Proceedings of I-KNOW '05
Graz, Austria, June 29 - July 1, 2005

Audio Spatial Organisation Tool to Support Audio Knowledge Work

Kirstin Lyon and Peter J. Nürnberg
(Aalborg University Esbjerg, Denmark
{kirstin,pnuern}@cs.aue.auc.dk)

Abstract: Extracting knowledge from audio recordings is frequently necessary in usability studies. At present, recordings are usually transferred to text before analysis begins. This is a long, tedious process, during which some information is lost. When analysing information, analysts work in text analysis tools that encourage them to organise ideas linearly. This is not always convenient to do. In this paper, we describe a tool that will allow users to organise audio recordings in a spatial environment without the need for full transcriptions.

Key Words: knowledge management, audio recording organisation, spatial hypermedia

Category: H.4 Information systems Application, H.5 Information Interfaces and Presentation

1 Introduction

Qualitative research is an important aspect of product design. It is used in the initial stages to gain insights before a product is created, throughout the implementation process to check user needs are being met, and as part of the testing and evaluation phase [1].

At present, most audio information gathered is transferred to text. Much information may be lost in this process, as it is not always possible to accurately convert speech into text. It is then organised linearly in a text analysis tool. Linearly-organised text may not be the most efficient method to view multi-dimensional information. In our work, we focus on the possibilities of importing and working with original recordings without the need for full transcriptions and allow implicit organisation within a 2D space.

This paper is organised as follows. Section 2 provides some background to the area of audio knowledge work. Section 3 describes a typical usability study scenario, and how audio data is analysed at present. Section 4 outlines our audio spatial organisation tool and how it is used. Section 5 describes our tool. Section 6 describes some related work. Section 7 concludes our paper.

2 Background

2.1 Audio as an Information Resource

Audio recordings are a useful knowledge resource, and may be gathered through various qualitative research techniques, such as interviews and focus groups. Obtaining audio information is relatively simple. When the interview is complete, the interviewer has a permanent record of what took place that can be revisited as often as required. This is a relatively unobtrusive method for collecting large amounts of data. Interviewers are able to concentrate fully on the interview instead of taking notes throughout. It is often easier to search through texts than to listen to them, as users may scan text very quickly for a particular word or phrase, whereas speech is presented linearly. Spoken words can convey more information than written text, with the meaning being different to the words that are used. The biggest disadvantage of capturing audio data is the length of the transcription stage. For every hour of tape, an experienced transcriber will need approximately six hours to transcribe it [2].

2.2 Computer-Supported Audio Knowledge Work

Computers may be used to extend users' capabilities. They are particularly useful at completing tasks that are automatic in nature, for example, sorting a list into alphabetical order. Analysis of audio recordings is user-intensive, with little that may be automated easily. There are a few ways to support an audio knowledge worker. These include speech recognition tools either used to transcribe recordings automatically, or to search for keywords within recordings [3]. This is difficult to achieve for many reasons, such as: tone of speaker; dialect; accent; rate of speech; number of speakers, etc. [4]).

Another approach that is being used successfully in knowledge management work is to use a human-centred approach; that is, to use a computer for easily automated tasks, and allow users to work with unstructured tasks that are difficult to automate. One research area that discusses how to extend user abilities is hypertext (or hypermedia). Organisation structures discussed within the hypermedia community include associative and spatial structures. Associative structures allow users to organise through linking items together. This takes advantage of our ability to remember objects through association. Early examples of associative hypertext tool include the "Memex" [5] and NLS [6]. They allow users to connect objects together without users needing to explain their choices. Associative "trails" can be built that could be viewed later.

Spatial hypertext takes advantage of users' visual and spatial intelligence. Users organise objects with spatial attributes, such as colour, size, location and proximity [7]. These tools allow users to build workspaces containing a variety

of documents and pictures. As with associative hypertext tools, users do not explain their choices. Often in information organisations, the relationships between information can be as important as the information itself. Knowledge of a problem space evolves over time. Spatial hypertext tools support users in creating structures when it is not clear how information should be organised.

3 Scenario

A focus group of 12 people discuss the bus network of Esbjerg. The chairperson has a list of issues to cover within the one-hour period, such as cost, timetable, efficiency, etc. Each attendee has opportunities to express their views about those points, as well as other issues that arise during the time. Once the meeting is finished, the chairperson gives a recording of the meeting to an analyst along with the original list of points.

3.1 Capturing Audio

In order to capture the audio, some form of recorder is used, either analogue or digital. The recorder is placed in the centre of the focus group where it can capture the participants voices. As the audio quality is important, it is necessary to test this before the participants arrive.

3.2 Transcribing

There are several strategies for transcriptions. Either the recording is transcribed word for word by a transcriber, or is imported directly into an analysis tool that supports audio files. The first option takes around one day. For the latter option, recordings are transferred directly to a computer, which takes a few minutes. Only the audio required is transcribed.

3.3 Analysing

An analyst first begins by looking at the key points that the chairperson wished to address. Starting at the beginning of the text, the analyst works through it, annotating areas that match with key points to get an idea of what general opinions were. This process continues, as it is not possible to get all the details during the first pass. This transcription is compared to other Esbjerg bus networks transcriptions to see if a consensus emerges. Knowledge develops over time. New transcriptions can bring out new issues that were mentioned briefly in previous transcripts.

Text analysis tools encourage list ordering. It is not possible to see all of the transcription at one time, so annotated areas can be collected into lists. Audio

sections can be placed between text entries for clarification purposes. A linear path through the transcriptions does not exist, so the analyst must remember some of the relationships between points. Once the analysis is complete, the findings are collected and presented in a report. The report is given to managers within Esbjerg's bus network.

4 Knowledge Work Tools in Audio

Present text analysis tools encourage users to organise linearly. For analysis purposes, this can be restricting, as one point in a discussion may lead to several others. Complex information can be difficult to represent as a list, as a file may belong in more than one place. Moving files takes time, and moving them may not be helpful. In this method, users are expected to find ways to remember relationships between themes. Users are often not able to see all of a transcript at one time. In this case, users are expected to remember, or to make a note of, common themes.

Some text analysis tools offer filters so themes and ideas can be grouped together into lists. However, when texts become heavily annotated, it can be difficult to distinguish between themes. Text analysis tools that are able support audio files simply insert audio clips into lists within the transcripts.

4.1 Spatial Organisation

Observations in offices have shown that users often use spatial placement to remind them where particular items are [8]. Using a spatial arrangement can help analysts to remember where particular themes are, instead of scanning lists. A difficulty of spatial organisation is becoming lost within the space. This happens when the workspace areas look similar to each other. In the physical world, users use landmarks and maps to navigate within an unfamiliar area. Our tool provides users with a 2D workspace. To help with navigation, we provide a map of the area, as well as landmark objects that aid in navigation.

4.2 Using Audio to Reduce Visual Overload

Visual overload is a known problem of spatial hypermedia tools. Most suggestions to decrease visual overload include using some form of visual cue [9],increasing the number of modalities used [10] or adding a third dimension [11]. However, information can be organised in many ways. Increasing the number of dimensions helps, but does not solve the problem entirely. Users have other abilities that may be taken advantage of, such as hearing. Our tool increases and develops the number of audio cues to reduce visual overload.

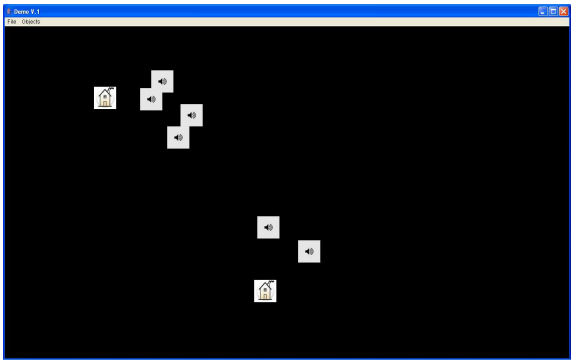

Figure 1: User interface showing the Viewport with landmarks and sound objects.

5 Current Work

5.1 Overview

Our tool is divided into three layers: user interface; model; and, storage mechanism (Fig. 1). We have implemented it as a monolithic system, but this structure could be divided into its three layers. Our tool is written in Java 1.5 using JOAL for our audio implementation. We used Java serialisation for our storage mechanism, but others may be used. The model is based on other spatial hypermedia tools, such as the Visual Knowledge Builder. Our interface also has elements of most spatial hypermedia tools.

5.2 Graphical User Interface

There are two parts to our graphical user interface; the viewport and the map. The viewport is the size of users monitors, and is their workspace(Fig. 2). Users are able to navigate outside of this area by dragging their cursor around their screen. From this view, users can create two kinds of object: a sound and a landmark object. A sound object is the audio file that is being organised. A landmark object can be used to help categorising the audio files.

To help users to maintain an overview of their workspace area, we have provided a map. The map shows each object in the workspace, as well as where the viewport is at present(Fig. 3). Sound objects have a different colour to that of landmark objects so that users may differentiate between the two. Users can navigate using either the map or the viewport.

5.3 Audio User Interface

In order to reduce visual overload, we have introduced audio cues into our tool. Visual overload occurs when the amount of visual information is too large for a

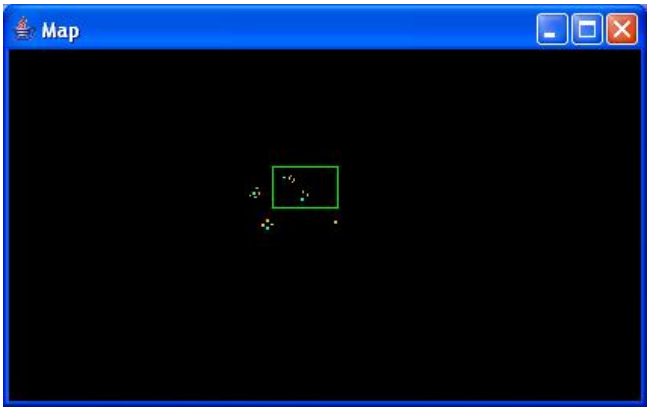

Figure 2: User interface showing the map. Landmark and sound objects have different colours. The viewport is also outlined.

user to make sense of. We suggest transferring some visual information to the audio domain in order to reduce visual overload.

Both the viewport and the map have audio cues attached to them. When users click on the map, the audio cue of the closest landmark is played. In the viewport, landmarks and sound objects have audio cues attached. Landmarks have a random short audio cue, users learn to associate the audio cue with the landmark. Sound objects have two sounds attached: the audio data, in this case and interview, and a composite of the closest landmark sounds. Relationships between sound objects can be built up in this way.

5.4 Implications for Scenario

A group of analysts first copies the audio recordings into their computers. Analysts listen to the first audio section, then place it somewhere in their workspace and move to the next audio file, which is easily compared with the first to see if some similarities exist between the two. Depending on the outcome, the second file is placed appropriately. Every time a new file is imported, the analysts listen to it and place it accordingly. This way the analysts can easily see what topics were covered more thoroughly, or if different themes are emerging. The analysts are able to listen easily to files and can quickly change placements of files. To get an overview of the workspace, the analysts look at the map; for finer detail, they consult the viewport.

6 Related Work

6.1 Media Analysis Tools

AnnoTape. AnnoTape [12] was developed with anthropological research in mind. It allows people to immediately transfer interviews into the system and only transcribe what is necessary. It is able to record, store and provide support for analysing files.

ATLAS/ti. ATLAS/ti is a powerful tool for analysing of large bodies of textual, graphical, audio and video data[13]. It offers a variety of tools for accomplishing the tasks associated with any systematic approach to qualitative data. It provides tools to help manage, extract, compare, explore and reassemble meaningful segments of large amounts of data in systematic ways.

Analysis. These tools organise sections into lists, with one section coming after another. It is difficult to organise information into a list, as one issue may lead to several issues, this is not easy to represent as a list. Our tool concentrates on audio organisation at present. Instead of list organisation, we use spatial organisation. Themes may be grouped together instead of following each other in a list. Audio files can be listened to easily, and moved in the space when appropriate

6.2 Spatial Organisation Tools

Visual Knowledge Builder (VKB). VKB is an example of a 2D spatial hypertext tool [14]. Users organise files spatially using colour, size, location and proximity to express relationships between files, or be grouped into collections. A spatial parser provides suggestions to users about possible relationships between files by looking at various visual attributes.

Data Mountain. The Data Mountain [11] is a 2.5D spatial organisation tool. 2.5D tools allow users full control over two axes, but the third remains discrete. Data Mountain was designed as an alternative to the bookmarking function found in Internet Explorer and similar applications. Users organise thumbnails of their webpages within a 2.5D environment. The further the thumbnail is from a user, the smaller it becomes.

Analysis. Spatial tools so far focus mainly on the arrangement of either text-based forms, or pictures. Audio formats are usually not supported and therefore it is not possible to use these tools for organising audio files. Spatial organisation has proven to be a successful way in which to organise documents, and has already been tested with users. We believe that it will be possible to extend these tools with audio abilities allowing audio to be interacted with them.

7 Conclusions

In this paper we described what audio knowledge work consists of and a possible tool to support audio knowledge workers. At present we focus on spatial placements and previewing of audio files. There are still difficulties in representing information that belongs in more than place. In the future, we will look into possible methods for showing other relationships, other than by placement. Possible solutions will be evaluated with users from our target groups.

References

1. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons, Inc. (2002)
2. Gauvain, J.L., Lamel, L., Adda, G.: Transcribing broadcast news for audio and video indexing. Communications of the ACM **43** (2000) 64–70
3. Furui, S.: Automatic speech recognition and its application to information extraction. In: Proceedings of the Thirty-seventh Conference on Association for Computational Linguistics, Association for Computational Linguistics (1999) 11–20
4. Shneiderman, B.: The limits of speech recognition. Communications of the ACM **43** (2000) 63–65
5. Bush, V.: As we may think. Atlantic Monthly (1945)
6. Engelbart, D.C.: Augmenting human intellect: A conceptual framework. Technical report, Direction of Information Sciences, Air Force Office of Scientific Research, Washington (1962)
7. Shipman, F.M., Marshall, C.C.: Spatial hypertext: an alternative to navigational and semantic links. ACM Comput. Surv. **31** (1999) 14
8. Barreau, D., Nardi, B.A.: Finding and reminding: file organization from the desktop. SIGCHI Bullit. **27** (1995) 39–43
9. Shipman, F.M., Marshall, C.C., LeMere, M.: Beyond location: hypertext workspaces and non-linear views. In: Proceedings of the Tenth ACM Conference on Hypertext and Hypermedia : Returning to our Diverse Roots, ACM Press (1999) 121–130
10. Brown, M.L., Newsome, S.L., Glinert, E.P.: An experiment into the use of auditory cues to reduce visual workload. In: CHI '89: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM Press (1989) 339–346
11. Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., van Dantzich, M.: Data Mountain: using spatial memory for document management. In: Proceedings of the Eleventh Annual ACM Symposium on User Interface Software and Technology, ACM Press (1998) 153–162
12. AnnoTape: (http://www.annotape.com)
13. Atlasti: (http://www.atlasti.com)
14. Shipman, F.M., Hsieh, H., Maloor, P., Moore, J.M.: The Visual Knowledge Builder: a second generation spatial hypertext. In: Proceedings of the Twelfth ACM Conference on Hypertext and Hypermedia, ACM Press (2001) 113–122