



## D5.1

# Prototype of Looping and Marking Modules

---

### Abstract

The EASAIER system provides the end user with the ability to query large multimedia archives and access the retrieved content directly. Upon retrieval of content, the user then has the option to interact with this content in many powerful ways. This interaction is provided by the enriched access tools currently being developed in WP5. Deliverable 5.1 describes the development of the looping and marking module, which gives the user the ability place markers at specific points of interest along the timeline of the audio/video content. The content between these markers can then be looped or repeated seamlessly. The user may also avail of automatic marking generated by the segmentation routines developed in WP4

---

Version 1.8  
Date: 2/1/2007  
Editor: DIT  
Contributors: DIT, QMUL, RSAMD

## Table of Contents

<b>TABLE OF FIGURES .....</b>	<b>3</b>
<b>1. EXECUTIVE SUMMARY .....</b>	<b>4</b>
<b>2. USER NEEDS AND REQUIREMENTS .....</b>	<b>5</b>
2.1. USE CASE SCENARIOS.....	5
2.2. LOOPING AND MARKING: EXISTING SOFTWARE.....	7
<b>3. RELATIONSHIP TO THE SYSTEM AND OTHER WORK PACKAGES .....</b>	<b>12</b>
3.1. WP3 – RETRIEVAL SYSTEM.....	13
3.2. WP4 – SOUND OBJECT REPRESENTATION.....	13
3.3. WP5 – ENRICHED ACCESS TOOLS .....	13
3.4. WP6 – INTELLIGENT INTERFACES .....	14
3.5. WP7 – EVALUATION AND BENCHMARKING .....	14
<b>4. LOOPING MANUALLY MARKED AUDIO SEGMENTS .....</b>	<b>15</b>
4.1. USER INTERFACE EXPECTATIONS.....	15
4.2. USER INTERFACE EXAMPLE AND WALK-THROUGH.....	15
4.3. INTERFACE AND IMPLEMENTATION DECISIONS AND NOTES.....	18
<b>5. AUTOMATIC SEGMENT LOOPING.....</b>	<b>19</b>
<b>6. CONCLUSIONS .....</b>	<b>21</b>
<b>REFERENCES.....</b>	<b>21</b>

## Table of Figures

Figure 2-1: <i>Audacity</i> – Selections are indicated by the darkened area of the plot .....	8
Figure 2-2: Loop play in <i>Audacity</i> requires the user to hold down the shift key and click play simultaneously.....	8
Figure 2-3: Two Individual annotation tracks beneath the waveform in <i>Audacity</i> .....	8
Figure 2-4: A section of the <i>Goldwave</i> interface; the second play button (yellow) can be configured to have multiple behaviours, the default being loop playback 10 times.....	9
Figure 2-5: Annotating in <i>Goldwave</i> ; right clicking on the cue bar reveals the annotation menu .....	9
Figure 2-6: Selection can be seen in white and a single button (arrow) induces looped playback .....	10
Figure 2-7: Annotating in <i>Cool Edit</i> ; Right mouse clicking above waveform reveals annotation menus..	10
Figure 2-8: <i>Wavelab</i> Interface; Looping requires clicking loop button followed by play button. ....	11
Figure 2-9: Multimedia manipulation tool used in HOTBED; annotation is not visualised on a timeline.	12
Figure 3-1: EASAIER system overview; Looping and Marking is a client side access tool. ....	13
Figure 3-2: WP 5 Enriched Access Tools; internal structure and input/output mapping. ....	14
Figure 4-1: Prototype looping and marking interface.....	15
Figure 4-2: Viewing a region of interest in a waveform.....	16
Figure 4-3: A selected region is shown in purple .....	16
Figure 4-4: An advanced visualisation, a spectrogram in this case, is displayed aligned with the waveform.....	17
Figure 4-5: Marking time instants which appear in purple.....	17
Figure 4-6: Restoring a previously selected area by clicking between boundaries automatically extends the selection to the left and right marker locations .....	17
Figure 5-1 Mock-up automatic segment looping example. Clicking on any coloured segment would automatically select that span in time.....	20

## 1. Executive Summary

Standard media file players such as Windows Media player or RealAudio player allow basic playback functionality along with standard features such as shuttling forwards and backwards through the content. They typically also allow the user to jump to any point in the media using a slider control. More advanced users such as music students or DJs for example often require more control over playback, usually for the purposes of analysis. To this end, the EASAIER system will provide functionality to allow the user to mark out and annotate particular points of interest within the media content. Furthermore, the user will be able 'loop' the playback between these markers. In fact, playback may be looped between any two arbitrary points within the file. This is of particular use when the user is only interested in a particular section of the content. The fundamental specification for the looping and marking module is that the user must be able to make selections within a given piece of media, and that the selection can be looped continuously and seamlessly.

Looping and marking is not a novel concept but rather a prerequisite for advanced users. The functionality already exists in many specialist audio applications. However, its implementation within EASAIER is complicated by the fact that the looping will have to be synchronised with other media types such as video, musical scores or advanced visualisations for example. To add to this, the user may also be availing of other enriched access tools which effect the content in real-time.

This report outlines the approach taken to develop the prototype looping and marking module for EASAIER. The initial prototype design is based on *Sonic Visualiser* [3], an open source audio analysis tool which has been developed by QMUL. Ongoing research will see the prototype refined as user group testing validates its functionality.

User needs and requirements will be addressed in section 2, which describes the necessity and usefulness of looping and marking using real world examples. A brief survey of looping and marking functionality within existing software is also presented in this section.

Section 3 will describe how the looping and marking module is related to the overall EASAIER system. The section will discuss any dependencies relating to the module including the input output structure associated with it.

Section 4 introduces the looping and marking prototype and describes its functionality by way of an example use case. The design and implementation of the prototype are also discussed in this section.

Section 5 describes how specific work on segmentation routines in WP4 will contribute to automatic marking and the advantages of such functionality.

## 2. User Needs and Requirements

This section discusses the need for a looping and marking module. Some real world scenarios have been provided along with an informal survey of existing software which also contains this functionality. In general, looping and marking will be of interest to those users who will use EASAIER for the purposes of education and analysis, and as such, this section focuses on the needs of this user community.

It has been noted in previous user studies, particularly those of the HOTBED [1] project that the ability to loop and mark material for quick, easy reference and repetition within an audio file, is a vital feature for those learning aurally 'by ear'. Combined with the EASAIER time/pitch scale modification functions, the looping and marking tools could have wider implications for disciplines passed on by oral tradition. For example, many forms of traditional folklore and music are passed on from an older, more experienced 'tradition bearer' to the members of the next generation without recourse to written material. The methods of learning mainly involve repetition of one sort or another. Oral traditions are by their nature defined by this manner of passing on material, indeed 'repeating' material from one generation to the next, and as such exponents of an oral tradition will place great significance on where and from whom they learnt a certain tune, song or poem. The inevitable change that comes with passing material on in this way is accepted as part of the fabric of an oral tradition.

With the advent of recording, materials previously only performed by tradition bearers in often-remote localities were now available to a larger audience. Digital technology has further expanded this audience. The tools discussed here have potential to reinvigorate the 'oral tradition' method of learning whereby the user is interacting with the digital interface in a manner similar to that prevalent in a discipline passed on by 'oral tradition'. Within such learning styles, material is learned by repeating and mimicking phrases and important elements rendered at a slower pace to enable a quicker grasp of details. Learning by listening to recordings is widely practised, if not deeply discussed. However, EASAIER's looping and marking function allows the user to engage with more dynamic learning disciplines, like those of an oral tradition.

### 2.1. Use case scenarios

#### Use-Case Scenario 1 - Music performance and 'aural learners'

Moira is a fiddler in the second year of the Scottish Traditional Music degree at the Royal Scottish Academy for Music and Drama. She learns a large number of her tunes by listening to recordings, and particularly enjoys listening to different fiddlers playing the same tune. Moira believes this not only gives her new ideas, but also demonstrates clearly and quickly the geographical and historical variations in traditional fiddling.

Before she started her degree, Moira wore out several CD players by repeatedly pausing and skipping back through tracks in order to learn a particular phrase or ornament within a tune. Having used the HOTBED [1] system a few times in her first year, she was impressed with the *ability to create 'loops'* in the recording [2]. After her Contemporary Scottish Music Studies lecturer spoke about a new Scottish music archive that employed the EASAIER software, Moira logged on during a practice session at home. She quickly found *the looping function*, and *created a loop of an ornamented phrase* within a James Scott Skinner rendition of 'The Laird Of Drumblair' that she had wanted to learn for her own performance of the same tune. Playing along with this loop, Moira realised how easy this manner of 'aural learning' was, eliminating the need to be continually pressing buttons on the CD player.

## Use-Case Scenario 2 - Theatre and 'aural learners'

As an actor, John's knack for quickly picking up accents served him well professionally. He had recently seen calls for auditions advertised for a feature film, where the actors were required to have old English 'West Country' accents. Usually, when he needed to pick up a new accent, John would call on one of his foreign friends with the appropriate accent, or study a suitable foreign actor playing an English-speaking role. However, none of his friends or colleagues had West Country accents, and John needed to learn this accent within three days.

He called up his old drama school lecturer, who recommended he look for an archive or museum in Devon or Cornwall who might be able to help him. Searching on the internet, John discovered the Bishopsteignton Museum of Rural Life. This museum had recently digitised 23 hours of audio material including local accounts of historical diary farming from the mid 1950's. This museum used the EASAIER software to provide access to this collection *via the Internet*. John listened to a selection of audio material and *created loops of spoken phrases* he thought would be most useful for learning the accent, repeating them back parrot style. After an evening's work John felt he had a firm hold of a Devonshire accent from two or three generations ago, and wrote an email recommending the EASAIER software to his old drama school, who he knew were digitising their collection of audio-visual teaching material.

## Use-Case Scenario 3 -

Marking function: Easy use of online digital material in presentations

Upon completing his Masters of Arts from Indiana University's Jacob School of Music, Christoph secured his first job lecturing undergraduate musicology students at the Sorbonne in Paris. During the course of his postgraduate study, he had access to Indiana University's Digital Music Library. One of the features of this digital access software was the ability to 'mark up' recordings for quick referral. Ryan was due to give a lecture regarding a John Dowland song, but found to his disappointment that he could not access this library outside Indiana University's campus.

Searching online, he found an English song library that had digitised an audio-visual record from the 1970's of well-known lutenist Robert Spencer performing four Dowland songs with an unnamed singer. As this archive used the EASAIER software, Ryan was able to *access and view this material*, and *put markers within the recording* at those places he thought might be useful for his lecture. Having *marked the beginning of the four songs and their verses*, he viewed the material again with a copy of the music to hand, and placed a *second set of markers at points where Spencer had improvised* on the Dowland composition. He also *placed markers where he thought the singer had perhaps missed opportunities for greater expression*.

As the material and his markings were saved online, Ryan only needed to connect his laptop to the Sorbonne's wireless internet network and a digital projector to use this resource in his lecture presentation. Using the EASAIER interface, he only needed to double-click on a specific mark to play the audio-visual example he wished to use.

**Use-Case Scenario 4 - Marking and Looping - Foreign Language Aid**

Luis has recently come back to Spain from London, where he spent the summer learning English. Even though Luis has strong writing skills, he was particularly frustrated by the fact that his manner of pronouncing certain words in the English language had caused him more than one unfortunate misunderstanding. On one occasion, Luis went to a grocery and asked the grocer: "Do you have a good pitch?". What he had actually meant to say was: "Do you have a good peach?". The following day, Luis told the story to his English teacher, who explained to him that the English language is comprised of certain phonemes made by long vowels that do not exist in the Spanish language. Consequently, it is very common to hear a Spanish person pronouncing "peach" as "pitch" or "sheep" as "ship", since the phoneme "i:" that represents the "ea" and "ee" sounds is too long to exist in the Spanish language.

Though these scenarios are hypothetical, they illustrate the needs that have already been documented in [1, 4, 5]. For, instance, In [1], user needs analysis for the music library project HOTBED concluded that, for both audio and video, the ability 'to repeat phrase-by-phrase and to loop phrases', was a desirable feature. They further generalized this as a need for real-time, user-defined mark-up and looping functions.

These use-case scenarios illustrate both the need for looping and marking functionality in a sound archive access system, and the various uses to which this functionality may be put.

Summarising and collating the results of user needs studies and use case scenarios, we have the following established requirements;

- Automatically mark-up recordings
- Manually manipulate and cycle through markers
- Loop key phrases in audio
- Tailor the looping and mark-up for both speech and music
- Implement functionality with an emphasis on simplicity and ease-of-use

In the following section, we will describe how various existing software packages address these requirements.

## ***2.2. Looping and marking: Existing software***

As mentioned previously, manual marking and looping is not a novel concept and does indeed exist in several specialist software applications. Previous user needs studies [1] have indicated that such functionality would be a necessity for certain user groups targeted by EASAIER. This section aims to discover how EASAIER can adapt and extend upon the most effective looping and marking tools available in existing software. What follows is an informal survey of several popular audio/video production and analysis software tools. Each application is tested for the presence and ease of use of 3 functionalities; Marking, Looping and Annotation.

### ***Audacity***

*Audacity* [6] is an open source standalone audio editor which gives users the ability to load audio files and process and interact with them in various ways. The audio can be visualized in time and frequency

domains allowing users to visually inspect waveforms. A selection tool allows the user to make selections by left clicking and holding down on the mouse at the intended start point; the user then drags to the intended end point of the selection and releases the mouse button. The selection boundaries can then be adjusted by dragging each boundary individually. This method of selection and minor variants of it account for almost every software application surveyed here. Figure 2.1 shows how a typical selection is displayed. The shaded area of the waveform plot is effectively the selected range. Audio playback will now be localised to this selection area.



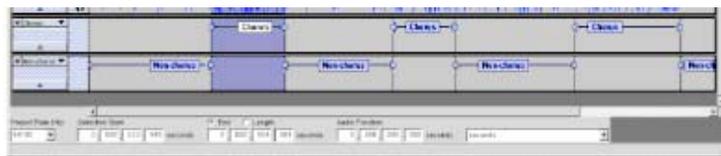
**Figure 2-1: Audacity – Selections are indicated by the darkened area of the plot**

In audacity there is no specific button for looping playback nor is it obvious how to do it initially but it is possible. Looped playback is achieved by holding the shift key and pressing the play button with the mouse. This is only indicated to the user by means of a small floating tool tip which only appears when the mouse pointer is positioned over the play button for more than 1 second as illustrated in Figure 2.2. Looping requires 4 mouse/keyboard operations to achieve looping, where 2 operations are required in order to select the audio segment.



**Figure 2-2: Loop play in Audacity requires the user to hold down the shift key and click play simultaneously**

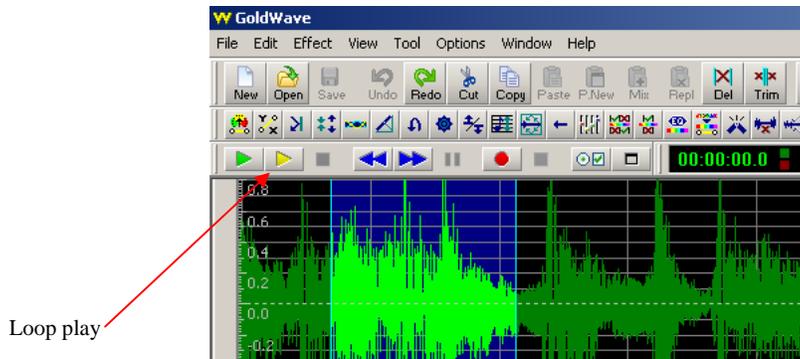
*Audacity* also allows the user to add an annotation track in which text can be placed along the timeline beneath the waveform on an ‘annotation track’. In order to place an annotation at a particular point the user must use the selection tool and click on a point in the annotation track, but no marker is placed at this point. The user must begin to type before there is any indication that a marker will be placed at this point. These markers can not be repositioned or deleted after they have been positioned. The annotation process is particularly tedious and inflexible in the standard version of *Audacity* but a useful annotation extension is provided in [7]. The authors provide a more intuitive way to place and visualise annotations in *Audacity*. Instead of visualising multiple annotations of different types on one annotation track, each type or class of annotation can be visualised separately. In figure 2.3, two annotation tracks have been created, one contains the boundary locations of choruses and the other contains the boundary locations of the verses.



**Figure 2-3: Two Individual annotation tracks beneath the waveform in Audacity**

## Goldwave

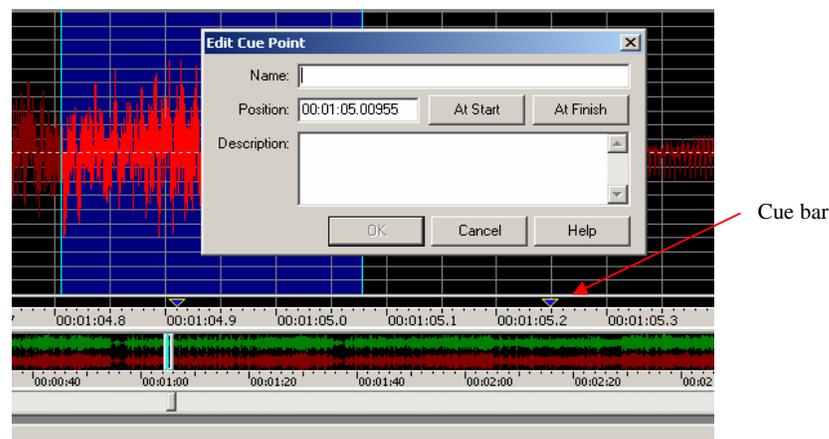
*Goldwave* [8] is a well known audio editor which has been available for over 10 years now. It has much the same general functionality as other editors. The method of manually selecting a segment in *Goldwave* is almost identical to that of *Audacity*, however, the task here is slightly simplified by the fact there is no requirement to select a particular tool before the user makes a selection. In *Audacity*, figure 2.2, you will notice that you must first select 1 of 6 pointer tool buttons before a selection can be made. *Goldwave* defaults to just one pointer tool which is used specifically for making selections.



**Figure 2-4:** A section of the *Goldwave* interface; the second play button (yellow) can be configured to have multiple behaviours, the default being loop playback 10 times.

Looping is then achieved by clicking on the yellow playback button which can be configured to have numerous behaviours including looping which is its default setting. Only 2 mouse operations are required to achieve looping in *Goldwave* but the looping is not infinite. The default behaviour is 10 loops which is hardly sufficient for applications such as learning a complicated passage of music. The number of loops can be changed, but only at the cost of going through several settings menus.

Annotation is also possible in *Goldwave* by right mouse clicking on the cue bar just below the wave form. A floating menu offers the opportunity to enter a name and details about the cue point (annotation point). Figure 2.4 shows 2 existing cue points and a new point being created. The details of these cue points, i.e. the annotation information only becomes visible when the mouse is positioned over inverted blue triangle representing the cue point. This makes finding a particular annotated point in the media quite difficult although the cue information can also be viewed as list also. These cue points can be repositioned in time or deleted quite easily.



**Figure 2-5:** Annotating in *Goldwave*; right clicking on the cue bar reveals the annotation menu

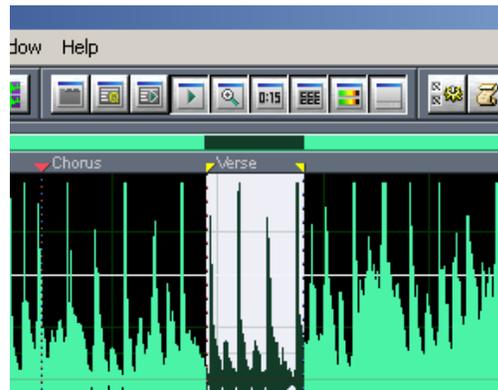
One draw back within *Goldwave*'s annotation functionality is that the cursor will not snap to annotation points. This feature would simplify interaction required by the user.

### ***Syntrillium Cool Edit / Adobe Audition***

*Cool Edit*, now known as *Adobe Audition* [9] is a popular 2 track and multi-track audio editor with newer versions also supporting video synchronization. *Cool Edit* also provides functionality for marking, looping and annotating. The methods for selection and looping are very similar to what has been described above but the interface in *Cool Edit* is much more intuitive. There is a single button for play and a single button for loop play (figure 2.5) which will invoke infinite looping until the user presses stop. The selection boundaries can also be modified as the audio is actively looping, which is particularly useful for fine tuning the boundaries in real-time.



**Figure 2-6: Selection can be seen in white and a single button (arrow) induces looped playback**

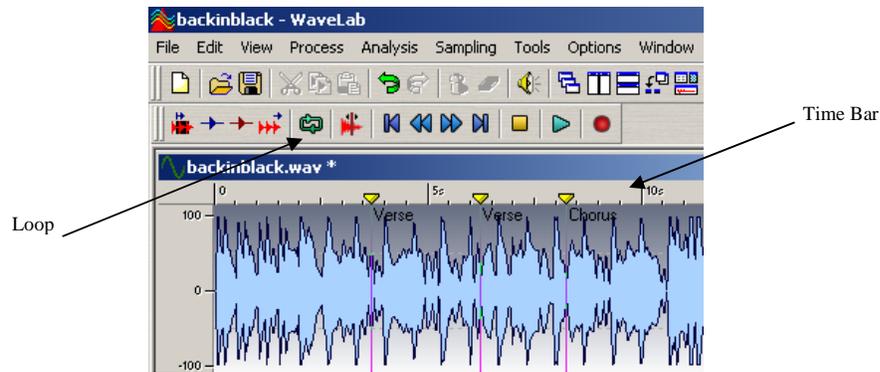


**Figure 2-7: Annotating in Cool Edit; Right mouse clicking above waveform reveals annotation menus.**

Annotating in *Cool edit* is slightly tedious and requires several mouse operations to complete a single annotation. To its advantage, however, the annotations are displayed along the cue bar above the waveform as in figure 2.6. The manual selection procedure can be configured to snap to cue points which is a useful feature. The marker points can also be repositioned and edited easily but the parameters of the text display such as font size are not editable which may be an issue for the sight impaired.

### ***Steinberg Wavelab***

Steinberg's *Wavelab* [10] is a specialist audio editing and restoration application. It includes functionality for marking, looping and annotation. Selecting a section of audio is identical to the methods described above except that the default selection behaviour is not standard and must be configured manually. Looping a selection is achieved by first selecting a button indicating that playback will be looped and then clicking on the play button.



**Figure 2-8: Wavelab Interface; Looping requires clicking loop button followed by play button.**

Placing markers and annotating is particularly easy and intuitive in *Wavelab*. A right click on the time bar opens a floating menu which allows you to insert a marker and annotate it all in one operation. Clicking anywhere between these marker will extend the selection from the marker left of the cursor to the marker right of the cursor. This is a particularly fast and easy way of making meaningful selections such as a verse or a chorus.

### ***Quick Time / HOTBED***

QuickTime is a multimedia editor developed by Apple [11], which is capable of playing various formats including video and audio recordings. The free version of this software is limited to the most basic playback operations. If the user wishes to perform more sophisticated operations such as looping a selected audio segment, the purchase of the QuickTime Pro version is required.

The HOTBED project [1] provides web-based access to sound and video materials. The system utilises QuickTime to play the retrieved digital recordings. In order to overcome the limitations of the free QuickTime version in looping and marking, the HOTBED system accesses Quick Time controls using Javascript [12]. The multimedia manipulation tool provides a transport bar and three buttons: “mark,” “loop” and “edit/save”. In order to create a marker, the user clicks the “mark” button. This will set a marker point in the piece at the current location, and a new button labelled with the marker number will automatically be displayed. Figure 2.8 illustrates an example of marking with the HOTBED system, where three markers have been generated. The user can navigate through the piece by clicking on the three marker buttons. In order to create a loop between two existing markers, the “loop” button must be first clicked. Then, the start and end of the loop is selected by clicking on two different markers respectively.

It should be noted that the audio waveform is not displayed. Thus, due to the small dimensions of the transport bar, the “mark” buttons do not visually provide the information of the time instant they represent. This information is provided by a new window that pops up when clicking on the “edit/save” button. The facility for annotating the marker names is also included in this window. In order to read the annotation, the user places the mouse pointer over the “marker” button. The noticeable difference with Hotbed, is that no visualisation of the audio is present and so the markers are not visible on a timeline. Timeline visualisations of markers are advantageous in that the user can see the distribution of marker points at a glance. This also gives the user a sense of distance between segments.



Figure 2-9: Multimedia manipulation tool used in HOTBED; annotation is not visualised on a timeline.

### ***Other – Windows Media Player / RealAudio Player***

Standard audio players such as the Windows Media Player [13] and RealAudio [14] player lack the functionality of looping and marking. This is understandable given the target audience and usage of these media players. Although these players lack marking functionality, some users have seen the need for it. This has resulted in creation of an independent software media player which uses the Windows Media Player SDK but incorporates the needed functionality [15]. This application is an alternative to the standard media players which allows the user to mark out sections of media for the player to ‘skip’ over. The author gives the example of marking out the commercials in a recording of a webcast in order to skip over them during playback.

### ***Comparison between Audio players***

Table 2.1 shows some basic findings about each of the applications that were surveyed. Each application was informally tested for the presence and ease of use of 3 functionalities, Marking, Looping and Annotating. The level of difficulty to achieve each of the functionalities was ranked as being simple, moderate or difficult. The number of mouse clicks and keystrokes to achieve each functionality is also provided in square brackets for each heading. The general scheme for achieving these functionalities is very similar in all cases but some implementations seem to be more efficient and intuitive to the general user. Table 2.1 also includes the results of the same basic tests on *Sonic Visualiser*, which as mentioned, is the template for our current looping and marking prototype. Although, in certain instances it requires more user operations to achieve a given functionality, it is significantly more flexible and detailed in this functionality. Section 4 will provide a detailed look at this functionality.

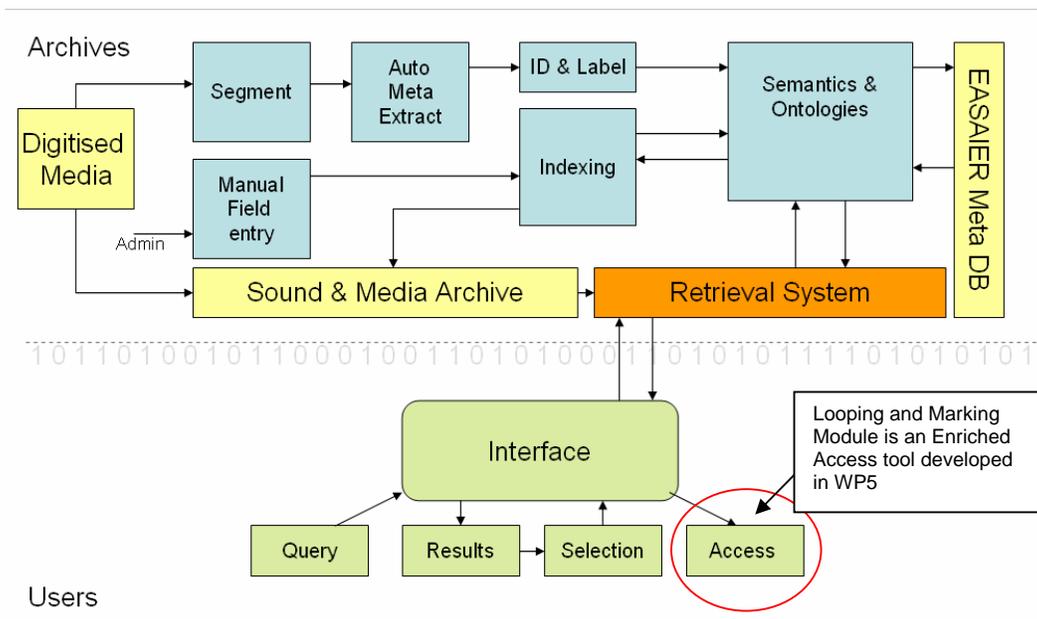
Software Package	Marking / Ease of use		Looping / Ease of use		Annotating / Ease of use	
<i>Audacity</i>	Simple	[2]	Moderate	[2]	Difficult	[3]
<i>Goldwave</i>	Simple	[1]	Simple	[1]	Moderate	[3]
<i>Cool Edit / Audition</i>	Simple	[1]	Simple	[1]	Difficult	[6]
<i>Wavelab</i>	Moderate	[2]	Moderate	[2]	Simple	[2]
<i>Quick Time/ Hotbed</i>	Simple	[1]	Moderate	[3]	Simple	[3]
<i>Windows Media</i>	No facility*		No facility		No facility	
<i>Real Audio</i>	No facility		No facility		No facility	
<i>Sonic Visualiser / EASAIER Prototype</i>	Simple	[2]	Moderate	[3]	Moderate	[3]

\*The user modified software includes a marking feature which requires 3 or more clicks to operate.

Table 2-1: The functionality, Ease of Use and number of clicks required to invoke functionality

## **3. Relationship to the system and other Work Packages**

This section aims to display the dependencies and relationship this deliverable has with other elements of the EASAIER system. The looping and marking module is an enriched access tool which resides exclusively in the client side application. It is first revealed to the user after he/she has selected media from a list of results returned by the retrieval system in WP3. The looping and marking module is the first opportunity the user has to interact with the retrieved media. Figure 3.1 illustrates where looping marking falls within the overall system.



**Figure 3-1: EASAIER system overview; Looping and Marking is a client side access tool.**

### 3.1. WP3 – Retrieval System

In order to have any media content to interact with, the user must provide a query for the retrieval system to resolve. The user is then presented with a text based ranked list of possible media which fit the query criteria. Upon choosing a result from the returned list, the advanced user is immediately transported to the enriched access tools interface which allows them to begin interacting with the media. This interface will allow the user to make selections and loop them immediately. Section 4 describes this in detail. Essentially, the selection process within the retrieval system is a prerequisite to looping and marking.

### 3.2. WP4 – Sound Object Representation

In the general case, many users will simply want basic marking, annotation and looping functionalities as described in section 2 but WP4 will provide a wealth of metadata which can also be used to help the user make useful selections automatically. The looping and marking module will avail of this metadata generated in WP4 to provide the user with several automatic mark-up possibilities, e.g. place markers at all points in the audio where a given speaker is talking. Section 5 of this document describes these automatic segmentation possibilities in detail.

### 3.3. WP5 – Enriched Access Tools

The looping and marking tool itself belongs to the suite of enriched access tools but it is particularly important in that it provides the first link between the retrieval engine and all other enriched access tools. All other processing within WP5 is dependent on the looping and marking module in that only the marked selection of media will be passed on to other processing tools such as the time scale modification tool or the noise reduction tool. Figure 3.2 illustrates how the processing chain in WP5 is structured. Firstly media content is delivered directly to the looping and marking module along with metadata which has been generated in WP4. The user can make selections manually or avail of automatic segmentation and play or loop these selections unaffected or with optional real-time processing. The user then has the option to process the selected audio in 4 ways;

- Sound source separation
- Time and Pitch Scale modification
- Equalisation

- Noise Reduction

Each of these processes is carried out in real-time in a non destructive manner meaning that the original file is never altered at source or even locally. The processes are only apparent on playback so they can be manipulated or toggled on and off as the user sees fit. In figure 3.2 it can be seen that the audio timescale modification is linked to and in fact controls the play back rate of the video if present.

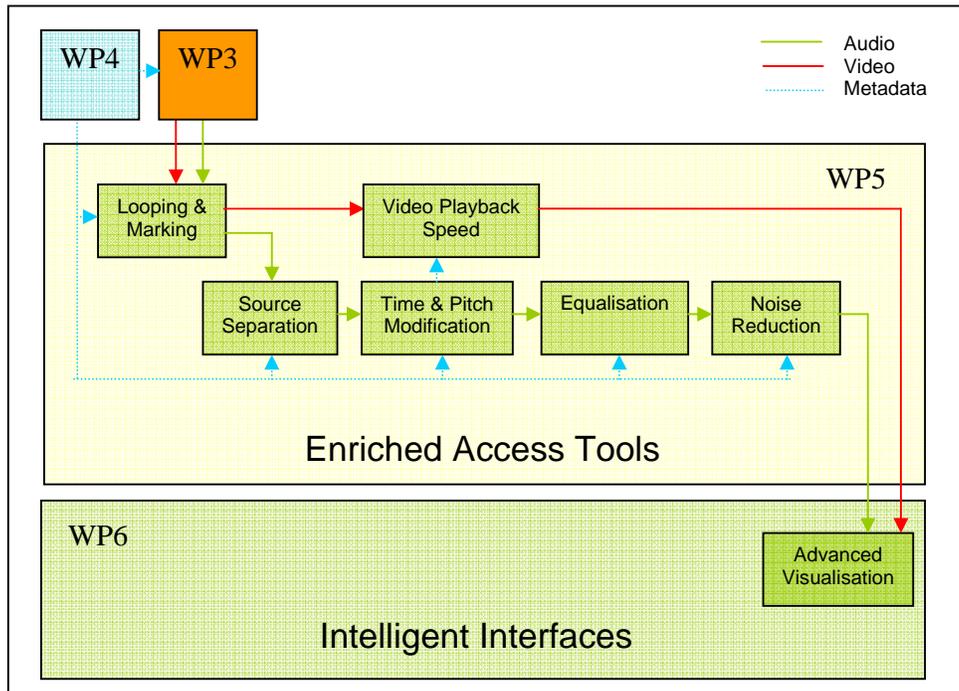


Figure 3-2: WP 5 Enriched Access Tools; internal structure and input/output mapping.

### 3.4. WP6 – Intelligent Interfaces

Almost all of the functionality of EASAIER is exposed to the end user through the client application interface. WP6 is responsible for displaying and integrating all the functionality of the enriched access tools. The prototype described in section 4 of this document adheres to the API described in the technical architecture document. WP6 will also provide the necessary synchronisation and alignment between the media and other elements such as a musical score or automatically transcribed text. The interface will be configurable for a simple or advanced user.

### 3.5. WP7 – Evaluation and Benchmarking

The functionality and usability of the looping and marking module as with all other modules will be validated in WP7. An iterative evaluation process will see the initial prototype presented in this document, developed into the final looping and marking module which will be integrated with the final EASAIER system.

## 4. Looping manually marked audio segments

### 4.1. User interface expectations

After content has been retrieved and accessed using the EASAIER retrieval system, the user will have the option to interact with the content in one of two ways: via a simple media player, or using the EASAIER enriched access tools. Looping and marking belongs to the suite of enriched access tools.

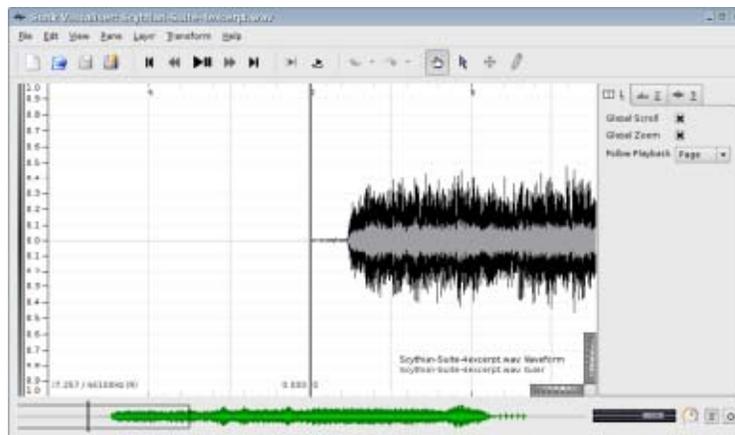
This section introduces the expected functionality and some implementation notes for the interactive aspects of the looping and marking module. This functionality is currently being prototyped through a standalone application, allowing us to test it independently of the EASAIER system. The application supports looping and marking on audio files stored locally or retrieved in their entirety from a remote server; it does not support streamed content. The design and implementation is in line with the initial API specification presented in the technical architecture document[16].

Expected capabilities for this module include:

1. The ability to view the audio file as a waveform that can be zoomed and panned
2. A simple mechanism for selecting a region of audio and playing it once or in a loop
3. The ability to adjust the extents of a looped region after selecting it
4. The ability to make a note of the extents of a loop and recall it for later study
5. Seamless looping (i.e. avoiding painful clicks at loop fly back points)
6. The ability to change the playback speed while retaining the correct loop points (and retaining synchronisation with any associated video or other material as appropriate).

### 4.2. User interface example and walk-through

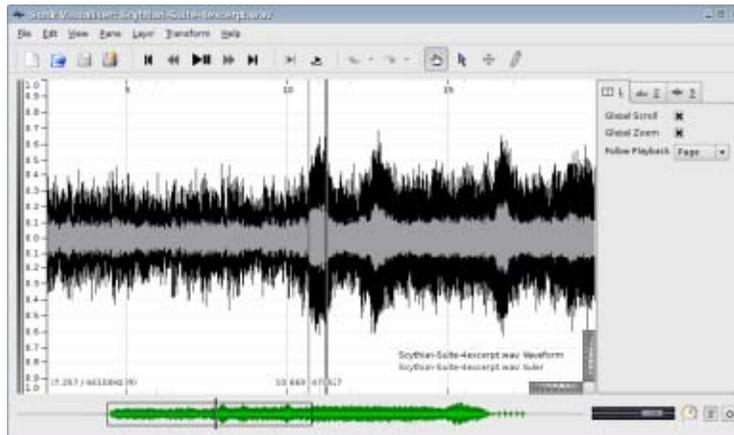
The following example shows some of these principles in action in the prototype application. The music shown in this example is an excerpt from the fourth movement of Prokofiev's Scythian Suite (Désiré Defauw / Chicago Symphony Orchestra 1945, transferred from RCA Victor 78 by Bill Anderson). This is a single-channel (mono) audio file; the prototype and the final EASAIER system (will) also support two channel (stereo) and optionally multi-channel files.



**Figure 4-1: Prototype looping and marking interface**

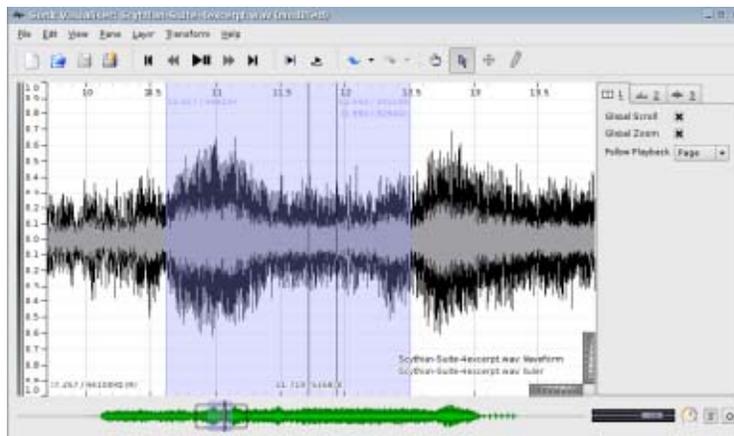
Figure 4.1 shows the prototype viewer with the audio file loaded and positioned at the start of the piece. Using the Play control at this point will cause the piece to play from start to finish and then stop, as no loop is marked.

In this file there are three “humps” visible on the waveform at around 11, 12.5 and 16.5 seconds that correspond to repetitions of a short wind motif. Figure 4.1 illustrates this:



**Figure 4-2: Viewing a region of interest in a waveform**

In this prototype, the user can select an area by clicking the Select tool on the toolbar, and then click-dragging across the waveform. See Figure 4.3, which also shows that the waveform can be zoomed in as a close-up.



**Figure 4-3: A selected region is shown in purple**

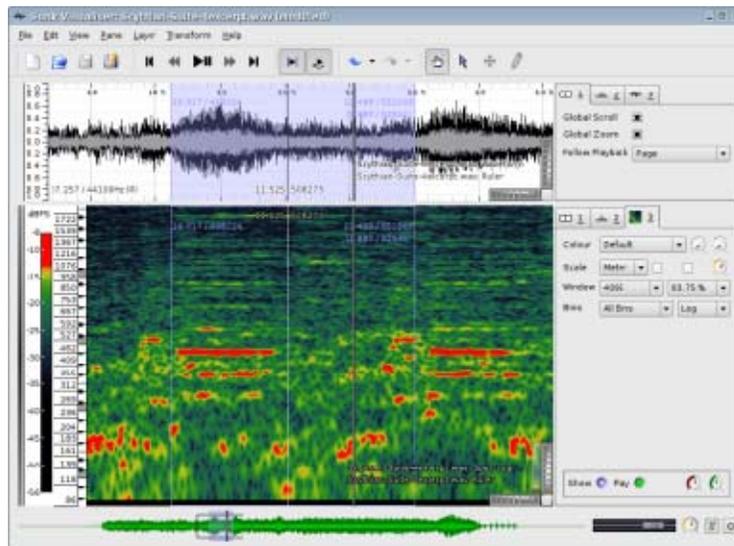
The interface also provides us with useful information such as the duration of the selection. In this example the selection covers a range from 10.617 seconds to 12.498 seconds, a total duration of 1.88 seconds.

The user can now play the selected region once by clicking the Play Selection toolbar button to activate “play selection” mode, then pressing play; or can play the selected region in a continuous loop by also clicking the Play Loop toolbar button to activate “loop” mode.

The extents of the selected region can be adjusted after the initial selection has been made, by dragging the left or right edge of the purple selection region.

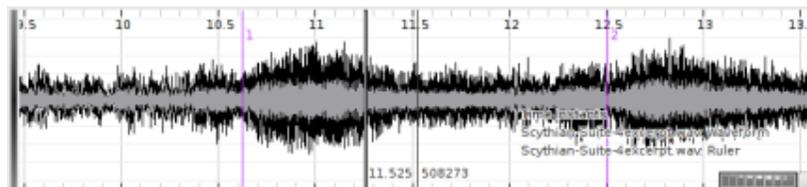
Other tools developed within WP5 such as time scale modification can be applied during playback to allow the content to be slowed down, or potentially transposed in pitch, separated into component instruments and so on.

If any further advanced visualisations are available, the viewer will display the same selected region across those. Figure 4.4 shows a spectrogram view aligned with the waveform.



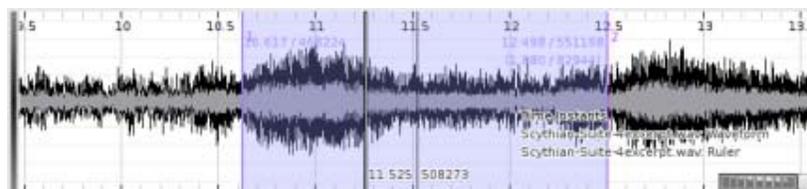
**Figure 4-4: An advanced visualisation, a spectrogram in this case, is displayed aligned with the waveform.**

In the prototype application, a selection only remains shown for as long as the user does not alter the selection boundaries. When the user selects a new region, any previous selection is forgotten. The looping and marking module also requires the ability to ‘fix’ locators and annotate time instants, so as to can recall a precise selection on demand. To this end, the prototype contains a menu function to place marks at the loop boundaries.



**Figure 4-5: Marking time instants which appear in purple.**

In Figure 4.5, the user has placed marks at the loop boundaries and cleared the selection. These numbered marks can now be stored and recalled later by the user. At any time later in the session, the user can re-select the same region as they has initially selected simply by clicking anywhere between the 2 purple time instants using the Select tool , as illustrated in figure 4.6.



**Figure 4-6: Restoring a previously selected area by clicking between boundaries automatically extends the selection to the left and right marker locations**

In the same way, selections will automatically “snap” to the positions of any other markers that have been added to the audio. This would make it easy, for example, to select a region that had been automatically identified by a segmentation algorithm as defined in WP4.

### ***4.3.Interface and implementation decisions and notes***

#### ***Mechanism for playing and looping a selection***

Several alternative arrangements for playback controls have been considered:

1. Offer an action button for Play, and toggle buttons for Play Selection and Play Loop. To loop a selection, first “switch on” Play Selection and Play Loop, and then press Play. This is the mechanism currently used in the prototype, and is similar to the mechanism used in *Wavelab*. It is an appropriate interface if the user is likely to want to play the whole piece even while a selection is active, and to switch between playing the selection and the whole piece without stopping playback.
2. Offer Play and Play Selection as separate action buttons, with Play Loop remaining a toggle. To loop a selection, first “switch on” Play Loop, and then press Play Selection. This would be appropriate if the user is likely to want to play the whole piece even while a selection is active, but is not likely to be interested in switching between playing the selection and the whole piece without stopping and starting.
3. Offer only one Play button, plus the Play Loop toggle. The Play button plays the selection if one has been made, or the whole piece otherwise. To loop a selection, “switch on” Play Loop, and then press Play. This would be appropriate if the user is unlikely to want to play the whole piece while there is a selection active.
4. Offer Play and Play Loop as separate action buttons. The Play button plays the whole piece; the Play Loop button loops the selection if one has been made, or the whole piece otherwise. This is similar to the mechanisms used in *Goldwave* and Adobe *Audition*. It would be appropriate if the user is unlikely to want to loop the whole piece.
5. Offer only one Play button (and no Play Loop). The Play button plays the selection if one has been made, or the whole piece otherwise. Playback loops if a selection is being played, and does not loop otherwise. To loop a selection, just press Play. This would be appropriate if the default behaviour satisfies most users and the simplest possible interface is sought.

Further user testing will be sought in order to identify the best option for the EASAIER system.

#### ***Click-free looping***

When a region is selected and played in a loop, playback necessarily goes straight from the final audio sample in the loop back to the first. If the loop ends with a sample that is significantly different from the one it started with, there will be a sudden voltage jump in the audio output, and this will be heard as a (potentially loud) click.

There are two standard ways to avoid this:

1. Place the loop boundaries at zero-crossing points, that is, points where the sampled voltage crosses from negative to positive or vice versa. This increases the likelihood that the difference between the last and first samples is small and so reduces the likelihood of an audible click. Ideally both boundaries should be at crossings in the same direction (positive to negative or negative to positive). This has the advantage of simplicity in most situations where only a single channel of audio is present. It is not possible for multi-channel audio, nor in some pathological cases for a single channel (e.g. if there is a significant DC offset in a quiet passage).
2. Cross-fade the ends of the loop; that is, make the loop ends overlap by a small number of audio

samples (ten or so is enough), and “fade out” the end of the loop during that period while “fading in” the start of the next loop. This method is less straightforward, particularly in real-world situations where real-time playback processing happens in fixed blocks of audio data, but it is more broadly applicable.

The EASAIER system will use method 2 and cross-fade at loop boundaries.

### ***Retaining time synchronisation of loop points***

It is important to ensure that loop boundaries remain the same regardless of any time-sensitive transformations (such as time-stretching) being applied to the music. Loop points should be identified internally in terms of the audio sample frame of the source audio material.

### ***Audio playback of markers***

Aside from looping for review and study of a section of music, it is also desirable to be able to mark moments in a piece for annotation and recall – for example to record where a particular section starts, or where the beats lie in a section with variable tempo. For this purpose a useful feature is the ability to play markers with a recorded sound such as a click. The EASAIER prototype application has this feature and it will probably be desirable to retain it in the eventual enhanced access application. The feature can be extremely simple, no more than mixing in a short wavetable of a tap sound at each marker point.

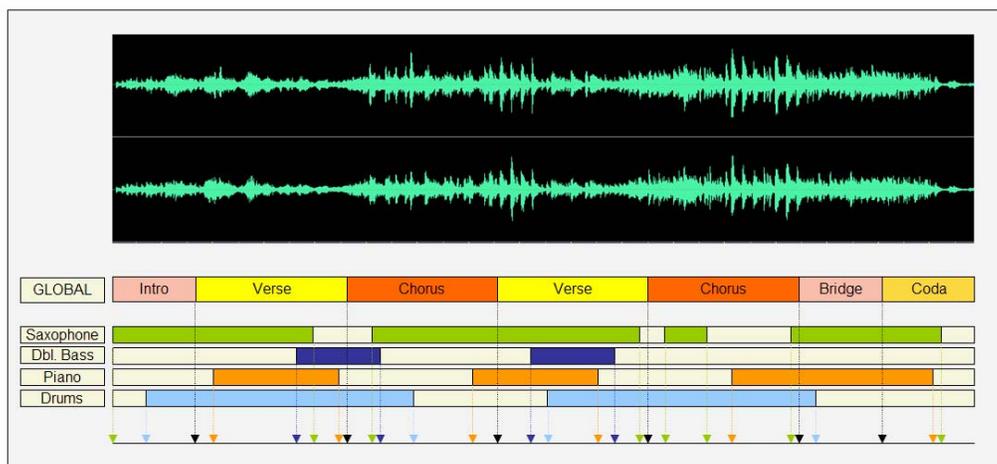
## **5. Automatic segment looping**

In Section 4, the functionalities of looping techniques between manually marked time instants were introduced. Section 5 covers the possibilities that automatic segmentation could offer, which would allow automatic access to specific relevant sections of a recording.

The EASAIER system allows the user to search and retrieve recordings in various formats. This includes audio recordings, which are comprised of speech, music and also audio from video recordings. WP4 will provide a segmentation routine which will first segment the audio signal into sections of music, speech and other. The information contained in these recordings is structured at multiple levels, where each level provides different information for automatic marking purposes:

1. **Music segmentation:** music has a well defined form; its structure can be decomposed into decreasingly smaller units, with each decomposition revealing finer levels of detail about the musical structure. The EASAIER system allows automatic segmentation at different hierarchical levels of the music structure, which are as follows:
  - **Segmentation based on onset times:** At the lowest level of the hierarchy are the individual musical note onsets, which provide information about when a new note starts. These onset marks will be automatically estimated by the onset detection algorithms that WP4 will provide [17]. As an example, a user fascinated by the ornamentation techniques of traditional Irish music, could navigate through the time onsets in order to investigate how a “cut” ornament momentarily raises the pitch of the onset part of the note.
  - **Segmentation based on beats:** WP4 will provide routines for estimating the tempo of a song [18], which will automatically produce marks at specific points in the song where each of the beats in the musical bar occurs. As an example, beat marking can assist a music student to analyse the metric structure of a song.
  - **Segmentation based on bars:** the estimation of the time signature of a song will be provided by WP4 [19]. This musical property combined with the previous estimation of the tempo provides the information required to calculate the time duration of the musical bar of the song. Thus, automatic marks will be generated at the beginning of each bar. By using this segmentation feature, a DJ can select the musical bars that constitute an exciting musical phrase that he/she wishes to include in a loop.

- **Segmentation based on music structural analysis:** A graphical example of the use of automatic segmentation to mark up the sections of an audio file is shown in Figure 5.1. In the top part of the figure, the stereo waveform of a popular song is depicted in green. Just below the waveform, the timeline of the global structure of the song, depicted as “GLOBAL”, is provided by the segmentation routines that WP4 will provide. Each different section of the song is depicted in a different colour, and is labelled according to the name of the section that it represents, e.g.: chorus or verse. These structural segmentation points will be created automatically by algorithms in WP4 but the annotation will most likely be manual. Following this, the user could manually navigate through the different automated marks by using standard step forward and backward buttons in the transport bar. In order to select a specific section of the song, the user can click anywhere in the part of the GLOBAL timeline labelled with the name of that given section. The selection will snap to the marker points in the song time line that corresponds to the beginning and end of the section respectively.



**Figure 5-1 Mock-up automatic segment looping example. Clicking on any coloured segment would automatically select that span in time.**

- **Segmentation based on instrument:** WP4 will also provide music separation routines [20], which allow the separation of each of the individual sources that the song is composed of. It should then be possible to show when each of these sources is active or not. In the example depicted in figure 5.1, the instruments are also labelled. Here, a saxophone, double bass, piano and drums are present, each depicted on a separate time line. Each instrument time line shows the activity (active/silence) of a different instrument at any time instant in the polyphony. Instrument identification algorithms provided by WP4 should facilitate this automatic instrument labelling procedure. In addition, by clicking at any point of an active segment of an instrument time line, the selection will also automatically snap to the first and last active points of that segment. Then, each selection can be played back or looped indefinitely.
  - **Segmentation based on music similarity queries:** the music retrieval system that WP3 will provide allows retrieval of audio recordings that contain specific melodic shapes, which can be inserted by the user as a query. In order to facilitate the user navigation experience within the retrieved music recordings, automatic markers will be placed at the points where the melodic shape from the query starts and ends along the time line.
2. **Speech segmentation:** speech has also a very structured form, where sentences are constructed by using words, and the latter can be subdivided by phonemes. The different levels of speech segmentation that the EASAIER system performs are as follows:

- **Segmentation based on words:** Within EASAIER, the user can query the system for a given word. Automatic segmentation points can be placed at the instants where the queried word appears allowing the user to navigate through the speech recording in order to find any instance of the word.
3. **Video segmentation:**
- **Video Segmentation based on audio content:** The cross-media nature of the EASAIER system allows query for video recordings and also the retrieval of video recordings related to speech and music recordings. The retrieved video recording will be segmented according to its audio content. Then, both audio and video will loop and play synchronously. As an example, HOTBED [1] mentioned a large visual element in the oral tradition of passing music between different players. As an example, a music student learning to play the fiddle, can use the EASAIER system to improve his/her skills by closely observing the finger technique of another fiddle player.

## 6. Conclusions

The prototype looping and marking module has been presented. It uses a combination of the most standard marking, looping and annotation techniques currently in existence. An informal survey was carried out on existing software packages to determine the most suitable functionality to adapt in the EASAIER prototype. Extensions to the state of the art include the ability to synchronise multiple visualisations of the media even whilst looping. The prototype's ability to avail of automatic mark up supplied as predefined metadata generated in WP4 is also advantageous. The user immediately has multiple options for navigating through the content according to many possible automatic segmentations. Added to this, the markers can, if desired by the user, be indicated during playback as audible clicks to indicate the transition from one segment to another. This prototype will form the foundation on top of which all other enriched access tools are predicated. User evaluation will validate the functionality and usefulness of the looping and marking prototype. The results of the evaluation phase and test user feedback will indicate whether any further modifications to the prototype are required.

## References

- [1] S. Barrett, C. Duffy, and K. Marshalsay, "HOTBED (Handing On Tradition By Electronic Dissemination),", Royal Scottish Academy of Music and Drama, Glasgow 2004, [www.hotbed.ac.uk](http://www.hotbed.ac.uk).
- [2] P. Cooke, "Feasibility Study," RSAMD Scottish Music Performance Resource Centre 2003.
- [3] C. Cannam, C. Landone, M. Sandler, and J. P. Bello, "The Sonic Visualiser: A Visualisation Platform for Semantic Descriptors from Musical Signals," presented at ISMIR, 2006.
- [4] R. Tucker, "Harmonised Access & Retrieval for Music-Oriented Networked Information (HARMONICA)," 1997 - 2000. <http://projects.fnb.nl/harmonica/>
- [5] E. Fønss-Jørgensen, "Applying Telematics Technology to Improve Public Access to Audio Archives (JUKEBOX)," Århus, Denmark, 1997. [www.statsbiblioteket.dk/Jukebox/finalrep.html](http://www.statsbiblioteket.dk/Jukebox/finalrep.html)
- [6] Audacity, *D. Mazzoni, Inc*, <http://audacity.sourceforge.net/>
- [7] B. Li, J. Ashley Burgoyne, and I. Fujinaga, "Extending Audacity for Audio Annotation," presented at 7th International Conference on Music Information Retrieval, Victoria, Canada 2006.

- [8] GoldWave, GoldWave *Inc*, <http://www.goldwave.com>
- [9] Adobe Audition Syntrillium Software Corporation *Inc*,  
<http://www.adobe.com/special/products/audition/syntrillium.html>
- [10] WaveLab Studio, Steinberg Media Technologies *Inc*, <http://www.steinberg.net/>
- [11] Quick time, Apple *Inc*, <http://www.apple.com/quicktime/win.html>
- [12] S. Barrett, "HOTBED Technical Development," Royal Scottish Academy of Music and Drama, Glasgow 2004, [www.hotbed.ac.uk](http://www.hotbed.ac.uk).
- [13] Windows Media Player, Microsoft *Inc*, <http://www.microsoft.com/windows/windowsmedia>
- [14] Real Audio, RealNetworks *Inc*, <http://www.real.com/player>
- [15] A. Kulp, "Non Destructive Media Edits," 2005.  
<http://msdn.microsoft.com/coding4fun/windows/article.aspx?articleid=1067234&title=Non-Destructive+Media+Edits>
- [16] L. Barthélémy, L. Bajard, D. Barry, C. Cannam, and C. Landone, "EASAIER Technical Architecture Report," Dublin Institute of Technology, Dublin September 7th 2006, <http://www.elec.qmul.ac.uk/easaier/files/technical/architecture/architecture.htm>.
- [17] C. Duxbury, J. P. Bello, M. Davies, and M. Sandler, "Complex Domain Onset Detection For Musical Signals," presented at 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK, 2003.
- [18] M. E. P. Davies and M. D. Plumbley, "Causal Tempo Tracking of Audio," presented at Proceedings of the International Conference on Music Information Retrieval. Audiovisual Institute, Universitat Pompeu Fabra, Barcelona, Spain, 2004.
- [19] M. Gainza and E. Coyle, "Time Signature Detection by Using a Multi-Resolution Audio Similarity Matrix," presented at Audio Engineering Society 122nd Convention, Vienna, 2007.
- [20] D. Barry, E. Coyle, and B. Lawlor, "Sound Source Separation: Azimuth Discrimination and Resynthesis," presented at 7th International Conference on Digital Audio Effects, DAFX 04, Naples, Italy, 2004.