ENHANCING LEARNING USING SYNCHRONISED MULTIMEDIA ANNOTATION

Mike Wald, Gary Wills, Dave Millard, Lester Gilbert, Shakeel Khoja, Jiri Kajaba, Yunjia Li, Priyanka Singh
School of Electronics and Computer Science, University of Southampton, Highfield, Southampton, SO171BJ, United Kingdom, M.Wald@soton.ac.uk

Keywords
Synchronized annotation, multimedia

1. EXECUTIVE SUMMARY

Audio and Video has become easier to record and store but while users can easily bookmark, search, link to, or tag the WHOLE of a recording available on the web (e.g. YouTube) they cannot easily find, or associate their notes or resources with, PART of that recording. This project therefore addresses the important user need of making multimedia web resources easier to access, search, manage, and exploit for learners, teachers and other users through developing Synote, a unique web based application that supports the collaborative creation of synchronised bookmarks called Synmarks, containing notes, tags and links. Synote has been developed using JAVA and the Google Web Toolkit and is able to play audio or video recordings stored by users on their chosen web accessible space. Since the Synote server only stores the text annotations, millions of hours of recordings can be catered for. The performance, efficiency and effectiveness of Synote was evaluated by using it with students and teachers for a number of undergraduate courses and this has shown that they like using Synote and want more recordings and lectures to be available in this way. Synote enables learners to: search text transcripts for specific topics and then replay recordings from that point; read captions to support learning style preference, deafness, or a second language; learn more easily from the colloquial style of transcribed text; insert a bookmark to be able to continue later from where they left off; link to sections of recordings from other resources or share these sections with others; tag and highlight sections of recordings/transcripts to revisit later for clarification; annotate recordings with notes and URLs of related resources to support revision; annotate recordings of group meetings to provide evidence for assessment. Synote enables teachers to: index their recordings using syllabus topic tags; provide synchronized slides and text captions to accompany podcasts; identify which topics need further clarification from learners' notes and tags; provide feedback on learner-created recordings of presentations; tag recordings with URLs of related resources; link to and re-use sections of existing multimedia without having to edit the recording. A typical example scenario that demonstrates the use of Synote is:

> "Mike records a narrated PowerPoint of his lecture and as the students are leaving his class he uploads this into Synote which automatically creates a clickable index from the slide titles and a transcript synchronised with the recording of his voice and the slides. He quickly adds some questions students should think about for next week as well as the URLs of some other resources (including a section of a recording he made the previous year) and synchronises these with the relevant parts of lecture. By the time Sally, one of Mike's students, sits down at a computer all Mike's material is available on the web allowing her to search the slides and text transcript and notes for specific topics and then replay the recording from that point, automatically highlighting currently spoken words. She annotates the recording at specific points with notes made from the text book to aid exam revision and also tags and highlights a section of the transcript she doesn't understand fully for Mike to clarify. Before going to lunch and logging off she prints out some information she wants to discuss with her group and inserts a synchronised bookmark into the recording to be able to continue later from where she left off."

PROJECT IMPLEMENTATION URL: www.synote.org

2. INTRODUCTION

This paper discusses how the web based application Synote[i] has been developed to meet the important and pervasive user need of making multimedia web resources (e.g. podcasts) easier to access, search, manage, and exploit for learners, teachers and other users through supporting the creation of accessible synchronised notes, bookmarks, tags, links, images and text captions. The synchronised bookmarks, containing notes tags and links are called 'Synmarks' and the synchronised Annotations will be referred to as 'Synnotations'. Figures 1 & 2 schematically show the system developed. Multimedia has become technically easier to create (e.g. recording lectures) but while users can easily bookmark, search, link to, or tag the WHOLE of a podcast or video recording available on the web (e.g. YouTube[ii]) they cannot easily find, or associate their notes or resources with, PART of that recording. As an analogy, users would clearly find a text book difficult to use if it had no contents page, index or page numbers. Therefore the growing amount of knowledge available in multimedia format has yet to achieve the level of interconnection and manipulation achieved for text documents via the World Wide Web and so realize the exciting opportunities for learning that can occur in 'Web 2.0' and 'social software' environments. The provision of synchronized text captions and images with audio and video enables all their communication qualities and strengths to be available as appropriate for different contexts, content, tasks, learning styles, learning preferences and learning differences. Text can reduce the memory demands of spoken language; speech can better express subtle emotions; while images can communicate moods, relationships and complex information holistically. Deaf learners and non-native speakers may also be particularly disadvantaged if multimedia involving speech is not captioned. The new generation of Web applications emphasise social interaction and user participation through social networks, collaborative content creation, and freeform metadata creation in the form of tagging. Synote is being developed and evaluated iteratively with the involvement of users and with the support of JISC[iii] and Net4Voice[iv] .
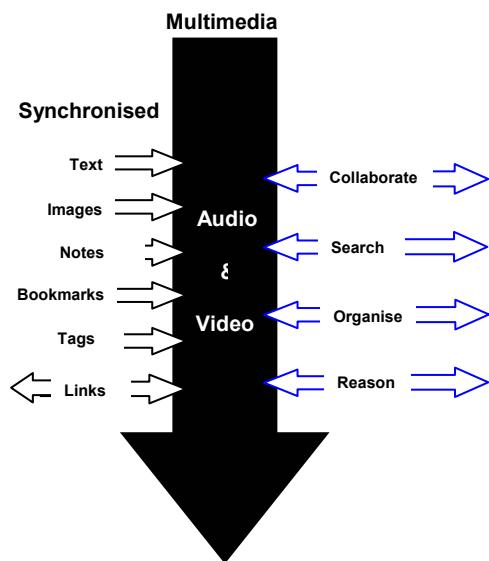

Figure 1. System overview   Figure 2. Time-based view of system
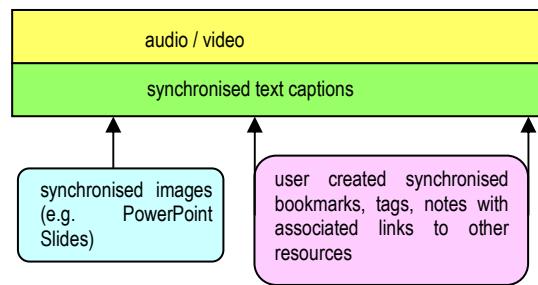
3. IDENTIFICATION OF USER REQUIREMENTS

User requirements were established through studies which showed how teachers and learners benefitted from making annotations to help search and manipulate recordings of lectures or classes to find and make use of information. Consultations within the University, including the podcasting user group, confirmed the user needs, interest in involvement in the concept and synergy with the University's road map to support and empower learners in their studies. Dr Wald in collaboration with the international Liberated Learning Consortium conducted a number of qualitative and quantitative user studies involving questionnaires, interviews, focus groups and observations with

learners and staff that established the importance of text captions for searching multimedia transcripts (Bain et al 2007). User needs analyses and evaluations were conducted with learners at the University using both low and high fidelity prototype technologies and interfaces and included reflection by learners on how they learn using recorded multimedia and how synchronized notes and tags could assist them. These studies confirmed the importance of captions for searching the recordings and reading the transcripts and the value of being able to personally annotate the recordings (e.g. bookmarks, notes and tags) and search these annotations. Other research also supports these identified user needs (e.g. Whittaker et al 1994, Wilcox et al 1997, Chiu et al 1999, Brotherton & Abowd, 2004). Annotating multimedia recordings in the ways described in this project can:

Enable learners …

- to search text transcripts for specific topics and then replay recordings from that point
- to read captions rather than listen to recorded speech to support learning style preference, deafness, or English as a second language
- to use the transcript of lectures to support their understanding of text books and academic articles if they find the more colloquial style of transcribed text easier to follow than an academic written style
- to insert a bookmark at a particular point in a recording so as to be able to continue later from where they left off
- to link to sections of recordings from other resources (e.g. documents, web pages etc.) or share these sections with others
- to tag and highlight sections of recordings/transcripts they don't understand fully so they can revisit them later for clarification
- to annotate recordings with notes and URLs of related resources (e.g. documents, websites etc.) at specific places in a recording to clarify issues and support revision
- to tag recordings using their own terms as a personal index. A community folksonomy 'index' can also be created from everyone's tags to help provide a shared understanding and a more detailed and complete index.

Enable teachers/lecturers …

- to index their recordings using syllabus topic tags
- to provide synchronized slides and text captions to accompany podcasts
- to identify which topics need further clarification from the pattern of learners' 'not understood' tags
- to provide feedback on learner-created recordings of presentations (e.g. 'good use of voice', 'good use of audio visual material')
- to ask learners to annotate recordings of their group meetings using unstructured or structured tags to provide evidence of their contributions as part of the module assessment
- to analyse unstructured tags learners use (folksonomy) to help create structured tags (ontology) for indexing future recordings
- to tag recordings with URLs of related resources (e.g. documents, websites etc.)
- to link to and use sections of existing multimedia without having to edit the recording

The importance of the identified user need of making multimedia accessible and searchable through the creation of synchronised transcripts and captions is supported by Google's encouragement for people to caption their multimedia before uploading it to Google[v] and their addition of a subtitling annotation facility to Youtube. Project Readon[vi], Overstream[vii] and dotSUB[viii] also have been developed to display manually created captions for videos that are however non-searchable. Manual captioning is time consuming and costly but speech recognition has been demonstrated to provide a cost-effective way of automatically creating accessible text captions and transcripts synchronised with audio and video and so allows audio visual material to be manipulated through searching and browsing the text. The appearance on the web of Everyzing[ix] which uses speech recognition to create a searchable transcript from podcasts or multimedia recordings provides further evidence for the importance of the identified user need and the value of this approach. However Everyzing does not display fully synchronised captions and so does not fully meet the identified user accessibility

needs. Dr Wald has collaborated with IBM and the international Liberated Learning Consortium since 1999 to investigate how speech recognition captioning can enhance learning and to develop ViaScribe[x] which uses speech recognition to create synchronized searchable transcripts, slides and captions from multimedia recordings of live or recorded speech or video and which can demonstrate the following features:

- The text can be highlighted and scroll automatically in time with the speech
- The controls can move backwards or forwards through the presentation or pause it
- The browser 'find' facility can search for text and play the multimedia from that position
- Selecting a slide thumbnail will move the presentation to that position
- Selecting a slide in the text frame will open it full size in a separate window
- The frames and slide thumbnails can be resized

Learners and teachers have found teaching and learning improved using this technology (Leitch & MacMillan 2003). Portable wireless systems can facilitate personalisation of the display and manual correction can be used when required to reduce speech recognition errors, especially if there are multiple speakers (Wald & Bain 2007).

4.   INVESTIGATIONS OF EXISTING ANNOTATION SOFTWARE

Investigations of existing annotation software (including trials and discussions with developers and users) have shown the widespread recognition of the importance of the identified user need of supporting the creation of annotations e.g.:

- The Memetic[xi] project's overall aim was to extend the functionality of the next generation AccessGrid[xii] collaboration environment with advanced meeting support and information management annotation tools that were developed in CoAKTinG[xiii] which drew on the extensive multimedia retrieval and open hypermedia knowledge and experience gained through ECS Southampton's earlier research (e.g. HyStream[xiv,] Microcosm[xv]).
- JISC and NSCF digital libraries programme funded the development of Project Pad[xvi] to assist learners in interacting with digital multimedia archive recordings.
- Transana[xvii] software for professional researchers analyzes multimedia data and the NXT [xviii] set of libraries and tools was developed for the manipulation, query and analysis of multimedia language data while Vannotea[xix] is a research tool based on the W3C Annotea project[xx].
- OneNote[xxi], Tegrity[xxii] and Panopto[xxiii] enable learners to take notes that are synchronized with recordings so that selecting a note replays the recording from that point.
- Proprietary non-standard formats[xxiv] requiring proprietary media players can be used to create enhanced podcasts by embedding markers in a multimedia file to jump to predetermined spots in a presentation.
- Video annotation tools proliferate (e.g. Youtube[xxv], PLYmedia[xxvi]) but don't support the full required functionality.

5.   SYSTEM DEVELOPMENT AND EVALUATION

No existing technology satisfied all the identified user needs which requires software that:

- Works in a web browser with web based multimedia and stores annotations separately;
- Synchronises captions, images, tags, links, notes and bookmarks;
- Enables users to add, and search for, annotations quickly and easily;
- Supports private or shared annotations;
- Is accessible.

Synote, an annotation system that meets these requirements was therefore developed to add to the synchronised multimedia captions, images, and slides demonstrated using ViaScribe, the facility for users to also create synchronised tags and notes and bookmarks and links (Wald 2007). The system design allows the future incorporation of other specific types of annotation such as audio, video, animations, drawing etc. although this can also be achieved with the current system through linking to external files or embedded videos (e.g. Youtube) or presentations (e.g. SlideShare). Synote has

built on the user interface knowledge gained in developing a real-time editing system (Wald 2006) (that enables an operator to select and correct errors in the speech recognition captions as they are displayed) and the experience with ViaScribe.

Figures 3, 4, 5 and 6 show actual screen displays of Synote. A tool has also been developed to automatically create Synote presentations from narrated PowerPoint files (see Figure. 7) and this tool also creates searchable Synmarks from the slide title, text, notes and timings and can create a synchronised transcript if this is typed into the PowerPoint slide notes.
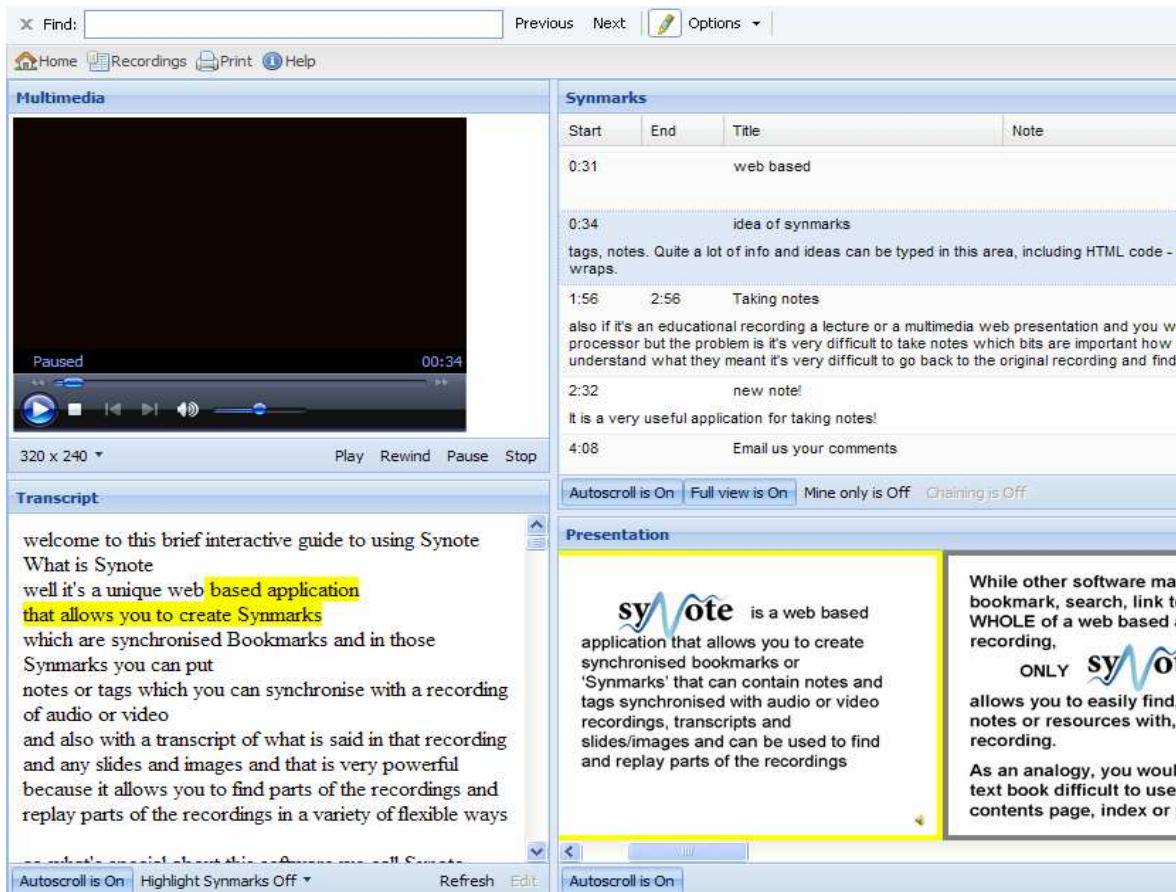
Figure 3. Synote Replay Interface

Figure 4. Synote Group Permissions Screen
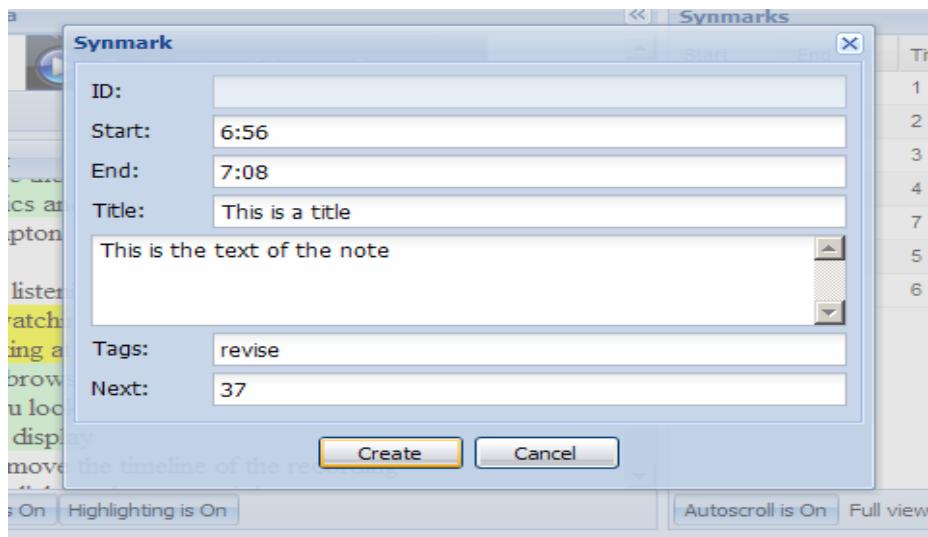
Figure 5. Synmark Creation Window

Figure 6. Synote Recording List Screen
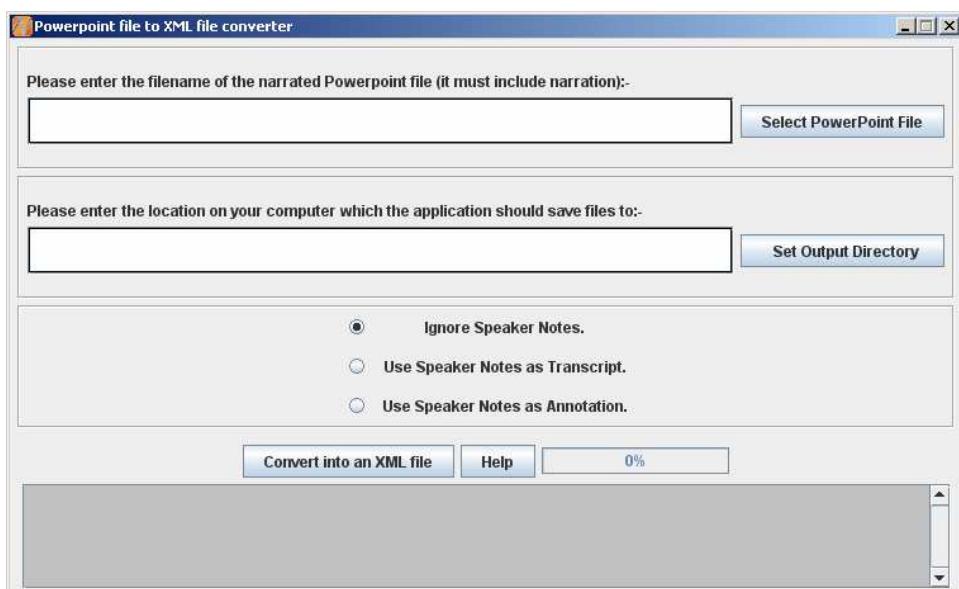
Figure 7. Powerpoint to Synote XML Converter

Synote also has a manual editing system for creating synchronised transcriptions and slides when ViaScribe or PowerPoint Narration with transcription in slide notes is not available and this system allows for both accurate synchronization by hand or approximate automatic synchronisation (See Figure 8.)
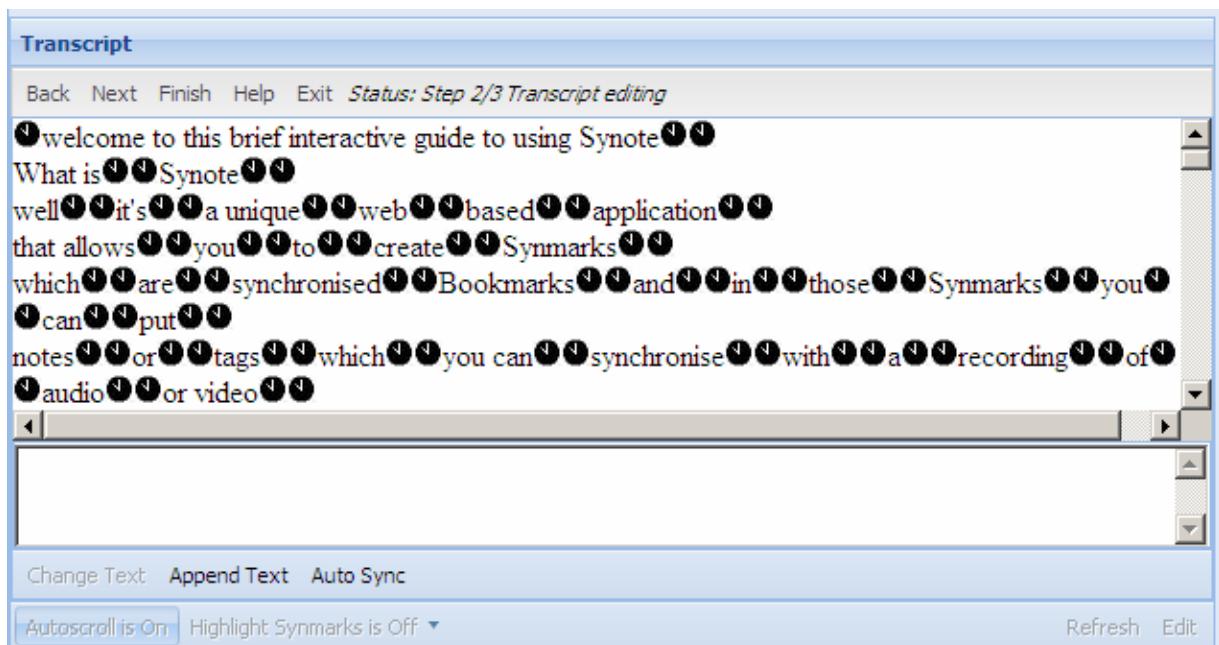

Figure 8. Transcript Editor Showing Synchronization Timing Points as Clock Icons

Synote can play most media sources through the use of automatic selection of Media Player or JW Player depending on the recorded video format. Synote is a JavaScript application that works on both PC and Apple MAC computers and has been developed using the Google Web Toolkit and tested using Internet Explorer, Firefox, Google Chrome and Safari browsers. Synote uses Grails, an MVC Web framework based on solid Open Source technologies such as Groovy, Spring and Hibernate which integrates seamlessly with Java. It aims to provide a Rails-like environment that is more familiar with Java developers to make their lives easier. The integration with major Ajax frameworks makes it suitable to implement better Web interaction experiences.

Recordings used in Synote can be created especially for teaching or can be audio and video recordings of actual lectures. Screen captured video recordings of animated PowerPoints can also be used where the images of the slides do not convey the information fully. A user permissions system is also included in Synote to allow users to specify who will be able to see or annotate their recordings. When the recording is played the currently spoken words are shown highlighted in the transcript. Selecting a Synmark, transcript word or slide/image moves the recording to the corresponding time. The browser 'Find' facility can be used to search the transcript or Synmarks and a search facility is also available for more advanced searching of all the media, transcripts and annotations stored on the database. Since the audio or video recordings are stored by users on their chosen web accessible space the Synote server only stores the text annotations and so millions of hours of recordings can be catered for. Toggling Autoscroll can ensure the current selection stays visible in the panel or allow the user to scroll manually. The displayed size of the slides and the video can be altered by the user, as can the amount of the screen display used to display the transcript or the Synmarks. A Synmark can have a Title, a note that can contain valid HTML code, tags separated by spaces and an ID of the next Synmarked part of the recording to jump to if required. The start time is automatically entered and is the time in the recording corresponding to when the Synmark was created. If some text in the transcript is selected before creating a Synmark then the corresponding start and end times are

entered automatically and the selected text can be highlighted if required. All of the content or just one line of the individual user's Synmarks or everybody's Synmarks can be displayed and Synmark information can be sorted and hidden. Users can edit or delete their own Synmarks. Information for the whole time period of the recording or only part of it can be printed and the user can choose to print information for the whole of the recording within this time period or only the Synmarked parts of it and select whose Synmarked parts. The user can also choose whether they want to print the transcript and/or the Presentation Slides/images and/or the Synmarks and select whose Synmarks and which information in the Synmarks.

Initial evaluations using questionnaires (see Figure 9) and interviews with 120 students in four classes have shown that students and teachers really liked using Synote and found the synchronised transcripts and note taking facilities very useful and wanted more recordings and lectures to be available in this way. Students who were not native speakers found the synchronised transcripts of Synote particularly helpful e.g.

> *"Synote gives a very clear understanding of module and lecturer's voice is well enough to understand the main ideas. It was very useful for me especially as I am a non English native speaker"*

Feedback from these students and the many other teachers who have tried Synote has helped to improve its usability and accessibility.

Figure 9. Synote Evaluation Questionnaire

6. CONCLUSION

A unique and innovative accessible multimedia annotation system 'Synote' has been developed to make multimedia web resources easier to access, search, manage, and exploit through supporting the creation of synchronised notes, bookmarks, tags, links, images and text captions. The system works with all the operating systems and browsers most used by students and staff and links to web based multimedia while storing the annotations separately. The new terms 'Synmark' and 'Synnotation' have been invented to denote the synchronised bookmarks and synchronised annotations created using Synote. Evaluations have shown that almost everyone has found the synchronised transcripts and note taking facility useful and wanted more recorded material made available for use in Synote.

7. REFERENCES

Bain, K., Hines, J., Lingras, P., Yumei. Q. (2007) "Using Speech Recognition and Intelligent Search Tools to Enhance Information Accessibility" *in Proceedings of HCI International In Lecture Notes in Computer Science series (LNCS)*

Brotherton, J. A., Abowd, (2004) G. D. Lessons Learned From eClass: Assessing Automated Capture and Access in the Classroom, *ACM Transactions on Computer-Human Interaction, Vol. 11, No. 2.*

Chiu, P., Kapuskar, A., Reitmeief, S., Wilcox, L. (1999) NoteLook: taking notes in meetings with digital video and ink, *Proceedings of the seventh ACM international conference on Multimedia (Part 1)* 149 – 158.

Leitch, D., MacMillan, T. (2003) Liberated Learning Initiative Innovative Technology and Inclusion: Current Issues and Future Directions for Liberated Learning Research. *Saint Mary's University, Nova Scotia.* http://www.liberatedlearning.com/

Wald, M. (2006) Creating Accessible Educational Multimedia through Editing Automatic Speech Recognition Captioning in Real Time. *International Journal of Interactive Technology and Smart Education: Smarter Use of Technology in Education 3(2)* pp. 131-142.

Wald, M. (2007) A Research Agenda for Transforming Pedagogy and Enhancing Inclusive Learning through Synchronised Multimedia Captioned Using Speech Recognition. *Proceedings of ED-MEDIA: World Conference on Educational Multimedia, Hypermedia & Telecommunications.*

Wald, M. Bain, K. (2007) Enhancing the Usability of Real-Time Speech Recognition Captioning through Personalised Displays and Real-Time Multiple Speaker Editing and Annotation. In Proceedings of HCI International: *12th International Conference on Human-Computer Interaction, Beijing.*

Whittaker, S., Hyland, P., Wiley, M. 1994 Filochat handwritten notes provide access to recorded conversations, *Proceedings of CHI,* 271-277.

Wilcox, L., Schilit, B., Sawhney, N. Dynomite, A. 1997: Dynamically Organized Ink and Audio Notebook, *Proc. of CHI*, 186–193

---

[i] http:www.synote.org

[ii] http:www.youtube.com

[iii] http://www.jisc.ac.uk/fundingopportunities/funding_calls/2007/07/circular0207.aspx

[iv] https://spazivirtuali.unibo.it/net4voice/default.aspx

[v] http://video.google.com/support/bin/answer.py?hl=en&answer=26577

[vi] http://projectreadon.com

[vii] www.overstream.net

[viii] http://dotsub.com/

[ix] http://www.everyzing.com/

[x] http://www.liberatedlearning.com/technology/index.shtml

[xi] http://www.memetic-vre.net/

[xii] http://www.accessgrid.org/

[xiii] http://www.aktors.org/coakting/

[xiv] http://www.iam.ecs.soton.ac.uk/projects/hystream/

[xv] http://www.mmrg.ecs.soton.ac.uk/projects/microcosm.html

[xvi] http://www.jisc.ac.uk/whatwedo/programmes/programme_dlitc/project_spoken_word.aspx

[xvii] http://www.transana.org/

[xviii] http://www.ltg.ed.ac.uk/NITE/

[xix] http://www.itee.uq.edu.au/~eresearch/projects/vannotea/index.html

[xx] http://www.w3.org/2001/Annotea/

[xxi] http://office.microsoft.com/en-us/onenote/HA101212541033.aspx

[xxii] http://www.tegrity.com/

[xxiii] http://panopto.com/

[xxiv] http://docs.info.apple.com/article.html?artnum=301880

[xxv] http://www.youtube.com/t/annotations_about

[xxvi] http://www.plymedia.com