

Applying Open Hypermedia to Audio



David DeRoure, Steven Blackburn, Lee Oades, Jonathan Read, Neil Ridgway

Multimedia Research Group
Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK
Tel: +44 (0)1703 592418

E-mail: {dder, sgb97r, lro96r, jnr95r, cnhr}@ecs.soton.ac.uk

ABSTRACT

We describe a set of tools to support navigational hypermedia linking within audio ('branching audio') and between media types including audio. We have adopted an open hypermedia approach, with a component-based architecture, and aim to be compliant with the emerging Open Hypermedia Protocol (OHP). Content-based navigation is supported and we have focused on speech and musical content for our case studies. Although our investigation concentrates on audio, many of the techniques are generic and therefore applicable to other temporal media.

KEYWORDS: Open hypermedia, content-based navigation, Open Hypermedia Protocol (OHP), branching audio.

INTRODUCTION

The open hypermedia approach is well established but there is little experience of its application to temporal media. We have built a set of prototype audio tools to demonstrate and explore the issues involved in applying the principles of open hypermedia to audio, and we extend our treatment to include content-based navigation. The prototype tools were demonstrated at ACM Multimedia 1997 in Seattle.

The design of our tools has been driven by a number of scenarios, two of which are described in the following section. This is followed by an overview of the tools, discussion of experiences and our plans for future work.

THE SCENARIOS

Perhaps the most familiar opportunity for links within audio ('branching audio') or video occurs in structured presentations. For example, lectures, documentaries or meetings, which typically commence with an overview. When playing back a recording links can be made available from within the overview to the corresponding sections of the presentation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HyperText 98 Pittsburgh PA USA

Copyright ACM 1998 0-89791-972-6/98/ 6...\$5.00

We are exploring this scenario through recordings of meetings held using conferencing facilities such as the MBONE tools, providing a source of structured multimedia content which users wish to navigate. We are also assisting historical researchers working with speeches, such as those of Winston Churchill. Here the particular requirement is for a close coupling between the digital audio and the text transcripts, and for finding occurrences of similar phrases by content-based navigation based on text.

Our second scenario involves musical content. Interaction with musical structure usually occurs in the composition, production and publishing of music and is perhaps less familiar to the average end user. Where multiple versions of a performance are required, our approach is to treat each as an alternative view on the performance; i.e. structure is 'first class'. Content based navigation has a role in facilitating the authoring of these structures, and plays a part in delivery for use by musical researchers and educators.

In addition to the techniques adopted with speech, we have employed the MIDI format as a useful abstraction of musical performance for our investigation: it is closely associated with a digital audio representation via time-stamps and position pointers. We can abstract pitch contours [3, 5] to facilitate various matching operations, and we can convert digital audio to MIDI by pitch tracking.

These scenarios raise a number of research issues. Presentations, such as a radio documentary, can be viewed as a guided tour through a branching structure, and by default this tour is 'pushed' as a stream; we propose that the user could interact in order to follow different routes. A close coupling is required between different media (e.g. audio and MIDI) and, together with structural and external links, the associations need to be preserved through conversion between media types (e.g. synthesis) and format conversion within a type (e.g. downsampling and compression). Audio information delivered as a stream may be transient.

TOOLS

We have developed a suite of tools to implement the scenarios. Tools included in the suite are:

- *Link Manager*. This component receives source endpoints from the other tools and resolves them using the link service (e.g. a local linkbase, or a remote OHP service), presenting available links to the user via an appropriate interface component.
- *Link Player*. This is a general purpose media player (essentially a wrapper for the OS multimedia capabilities) which can send source endpoint information to the Link Manager, automatically or as a result of user interaction.
- *Content based retriever*. This tool performs matching of features against a database, providing both match position and relevance information. Our prototype tool used pitch contours [3, 5].
- *Audio Linker*. This is an authoring tool which facilitates the creation of links between speeches and their transcripts, controlled using a Web browser.
- *Sequence Player*. This tool produces a linear sequence according to a description consisting of a series of media fragments, as required in producing a synopsis of an existing presentation for linking purposes.
- *Link Hider*. This tool embeds endpoint information directly within digital audio data. This is useful where transport is restricted to one digital audio stream, and for editing with standard audio software.
- *Streaming soundviewer*. This is an evolution of the Microcosm SoundViewer [4], designed to use RealTime Streaming Protocol (RTSP).
- *RTSP server*. This server deals with endpoint and link information as well as multimedia data streams.

The communication of endpoints between the components is very simple (compliant with URL syntax). When linking on content rather than position, the position information is used to identify content, and a feature extracted from this; it is sometimes useful to deal directly with content (e.g. with streams and 'unaware' applications) but this may lose context.

We have adopted a component-based approach to the development of the tools as this allows individual components to be replaced with alternative implementations. The system is easily extended to inter-operate with other open hypermedia systems by writing a component which communicates with the other system. The system is designed to be compliant with the Open Hypermedia Protocol (OHP) [2]; our design also addresses interoperability with an existing system, the Distributed Link Service (DLS) [1].

EXPERIENCE

Our first experiments with the prototype tools included production of branching presentations by editing (linear) recordings of lectures, linking a historical speech to its transcript and to external documents, production of alternative views of musical performances, and demonstration of content based retrieval and navigation. We have included video in our demonstrations by linking with the audio track. These experiments have provided proof of concept.

The tools have evolved in response to user feedback. For example, we have introduced file history to allow similar functionality to a web browser.

The requirements for the feature matching algorithm vary according to the nature of the data and activity. For example, a search of the database using an error-prone query, such as a contour extracted from humming, requires a different parameterization of the matching algorithm to searching for a selection in a linkbase (both of which are typically high quality).

CONCLUSIONS AND FUTURE WORK

We have shown that open hypermedia principles can be applied to audio and that scenarios exist where this can be useful.

The tools will continue to evolve through the various case studies, with a richer set of abstractions from content (such as alternative contour representations). The tools will continue to be adapted to track the emerging OHP.

There are architectural implications to feature matching in the context of OHP and the Distributed Link Service: where link resolution occurs remotely, how is the matching algorithm provided and parameterized? There are also open interface issues relating to the presentation and selection of available links.

We also plan to employ the tools in novel situations, to test the ubiquity of branching media.

REFERENCES

1. Carr, L., DeRoure, D., Hall, W., and Hill, G. The distributed link service: A tool for publishers, authors and readers, in Proc. Fourth International World Wide Web Conference: The Web Revolution, Boston, Massachusetts, USA, December 1995.
2. Davis, H., Lewis, A., and Rizk, A. OHP: A Draft Proposal for a Standard Open Hypermedia Protocol, in Proc. 2nd Workshop on Open Hypermedia Systems. UCI-ICS Technical Report 96-10, University of California, Irvine, pp. 27-53.
3. Ghias, A., Logan, J., Chamberlin, D., and Smith, B. C. Query by humming - musical information retrieval in an audio database, in Proc. Multimedia'95 (San Francisco, California, November 1995).
4. Goose, S., and Hall, W. The Development of a Sound Viewer for an Open Hypermedia System, in The New Review of Hypermedia and Multimedia, vol. 1, pp. 213-231, 1995.
5. McNab, R. J., Smith, L. A., Witten, I. H., Henderson, C. L., and Cunningham, S. J. Towards the digital music library: Tune retrieval from acoustic input, in Proc. DL'96, 1996., pp. 11-18.