

# Synote: Multimedia Annotation ‘Designed for all’

XXXX  
XXXXXXXXXX  
XXXXX  
XXX

**Abstract-** This paper describes the development and evaluation of Synote, a freely available web based application that makes multimedia web resources (e.g. podcasts) easier to access, search, manage, and exploit for all learners, teachers and other users through the creation of notes, bookmarks, tags, links, images and text captions synchronized to any part of the recording. Synote uniquely enables users to easily find, or associate their notes or resources with any part of a podcast or video recording available on the web and the students surveyed would like to be able to access all their lectures through Synote.

## I. INTRODUCTION

Synote<sup>1</sup> has been developed to overcome the problem that 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 [1]. As an analogy, users would clearly find a text book difficult to use if it had no contents page, index or page numbers. Synote [2] is freely available and uniquely can use speech recognition to synchronise audio or video recordings of lectures or pre-recorded teaching material with a transcript, slides and images and student or teacher created notes. In 2009 Synote won the EUNIS International E-learning Award and was described by the awards panel as the most exciting educational innovation they had seen for many years. The system is unique as it is free to use, automatically or manually creates and synchronises speech transcriptions, allows teachers and students to create real time synchronised notes or tags and facilitates the capture and replay of recordings in any media format and browser and stored anywhere on the web. Synote has been developed and evaluated with the involvement of users and with the support of JISC<sup>2</sup> and Net4Voice<sup>3</sup>. Figs. 1 & 2 schematically show the Synote system while figs. 3 and 4 show screen displays of its interface. Fig. 5 shows the manual transcript editing interface with clock icons indicating synchronization points in the text. Fig. 6 shows an example of a print preview. Synote can also automatically create synchronized and searchable audio, transcripts and slides (including titles, text and notes) from Narrated PowerPoint Slides. The synchronised bookmarks, containing notes tags and links are called Synmarks. When the recording is replayed 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 provision of text captions and images synchronized with

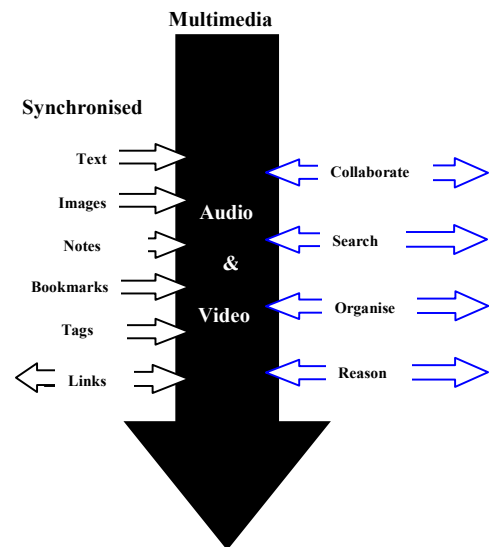


Fig. 1. Synote overview

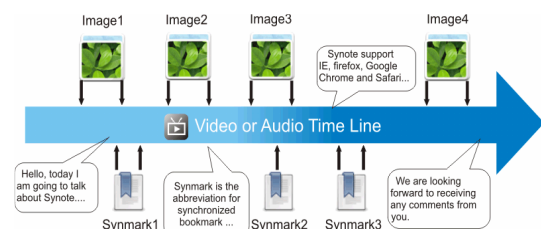


Fig. 2. Time-based view of Synote's features

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. Synote's synchronised transcripts enable the recordings to be searched while also helping support non native speakers (e.g. overseas students) and deaf and hearing impaired students understand the spoken text. The use of text descriptions and annotations of video or images help blind or visually impaired students understand the images.

<sup>1</sup> <http://www.synote.org>

<sup>2</sup> <http://www.jisc.ac.uk>

<sup>3</sup> <http://www.net4voice.eu>

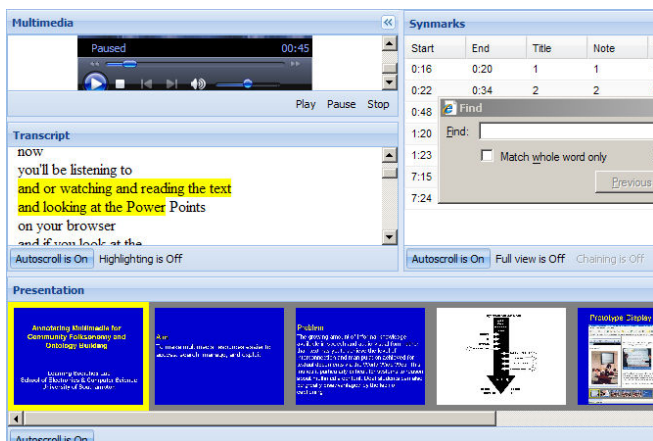


Fig. 3. Synote interface

## multimedia

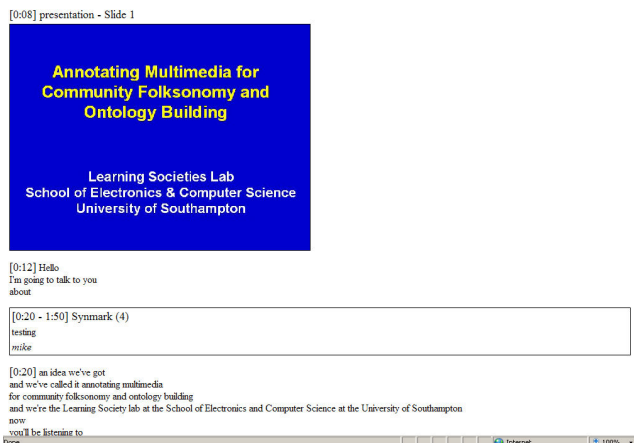


Fig. 6. Synote Print Preview

## II. ASSESSMENT

It is very common for university students to be assessed on their presentation skills. Typically they will make a short live presentation using slides for which they may be given a mark and some oral and/or written feedback. Presenting live in class is however very time consuming. For example, it could take 10 hours if each student in a class of 60 students presented for 5 minutes, allowing for questions and change over time. It could take 15 hours if teachers and other students provided oral feedback immediately after the presentations. Also while feedback can be given for the WHOLE of the presentation it is not easy to associate feedback with a specific PART of that presentation such as a critical moment(s). If however students record presentations Synote allows teachers and peers to provide feedback linked to the appropriate and critical moment(s) (for example 'you were reading from notes rather than making eye contact with the audience'). The written feedback can be created live or after the event and the student can read it afterwards synchronised to the recording. It is possible also to provide the feedback in audio or video format (e.g. sign language)

It is difficult to assess students undertaking practical tasks (e.g. student nurses in real or simulated wards or student teachers giving real or simulated lessons). Again, while feedback can be given for the WHOLE of the task it is not easy to associate feedback with a specific PART of that task such as a critical moment. If performance on tasks is recorded then Synote allows teachers and peers to give feedback linked to the appropriate and critical moment(s). Written feedback can be created live or after the event and the student can read it afterwards synchronised to the recording. It is possible also to provide the feedback in audio or video format (e.g. sign language)

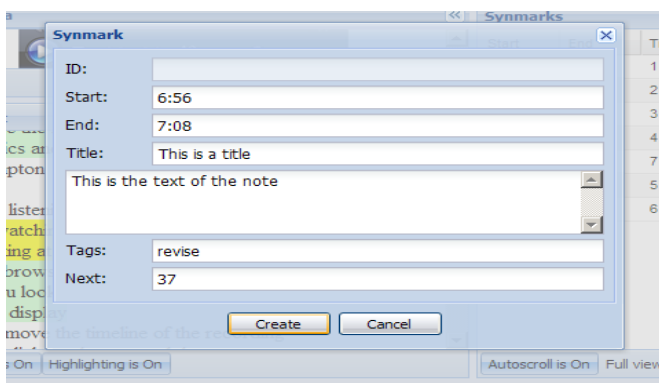


Fig. 4. Synmark creation

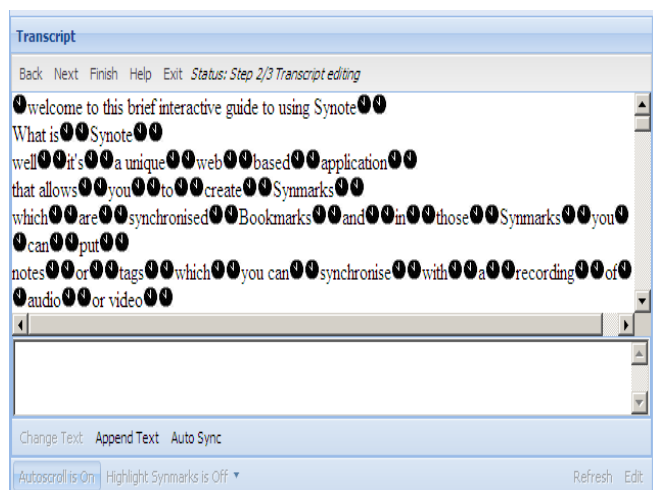


Fig. 5. Transcript Editing and Synchronisation

It is very common for university students to be assessed on their work in a group. Typically they will be asked for evidence for their contribution to a group. It is however

difficult for students to provide evidence for their contributions at face to face meetings as typically minutes only record agreed actions. If meetings are recorded students can annotate the recordings at appropriate and critical moment(s). The synchronised annotations can be made live or after the event and the assessor/evaluator can read them afterwards synchronised to the recording and also provide feedback linked to the appropriate and critical moment(s). It is possible also to provide the feedback in audio or video format (e.g. sign language)

It is difficult for students to really interact in large lectures as only a few students will be able to ask or answer questions. 'clickers'/'electronic voting systems' [3] are sometimes used to give feedback or answer in-class tests but this requires students to buy these at the beginning of their studies and carry them with them at all times or they can be given out at the start of the class and collected in at the end of the class, which can be very time consuming for the teacher/lecturer. If lectures are recorded, using Synote teachers can pose questions at points in a recording and provide feedback to students' answers. Students can answer after class or use twitter to answer live in class, with these answers automatically synchronised with the recording. Future developments could include the use of synchronised multiple choice questions.

### III. SPEECH RECOGNITION TRANSCRIPTION

Synote builds on 10 years of Dr Wald's work on the use of speech recognition for learning in collaboration with IBM, MIT and other organisations in the international Liberated Learning Consortium of which he was a founder member. Google announced in November 2009 that they plan to use speech recognition to caption YouTube videos , a feature available for nearly two years for students using Synote. The availability of 'continuous' speech recognition dictation systems that didn't require the speaker to pause between words led to the Liberated Learning Consortium being formed in 1999 to trial real-time transcription for deaf students in classrooms [4]. All students liked using the corrected transcripts after the class but the accuracy from spontaneous speech was lower than from dictated speech and the text was difficult to read without the dictation of punctuation. A few speakers achieved 90% accuracy speaking clearly at an even pace. Readability was improved by developing software to automatically break up the text using new lines or punctuation triggered by pauses in the speech. Further developments in collaboration with IBM included automatically synchronising audio, text and PowerPoint slides (which enabled recordings to be searched) and training the system using spontaneous speech rather than scripts. IBM's ViaScribe [5] also provided a programming interface for the speech recognition output to be enhanced through interoperability with systems that; displayed the recognised text on personalized displays on networked laptops in ways users preferred (e.g. colour, font, size etc.); allowed the recognition errors to be correct in real time by one or more people; and merged the text output from

any number of speakers each using ViaScribe. IBM more recently developed the speaker independent 'Attila' system that can be hosted on the web , interfaced with other applications and used to transcribe, edit and display recordings created in a wide range of multimedia file formats. Synote can add user created notes and tags synchronised to the recordings, transcripts and slides provided by ViaScribe and Attila. Integration of the IBM Hosted Transcription System with Synote has simplified the process of transcription.

### IV. CLASSROOM USE AND EVALUATION OF SYNOTE

Synote is currently being used by teachers in universities in the UK, Italy, Germany, Pakistan, Australia, US and Canada. Since 2007 Dr Wald has used Synote with over 40 recordings of his lectures with synchronised transcripts and slides for his teaching of approximately 200 students on 5 undergraduate and postgraduate modules in The University of Southampton's School of Electronics and Computer Science (ECS). At the time of writing there are over 420 recordings publically available on Synote (most with synchronised transcripts) for students to use for their learning. These include lectures of colleagues and guest lectures presented by leaders in their field. The provision of a verbatim synchronised transcript enables students to concentrate on learning and take only brief synchronised notes in Synote (e.g. 'revise this section for exam', 'I don't understand this fully' etc.). This feature is greatly valued by all students, not only by deaf students who need to lipread or watch a sign language interpreter and so can't take notes or dyslexic students or non-native speakers who find it difficult to take notes. The fact that Synote is used and valued by all students means that non native speakers and disabled students feel more included and do not have to use special technology. They have said that they do not like standing out from the crowd by having to walk to the front of the class to ask the teacher if they can record the lecture on their personal digital recorder. Also the quality of recording from a teacher's wireless headworn microphone is significantly better than from small personal digital recorders placed at the front of the class. Students did not like retyping handwritten notes they had taken in class into Synote after the recording had been uploaded and so Synote has recently been enhanced so that synchronised notes taken live in class on mobiles or laptops using Twitter can be automatically uploaded into Synote. This process is shown in Figs. 8, 9 and 10.

5 ECS classes (two 1st year classes, one 3rd year class and two Masters classes) with approximately 200 students were surveyed, with 101 students filling in questionnaires about their experience with Synote. The questionnaire results showed that Synote's design to support a wide range of browsers was justified and appreciated with 54% using Internet Explorer, 31% using FireFox, 12% using Google Chrome and 3% using Safari. The results also confirmed that Synote was easy to use as over 80% of the respondents didn't need to read the Synote guide and rated Synote as 4 or 5 on a 5 point scale of ease of use with the remaining percentage rating

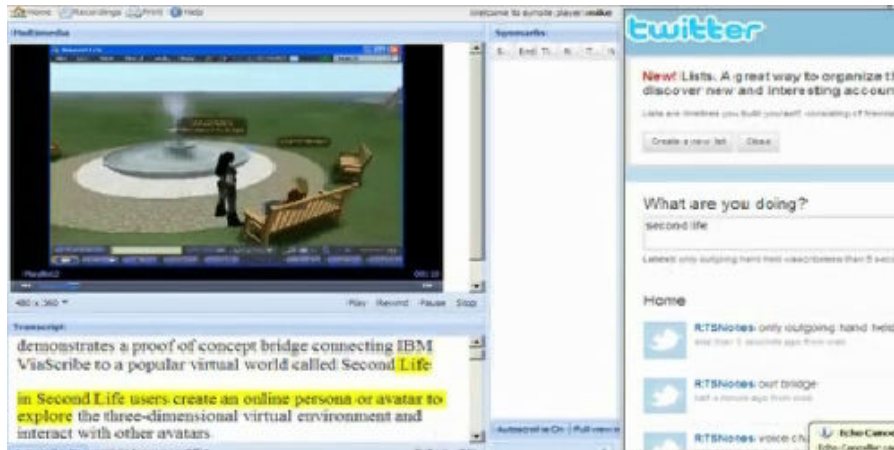


Fig. 8. Twitter live notes



Fig. 9. Synote's Twitter interface

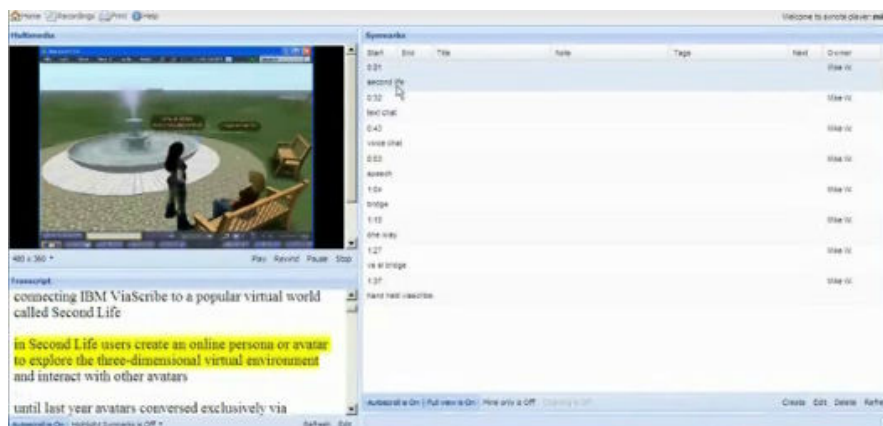


Fig. 10. Twitter notes displayed in Synote

it as 3. The design of Synote to provide synchronised slides, video, audio and transcripts was also appreciated with over 80% respondents rating Synote as 4 or 5 for being useful overall as well as its slides, video and audio and transcript also each being useful with average ratings of between 4.1 and 4.5. A slightly smaller percentage rated the Synmark option and the print out option as being useful with an average of 3.7 and

3.3 respectively on the 5 point scale used. One student commented that if Synote was used regularly then it would become second nature to use all these facilities. 97% of the students wanted all their lectures to be presented on Synote. 44 students in two of these 5 classes identified in more detail how using Synote affected their learning by indicating whether

Synote improved or reduced or provided no change in the aspects of their learning as shown in Table 1.

Table 1.  
Effect on Aspects of Learning

	improve	no change	reduce
learning	95%	5%	0%
attention	61%	34%	5%
motivation	50%	50%	0%
efficiency	77%	23%	0%
enjoyment	66%	32%	3%
results	69%	31%	0%
notetaking	53%	30%	18%
attendance	13%	45%	42%

A few students commented that some students' notetaking skills might not develop if all classes were recorded and presented on Synote. Some students commented that if they were ill or had serious problems they would normally still try and come in to classes because they were worried they would get behind in their work and would be unable to catch up, whereas with Synote they would be able to stay at home and get well secure in the knowledge that they would not miss anything as they could learn using the Synote recording.

Students stated that it was important that that ALL lectures were recorded so they didn't find that the one lecture they missed hadn't been recorded. Of these 44 students 37% identified themselves as native speakers, 26% as fluent, 28% as having good language knowledge and 9% as having little language knowledge. Non native speakers in particular commented how valuable Synote was for them as it was sometimes difficult to understand lecturers' speech and notetaking in a foreign language was very difficult for them. One commented that they could get words not understood in the transcript translated by Google. Of these 44 students 7% identified themselves as having hearing disabilities, 2% visual disabilities, 11% learning disabilities and 7% other disabilities. Two overseas students wrote:

*"Synote gives a very clear understanding of module ...It was very useful for me especially as I am a non English native speaker"*

*"I think Synote is a very good way to listen to lectures. If for example we miss the lecture we can actually listen to it in our own time or if we didn't really understand the lecture we can go back to it and listen to it carefully. I also like the highlighted part whenever the lecturer speaks on the text so we can't actually get lost within long texts."*

Other students wrote

*"Synote is very useful for students in general, I think at present all they need is to get used to something like Synote. It will then become second nature"*

*"Synote is the best system I have ever seen for assistive technology it is very useful for me to understand what the lecturer taught after class I hope all school majors could integrate this system thanks"*

## V. ERROR CORRECTION

Commercial rates for manually transcribing and synchronising a lecture recording are typically around £2/minute (rates vary dependent on quality and quantity) and so it would cost about £90 for transcribing and synchronising a 45 minute class. For Speech recognition to be used it must therefore cost less than this. The cost of the speech recognition will involve licensing costs as well as server and maintenance costs. Manual correction of errors will also be required if 100% accuracy is to be achieved. The Attila IBM speech recognition system used with Synote has only available a US and not a UK English voice and language model and has no facility for adding vocabulary or training to individual users' voices and so typically gives word error rates between 15% - 30% for UK speakers using headset microphones. This compares well with the National Institutes of Standards (NIST) Speech Group reported WER of 28% for individual head mounted microphones in lectures [6]. It is important to note that no student themselves ever correctly recognises 100% of a spoken lecture and that many speech recognition 'errors' don't seriously affect readability (e.g. 'a' instead of 'the') or search results. However if improved accuracy is desired then manual editing is required. Analysis of the data shows that typical lecturer speaking rates varied from 120 - 160 wpm with an average around 140 words per minute. With 30% error rates, editing at an average of 14 corrections/minute takes approximately 3 times as long as the lecture itself whereas editing lectures that have a 15% error rate takes twice as long as the lecture. Synote allows transcripts to be created and synchronised manually as well as through speech recognition and this approach could be quicker than correcting speech recognition errors if these are greater than 30%. In the UK, disabled students (e.g. Deaf students) requiring transcription of recordings can pay for this from their Disabled Student Allowance. Universities do not however routinely transcribe or caption all recordings as they claim it is too costly for them to do so. One possible sustainable approach to obtaining accurate transcriptions could be to devise a system where students in the classes themselves corrected errors they find in the transcript, either voluntarily or through being paid or through being given academic credit. For speech recognition systems to be able to transcribe lectures more accurately in addition to having 'local' (e.g. UK) language and voice models they need to be designed for education rather than for dictation. For example they need to: Be speaker independent with a speaker dependent training facility; Be customisable for different subject domains; Cope with low quality speech signals and background noise; Recognise or ignore partial words (e.g. hesitations) or 'fillers' (e.g. 'um', 'er')

## VI. CONCLUSION AND FUTURE WORK

Synote has been shown to provide accessible and very well received enhancements to web based teaching and learning for all students and to integrate well with other applications including PowerPoint, Twitter and Speech Recognition Software. While Synote has much to offer even without

synchronised transcripts, there could be a huge educational demand for speech recognition transcription if it can be made sufficiently accurate and affordable. Further developments of Synote currently being investigated include methods to enable multiple users to simultaneously edit the transcript.

#### REFERENCES

- [1] Whittaker, S., Hyland, P., Wiley, M. Filochat handwritten notes provide access to recorded conversations, *Proceedings of CHI* 1994 271-277.
- [2] Wald, M., Wills, G., Millard, D., Gilbert, L., Khoja, S., Kajaba, J. and Li, Y. Synchronised Annotation of Multimedia. *Ninth IEEE International Conference on Advanced Learning Technologies*. 2009 594-596
- [3] MacGeorge, E., Homan, S., Dunning Jr, J., Elmore, D., Bodie, G., Evans, E., Khichadia, S., Licht, S., Feng, B., Geddes, B. Student evaluation of audience response technology in large lecture classes. *Education Tech Research Dev* 2008 56:2 25-145
- [4] 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/>
- [5] 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.
- [6] Fiscus, J., Radde, N., Garofolo, J., Le, A., Ajot, J., Laprun, C., The Rich Transcription 2005 Spring Meeting Recognition Evaluation, *National Institute Of Standards and Technology*