

Synchronised Annotation of Multimedia

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

Abstract

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 they cannot easily find, or associate their notes or resources with, PART of that recording. This paper describes the development of a web based application that makes multimedia web resources (e.g. podcasts) easier to access, search, manage, and exploit for learners, teachers and other users through the creation of notes, bookmarks, tags, links, images and text captions synchronized to any part of the recording.

1. Introduction

This paper discusses how to make multimedia web resources (e.g. recorded lectures) easier to access, search, manage, and exploit for learners, teachers and other users through developing a technology that supports the creation of synchronised notes, bookmarks, tags, links, images and text captions. Figure 1 schematically shows the Synote system which has been developed using the Google Web Toolkit. The synchronised bookmarks, containing notes tags and links are called Synmarks and Synchronised Annotations called 'Synnotations'. Multimedia has become technically easier to create but while users can easily bookmark, search, link to, or tag (i.e. classify) the WHOLE of a podcast or video recording available on the web 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. 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 speech is not captioned. Synote is being developed and evaluated iteratively with the involvement of users and with the support of JISCⁱ and Net4Voiceⁱⁱ.

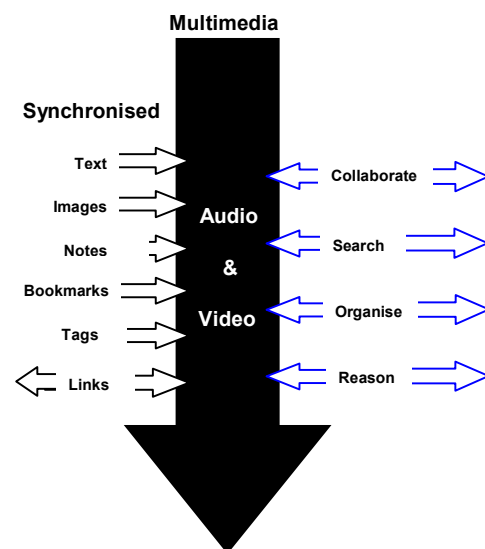


Figure 1. System overview

2. Identification of User Needs

User studies have established the importance of synchronized text transcripts for searching and reading and the value of being able to personally annotate the recordings and search these annotations ([1] [2] [3] [4] [5]). Synote enables learners or teachers:

- to read and search text transcripts and slides and replay recordings and slides to support learning style preference, deafness, disability or English as a second language

- to bookmark, tag and highlight sections of recordings/transcripts for indexing, revision, clarification or feedback
- to collaboratively annotate recordings with notes and URLs of related resources
- to link to or from sections of multimedia

Manual captioning is time consuming and costly (Googleⁱⁱⁱ, Project Readon^{iv}, Overstream^v, dotSUB^{vi}) but speech recognition can automatically create accessible text captions and transcripts synchronised with audio and video (Everyzing^{vii}, ViaScribe^{viii ix}) and learners and teachers have found teaching and learning improved using this technology [6]. 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 [7].

3. Investigations of Existing Annotation Software

Investigations of existing annotation software have shown the widespread recognition of the importance of annotations e.g.: Memetic^x, CoAKTinG^{xi}, HyStream^{xii}, Microcosm^{xiii}, Project Pad^{xiv}, Transana^{xv}, NXT^{xvi}, Vannota^{xvii}, Annotea project^{xviii}, OneNote^{xix}, Tegrity^{xx}, Panopto^{xxi}, Proprietary non-standard formats^{xxii}, Youtube^{xxiii}, PLYmedia^{xxiv}.

4. System Development and Evaluation

No existing technology satisfied all the identified requirements and so it was necessary to develop Synote to: work with web multimedia and store annotations separately; synchronise captions, images, tags, links, notes and bookmarks; enable users to add, and search for, annotations quickly and easily; support private or shared annotations; be accessible. Synote adds to the synchronised multimedia captions, images, and slides demonstrated using ViaScribe, the facility for users to also create shared tags, notes, bookmarks and links [8]. Figures 2 and 3 show actual screen displays of the Synote prototype which can also create searchable synmarks from slide titles, text, notes. A user permissions system allows users to specify who can annotate their recordings and read their annotations. Synote can create and edit synchronised transcriptions and slides and can play most media sources through the use of Media Player and JW Player. 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.

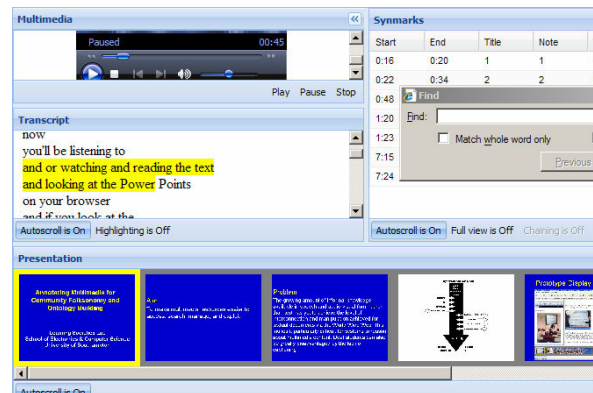


Figure 2. Synote interface

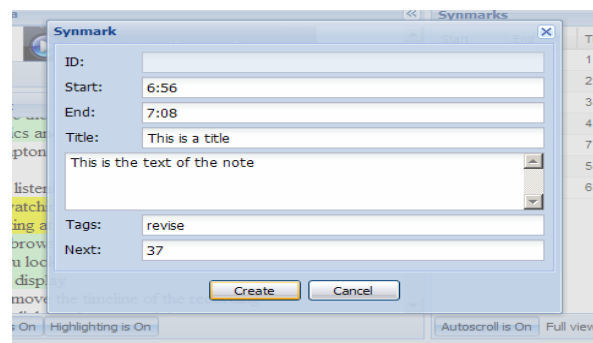


Figure 3. Synmark creation

The browser 'Find' facility can be used to search the transcript or Synmarks and more advanced searching of all the information stored on the database is available. Since the audio or video recordings are stored by users on their chosen web accessible space, the Synote server need only store the text annotations and so millions of hours of recordings can be catered for. 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. Any part of the Synmark can be sorted and displayed and 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 with students in three classes using both recordings of actual classroom lectures and pre-recorded material have shown that students like using Synote, found the synchronised transcripts and note taking facility useful and want more recordings and lectures to be available in this way. Further evaluations are in progress and the results will be presented at the conference.

5. Conclusion

The unique multimedia annotation system ‘Synote’ makes multimedia web resources easier to access, search, manage, and exploit through supporting the creation of notes, bookmarks, tags, links, images and text captions synchronized to any part of the recording. Evaluations have shown that almost everyone found the synchronised transcripts and note taking facility useful and wanted more recorded material made available on Synote.

6. References

[1] 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 2007, In Lecture Notes in Computer Science series (LNCS)

[5] Brotherton, J. A., Abowd, G. D. Lessons Learned From eClass: Assessing Automated Capture and Access in the

Classroom, ACM Transactions on Computer-Human Interaction, 2004, Vol. 11, No. 2.

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

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

[9] Wald, M. 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 2006 3(2) pp. 131-142.

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

[7] Wald, M. and Bain, K. 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 2007: 12th International Conference on Human-Computer Interaction, Beijing.

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

[3] Wilcox, L., Schilit, B., Sawhney, N. Dynamite: A Dynamically Organized Ink and Audio Notebook, Proc. of CHI 1997, 186–193

ⁱhttp://www.jisc.ac.uk/fundingopportunities/funding_calls/2007/07/circular0207.aspx

ⁱⁱ<https://spazivirtuali.unibo.it/net4voice/default.aspx>

ⁱⁱⁱ<http://video.google.com/support/bin/answer.py?hl=en&answer=26577>

^{iv}<http://projectreadon.com>

^vwww.overstream.net

^{vi}<http://dotsub.com/>

^{vii}<http://www.everyzing.com/>

^{viii}http://www-306.ibm.com/able/solution_offerings/ViaScribe.html

^{ix}<http://www.liberatedlearning.com/technology/index.shtml>

^x<http://www.memetic-vre.net/>

^{xi}<http://www.aktors.org/coacting/>

^{xii}<http://www.iam.ecs.soton.ac.uk/projects/hystream/>

^{xiii}<http://www.mmrg.ecs.soton.ac.uk/projects/microcosm.html>

^{xiv}http://www.jisc.ac.uk/whatwedo/programmes/programme_dlit/pr object_spoken_word.aspx

^{xv}<http://www.transana.org/>

^{xvi}<http://www.ltg.ed.ac.uk/NITE/>

^{xvii}

<http://www.itee.uq.edu.au/~eresearch/projects/vannotea/index.html>

^{xviii}<http://www.w3.org/2001/Annotea/>

^{xix}<http://office.microsoft.com/en-us/onenote/HA101212541033.aspx>

^{xx}<http://www.tegrity.com/>

^{xxi}<http://panopto.com/>

^{xxii}<http://docs.info.apple.com/article.html?artnum=301880>

^{xxiii}http://www.youtube.com/t/annotations_about

^{xxiv}<http://www.plymedia.com>