Using Sonic Hyperlinks in Web-TV

Norbert Braun

Fraunhofer Institute for Computer Graphics (Fraunhofer IGD)
Darmstadt, Germany

Ralf Dörner

Fraunhofer Institute for Computer Graphics (Fraunhofer IGD)
Darmstadt, Germany

Abstract

The transfer of Hypermedia features to audio in an audio-visual environment is discussed, introducing sonic hyperlinks. Sonic hyperlinks are links annotated using sound within an audio stream that lead to arbitrary multimedia content. As an example application, sonic hyperlinks have been integrated in interactive Web-TV which is broadcasted via the Internet. A system architecture and implementation relying on commercial WWW technology like RealMedia is presented. The system includes an authoring tool, as well as the necessary presentation plugin for an Internet browser.

1 Motivation and Goals

Hypermedia features are well-known and investigated, there are a lot of applications using hyperlinks, e.g. online help tools or WWW browsing systems. These applications use hyperlinks in a graphical representation. This does not reflect the important role of the audio channel within an audio-visual presentation.

There are few graphic systems like [10] utilizing the audio channel. The sonification systems of [1, 4, 5] add sonic information on graphical interfaces. Systems for blind users like [3, 9] translate graphical and textual information, even textual hyperlinks, to sonic information, but essentially these systems do not add hypermedia features to audio information. Systems without graphic output like [11] use the audio channel to give information to the user. However, all these systems do not integrate and represent hyperlink annotation of the audio stream in the audio stream itself.

Enhancing the audio channel with an intramedial hypermedia structure offers an alternative to the visual channel. This is especially important if information is not redundant in the sense that it is only provided on the audio but not on the visual channel. In some hypermedia systems, a somewhat artificial solution for this problem is to show additional, isolated textual hyperlinks [10]. Relating audio information to textual hyperlinks, however, is not easy for a user to perform and distracts his attention from the actual information. Using hyperlinks represented in audio offers the advantage of a close relationship between information and annotation, because they stay within the same medium. Sonic hyperlinks are expected to be more intuitive as they are closer to natural communication in which it is common to inquire about additional information within a conversation. Further, all users are able to listen to different audio streams, and tell audio information from hypermedia annotation by taking advantage of the 'cocktail party effect' [2]. Annotated audio streams are a non-used interaction potential in most audio-visual presentations.

2 Users and Application Background

In the near future, Broadcast TV and the WWW will grow together. An important feature of the WWW is interactivity via hyperlinks in text, images and video. This implicates new interaction facilities for television on the Web in order to follow video hyperlinks or present synchronized Web pages.

The Web-TV market is growing rapidly; currently there are several well-known media content providers that distribute audio-visual presentations through the WWW, e.g. the Encyclopædia Britannica Inc., the German ARD TV stations or the US TV station CNN. As image quality is not very high due to limited bandwidth, the focus of applications lies not in the movie sector but in the area of information supply (news services, business TV), entertainment (video clips), education (museum TV, product training) and marketing (advertisements, commercials, product presentations, promotion). Accordingly, the users of Web-TV stem from a widespread area and are especially interested in information access and presentation. This can be accomplished by using hypermedia methodologies where sonic hyperlinks are a field that needs further exploitation.

3 Sonic Hyperlinks

In this section, we deal with the question of what constitutes a sonic hyperlink, i.e. hyperlinks annotated within an audio stream that lead to arbitrary multimedia content, and which requirements a system that supports sonic hyperlinks has to fulfil.

A sonic hyperlink works as follows: the author integrates the hyperlink in the audio stream and defines what happens if the user would like to follow the link. This additional information is transferred via the WWW to a user. When the hyperlink is presented to the user, a certain sound is played in parallel to the regular audio. If the user reacts using a certain interaction metaphor in a certain time interval, a system reaction occurs as defined by the author.

The following major components for a sonic hyperlink can be identified:

- The *sensitive area* of the audio information that is annotated; this is the hyperlink start- and stop-time: The sonic hyperlink's sound should be played during this time interval. Even if this is the main time interval for the sonic hyperlink, this does not mean that this interval is the user's only time interval for interaction; see point 'time delay'.
- The *time delay* for the user reaction (delay after hearing the sonic hyperlink): The time delay specifies the time interval in which the user's reaction is awaited by the system, because like other continuous media, a hyperlink on audio information can pass by with the user not reacting fast enough. The reaction interval should be longer than the interval represented by the hyperlink start- and stop-time, because of the announcement effect of a finished sonic hyperlink sound: users want to react, because they just heard the sound disappear. Figure 1 shows the time relations.

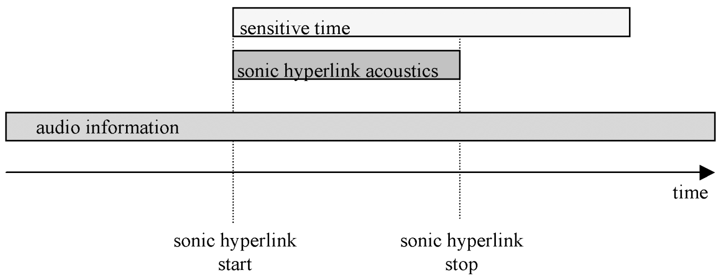

Figure 1: Sensitive Time of a Sonic Hyperlink

- The *signal sound* that represents the sonic hyperlink (e.g. earcons [5]), eventually indicating different kinds of information: The sonic hyperlink's sound is the signal for the user that there is something to react to. The sound should not disrupt the user's attention to the audio information, but should tell him in a comprehensible way that there is a hyperlink he can use.
- The *link-target* of the hyperlink (e.g. URL): The link-target of the hyperlink could be a web-site or another part of audio-visual content. Within the scenario of Web-TV, it would typically be another related audio-visual stream.
- The *user interaction metaphor* (e.g. keyboard stroke, mouse click, voice entry): The interaction metaphor should be oriented to the user's expectations.
- The *hypermedia navigation metaphor* (e.g. button, voice commands).
- The *system reaction* (e.g. display location, presentation break, rewind of the presentation): The systems reaction to the user when she interacts with a hyperlink is dependent on the application scenario.

We distinguish between three different application scenarios: If the scenario is a non-linear Web-TV presentation with a network-structured number of presentation streams, then the user may not want to get back to the point of the presentation where he initially has reacted. Hence, this presentation is stopped and the user listens to the (via sonic hyperlink selected) next presentation; see Figure 2.

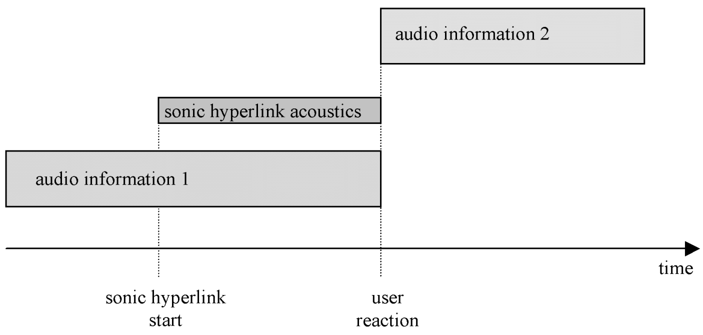

Figure 2: Network-like Presentation

If the scenario is a linear Web-TV presentation with some details concerning specific audio information, then the user might want to get back to the starting point of his reaction and especially if he wants some more information on where he is started. This leads to a presentation break in the main information stream with some rewind of the main presentation stream upon resuming the stream. This is just like a commercial break on normal TV: when the TV-programme continues, the presentation picks up a little bit before the commercial break. Figure 3 illustrates the time dependencies.

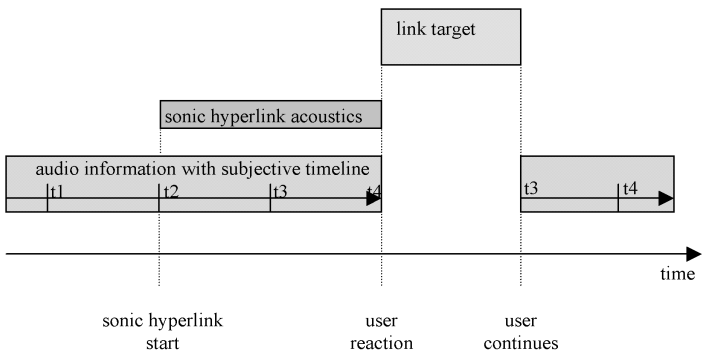

Figure 3: Subjective Timeline of Audio-visual Presentation After Break

If the presentation is live, there is no way of resuming. In this case, the requested target of the hyperlink is shown in an extra window (if there is more than one hyperlink activated, the targets could overload each other or be shown each in its own window). This is illustrated in Figure 4.

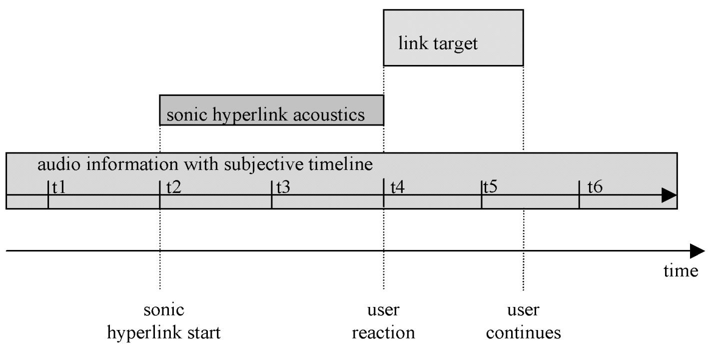

Figure 4: Live Stream Continues While the Hyperlink's Target Is Explored by the User

4 System Architecture

An environment that supports sonic hyperlinks has to take care not only of the conception of a sonic hyperlink itself but also of the authoring process, the transmission of the hyperlinks via the WWW, and the presentation component. It should allow the further definition or adjustment of the parameters of each sonic hyperlink component by the author or user, respectively.

An architectural decision to be made concerns the question whether the sounds for the sonic hyperlinks are integrated in the audio stream, send in a separate audio stream or whether timing information is sent, see Figure 5, 6 and 7.

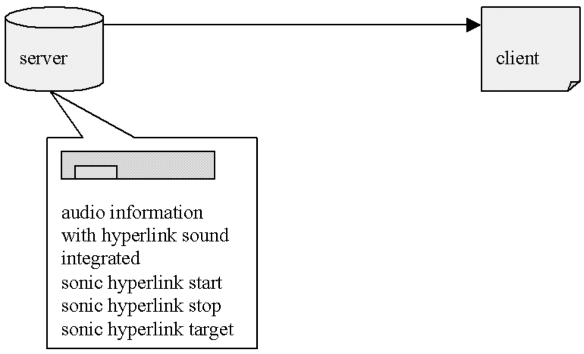

Figure 5: The Sonic Hyperlink Sound Is Mixed with the Audio and Stored on the Server

- In the first solution, see Figure 5, the audio information and the sonic hyperlink sounds are merged during the authoring process, then stored on a server. The user got to listen to both, without the possibility of changing or disabling the sonic hyperlink sound.

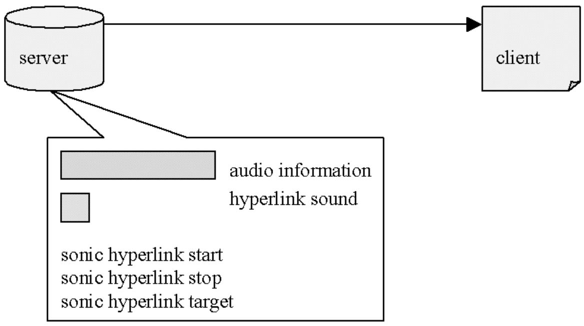

Figure 6: The Sonic Hyperlink Sound and the Audio Are Stored Separately on the Server

- The second solution, see Figure 6, potentially allows the user to modify the sonic hyperlink representation to the sounds offered by the author, since they are stored separately on the server. In case he does not want to listen to the sonic hyperlink sound, he could disable the sound and listen to the pure audio information.

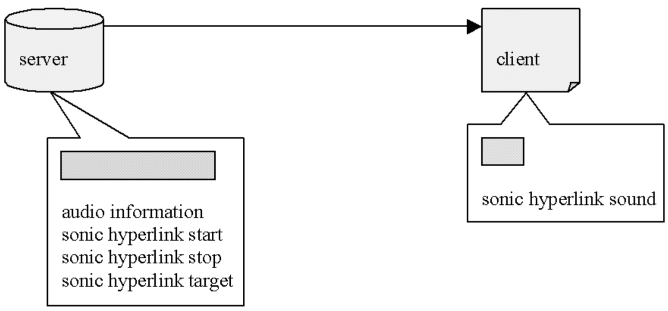

Figure 7: The Sonic Hyperlink Sound Is Stored by the Client

- The third solution, see Figure 7, allocates the responsibility for the sonic hyperlinks' sounds exclusively to the user on the client. He can pick out sounds or disable the sonic hyperlink feature as he chooses.

It seems that the union of the second and third solution provides a maximum degree of freedom to both, the author of an interactive audio-visual presentation and the listener. Currently, however, WWW video servers and clients only support the first solution.

4 Implementation of the System

The implementation of the system providing the sonic hyperlink is done using standard software like video server and web browser standard technology. We chose this solution as our aim was to use a commercial Web video server with an API. Here, we selected the RealMedia Web video server [8]. RealMedia has established itself as a de facto standard on the World Wide Web. A major benefit of this Web video server is that it also provides visual hyperlinks on the video stream, synchronized Web pages and a Java application programmers interface. Therefore, programming is done in Java as an Applet running in Netscape Navigator, Version 3.0 or later, using RealPlayer, Version 4.0 or later, from RealMedia for streaming audio/video and sonic hyperlinks.

In this section we describe the authoring characteristics, the service of an annotated audio-visual presentation and the user's reception tool.

4.1 Authoring of the Audio Stream

For authoring, we developed a tool that requires the original audio, sonic hyperlink start and stop times, link URLs, hyperlink sounds and volumes from the author. The tool is implemented using the RealMedia authoring API and tools. It adds Web page push events at a push-times relative to the presentation time to an audio-visual presentation. These Web page events are marked as sonic hyperlink target URLs using RealMedia's Synchronized Multimedia events. (RealMedia Synchronized Multimedia events consist of 3 parts: time, link URL and link target.) For each sonic hyperlink, we insert two events:

- The first event at the start time of the sonic hyperlink using the link target AHL_START
- The second event at the end time of the sonic hyperlink using the link target AHL_STOP.

This link target information is stored in a RealMedia stream. The sound of the sonic hyperlink is mixed with the audio stream exactly at the sonic hyperlink's sensitive time (excluding the user reaction time delay); the result is stored in a second RealMedia stream. These two streams are stored together in a third RealMedia stream which represents the annotated audio stream; see Figure 8. This annotated audio stream may be linked to a multimedia presentation with a video stream or pushed Web pages.

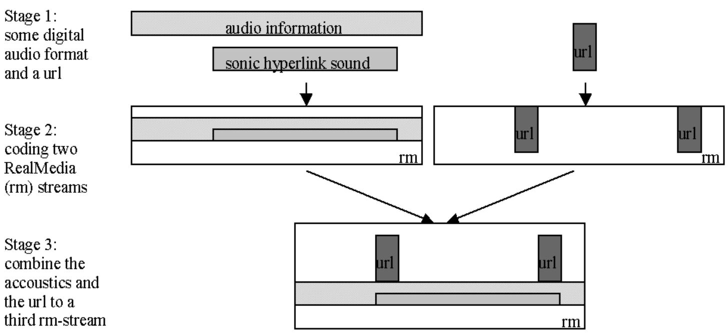

Figure 8: Assembly of Audio Information, Sonic Hyperlink Sound, Start, Stop and Target

The user can access the audio-visual presentation through the WWW. This means a Java applet must be placed in a web page together with a RealPlayer plugin object; see paragraph 4.3. The author can specify the presentation parameters and the elements to be displayed in the user interface in the applet's PARAM tags. This way, the author determines the degree of personal customisation for the user.

4.2 Service of the Annotated Audio-visual Stream

For transmission, the multimedia audio-visual presentation is served through the Internet via a non-modified RealMedia Web video server.

4.3 User client

The user perceives the audio-visual presentation within a web site through a web browser with a RealMedia client controlled by an applet. The applet checks Synchronized Multimedia events for a sonic hyperlink mark. If it detects one, it delays the event, waiting for a user reaction. If the user reacts, the applet pushes the event's link to the web browser. The applet stops the presentation (if adjusted) and resumes the presentation as specified.

The applet's functionality is divided into four modules:

- the main applet

- the user interface
- the RealPlayer connection module – depends on chosen media server
- the user reaction module – depends on desired user interaction possibility

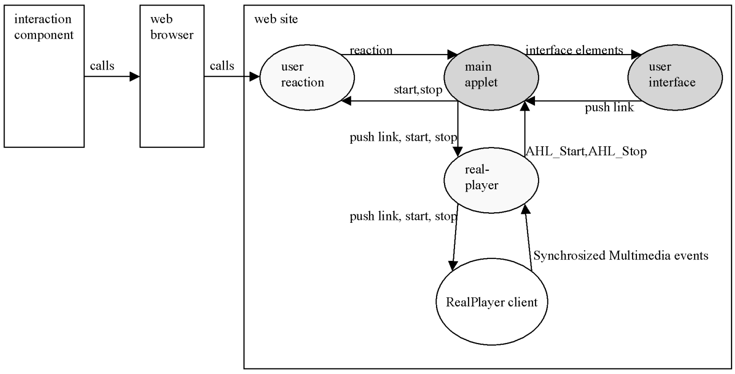

Figure 9: The Client's Structure

Figure 9 shows the structure of the modules and their interaction with the RealMedia client, the web site and the web browser. In the following, a brief description of these modules is provided.

Main Applet

The main applet must be started in a web page together with a RealPlayer plugin object. It creates one instance of each individual module.

If the applet receives a new sonic hyperlink (a Synchronized Multimedia event marked with AHL_START), it starts the user reaction module, stops (if adjusted) the media stream and tells the user interface to add the new sonic hyperlink to its list.

When the applet receives the end of a sonic hyperlink (Synchronized Multimedia event marked with AHL_STOP) it stops the user reaction module which then delays the actual stop of the user interaction by the specified time.

If the main applet gets a reaction call by the user reaction module or a link from the user interface, it pushes the sonic hyperlink target, means the Synchronized Multimedia event's link, to a destination window of the web browser, as specified in the user interface module.

User Interface

The user interface may be customized through applet PARAM tags in the web page. The user interface is able to display the following elements:

- A list of available sonic hyperlinks received thus far
- A choice of what to do upon user reaction (stop the media stream, continue the media stream)
- A choice of where to push the followed sonic hyperlink
- A text field to set the user reaction delay

In addition to the interaction module, a user can follow an (expired) hyperlink by doubleclicking an entry of the user interface's sonic hyperlink list. The user interface then gives the link to the main applet.

RealPlayer Connection

The RealPlayer connection module connects the applet to the RealPlayer object of the web page. It can start, stop and pause the RealPlayer on command by the main applet. Most importantly, it receives the Synchronized Multimedia events of the RealPlayer and informs the main applet.

User Reaction

The user reaction module implements the possibility for the user to interact. It gets activated and stopped by the main applet to allow user reaction (by keyboard, voice) to sonic hyperlinks. If the user reacts during the specified time, it calls the main applet to push the link to the web browser.

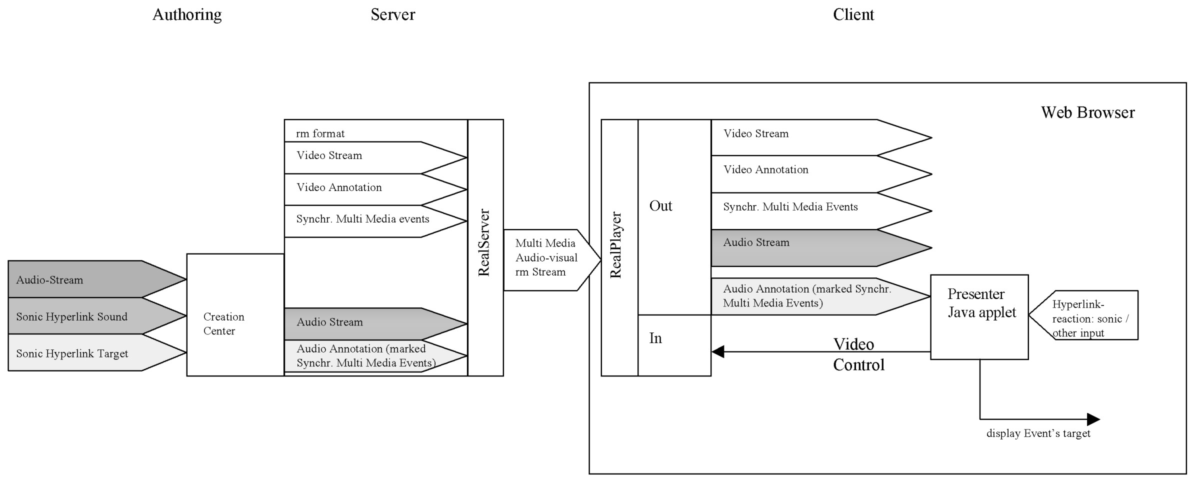

Figure 10: Sonic Hyperlink System Diagram

In Figure 10 a diagram of the system providing sonic hyperlink is shown. The coloured parts show the flow of audio information through the three components authoring, service and reception.

Summary

For presentation, the multimedia audio-visual presentation is received by the RealMedia Web video client. This client is controlled by a Java applet. With this approach, it is possible to adjust the following parameters at authoring time:

- The sensitive area of the audio information.
- The link-target of the hyperlink.
- The sound that annotates the audio information.

At presentation time, the user is able to modify the following parameters while every parameter is preset at authoring time:

- The user reaction time delay.
- The system reaction on a chosen sonic hyperlink: new link-target window (every target is displayed in the same window); new link-target windows (new window for every target), same window (reload the presentation window with the target).
- The system reaction upon selection of a sonic hyperlink: presentation break or presentation carry on.
- The system reaction concerning the resumption of the presentation: rewind-time of the presentation.

Graphical display of the URLs of the sonic hyperlink.

4.4 Implementation problems

RealMedias Synchronized Multimedia events are specified by start time, stop time, URL and target. The event is delivered to the applet only at the start time of the event, but not at the end time. For this reason we use two Synchronized Multimedia events for one sonic hyperlink, one at the start time, one at the stop time.

With the current technique of sending a sonic hyperlink start and a sonic hyperlink stop event problems can occur when the user is able to alter the timeline of the presentation, e.g. if he can skip forward or backward arbitrarily. There are two cases where problems occur:

- Skip the presentation forward or backward within the sensitive time of a sonic hyperlink, means 'jumping out'.
- Skip the presentation forward or backward the presentation into the sensitive time of a sonic hyperlink, means 'jumping in'.

When 'jumping out', the main applet can not receive the end of a sonic hyperlink, therefor it won't stop the user reaction module.

'Jumping in' leads to a bigger problem: How can the applet know a sonic hyperlink is active if it has no chance to receive the corresponding AHL_START event? For this reason we disabled the RealPlayer controls so the media stream can only be (re)started, paused and stopped.

5 Evaluation and System Demonstration

First evaluations of our tools showed promising results. Although additional information needs to be transferred, sonic hyperlinks do not cause any decrease in the transfer rate and the quality of service. Usability evaluations showed that although users are not familiar with sonic hyperlinks, they quickly get used to them. In fact, users seem to prefer sonic hyperlinks to video hyperlinks if no textual hyperlinks are provided. Our explanation is that with decreasing bandwidth, video quality is reduced by a higher amount than audio quality. This is because video servers have a higher priority for audio than for video, as users tolerate low video quality rather than low audio quality.

The sound of the sonic hyperlink is of significant importance. We conducted several tests with earcon-like sounds: The volume of the sonic hyperlink sound should be a little softer than the original audio information. On tests with stereo sound and 3D-sound showed that the sonic hyperlink sounds should come from the same direction as the annotated audio information, otherwise the user seems to be confused.

We achieved good results when we were not changing the user's information channel. This means, if the user hears audio information and sonic hyperlink annotation, he should be able to react with his voice. This leads us to intramedial interaction. In contrast, [10] shows an intermedial approach to the user's interaction: even if the audio information is annotated, the annotation is via text. This leads to a break in the user's natural conception of reacting to audio information.

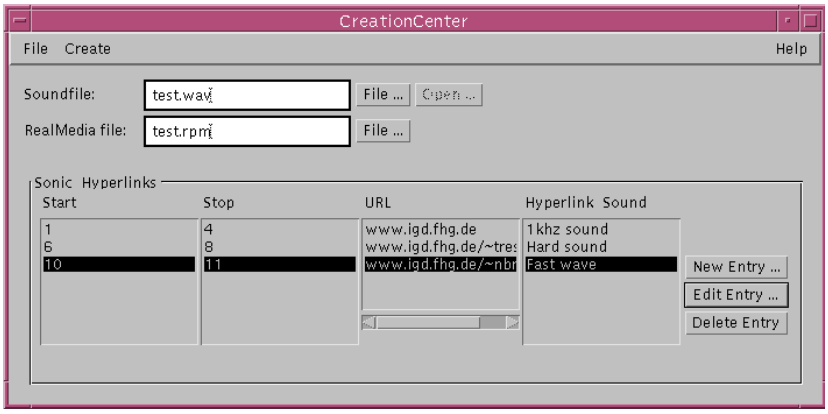

Figure 11: CreationCenter Main Window

In a demonstration, we will show the CreationCenter, our authoring tool for defining audio hyperlinks. The CreationCenter main window is depicted in Figure 11 while the edit window for specifying the adjustable parameters for one single sonic hyperlink is shown in Figure 12.

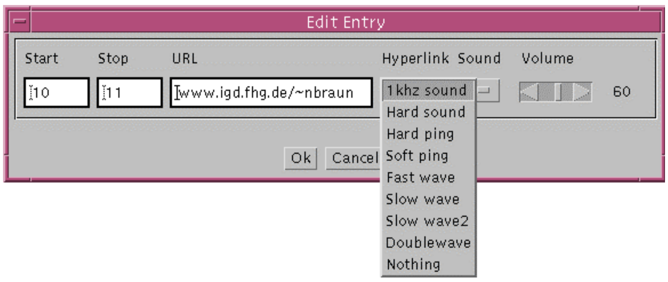

Figure 12: CreationCenter Edit Entry Window

Furthermore, it is shown how the annotated audio stream can be served through the Web to a RealMedia Web video client and how the sonic hyperlinks are accessible via the presenter Java applet.

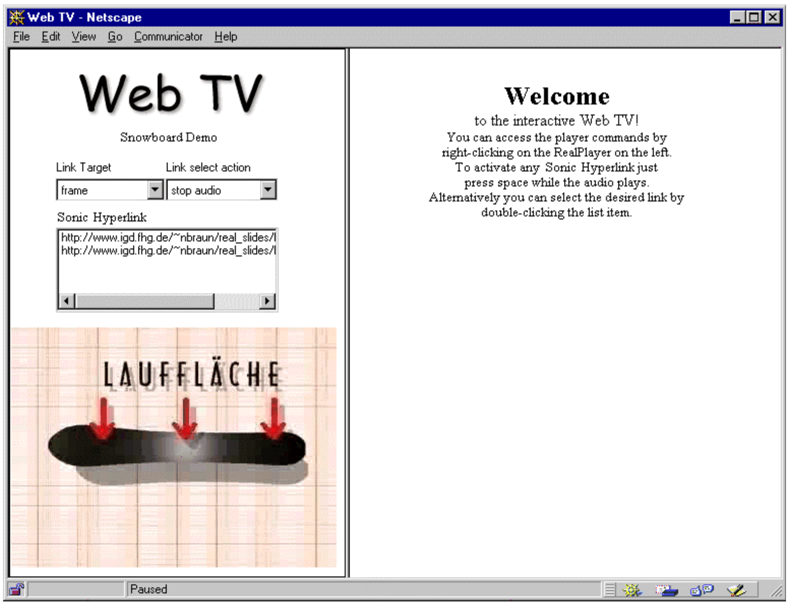

Figure 13: The Presenter Window within the Web-Browser

An example of the presenter window within the browser is shown in Figure 13. The example shows the menu 'Link target' for the selection of the link-target windows, the menu 'Link select action' for the system reaction upon selection of a sonic hyperlink and the menu 'Sonic Hyperlink' that displays the URLs of the hyperlinks after they have been played. The example is a lecture about snowboarding. There are several other examples like annotated news services or Web-radio.

6 Conclusions and Future Work

The transfer of Hypermedia features to audio in an audio-visual environment has been discussed introducing sonic hyperlinks. As an example, application sonic hyperlinks have been integrated in interactive Web-TV. A system architecture and implementation relying on commercial WWW technology like RealMedia is presented. The system includes an authoring tool, as well as the necessary presentation plugin for an Internet browser.

At first glance, the overall problem of annotated audio-visual presentation services like Web-TV seems to be of a technical nature. The software technique must be refined for annotation of audio-visual MPEG 4 streams and description possibilities of sonic annotation.

A second look reveals that the major challenge lies in the field of ergonomics, like suitable sounds for sonic hyperlinks, the types of sonic information that are useful to annotate and application-specific user reaction scenarios.

As future work we will investigate how sonic hyperlinks may be applied to interactive (digital) storytelling. Sonic hyperlinks could contribute a solution to numerous digital storytelling questions, like how to create content with a non-linear story line since traditional production methodic focuses on a linear way of viewing the content of a TV presentation.

7 References

1. Albers, Michael C., and Eric, Bergman, *The Audible Web: Auditory Enhancements for Mosaic*. CHI 95 Conference on Human Factors in Computing Systems Proceedings, 1995.

2. Arons, B., *A Review of the Cocktail Party Effect*. Journal of the American Voice I/O Society. 1992.

3. Asakawa, Chieko, and Takashi, Ito, *User Interface of a Home Page Reader*. Conference on Assistive Technologies ASSETS Proceedings, 1998.

4 Boelke, L., and P. Gorny, *Direkte Manipulation von akustischen Objekten durch blinde Rechnerbenutzer*. D. Böcker (Hg.): Proc. Software-Ergonomie '95. Fachtagung des German Chapter ACM und der GI.Darmstadt Feb. 1995. Stuttgart, 1995: Teubner, 93-106.

5. Brewster, S.A., and C.V.Clarke, *The Design and Evaluation of a Sonically-Enhanced Tool Palette*. ICAD97 Proceedings.

6. Buxton, W. , *Speech, Language & Audition*. Chapter 8 in R.M. Baecker, J. Grudin, W. Buxton and S. Greenberg, S. (Eds.), 1995.

7. Campbell, R.H., Chen, Z., Tan, S.M., and Y. Li, *Real Time Video and Audio in the World Wide Web.* Fourth International World Wide Web Conference Proceedings.

8. Hefta-Gaub, B., *The RealMedia Platform Architectural Overview*, Real Networks Conference, 1997.

9. James, F., *Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext*. ICAD96 Proceedings, Session Audio and the World Wide Web.

10. Martel, R., *A Distributed Network Approach and Evolution to HTML for New Media*. In: Hoschka, P. (ed) Real Time Multimedia and the World Wide Web. 1996, No. 25.

11. Stiefelmann,L.J., *Augmenting Real World Objects: A Paper-Based Audio Notebook*. Short Paper in CHI 96 Proceedings, ACM, 1996.