

Automated Podcasting System for Universities

<http://dx.doi.org/10.3991/ijet.v8i1.2363>

Y. Grigoriadis, C. Stickel, W. Nagler, M. Ebner and M. Schön
Graz University of Technology, Graz, Austria

Abstract—This paper presents the results achieved at Graz University of Technology (TU Graz) in the field of automating the process of recording and publishing university lectures in a very new way. It outlines cornerstones of the development and integration of an automated recording system such as the lecture hall setup, the recording hardware and software architecture as well as the development of a text-based search for the final product by method of indexing video podcasts. Furthermore, the paper takes a look at didactical aspects, evaluations done in this context and future outlook.

Index Terms—automated podcast, didactics, OCR, university

I. INTRODUCTION

Audio and video media go along with us as normal as waking up in the morning. Anybody carries some kind of player or receiver whether it is an old fashioned radio, an iPod, or a smartphone is able to listen to podcasts anywhere and anytime. Though a podcast (former called audioblogging) by definition is an audio file distributed automatically via RSS (Really Simple Syndication) to its consumers, here the term is used in a broader sense as term for any kind of audiovisual media offered or broadcasted over the Internet. Podcasts in that sense have become a common part of our life. Even more; to publish a video or audio file on the Internet is no big deal any longer, it can be done in a breath by anybody using social media sharing platforms such as YouTube. Podcasts are an essential part of the Web 2.0 phenomena, as in [19].

It is negligent to assume that podcasts will stop at school's or university's doors. It has become usual that students record lectures with their mobile devices not only for private usage. Very often they do that without permission of the lecturer. Even more, they distribute their private recordings among their study fellows. At least, because of that legal aspect it has become evident for an education institution to offer recording services. Clearly, the documentation and distribution of lectures by assistance of media such as audio, video, on-screen interactive systems or similar gives a peerless advantage to students of any curriculum. Though recording lectures already has become a quite established practice in university's daily routine there are still some nuggets to earn due to the fact that most of the systems in use are either too expensive for university wide usage, fault-prone, or semi-automated.

In practice, an advantage of the method of podcasting lies in the possibility for working students to participate in the lecture activities without physical presence and to enable them to complete in some cases the lectures with good results even without visiting the classroom. Even more, the personal oral evaluations result in positive effects on the learning outcome. But it must be definitely

pointed out that it is not a goal to replace attendance of students at university by offering recorded lectures; TU Graz is not a distance university. Considering these processes TU Graz decided to work on the establishment of a fully automated system for recording activities in 2010. Furthermore, the recording services done by the Department of Social Learning (DSL) for the entire TU Graz blast its personal and technical capacities. At least, the constantly increasing recording activities done or assisted by three staff members of DSL called for automating the service (compare fig. 1).

In the early two years of podcasting DSL was very interested in the evaluation of its new service, as in [7] [20]. For that reason students attending lectures recorded have been polled. After two years a total of 700 hours had been evaluated by n=217 students. The results were important for the continuity of the project. One-fifth of all polled students did not exercise the option of that new service. Surprisingly a quarter of all students listened to all recordings of their relevant lectures. Only 14% tried the offer once but did not stuck to it longer. The most stated reason for not using the offered podcasts had been technical problems (51%) or ignorance of the service (35%). Improvement of both the technical part and better information are problems of the past by now. Mainly podcasts are seen as a supplement to the general teaching and learning content offered (37%) and to have a possibility to go through the lecture again (25%) as well as to have a backup (21%). For learning efforts 10% used the podcasts intensively. The quality of the podcasts has been rated "well" by more than half of all students (54%) in means of technical quality but not of the content. After three years of enhancement we started further evaluations about the podcasting services in study year 2011/2012. For the number of recordings has grown since then only a couple of selected and exemplary lectures have been polled; a

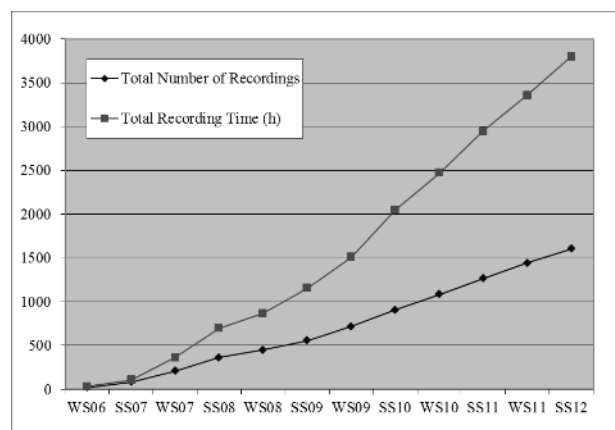


Figure 1. Total number of recordings and recording time from 2006 till 2012 including recording and streaming activities

mass lecture, a small lecture, an obligatory lecture and a lecture that has not been recorded before. A first glance at the results encourages our efforts. The acceptance for recordings has turned into a demand. Even those students that have not been in touch with this service answered very optimistically throughout. For not all results have been yet analyzed this will be subject of future publication.

II. CORNERSTONES OF THE DEVELOPMENT OF AN AUTOMATED RECORDING SYSTEM AT TU GRAZ

A. *Established Recording Services at TU Graz*

At TU Graz recording services started as a centralized service in autumn 2006 for the entire university in the course of the formation of DSL as a part of the Information Technology Services (ITS), as in [7] [20]. After a period of trial-and-error a simple standard was concluded recommending at least a minimum of didactical settings, as in [10]. Other research works corroborate this didactical need, as in [1] [2] [4] [6] [11] [12] [18].

At that time all recordings were screencasts using the screen-capturing software Camtasia by TechSmith. The service was predominantly carried out or assisted by staff members of DSL. Only a few teachers did recordings on their own. At least post-processing tasks as well as publishing of the final media files were always done by DSL. The media files were published on the learning management system (LMS) of the TU Graz called TU Graz TeachCenter (TUGTC). TUGTC is a home-made system developed at the Institute of Information Systems and Computer Media (IICM) of TU Graz. It bases on the platform called WBTMaster that has been developed and used by the team of Prof. Nikolai Scerbakov at IICM since the later 1990s, as in [3] [5] [17]. Since autumn 2006 it has been the central LMS for the entire university. As one result of continuous evaluations the offered formats of published files often changed. At present a MP4 file is being generated as standard output format. It can be consumed and download as simple MP4 file or opened online within a flash environment using any browser supporting flash. The flash environment comes with navigation and offers a text based search too. The development and functionality of this flash environment will be subject to this paper too.

The recording services then were enriched by the installation of a portable live streaming server in autumn 2008. Core of that service is the streaming server from the Canadian company desire2learn (former called ePresence (server v.4.1)) with its web-interface for displaying the live casts and as an archive of them afterwards, as in [15] [16]. Besides that the whole system came with two software packages; the ePresence Presenter for screening the lecturer's laptop and the ePresence Producer for (post-) recording and publishing procedures. Furthermore a portable media station is linked over Internet to the server to manage the recording easily. Unlike the screening setting the streaming always includes a filming, an ordinary camcorder can do. During the live streaming attendees of the cast can follow easily using any browser. They may discuss the ongoing activities within a chat, may take notes to several slides presented or comment them online. An archived cast (the archiving is done by the software too) can be searched through by words and is automatically structured subject to the slides of the PowerPoint

presentation. Furthermore a code is offered for embedding the cast into one's own web interface. Fig. 2 shows the production process of the ePresence system in general.

Main area of application is a relief of lectures with a high number of students registered. The fact that the largest lecture hall of the TU Graz did not cover the biggest lectures has always been a problem of the past. Secondly streaming is applied for lectures or speeches with special interest to the public. The increasing request for distance education or blended learning scenarios has focused the discussion on streaming media and technology latest. Though the access of the service is generally restricted to university usage, a couple of those recordings done with the streaming server are of public interest too. Such recordings can be accessed for free and without login. In winter term 2011 the ePresence server has been updated to the current standard desire2learn offers.

In spring 2007 Apple opened an academic version of it called iTunes U to manage, distribute, and control access to educational audio and video content and PDF files for students within a college or university as well as the general public on the Internet. In November 2009 TU Graz was pleased to launch its own portal on Apple's iTunes U platform primarily for publicity and promotion purposes for free. It was the first iTunes U platform of a university for technology in Austria.

To enable anybody with Internet access but regardless to the system used (Windows or Mac OS), the content is offered once on iTunes U platform and on a corresponding TU Graz variant of it (<http://itunes.tugraz.at>). Main target is to have a collection of multimedia content that represents all subjects of the TU Graz; research and teaching, news and stories, conferences and events as well as general information about Graz. For TU Graz is the first academic partner of Apple in Austria to develop iPhone applications officially there is a special series about that subject as well. Since then over 500 media files had been offered resulting in a total of more than 688000 full downloads. Though iTunes U is usually associated with video or audio files 55% of the files offered by TU Graz are of format PDF taking 66% of all downloads. MP4 holds 33% of all files with a download rate of 26%; the rest goes to MOV and MP3 files. Though the number of offered files has not been increased during the last year that much the number of downloads has doubled from 360.000 downloads since autumn 2011 which is an unmistakable argument for the advertising power of iTunes U.

All recording services offered by DSL can be seen in Fig. 3 as far as it has been executed before the project of automated recordings. The left column lists the different recording services beginning with standardized recording services followed by special shootings which are also done by other departments of the ITS.

In case the special shooting targets teaching aspects or takes place in the course of lectures it is assisted by DSL for free. Streaming by DSL comprise any formal and informal streaming activities in this context done by DSL as a department or by any of its members. For informal individual streaming any mobile phone connected to a Web 2.0 online video-sharing portal can do such as Qik, vimeo, Ustream and AudioBoo. In November 2009 DSL (in cooperation with iUNig – Initiative for New Media of Graz Universities, as in [9] – and German EduCamp founder) organized the first EduCamp in Austria. It was

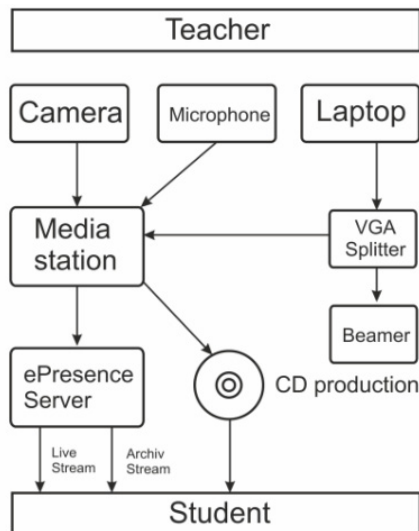


Figure 2. Production process with the ePresence system

the first German speaking, educational camp to be completely broadcasted live on Internet. All needed for that was a set of cameras and microphones each connected to ordinary laptops that have Internet access to the DSL-account on Ustream. The column in the middle covers the platforms in use: TUGTC as main platform for distributing recorded lectures to the students, the so called curry-server for all streaming events and the iTunes U platform for public relational aspects as well as personal accounts at informal Web 2.0 platforms. The right column lists the main interaction possibilities to connect and take part with the content presented and deals with both hardware as well as software components. The connection between the middle and the right columns signifies the multiple ways of interacting with the media and is meant to be bidirectional. Interaction and communication anywhere anytime by anybody apart from simple consummation are qualities of Web 2.0 applications.

B. Automated Recording – Framework of the Project

Fortunately the request for recordings has constantly increased and the service could be benchmarked as highly successful. Due to the fact that a lot of steps of the recording-workflow were serviced manually by DSL it was only a question of time when we hit the limit in doing so. Because of resource bottlenecks the need for a fully automated recording system constantly increased. We therefore decided to go this way. Comparable projects at Austrian universities did not help us much to find an appropriate way for fully automated recording systems. Only a few efforts had been reached so far at that time (autumn 2010). Whether they too were just in their beginnings or the setup did not match with our requirements sufficiently.

As a result of that first glance we have started a project for enabling fully automated recordings of lectures on our own. The project is carried out in cooperation with TU Graz internal departments – the Office for Lifelong Learning (LLL) as well as the Department for Information Design & Media (DIDM) – and with the external team of auphonic, a research group with business partners in the field of audio processing. The project is part of a Lifelong Learning initiative supported by the university itself as well as the Austrian Federal Ministry for Education, Arts and Culture. In general, it aims to investigate the use of

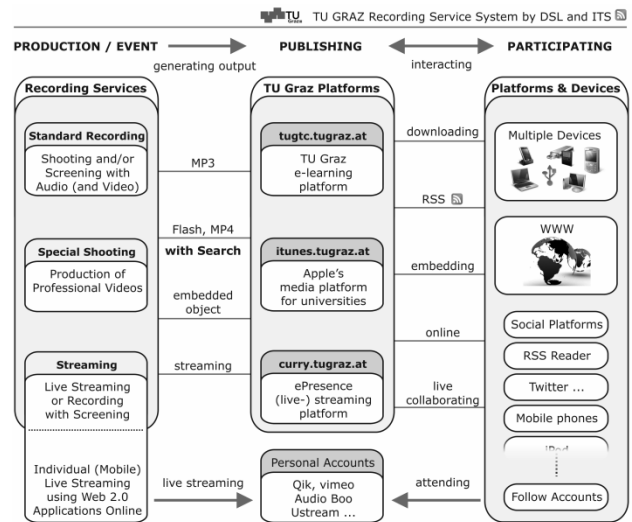


Figure 3. Overall recording services of DSL except automated recording

lecture recordings in the field of adult education as well as advanced trainings. The project will be finished in autumn/winter 2012/13. The first big steps could be reached by enhancing the post-recording procedure with a method that results in a text based search of the video recorded on base of indexing video. A second step was finalized in winter 2011/12; the prototype of the automated recording system installed in one of the most modern lecture halls at TU Graz. Both achievements will be focus of the following paragraph. The final step of developing a customized user friendly interface for managing recording tasks and publishing process needs to be implemented.

Last but not least, it is of highest importance to the success of any service offered to have didactical scenarios come along with the service. In the course of the project a lot of didactical background research and work has been done too. For instance, a series of questionnaires upon teachers and students has been carried to optimize the service. Results of the questionnaires will be argued to complete an overall insight into the project. They underline the fact that providing podcasts is a strongly requested goody indeed, but without didactical settings that take care of the new medium the effective benefit may not be seen instantaneously.

C. Achievements of the Project: Indexing Video

For the first aspect – enhancement of the post-recording procedure – it must be pointed out that there is a demand for providing search functionality for podcasts or so to say, for indexing podcasts. Search functionality on base of text and speech analysis within the podcast can help students to find specific terms, topics, or parts of the lecture instead of browsing more or less randomly the podcast spending valuable time. There are different ways of indexing a video podcast; thumbnail generation, audio transcription, text extraction from video frames, or even user tagging. A couple of these are very time consuming and cannot be carried out fully automatically at all. Since it is the project's aim to have a completely automated recording system installed we have been working on a custom method for extracting text from a screen-capture video as well as thumbnail generation. It was not the first time DSL was engaged into a project of extracting text from a screening. During 2006 and 2008 DSL was partner

of the so called TeleTeachingTool (TTT) project of the University of Trier by Dipl. Inf. Peter Ziewer, as in [13] [14]. It based on Java technology, offered text search functionality on base of XML-files generated with Optical Character Recognition (OCR) software, as well as automated positioning of captions and an own TTT-player. Although TTT had a couple of outstanding efforts it finally had not been further continued due to different reasons (no recording longer than 45 min, possible problems of synchronizing audio and video, problems with Flash converter).

Nevertheless, key technologies for text extraction and indexing are OCR and the capabilities of the Advanced Video Codec (AVC - or H.264-codec). It is thus meaningful to say that our indexing procedure is smart integrated in the main editing/encoding routine of post-processing a recording, the podcast production itself. The main editing/encoding routine consists of a Perl script that automatically carries out most of the tasks of post-processing; coordinating the tools used as well as introducing inherent algorithms for text filtering or frame sorting. Fig. 4 shows the general workflow scheme of the automated routine. The input file (at the top of Fig. 4, called "Recording AVI") is produced with screen-capturing software Camtasia Studio by TechSmith (for Microsoft Windows systems) and iShowU (for Apple Mac systems) delivering high-quality videos encoded with more or less lossless codecs. The audio signal usually is an uncompressed Puls-Code-Modulation (PCM) formed signal with 16 bit, 44100 Hz, mono character. Typical settings for the recorded file are 5-10 fps (frames per second), CQ setting (Constant Quality) 32, 40 kbps audio, resulting in podcasts with a good quality to size ratio. The final video itself is an AVC encoded video with Advanced Audio Codec (AAC) audio in an MP4-container. It is offered within a HTML page using Adobe Flash Player for controlling the play-back as well as featuring a Table of Contents (ToC) and the search functionality for tracking specific terms within the video. By linking the HTML page to TUGTC students can use the service.

The post-processing of the audio file is totally managed on behalf of the cooperation with auphonic. Auphonic offers an API for automated transfer and processing of audio files. The quality of the audio files can be optimized in that way very easily. Before that a homemade processing was in use that had to be started manually for each audio file. In case the service of auphonic is out of order due to network problems or internal ones the original recorded file will be used for further finalizing the processing. A local version of auphonic to be independent from such possible sources of error is already in test. By now the post-processing runs fully automated.

It must be mentioned that the procedure for indexing videos only works with not handwritten or drawings screen-captures. The general recording provides a recording of the lecturer's spoken word using wireless microphones (Sennheiser EW 100 3G) and a screen-capturing of the lecturer's laptop which by default is similar to the projector's output for the students. Very often the presentation is a sequence of prepared slides (Microsoft PowerPoint, Apple Keynote, or comparable software) but not exclusively. In case of handwritten content or drawings by using a tablet's surface the automated indexing will not work.

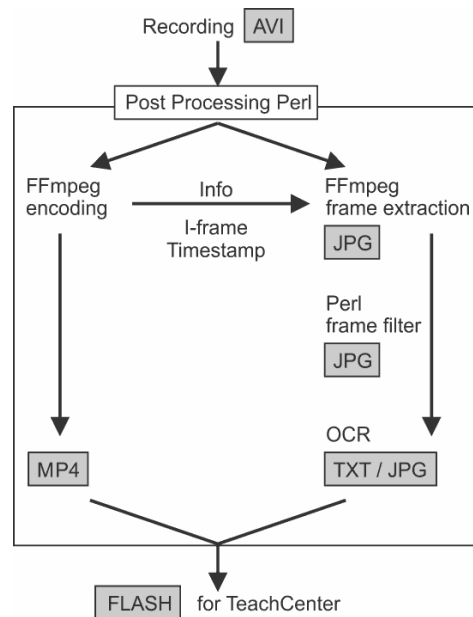


Figure 4. Workflow scheme of the editing/encoding routine

Actually this is a tightrope walk: ready-made presentations can be used for indexing, procedural content that evolves during the lesson held by writing and drawing on a tablet's surface is less useful for indexing but of greater didactical value. Nevertheless, both are of interest to be serviced as best as possible. The following paragraph goes into details about the indexing of video records to understand this problem better.

As mentioned above, OCR refers to methods for extracting text information from image files such as formats TIFF and JPEG. Several OCR programs offer different levels and complexities; there are for example plain text extractors with limited features (usually included with a new scanner), simple online (server-based) OCR tools, but also more elaborate document converters which can extract text in multiple scripts and languages, recognize graphics, and even fully reproduce the layout of a document. Unfortunately, none of open-source or commercial software that can directly extract text from video files was effective enough to be used for our purposes. A custom way of applying OCR had to be found. The solution was to extract frames from videos with the help of another software that results the frames in a form compatible with OCR software (e.g. JPG, PNG, TIFF) and then apply OCR to these image files. On the one hand such software must enable command line for to use scripts but also had to have high-performance due to the fact that extracting one frame per second of a video file and then filtering out frames with identical content consumes very much time and processing power. As a result FFmpeg has been chosen. FFmpeg is open-source versatile audio/video manipulation software including libavcodec, which is the leading audio/video codec library used by popular multimedia players such as MPlayer and VLC media player. FFmpeg has the ability to extract frames from a video file in many different formats and it is a command line tool too. Therefore, for the video/audio encoding, as well as the frame extraction for the indexing procedure, FFmpeg is used; one program for carrying out several tasks. Here are two examples of the usage of FFmpeg according to our tasks:

- Encoding a video file:
`$ ffmpeg -i <inputfile> -ac 1 -ab 40k -vcodec libx264 -fpre <codec_preset> -crf 23 -vstats_file <outputfile>`
 -i: name of the input video file
 -ac: number of audio channels
 -ab: audio bitrate
 -vcodec: video codec library
 -crf: constant rate factor
 -vstats_file: generation of -vstats file
- Extracting a specific frame from a video file:
`$ ffmpeg -ss <offset> -i <inputfile> -an -vframes 1 -qscale 1 <outputfile>`
 -ss offset: (time of frame to be extracted) in seconds
 -an: no audio
 -vframes: number of consequent frames to extract
 -qscale: quality factor (1[best] to 31[worst])
 Be sure that the output file ends with “.jpg”.

After extracting frames from the recorded video with the help of FFmpeg the frames need to be filtered according to their “useful information”. By term of “useful information” we mean a rate of difference between a frame (f[n]) and its preceding or following frame (f[n-1] or f[n+1]). The frames that need to be extracted are those with significant change in comparison to their preceding frames content. It must be outlined that a video with duration of 90 minutes and a frame rate of 5 fps contains 27.000 frames. To extract and apply OCR to all of them would be highly impractically. Fortunately, only a small fraction of the total frames contains useful information and is therefore actually needed, since most of them are identical or differ inconsequential. To select frames of interest modern MPEG-4 video codecs accommodates with our aims. Video files consist of a series of images; displaying them in sequence at a constant rate, the effect of motion is achieved. Due to the large number of images that a video file usually contains, its size can become extremely large. In order to make video files smaller, video codecs with algorithms for reducing the information without severely degrading the quality have been developed. These codecs take advantage of the fact that a great amount of information between neighboring frames is identical or very similar; the general idea of such algorithms is to discard the repetitive information while keeping the essential content. In the family of MPEG-4 codecs the Group of Pictures (GOP) model was introduced. These codecs produce video files with different types of frames: the I-frames, which are frames that keep their full initial information intact, as well as P and B frames, which are frames that contain only a part of their initial content, subject to their preceding and possibly following I-frames for their correct reproduction. Fig. 5 points out the principle and Fig. 6 a detailed example (compare types of the frames listed). In the process of encoding a video file these codecs compare the content of consecutive frames to detect significant alterations and, in that manner, decide which frames should become I-frames. It is this feature that is exploited in order to extract only the useful frames for OCR and indexing.

Remember, FFmpeg features the option “-vstats_file” (see example “Encoding a video file”) that allows the dumping of encoding information (such as frame types, size, timestamps etc.) in a text file during the encoding

process of a video file by the use of AVC. After the encoding is completed, this text file contains all the needed information for extracting the frames which are useful for OCR and indexing. Since a maximum distance between two I-frames needs to be set, it may occur that I-frames are identical when this distance is reached a new I-frame is generated regardless of its content. But there is still the possibility of comparing the sizes of the I-frames to determine whether they are actually different from each other. This comparison is made with Perl and is one of the custom algorithms that are used in the indexing procedure.

Another one is the search for I-frames that are too close to one another by comparing their timestamps; if for example a lecturer searches for a specific slide going through several slides in a relatively fast tempo like zapping TV channels, only the last I-frame will be extracted, since the others are not useful for indexing. The procedure can be described as follows.

The frames can be thought of as a sequence:

$$\dots, f[n-1], f[n], f[n+1], \dots (1)$$

IF

$$|fs[n-1] - fs[n]| < S \quad (2)$$

OR

$$ft[n] - ft[n-1] < T (3)$$

THEN

$$\text{discard the current frame } f[n] \quad (4)$$

with

n: number of the frame

fs: size in bytes

ft: time in ms

S: deviation parameter for the size

T: deviation parameter for the time

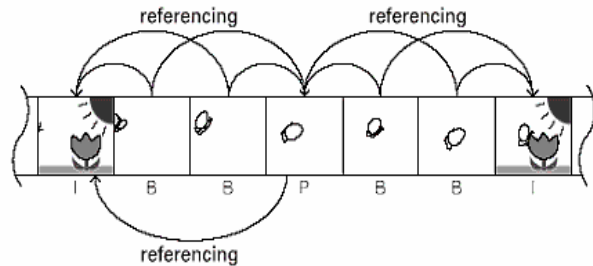


Figure 5. A Group of Pictures (GOP)

frame= 61	q= 31.0	f_size= 21139	s_size= 21kB	time= 4.067	br= 2536.7	kbits/s	avg_br= 41.6	kbits/s	type= I
frame= 62	q= 31.0	f_size= 642	s_size= 21kB	time= 4.133	br= 77.0	kbits/s	avg_br= 42.2	kbits/s	type= P
frame= 63	q= 31.0	f_size= 96	s_size= 21kB	time= 4.200	br= 11.5	kbits/s	avg_br= 41.7	kbits/s	type= B
frame= 64	q= 31.0	f_size= 99	s_size= 21kB	time= 4.267	br= 11.9	kbits/s	avg_br= 41.2	kbits/s	type= B
frame= 65	q= 31.0	f_size= 100	s_size= 22kB	time= 4.333	br= 12.0	kbits/s	avg_br= 40.8	kbits/s	type= B
frame= 66	q= 31.0	f_size= 104	s_size= 22kB	time= 4.400	br= 12.5	kbits/s	avg_br= 40.3	kbits/s	type= B
frame= 67	q= 31.0	f_size= 103	s_size= 22kB	time= 4.467	br= 12.4	kbits/s	avg_br= 39.9	kbits/s	type= B
frame= 68	q= 31.0	f_size= 101	s_size= 22kB	time= 4.533	br= 12.1	kbits/s	avg_br= 39.5	kbits/s	type= B
frame= 69	q= 31.0	f_size= 94	s_size= 22kB	time= 4.600	br= 11.3	kbits/s	avg_br= 39.1	kbits/s	type= B
frame= 70	q= 31.0	f_size= 83	s_size= 22kB	time= 4.667	br= 10.0	kbits/s	avg_br= 38.7	kbits/s	type= B
frame= 71	q= 31.0	f_size= 69	s_size= 22kB	time= 4.733	br= 8.3	kbits/s	avg_br= 38.2	kbits/s	type= B
frame= 73	q= 31.0	f_size= 68	s_size= 22kB	time= 4.867	br= 8.2	kbits/s	avg_br= 37.4	kbits/s	type= B
frame= 74	q= 31.0	f_size= 67	s_size= 22kB	time= 4.933	br= 8.0	kbits/s	avg_br= 37.0	kbits/s	type= B
frame= 75	q= 31.0	f_size= 71	s_size= 22kB	time= 5.000	br= 8.5	kbits/s	avg_br= 36.6	kbits/s	type= B
frame= 76	q= 31.0	f_size= 71	s_size= 22kB	time= 5.067	br= 8.5	kbits/s	avg_br= 36.3	kbits/s	type= B

Figure 6. A frame sequence record

By the way, the timing information (essential for indexing) is obtained by the “-vstats” option of FFmpeg too. During frame extraction it delivers information about the encoded frames which includes timestamps. The timestamp of each extracted frame is saved within the name of the image file created. For example: “120126_432002_fickert_0324400.jpg” with “0324400” is the timestamp in msec.

Fig. 7 and fig. 8 show the result of a recording post-processing file offered to the students via TUGTC including search functionality. The recording includes screening the lecturer’s laptop (main colored part) as well as shooting the blackboard activity with a camera (PIP). On the left side of fig. 7 (black area) the navigation as well as the search functionality as a result of the video indexing process can be seen. Fig. 8 zooms in on a detail of fig. 7, the search results. The German search term in this example was “Leitung”, the search results are marked turquoise. Furthermore the search text field as well as the name of the file can be seen.

A future version of the offered output media file will be embedded in a HTML5 environment enabling a lot of additional possibilities. First design drafts just have been started.

D. Achievements of the Project: Setup of Prototype

The second main achievement of the project so far was to build a prototype of the system’s total setup at one of the most modern lecture halls of TU Graz. The challenge was to implement new components to the already existing infrastructure of the well-equipped lecture hall. Furthermore, the setup should be as easy to integrate into other lecture halls as possible without big investments. We decided for a server-sided solution at low cost of commercial products.

The general setup of the hardware can be seen in Fig. 9. The hardware setup of the recording system at the lecture hall consists of the lecture recorder X2 from Epiphan Systems Inc, a SANYO network camera and a custom touch interface to control the recording (Crestron media control panel). The media control panel allows controlling the lecture recorder X2 via RS-232 API. Currently a button to start and stop a record is implemented. By default the media control panel only works with a hardware key that is given to lecturers. Additionally and due to data privacy protection the record function can be set password protected.

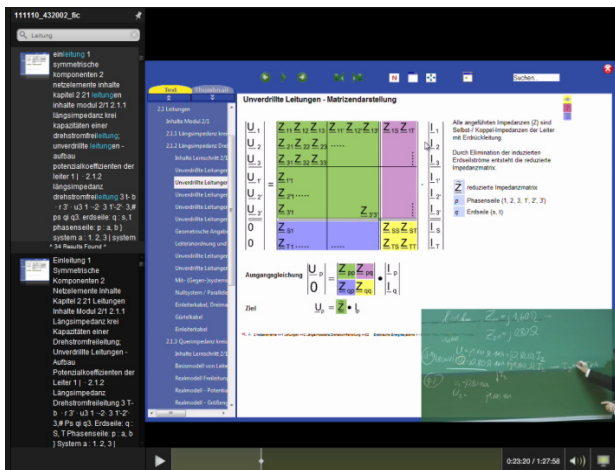


Figure 7. Example of recording with search functionality on the left

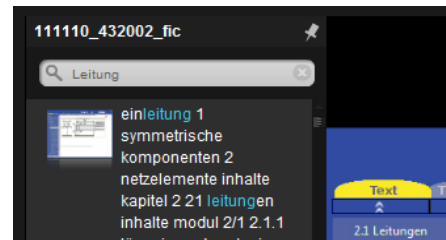


Figure 8. Zoom of Fig. 7 showing the search results marked

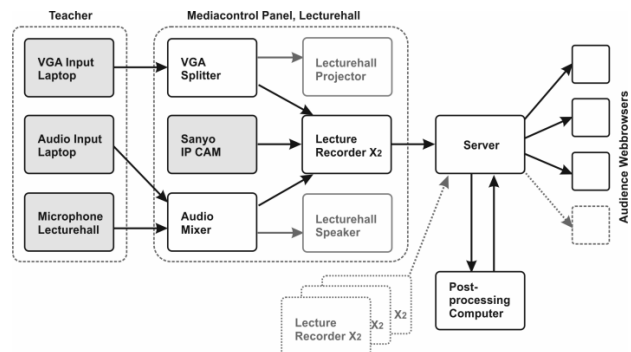


Figure 9. Hardware setup for server sided recording, input hardware is marked grey

All inputs to the system are standard cinch and VGA jacks. The lecture recorder device was integrated into the existing technical equipment of the lecture hall, using the audio input from the mixer and the VGA signal connected to the lecture hall projector. It is capable of recording or streaming a synchronized signal containing audio, a screen capture from VGA, DVI or HDMI, as well as a camera input which can be either a direct S-video or composite source or, as in this case, a remote network camera. The audio signals are amplified and mixed in the audio installations of the lecture hall. The output volume of the audio channels can be adjusted with the media control panel by the teacher. The external camera can be either a network camera or a camera that is directly connected to the lecture recorder X2 via S-Video. For the described setup, the SD video of the lecturer comes from a network-camera; the S-Video channel is not used. The incoming VGA signal is split and amplified, and then it goes to the projector as well as the lecture recorder. Due to restrictions of the projector the resolution of the laptop must not exceed 1280x960. To remove this limitation for the recording system it is planned to detect and scale the resolution of the laptop to the appropriate size for the projector (1280x960) by a scaler, while the lecture recorder is able to perform a real time scaling of the incoming source resolution on its own. VGA will be then scaled up to HD 1080p (1920x1080).

The lecture recorder is able to stream and record at the same time. The stream/record consists of two independent streams and an audio signal. The audio signal is a single channel mix-up from the audio mixer, while the SD channel is fed by the SANYO IPCam and the HD channel by the laptop video signal. HD, SD streams and audio are saved to separate tracks in a multi-track AVI file. Such AVI files are automatically forwarded to a FTP-server, where the data is stored for automatic post-processing and distributing. For the prototype actually the post-processing is done by an automated script running on an iMac. For the actual running system it is planned that the post-processing too should be done on the streaming server,

which delivers the video-on-demand streams and also the live streams. Furthermore it will be able to operate recordings from several lecture halls concurrently. This hardware setup allows a large scale implementation of an automated recording system.

For the user the workflow is as simple as possible: The teacher connects a laptop as usually to the system via VGA (video for projector) and cinch (audio by wireless microphone); additionally a lecture hall microphone may be used according to the size of the lecture hall. After that only one click on a “record”-button offered on the lecture room’s media control panel is needed to start (and stop) the recording. During all time the teacher can see a live-feed of the SANYO IPcam on the display. The rest is done automated up to the step of publishing the final video on TUGTC, which will be the final effort to do. By choice the recording may be live-streamed too in the near future.

E. Further Steps and Aspects of the Project

First recordings have shown that the system’s setup installed worked in principle. An AVI-container including two video files (codec: h.264) and one audio file (codec: pcm_s16le) is sent to the server and then post-processed by the script discussed before. Further adjustments and optimizations affect the scaling problem of the VGA signal and the control panel. Fig. 10 is an excerpt of metadata sent from X2 to the server during the test phase.

Fig. 11 shows the result of an automated recorded regular lesson. On the left side you see the video of the SANYO IPcam which shooting in a fixed frame position with no automated motion tracking. Above that video the title of the recording and the name of the lecturer are displayed. In Fig. 11 both as well as the head of the lecturer have been covered with bars for privacy reasons. On the right side there is the VGA signal video of the screening of the lecturer’s laptop with opportunities to switch between four different view styles as well as full screen modus. At the bottom of the frame there are player functionalities such as start/stop and volume control. Furthermore a field for text search is realized as well. The search will base on the outcome of the OCR analysis during the post-processing.

Next, a connection between the curry-server and the Epiphan-server has to be accomplished for better multi-casting of the (live-) streaming.

To sum up, besides technical aspects the implementation of such a new service and infrastructure leads also to new service levels the university must address in general. Automated recording of lecture has to be conducted by software to guarantee a solid management in advance. Teachers would like to mark their lessons online as to be recorded. For any lecture hall of the TU Graz is connected to the TU Graz information and administration system called TUGRAZonline it is planned to install a kind of online booking service for automated recordings. Teachers then will be able to assign for a recording using the calendar functionalities in combination with the lecture hall options on TUGRAZonline. The recording system therefore will “know” which lecture is being recorded and sends this additional information to the recording server. Fig. 12 shows first preliminary thoughts for the architecture of such administration software. In the middle of the map the server is visualized that manage all the tasks concerning a full automated process for the whole university.

```
Input #0, avi, from 'video_Feb27_15-26-34.avi':
Metadata:
encoder      : Lavf52.32.0
Duration: 00:01:56.33, start: 0.000000, bitrate: 2448 kb/s
Stream #0.0: Video: h264, yuvj420p, 1280x800 [PAR 1:1 DAR 8:5], 30 fps, 30 tbr, 30 tbn, 30 tbc
Stream #0.1: Audio: pcm_s16le, 22050 Hz, 2 channels, s16, 705 kb/s
Stream #0.2: Video: h264, yuvj420p, 720x576 [PAR 1:1 DAR 5:4], 30 fps, 30 tbr, 30 tbn, 30 tbc
[STREAM]
index=0
codec_name=h264
codec_long_name=H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10
codec_type=video
codec_time_base=1/30
codec_tag_string=H264
codec_tag=0x34363248
width=1280
height=800
has_b_frames=0
sample_aspect_ratio=1:1
display_aspect_ratio=8:5
pix_fmt=yuvj420p
r_frame_rate=30/1
avg_frame_rate=30/1
time_base=1/30
start_time=0:00:00.000000
duration=0:01:56.333333
nb_frames=3490
[/STREAM]
[STREAM]
index=1
codec_name=pcm_s16le
codec_long_name=PCM signed 16-bit little-endian
codec_type=audio
codec_time_base=0/1
codec_tag_string=[1][0][0][0]
codec_tag=0x0001
sample_rate=22.050 KHz
channels=2
bits_per_sample=16
r_frame_rate=0/0
avg_frame_rate=30/1
```

Figure 10. Excerpt of metadata sent from recording X2 to processing server



Figure 11. Screenshot of an automated recorded lecture of a regular lecture held; be sure that the head of the lecturer (on the left side) and his name have been covered with a bar for privacy reasons

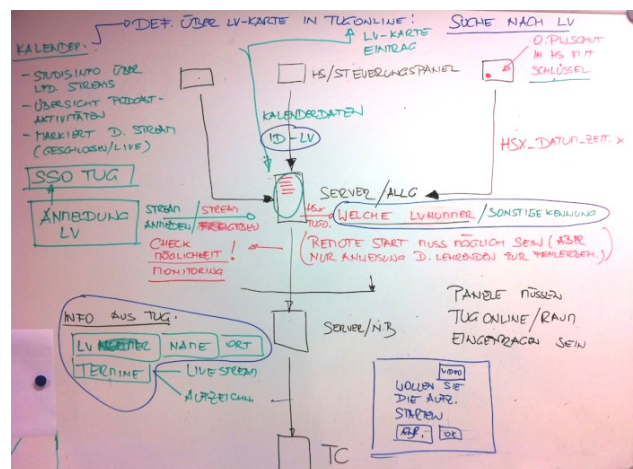


Figure 12. Brainstorming including architecture and tasks of a university wide automated recording management system

Another very important component is the legal aspect of automated recordings of lectures. We have to make sure that anybody being recorded (teachers as well as students) knows that a recording is going on and has given an affirmation to it. A discussion is going on about the phrasing of this affirmation as well as the point of time and place and kind of teacher's and student's affirming. Especially according to live streaming events the legal correctness must be taken carefully into account. Teachers must be clarified very well about copyright and distributing aspects of recorded lecture content too. Nevertheless, this has been subject to any kind of recording in the past too but had been done face to face in the course of personal assistance by the team of DSL.

III. CONCLUSION

This paper outlines the consequent development and enhancement of recording services at TU Graz to become an integrated part of didactical methods for teaching and learning purposes; from its beginning to the ongoing project aiming to automate the entire workflow. One essential step forward has been achieved by implementing search functionality to the output videos offered to the students. Other disciplines for indexing are planned to be researched and finally integrated in the workflow. One of these disciplines is speech-to-text applications or the introduction of interactive elements like user tagging. Speech-to-text is a rapidly advancing field which has lately shown positive signs, when a few years ago it seemed unusable or very hard to realize with a lot of training the system. As for user tagging, it is a way of having the automatically generated index controlled, enriched, and even corrected by the users themselves, which underlines the power of the Web 2.0 phenomenon and helps e-learning to constantly evolve, enhancing the quality of the learning content too. A second big step has been reached by a prototype for an automated recording system installed in a modern lecture room of TU Graz. First results underline the setup chosen. On base of the presented prototype six further lecture halls just have been enhanced by beginning of winter term 2012. Even more, these lecture halls are connected to each other in the way that current lectures held in one room can also be streamed and displayed via data projector to the other rooms. This is a very big step forward since a couple of lectures at TU Graz are attended by more students than the lecture halls offer seats. On demand the recording can be streamed to other lecture halls by easily using the control panel by the teacher.

It can be concluded that lecture recordings are an important step towards more quality in education due to the fact that it is a proven worthwhile help to students' learning process. The automation of the process preferentially in big lecture halls will help to increase the number of recordings but to keep personal and financeable aspects affordable at once. Also new didactical aspects shall be addressed by e.g. upgrading some face-to-face teaching through exclusively live-streaming for to gain more interactive character even in mass education.

ACKNOWLEDGMENT

We express our gratitude to the Department for Information Design & Media at TU Graz for their support during the installation of the hard- and software in the

lecture hall and their patience within manifold test scenarios. We are equally indebted to the Department of Life-Long Learning for supporting our ideas and helping us to work on the future of education.

REFERENCES

- [1] A. Almonte, and K. Gilroy, "Podcast for Learning," Cambridge, 2005, MA: The Otter Group.
- [2] C. Dale, "Strategies for Using Podcasting to Support Student Learning," *Journal of Hospitality, Leisure, Sport and Tourism Education*, vol. 6(1), pp. 49-57, 2007. <http://dx.doi.org/10.3794/johlste.61.155>
- [3] D. Helic, H. Maurer, and N. Scerbakov, "Knowledge Transfer Processes in a Modern WBT System," *Journal of Network and Computer Applications*, vol. 27(3), 2004, pp.163-190. <http://dx.doi.org/10.1016/j.jnca.2003.09.001>
- [4] G. Campbell, "There's Something in the Air - Podcasting in Education," *Educause*, vol. 40(6), pp. 32-47, 2005.
- [5] H. Maurer and N. Scerbakov, "Multimedia Authoring for Presentation and Education: The Official Guide to HM-Card", Bonn: Addison-Wesley. February 1996, pp. 250.
- [6] M. Blaisdell, "Academic MP3s: Is it time yet?," *Educause* 2006.
- [7] M. Ebner, and W. Nagler, "Lifelong learning and Doctoral Studies - Facilitation with Podcasting Techniques," *Computers in Education, Conference Proceeding IV, MIPRO 2007, Opatija, Croatia*, pp. 280-283.
- [8] M. Ebner, W. Nagler, and A. Saranti, "TU Graz goes Podcast," *Micromedia and Corporate Learning - International Microlearning Conference 2007, Innsbruck*, pp. 221-233.
- [9] M. Ebner, W. Nagler, and A. Saranti, "Lifetime Podcasting," *Graz University of Technology*, 2008, Graz, Austria
- [10] M. Schiefner, "Audio Distance Learning Konzeption und Erstellung eines Audio-Kurses zum Thema «Konstruktivismus und Lernen»," Unpublished Master Thesis, 2004, University Saarbrücken.
- [11] N. Townend, "Podcasting in Higher Education," *Media Onlinefocus 22, British Universities Film & Video Council*, 2005.
- [12] P. Edirisingha and G. Salmon, "Pedagogical models for podcasts in higher education," *EDEN Annual Conference 2007, Naples*.
- [13] P. Ziewer, and H. Seidl, "Transparent TeleTeaching," *Proceedings of ASCILITE 2002, Auckland, NZ*.
- [14] P. Ziewer, M. Ebner, C. Safran, and W. Slany, "Searching for visual content in electronic lectures," *Proceedings: IEEE International Symposium on Multimedia 2007 (ISM2007)*.
- [15] R. M. Baecker, D. Fono, and P. Wolf, "Towards a video collaborative," In (Eds. Goldman R., Pea R., Barron B. & Derry S.), *Video Research in the Learning Sciences*, 2007, Mahwah, NJ: Laurence Erlbaum, pp.461-478.
- [16] R. M. Baecker, "A principled design for scalable internet visual communications with rich media, interactivity, and structured archives," *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research, October 06-09, 2003, Toronto, Ontario, Canada*, pp.16-29.
- [17] T. Dietinger and H. Maurer, "GENTLE - General Network Training and Learning Environment," *ED MEDIA 1998 / ED-Telecom 1998, Freiburg, Germany*, pp. 274-280.
- [18] T. Y. Huann and M. K. Thong, "Audioblogging and Podcasting in Education," 2006, *Edublog.net*, Available from: <http://edublog.net/astinus/mt/files/docs/Literature%20Review%20on%20audioblogging%20and%20podcasting.pdf> [Accessed April 17th 2012]
- [19] T. O'Reilly, "Web 2.0: Stuck on a name or hooked on value?" *Dr. Dobbs Journal*, 2006, 31(7), pp. 10-10.
- [20] W. Nagler, A. Saranti, and M. Ebner, "Podcasting at TU Graz - How to Implement Podcasting as a Didactical Method for Teaching and Learning Purposes at a University of Technology," *ED-MEDIA, 2008, Vienna*, pp. 3858-3863.

PAPER
AUTOMATED PODCASTING SYSTEM FOR UNIVERSITIES

AUTHORS

Y. Grigoriadis is with the Department for Social Learning, Graz University of Technology, Graz, Austria, (e-mail: ypatios@gmail.com).

C. Stickel was with the Department for Social Learning, Graz University of Technology, Graz, Austria. He is now self-employed (e-mail: stickel@tugraz.at).

W. Nagler (Mag.rer.nat.) is with the Department for Social Learning, Graz University of Technology, Graz, Austria, (e-mail: walther.nagler@tugraz.at).

M. Ebner (Dipl.-Ing. Dr.techn. Univ.-Doz.) is with the Department for Social Learning, Graz University of

Technology, Graz, Austria. He is head of this Department since its founding in 2006, (e-mail: martin.ebner@tugraz.at).

M. Schön (Dipl.-Päd.) is with the Office for Lifelong Learning, Graz University of Technology, Graz, Austria, (e-mail: martin.schoen@tugraz.at).

This work was done in assistance with the Department for Information Design & Media of the ITS of TU Graz. It is an extended and modified version of a paper presented at the International Conference on Interactive Collaborative Learning (ICL2012), held 26 - 28 September 2012, in Villach, Austria. Received 15 November 2012. Published as resubmitted by the authors 27 February 2013