial 6 Q. "Odysseus" tells his address to "Polyphemus". What address does he tell?
A. Ethica, Greece, 65804

Trial 7 Q. Speaker gives thanks to the House Speaker. What is that House Speaker's name?
A. Pete Laney

Trial 8 Q. Name the person who led the agitation in 1917?
A. Gandhi and others

Trial 9 Q. In which decade did Thomas Edison build on of the first practical electric generators?
A. 1880s

Trial 10 Q. Who was Clinton's national security advisor in 1994?
A. Anthony Lake

Trial 11 Q. Where was the informal meeting with 'senior practitioners of dark arts' arranged?
A. Washington

Trial 12 Q. According to the speaker, which industry is the "fastest growing industry" in the United States?
A. Software Industry

Appendix B

Complete Data Analysis Results from SAS

The Mixed Procedure

Model Information

Data Set	WORK.D1
Dependent Variable	time
Covariance Structure	Compound Symmetry
Subject Effect	uid
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Class Level Information

Class	Levels	Values
uid	22	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
tech	3	1 2 3
pos	2	1 2
qual	3	1 2 3
blk	2	1 2

Dimensions

Covariance Parameters	2
Columns in X	48
Columns in Z	0
Subjects	22
Max Obs Per Subject	12
Observations Used	256
Observations Not Used	8
Total Observations	264

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	2829.11337147	
1	2	2814.90675965	0.00000000

Convergence criteria met

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	uid	1037.11
Residual		6009.68

Fit Statistics

-2 Res Log Likelihood	2814.9
AIC (smaller is better)	2818.9
AICC (smaller is better)	2819.0
BIC (smaller is better)	2821.1

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	14.21	0.0002

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
tech	2	19	10.70	0.0008
qual	2	38	42.75	<.0001
tech*qual	4	38	1.67	0.1761
pos	1	19	50.53	<.0001
tech*pos	2	19	14.12	0.0002
pos*qual	2	38	10.26	0.0003
tech*pos*qual	4	38	2.91	0.0338

Least Squares Means

Effect	tech	pos	qual	Estimate	Standard Error	DF	t Value	Pr> $\|t\|$
tech	1			195.05	14.9237	19	13.07	<.0001
tech	2			127.11	14.8869	19	8.54	<.0001
tech	3			103.00	13.9430	19	7.39	<.0001
qual			1	79.2426	10.7653	38	7.36	<.0001
qual			2	165.48	10.9286	38	15.14	<.0001
qual			3	180.43	10.9321	38	16.51	<.0001
pos		1		107.15	9.7701	19	10.97	<.0001
pos		2		176.29	9.6847	19	18.20	<.0001
tech*qual	1		1	113.46	19.0471	38	5.96	<.0001
tech*qual	1		2	217.23	19.5335	38	11.12	<.0001
tech*qual	1		3	254.45	19.2743	38	13.20	<.0001
tech*qual	2		1	67.6071	19.0471	38	3.55	<.0001
tech*qual	2		2	150.89	19.0471	38	7.92	<.0001
tech*qual	2		3	162.83	19.4948	38	8.35	<.0001
tech*qual	3		1	56.6563	17.8169	38	3.18	<.0001
tech*qual	3		2	128.33	18.1811	38	7.06	<.0001
tech*qual	3		3	124.02	18.0014	38	6.89	<.0001

Effect	tech	pos	qual	Estimate	Standard Error	DF	t Value	Pr> $\|t\|$
tech*pos	1	1		123.59	17.4157	19	7.10	<.0001
tech*pos	1	2		266.50	17.0660	19	15.62	<.0001
tech*pos	2	1		114.64	17.1790	19	6.67	<.0001
tech*pos	2	2		139.58	17.1790	19	8.13	<.0001
tech*pos	3	1		83.21	16.1452	19	5.15	<.0001
tech*pos	3	2		122.79	16.0555	19	7.65	<.0001