EDUCATIONAL VIDEO

Case Häme University of Applied Sciences Riihimäki campus

ABSTRACT

The aim of the thesis was to provide a comprehensive view to educational video and its functional parts. Topic was further narrowed to apply for Häme University of Applied Sciences Riihimäki Campus and the degree programmes operating out of the campus. Basis of the thesis was to gain knowledge and practical workflow, with descriptions of the tools, to create education videos. This thesis presents two basic scenarios and workflows for creating videos and video content.

Thesis includes demo system for delivering the video and adapting it to different user end devices. Thesis also explains different video content creation methods at use at HAMK campus. Methodology used was design science based process. Thesis describes common methods and techniques used to produce educational video. Thesis also describes application of common methods applied to specific infrastructure of software and hardware. Final part of the thesis describes a delivery process and concludes thesis with common workflow for both prerecorded video and live broadcasted video. Result of thesis is presented at conclusion as an decision tree that guides the production of educational video and forms a basis for workflows on both basic scenarios.

Key words: Educational video, video production, streaming

JÄRVENPÄÄ, TEEMU:	Koulutusvideot, Case Hämeen ammattikorkeakoulu Riihimäen kampus

Tiivistelmä

___

Opinnäytetyön tarkoituksena oli laatia näkökulma opetusvideon tuotantoon ja sen toiminnallisiin osiin. Opinnäytetyön aihe rajattiin kattamaan Hämeen Ammattikorkeakoulun Riihimäen kampuksella toimivia opinto-ohjelmia. Työn keskittyy antamaan tietoa ja työkaluja opetusvideon tuottamiseen ja muodostamaan videotuotantoa varten työnkulun. Opinnäytetyö esittelee kaksi erilaista työnkulkua jotka pohjautuvat erityyppisiin jakelupolkuihin.

Opinnäytetyö esittelee myös demo ympäristön suoratoistettavan videon jakeluun, joka pystyy mukautumaan käyttäjän päätelaitteeseen ja ympäristöön. Opinnäytetyössä käsitellään HAMK Riihimäen kampuksella käytössä olevat videon luontiin ja jakeluun tähtäävät laitteet ja käytänteet. Opinnäytetyön tekemiseen käytetty metodologia toteuttaa Design science mukaista tutkimusmenetelmää. Opinnäytetyössä esitetään päätöspuu, joka tähtää tutkimusvaiheessa esitettyjen komponenttien käyttämiseen työnkulussa. Tämä muodostaa lähtökohdan kummankin jakelupolun tarvitsemalle työkululle.

Asiasanat: Opetusvideo, videotuotanto, suoratoisto

TABLE OF CONTENT

1   INTRODUCTION

For higher education, world is changing. Demand is to have more graduating students in shorter timeframe. This demand presents itself on methods to give out education. One of these is to use video based platforms. Using video as a medium for education is one of the essential components of most online courses offered. This presents a question of how to make these videos and what type of tools to use.

1.1   The subject of the thesis

The Häme University of Applied Sciences (later HAMK) is one of the applied science universities in Finland and is currently undergoing changes in its strategy and methodology how teaching and learning are done. University has seven campuses with around 625 faculty members and around 7200 students. University offers education in 31 degree programs in multiple fields of study. (HAMK 2017.)

HAMK has devised a strategy for how to offer higher education. Strategy has been planned to reach for 2020. This strategy specifies three concepts for studies. These are 8-16, 24/7 and 18-100. In short, the main concept of 8-16 offers a work week for the student equivalent to working at a traditional job. Idea is to offer constant guidance and help student gain credits weekly. This forms a nucleus for education offering. To help the student reach degree earlier, 24/7 concept comes into play. This offers more independent studies and takes into consideration of learning in the workplace. Last concept 18-100, is a nod towards the ideology of life-long studies. This concept is meant for students that can apply their everyday work for their studies and education. (Häme University of applied sciences Ltd. 2015.)

HAMK strategic concepts are driven forward using concepts that Siemens et al. (2015) explain in their paper as 'the process of 'stitching' together distributed interactions' with learners who control their preferred toolsets. Video in education is one element of those 'distributed interactions' and

the role that video plays in education, and how that role develops will be explored further. (Siemens et al. 2015, 206.)

Focus of this thesis is on providing workflow schemas for video production and delivery. These schemas will support strategy defined concepts and provide basic instructions with procedures to publish lectures in video format. This thesis will answer basic questions like how to produce a recording of the lecture, how to deliver this recording to students and what are the practices used at HAMK Riihimäki campus. The thesis will discuss video delivery from the point of view of the single degree programme but should be applicable to any other degree programme at HAMK.

In point of the thesis, the video is a digitally recorded content. Content that can be stored or delivered live. Video, that may or may not have the lecturer visible and can vary from computer screen capture to using multiple audiovisual devices to capture an event.

Thesis research methodology will follow design science as a framework. This allow thesis to grow on top of already reached material, and apply it to real world target.

The structure of this thesis will be as follows. The thesis starts with a basic explanation of video usage in higher education, specifies this with practices used at HAMK Riihimäki and continues with explanations of used techniques. These will form a theoretical basis for the thesis. Next part will deal with actual workflow examples with basic cases of lecture capture. This part will also be used to demonstrate and explain the usage of video production studios, designed for HAMK Riihimäki specifically to provide workspace for lecturers preparing and giving video-based lectures for distance learners. Finally, the thesis will explain a case of delivery of complex video production using open source tools and platforms available through HAMK infrastructure. Conclusions and future development topic discussion concludes thesis.

1.2   Environment description

Work for this thesis has been completed at HAMK University of Applied Sciences campus at Riihimäki. The campus provides basic facilities such as lecture rooms and halls. Basics today include internet connection as well as computer infrastructure. The environment could be described as traditional classroom teaching environment, where until very recently lecture capture did not happen. Against this physical environment, the recent development of consumption of video on the internet is many magnitudes larger. This is showcased in many different venues, but YouTube presents this in their statistics:

> *"and everyday people watch hundreds of millions of hours on YouTube and generate billions of views." (Youtube 2016)*

And

> *"Once users are on YouTube, they are spending more time per session watching videos. On mobile, the average viewing session is now more than 40 minutes." (Youtube 2016)*

YouTube statistics represent near linear correspondence with Cisco forecast of global IP video traffic growth rate. Cisco forecasts that IP video traffic will grow 3-fold from 2015 to 2020, a compound annual growth rate of 26%. And that consumer IP video traffic will be 85% of consumer IP traffic in 2020, up from 77% in 2015. (Cisco 2016.)

There is a report made by Kaltura, a video technology provider, called "State of Video in Education" report. This report states that video is permeating educational institutions and that harnessing video will help to achieve improved outcomes by more effectively transferring knowledge. And continue to explain that video is one of the key aspects towards digital literacy, giving out considerable benefits to educational institutions: increased knowledge retention and improved learning outcomes. (Kaltura 2016.)

Over the last few years, video in education has changed dramatically; from how video is consumed to how it is created. This represents a change for the lecturers as well as for the students. For the teachers using analog video, from celluloid tape or VHS tape format. Presenting video meant preparation and planning. Finding a method to procure the video. And to ensuring lecture hall or classroom had adequate equipment available. Figure 1 shows a timeline adapted from Greenberg & Zanetis (2012).
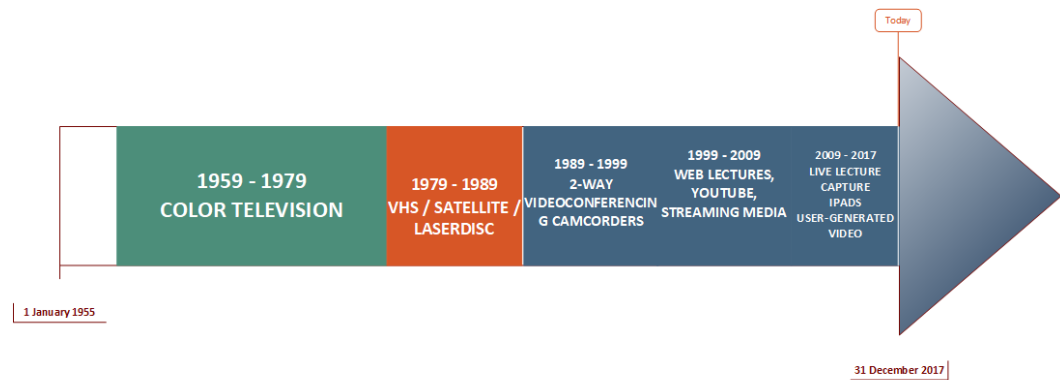
FIGURE 1. Timeline of video technology in education

As timeline shows, using video has become easier and at the same time more involved for the learner. From point of view of the user, this is a welcome change, making delivery of videos and materials contained in videos easier. Pattern represented here is parallel in the timeline of technological advances. Most of the video production technologies have progressed from technologies available only to professionals to technologies available to anyone. Change is seen in form of mobile phones or other hand-held consumer devices. This evolution can be seen in demand and in the availability of the video content. It makes content creation an everyday occurrence and as timeline shows, the emphasis is moving toward live production and delivery with a recording made available.

1.3 Research question – thesis topic

This research aims to find out answers to the following questions:

1. What is the workflow to produce quality educational video and what are the requirements of such a video?
2. How to produce quality video using equipment and facilities offered at HAMK Riihimäki campus?
3. What are the delivery options for educational video available at HAMK Riihimäki and considerations to use for each of these?

Finding the answers to these questions, this thesis suggests a workflow model for each of the video production cases. Selection between workflow options is done by evaluating the demands of the video and its projected audience.

1.4 Methodology

Thesis was based on framework of design science research methodology. Design science research methodology (DSRM) is a Design science research process model where there are five steps to each process. These steps are show on Figure 2. This thesis follows these steps loosely by combining proposal and tentative design steps. Then performing a development and evaluation phase. Conclusions are presented as logical steps combining knowledge from artifact steps. (Vaishnavi & Kuechler 2015)
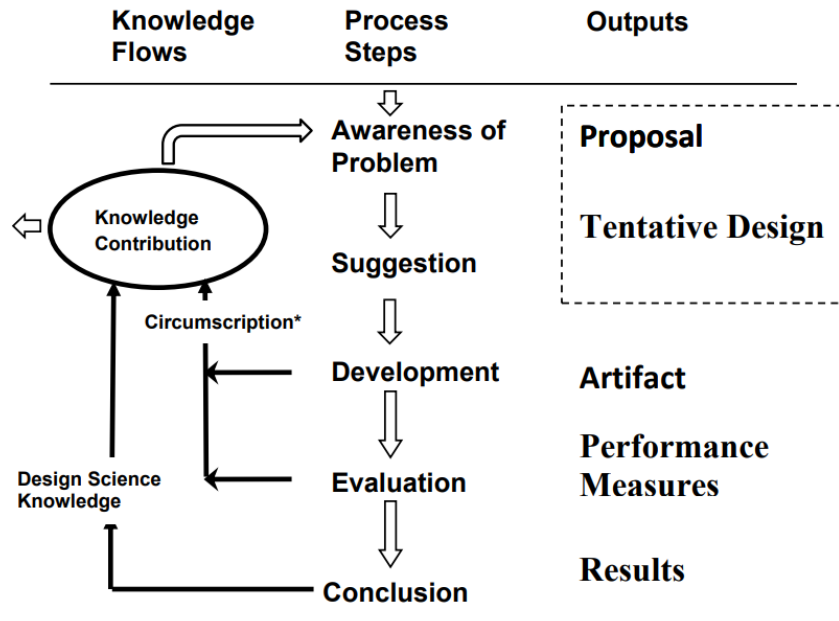
FIGURE 2. Design science research process model (DSR Cycle) (Vaishnavi & Kuechler 2015)

Design science research is a process where knowledge is accumulated through action (Vaishnavi & Kuechler 2015). This methodology describes process of this thesis accurately because most of the processes described on this thesis are accumulations of experimentations made during the writing of thesis.

Thesis can be divided on three parts. First part or phase of the thesis is knowledge gathering and technology overview. Second part of the thesis is DSR artifact phase where current systems are explained. Third and last part of the thesis is evaluation and conclusion phase where workflow steps and considerations are presented.

2   TECHNOLOGY OVERVIEW

Making lecture capture or event capture needs some varied technologies. These can be divided into categories, by their purpose or by functionality. Making the lecture capture can be done with multiple different methods or styles. Hansch et al. (2015) defines 18 production styles, represented in Figure 3. These include presentation slides with voiceover, screencast, classroom lecture and live video, that are main styles of lecture capture used at HAMK Riihimäki. The technology used in these is all based on making a compressed video in a specific format and then delivering it to students with either video-on-demand method or live. Delivery methods vary per capture method, the number of used cameras or desired method of subject presentation. This overview will present basic video and audio technologies and their starting points. On capture devices, delivery options and end-user players, there are large variety of choices. This thesis is only presenting those selected for HAMK usage by HAMK ICT department and education platform development working group within HAMK. This limitation is necessary because multitude of options available to everyday user.
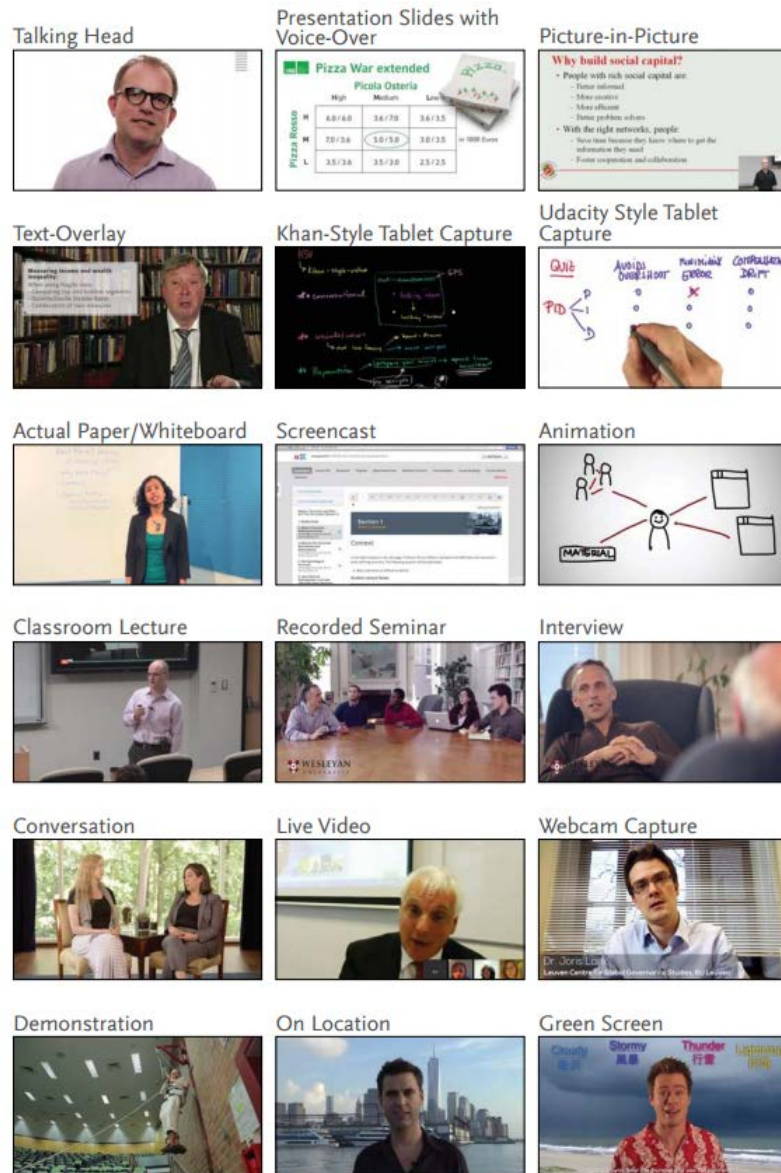
FIGURE 3. Production styles (Hansch et al. 2015)

For making effective educational videos, there are three elements, according to multiple studies (e.g. Kay 2012; Allen and Smith 2012; Lloyd and Robertson 2012; Rackaway 2012; Hsin and Cigas 2013), that instructor has to consider. These elements consist of cognitive load, engagement, and active learning. For making any educational materials, including video, one of the primary considerations is a cognitive load. Cognitive load theory (Sweller et al. 1994) suggests that memory has several components. Components include verbal or auditory channel

component and visual or pictorial channel component. These components together form channels for information acquisition and processing. According to Mayer & Moreno (2003), these can be described as a visual channel and auditory channel for pictorial and verbal processing. These components are presented in Figure 4.
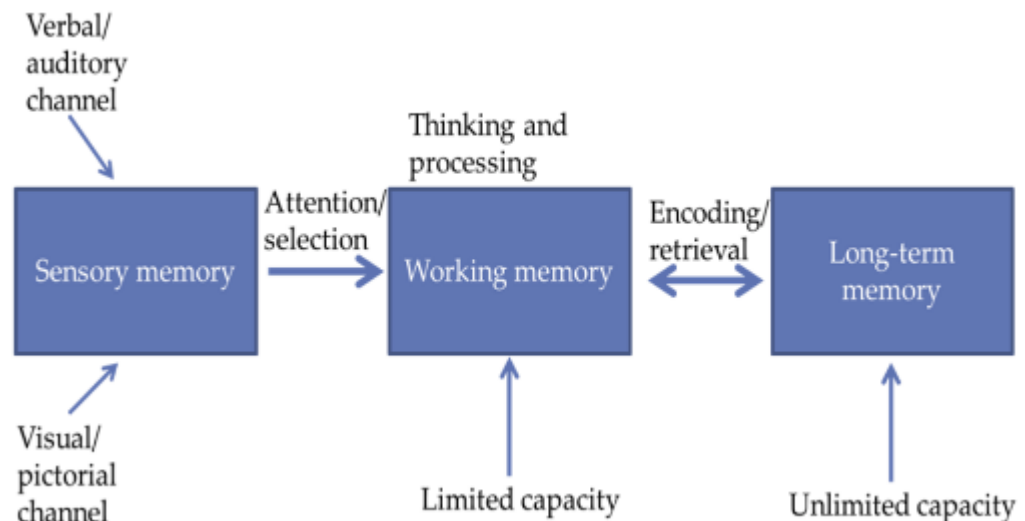
FIGURE 4. Sensory channels (Mayer 2005) and (Mayer & Moreno 2003)

To fully utilize these memory channels, instructor or educator needs to make use of four practices. These practices are signaling: to use on-screen text or symbols to highlight important information, segmentation: to make information chunks small enough to enable the learner to engage with the information. Weeding is a practice to eliminate interesting but extraneous information from the video and matching modality is a process to make both visual and audio channel carry new information. These practices have their basis made in multiple different studies. Mayer and Johansson (2008) explain the merits of the signaling in their study. Their study shows how redundant information in the sensory channel can increase the results of retention of studied material. Signaling strongly reflects the nature of online video or multimedia-based learning and should be considered one of the prime habits to have when producing the educational video.

Guo et al. (2014) make a strong correlation between length of video and student engagement. This study demonstrates that shorter videos are more engaging than longer pre-recorded classroom lectures (Guo et al. 2014). Following this principle, making a video in segments or with functionality to differentiate topics within longer video session, can make the online educational video more functional and engaging with students. Segmenting video can also be seen as a way to make video content interactive with the viewer. This makes a video with specific segments, more effective tools for the learner. (Dongsong et al. 2006) Segments can be by topic or with time.

According to Ibrahim et al. (2015), weeding the educational video can make the video more focused on essential content. This, in turn, can effectively decrease learners' self-reported mental effort and improve knowledge acquisition (Ibrahim et al. 2012).

Modality matching in practice means giving concurrent, redundant information from multiple channels, using the auditory channel to describe content while using animation or video to visualize content. This practice uses both visual and auditory channel to convey information and has been showing to increase students' retention of the taught subject. (Mayer & Moreno 2003.)

Technologies to allow following principles presented earlier are mostly based on mainstream video and audio technologies, and as such, not hard to procure. From a technology point of view, these functions can be divided into few categories, and furthermore dealt with in smaller increments. For this thesis, technology is dived in four different main categories: video, audio, delivery, and capture. Figure 5 shows relations of these categories.
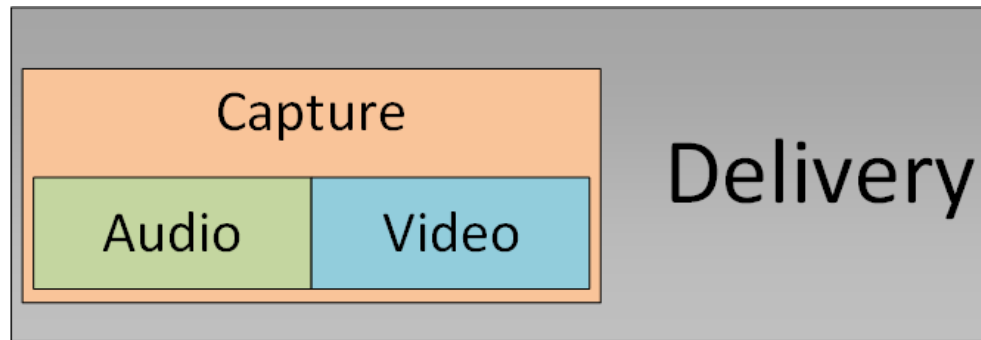
FIGURE 5. Thesis technology categories

To make increments stack as a coherent whole, most of the technologies are tied to psychology and to the needs recognized in making an educational content.

2.1 Video

Video, or more closely digital video has experienced drastic changes during the last two or three decades. These changes have revolutionized the way video is been consumed and produced. Today's video is in comparison very accessible and easy to use. Video copying used to mean a direct copy of celluloid film and thus consequent reduction of video quality. Analog video film material was and is a very easily damaged, it needs a highly-controlled environment where it can be edited and used as a source of video material. This functionally prevented video usage as an individual source of information in a large scale. Digital video in comparison means to end user an easy to consume, readily available information source. During last two decades one has seen the rise of online video and video delivery platforms.

For educational video, digital video means a convenient method of delivery, consumption and a channel of interaction through the use of online conferencing. To get to this point, video technology must have solved multiple delimiting factors. These factors include delivery of video frames, in the most basic form a series of still images, synchronization with audio channels and endpoint compatibility.

The basic digital video signal is almost always comprised of color components. These components are most of the time representatives of red, green and blue components of each individual image seen by the camera. Camera translates all of the color spectrum in relation of signal luminance and chrominance. Camera produced signal is there for called YUV component signal. Components are usually presented in 4:2:0 ratio, where Y is luminance and U and V are chrominance components presented at perpendicular vector axles representing blue and red color spectrums. First number of the ratio presents number of luminance samples taken. Second and third numbers present numbers of chrominance samples per even and odd sample point. (Bing 2015.)

Video image size, or number of horizontal and vertical picture elements found in each of frames of video, are called pixels. Comparing this to analog video, pixels then represent the number of horizontal and vertical line intersects per frame. When these pixels are being recorded, i.e. what actually is been saved, is a representation of color in each of these points. This representation is encoded in bits and it defines how many bits have been user per pixel. This defines the color depth of video. Usual color depth today is 24 bits per pixel (8 bits per color channel) that allows 2^{24} colors or approximately 16 million colors. High-end cameras can produce a digital video of the color depth of 32 bits that would equal 2^{32} colors per pixel or 68 billion colors per pixel. Commonly accepted perception capability of the human eye is up to 10 million colors. (Bing 2015.)

The number of vertical and horizontal pixels defines the resolution of the video. Most commonly, this refers to video size when it is been viewed, although this is misleading. When talking about resolution, a distinction must be made between the screen and video resolutions. Video resolution while it is been played relates to a number of pixels shown on physical display or monitor. Video resolution, or frame resolution, is always the number of pixels in one single video frame. The actual video is often resampled to fit physical monitor or screen, and rarely represent actual video resolution. This means that subjective video quality lowers and

becomes distorted. This is more prevalent when the video is being scaled to a higher screen resolution. (Bing 2015, 10-11)

While producing video, the aspect ratio of the frame must be considered. Produced video resolution can be varied by adjusting the pixel aspect ratio of the screen. Pixel aspect ratio is a ratio of the length to the width of the pixel on display. Screen aspect ratio is a direct result of original video resolution times pixel aspect ratio of the screen. This would mean that to obtain a 1280/720 resolution with an aspect ratio of 16/9, pixels would have to square with 80/80 aspect ratio when in comparison to screen. If this kind of frame would be scaled down to standard television resolution of 720/576, pixels would have to be transformed down to the aspect ratio of 64/45 and the resulting image would look stretched or squashed in the horizontal or vertical directions. This would present itself with symmetrical shapes. For example, circles would seem ovals. (Bing 2015, 11)

For educational video, aspect ratio and resulting image quality should play a significant part. Making content with text or in the case of mathematical subjects, easily interpreted footage is a definite requirement.

List of common video resolutions is shown in Table 1. This list comprises of commonly used video resolutions starting from lower resolutions (quarter common intermediate format (QCIF), common intermediate format (CIF) and standard definition (SD)) and ending with future high-resolution formats (Ultra High Definition (UHD-1) and Ultra High Definition 2 (UHD-2)). There is a progression shown here, where lower resolutions, that have been largely superseded, and higher resolutions roughly follow the pattern of enlarging shown resolution by a factor of four.

TABLE 1. Common Video Resolutions (Bing 2015)

Format	Resolution (pixels)
UHD-2 4320p (progressive)	7680 x 4320 (16:9 aspect ratio)
UHD-1 extra wide (progressive)	5120 x 2160 (21:9 aspect ratio)
UHD-1 2160p (progressive)	3840 x 2160 (16:9 aspect ratio)
High definition (HD) 1080p (progressive)	1920 x 1080 (16:9 aspect ratio)
High definition (HD) 1080i (interlaced)	1920 x 1080 (16:9 aspect ratio)
High definition (HD) 720p (progressive	1280 x 720 (16:9 aspect ratio)
Standard definition (SD)	720 x 480, 720 x 756
Common Intermediate format (CIF)	352 x 240, 352 x 288
Quarter common intermediate format (QCIF)	176 x 120, 176 x 144

Resolution plays an important role, especially when the resulting video is intended to be shown on multiple different screens. Choosing production resolution so that resulting resolution on delivery maintains correct pixel aspect ratio, makes resolution one of the biggest influencing factors to quality video.

2.2   Audio

For education purposes, audio is arguably more important media than video. This means that more attention should be spent on audio recording and post processing. The audio itself is easily produced with inexpensive digital video technology. This is one of the biggest changes in video production in last decade.

According to Holman and Baum (2013) audio production has four basic attributes that affect sound quality. These are frequency range, dynamic

range, space and time. Frequency range is a range of frequencies captured in sound production. This is limited to the human hearing range starting with approximately 20 Hz and ending with around 22 kHz. This frequency band is mostly referential because human hearing loses its frequency sensitivity when human ages.

Also, individuals experience audio in different ways, depending on multiple factors like hearing damage suffered, because prolonged hearing of loud sounds or individual's earlobe shape is radically different so subjective sound that reaches ears bone structure is maybe in a different phase. (Monteiro de Castro Silva & Feitosa 2006.)

Audio production is mostly done using microphones attached to computers or handheld devices such as tablets or phones. Most of these mobile devices are equipped with in-build microphones from the factory. The problem inherited from this starting point is that most of these devices are designed to have a passable quality in any environment. This effectively works against producing good quality sound in quiet, almost studio-like environment.

To produce a good quality audio recording, selection of microphone and placement of microphone should be done so that pattern of sound wave propagation is directly pointed at the recording device. Patterns of sound in different frequencies are displayed in Figure 6. Sound pattern and subsequent sound level reduction is then dependent of frequency of the sound. Figure 6 shows only vertical sound fields but similar sound reduction is also present on horizontal axis. For optimal placement the obvious position of the microphone is directly in front of the speaker. For the stationary speaker or lecturer, this is not a problem but when there is movement or multiple sources for the sound, microphone setup grows more complex.
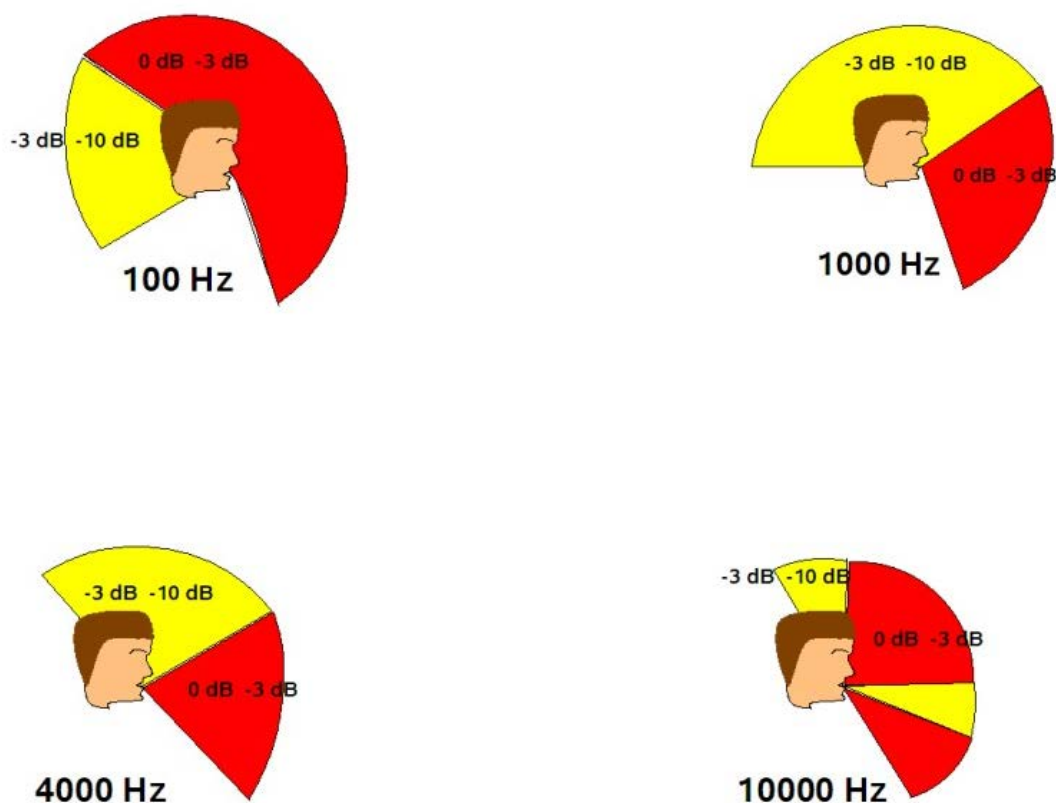
FIGURE 6. Sound patterns for speaker (Korpinen 2005)

Regardless of the microphone setup, to create a correct and good quality sound recording, the digital sound should be recorded in frequency that is at least twice to that of the produced sound. This is a derivate of so-called Nyquist or Shannon sampling theorem that states that "The minimum sampling frequency must be at least twice that of the highest frequency component present in the original signal". (Leis 2011, 82.)

For the everyday application, this means that the sampling frequency should be twice of the human hearing range, i.e. 44 kHz. Most commercial sound recorders use sampling rates of 44.1 kHz or 48 kHz. This is more than enough regarding sound quality needed when recording speech. Highest sounds in human spoken voice are s-sounds that commonly reach frequencies up to 10 kHz. (Korpinen 2005.)

Furthermore sound recording quality lays in the precision of sound sampling. Commonly used sound sample precision is 16 bits per sample. This gives word length of 1048576 digits (2^{16}) for each sample. For audio, this means that each bit of word length delivers approximately 6 dB of dynamic range, giving 16-bit recording precision a 96 dB range. This is a bit less than what human hearing is capable of. For most situations and recording environments, the background noise level is much higher than 0 dB. At the same time, loudest noises needed to capture will not approach 120 dB level giving that dynamic range of 96 dB will be more than sufficient. (Holman & Baum 2013, 19-20.)

2.3 Compression

Video compression is a topic that gets its roots from the need to deliver video through narrow bandwidth originated from the technology used at the start of 1980's. Video compression seeks to reduce or remove redundant information. Removal is done either in temporal or spatial sense, so that video can be stored or sent over a network as efficiently as possible. This is usually achieved by using an algorithm to eliminate the excess information from the picture stream.

There are multiple different compression methods. All the compression methods are reliant on processes called encoding and decoding. Interoperability of these processes forms a basis of a video codec. Video codec is nowadays usually a piece of software, but it can also be a hardware or combination of both. (Bing 2015.)

Mostly used video compression standards in today's internet driven landscapes are Motion Pictures Element Group standard 4 with advanced video coding (MPEG-4 AVC / H.264) or H.265, High-Efficiency Video Codec (HEVC) and VP9. All of these coding methods rely on basic processes of removing non-essential parts of the raw video. This aims to reduce redundancy, in order to reduce or compress video to its smallest size. Figure 7 presents a visual presentation of redundant information removal. Top of the figure presents non-compressed video stream and

bottom part presents a compressed stream. Frames 1 and 5 marked here are so called anchor frames where all the image information is retained. Frames in between are so called inter frames, that only contain parts that cannot be accurately predicted. Prediction here is done in both spatial and temporal domains.
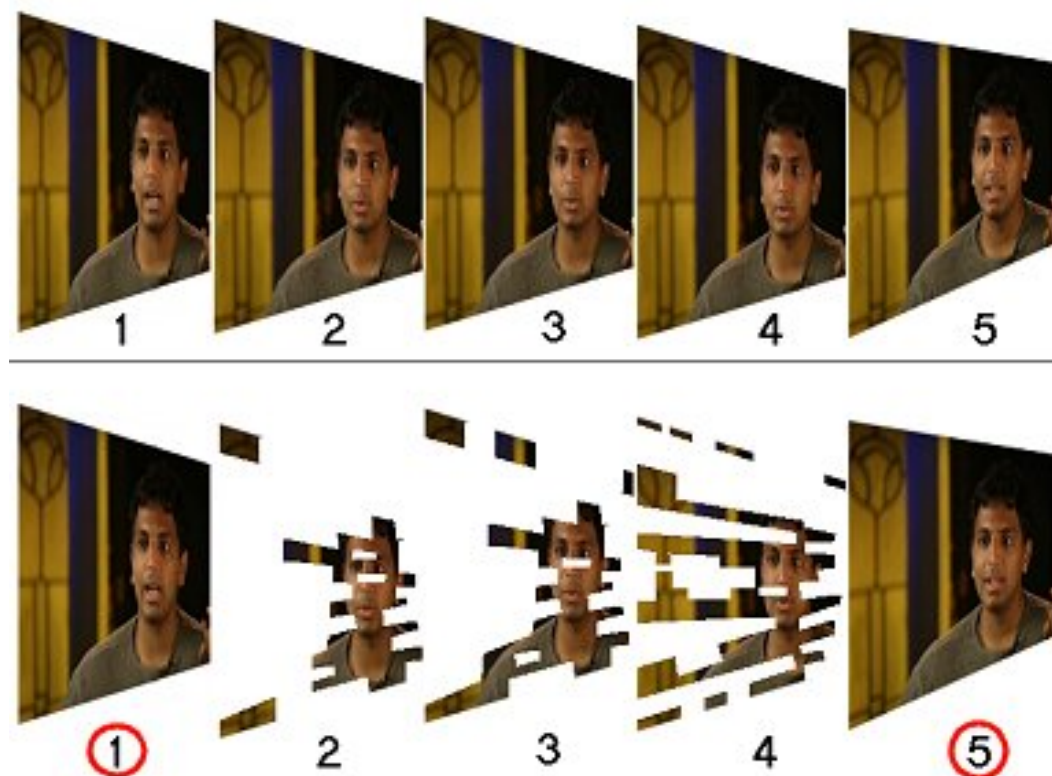
FIGURE 7. MPEG example of reduction (Jagetiya & Sharma 2015)

For educational video, original video file size does not present a critical issue. This is based on the fact that most educational videos are consumed by using some kind of dynamic video streaming service. This means that from the perspective of the consumer, the video will always be presented in a format that conforms to users device, bandwidth, and application. Figure 8 shows this progress.
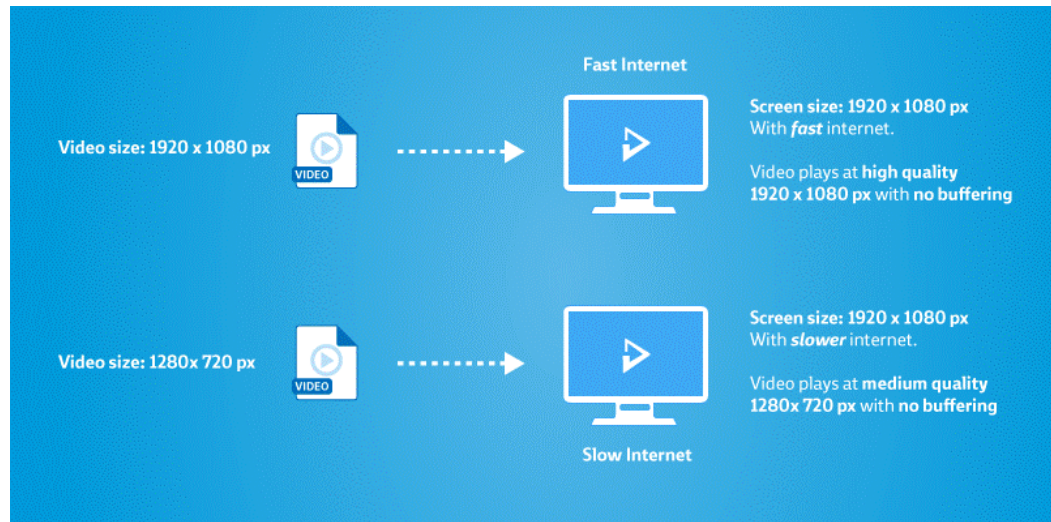
FIGURE 8. Adaptive video delivery (Bitmovin inc. 2017)

To allow this adaptation to happen, the video is usually transcoded to multiple different resolutions and bitrates. This allows the video delivery platform or the video player to choose the best-fitted video for the user. Or in the best case, to make selections between adaptations during the playback. Figure 9 depicts adaptation process, in regards of connection changes. This also applies to different users in different kinds networking environments.
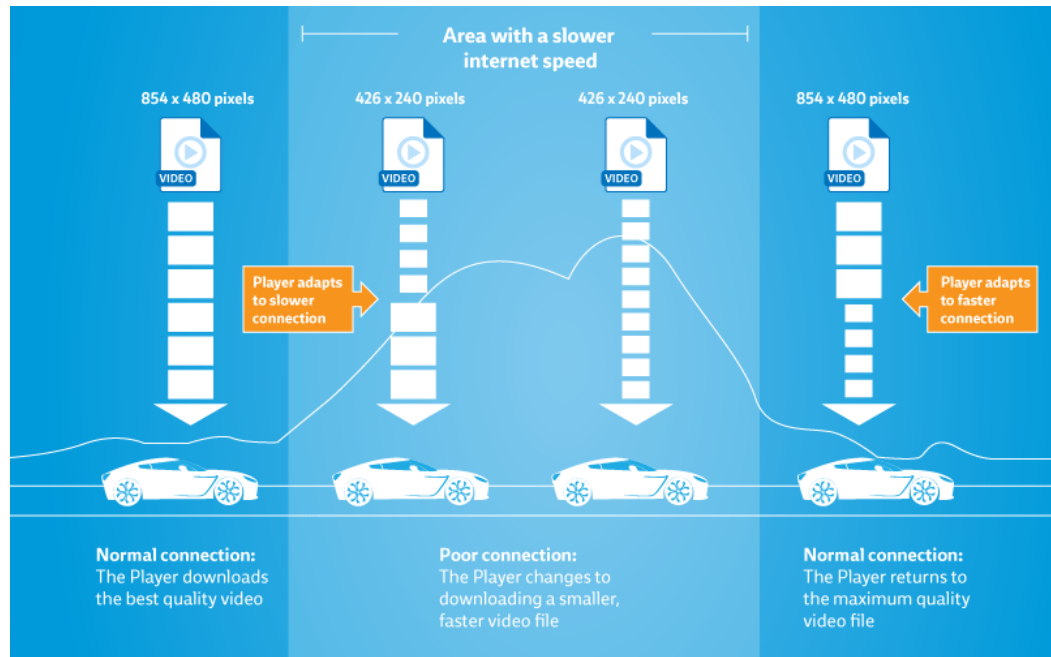
FIGURE 9. Adaptive streaming (Bitmovin inc. 2017)

Audio compression is also a process to reduce or remove non-essential elements from sound. This is usually done in regards of psychoacoustics, a science that studies a sound perception. This study has resulted in knowledge of hearing range and threshold of hearing. Figure 10 shows human hearing range in regards of intensity of sound and its frequency. Knowledge of human hearing range allows for sound compression algorithm to remove sound outside of human hearing range.

Voice compression has evolved at the pace of video compression. A basic perceptual audio codec uses this information as a perceptual model. At the same time codec uses a function called analysis filter bank, to decompose inputted signal into subsampled components. These components are results of time and frequency domain division. Together with perceptual coding model, these spectral components are then used as quantization inputs and coded into a bitstream. Aim of this process is to keep noise below masking threshold. This limit is introduced as threshold of hearing. (Brandenburg 1999.)
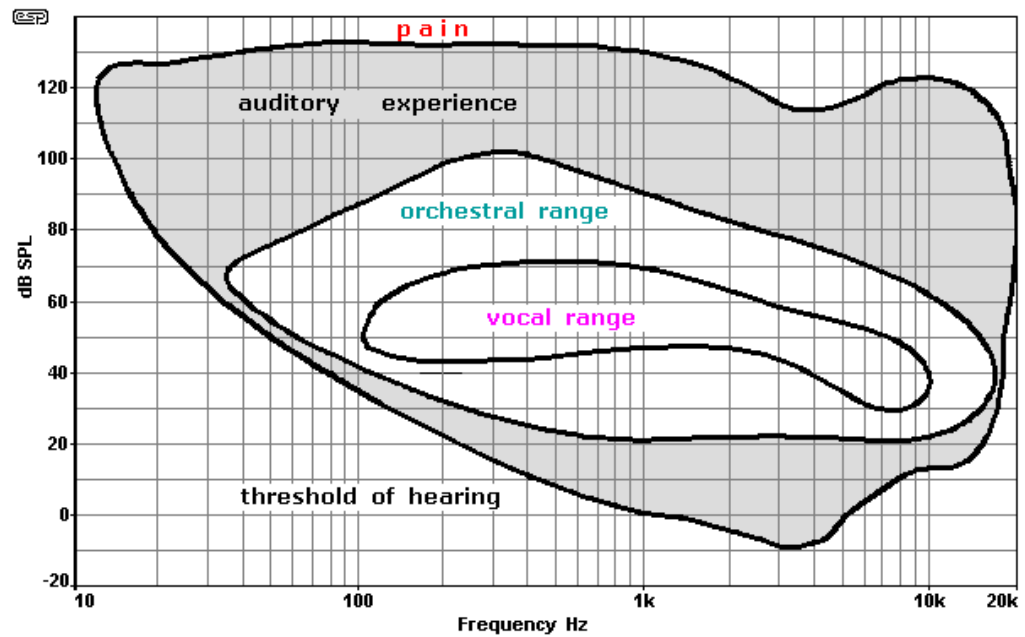
FIGURE 10. Perception of human hearing (Pandit & Speranza 2014)

2.4   Capture devices and software

There are multiple ways to capture video. Whether the video is made
using screen capture software or recorded using more traditional camera
equipment. There is always the question, how to deliver video to the
audience. Using capturing devices such as professional video capture
cards, USB dongles or dedicated capture devices with build-in monitoring
system versus using the software-based method of capture, will be
primarily decided by subject matter presented and method of gathering
materials. For this thesis, two of the professional capture cards, one USB
based dongle and dedicated capture device will be presented. Also, there
will be a view to video capturing, mixing and streaming software that is in
use at HAMK Riihimäki campus. Video capture can also usually be done
using video conferencing software such as Microsoft Skype for Business,
Adobe Connect Pro or Cisco WebEx system. Also, some of the video
publishing systems incorporate a screen capture system or software.
HAMK Riihimäki uses Kaltura screen recorded called CaptureSpace lite
for Mac OS and Windows systems. HAMK also employs YouTube for

Education and Telestream's Wirecast for YouTube system to make short live streaming sessions. There is also an app, called YouTube recorder for use with Adobe iPad devices. This app allows the user to record using the iPad's camera and lightly edit the video before uploading it to YouTube service.

Main capture devices used on distance teaching workstations are Magewell AIO 4K Plus card and Matrox vs4 Quad HD, Capture Card. Both of these cards are capable of capturing at least full HD resolution at 50 frames per second in full 4:4:4 color mode. Both cards also support Microsoft DirectShow filters thus making them directly usable with Microsoft Windows-based conferencing clients such as Skype for Business. (Matrox 2017; Magewell 2017.)

Both Matrox and Magewell capture cards provide a function called multiple streaming that allows card input to be simultaneously streamed to multiple applications. This among other functionalities, allows the hypothetical situation to occur where live camera feed can be used by both dedicated capture software and video conferencing software without the need for video signal splitters or multiple PC machines.

Video capture can also be done using dedicated capture device as mentioned earlier. This method of video capture or video recording usually happens when using multiple video image sources and video mixer along with an audio mixer. This kind of systems will be presented later in this thesis. The device in use at HAMK Riihimäki is Blackmagic Design's Video assist 4K. This device is simultaneously usable as video monitoring system and video recorder. The device records video feed from either SDI or HDMI input source up to UHD resolution. Using separate video recorder allows for independent selection of recording format as well as monitoring of produced video. This method allows higher material quality, and allows post-production editing without quality problems. (BlackMagic Design 2017.)

2.5   Video delivery

Video can be delivered through multiple different distribution channels. HAMK uses multiple delivery channels based on availability, functionality and source system. These are different video based conferencing systems such as Microsoft Skype for business, Cisco WebEx system and Adobe connect as well as purely video delivery systems such as Kaltura VPaaS (Video Platform as a Service) and Google YouTube for Education or Microsoft Stream. Video delivery can also be done using a purely file-based system such as Microsoft SharePoint publishing or filesystem publishing service from HAMK intranet. Using the file-based method of publishing bypasses all of the dedicated video delivery functionalities build into video delivery systems. Same extinction can be applied to both WebEx and Adobe Connect based systems but in a diminished capacity. Conferencing systems are more attuned to delivering one session of multimedia content to multiple recipients regardless of the subjective quality of the video. This is commonly achieved by using lower framerate, smaller resolution or both. Both of the independent conferencing systems have a functionality to share a recording of a conference held on the system. And allow the user to export conference recording in more common MP4 wrapper using H.264 codec format. Problem with these exports is inherent in the design of conferencing system: the basic requirement is to have a successful conference without connection loss or missed content. This results in lower quality recording and a lower quality video record.

Selection regarding what video platform to use, if often case of considering target audience. Microsoft published their new video streaming platform at the time of writing this thesis and thus it has not yet gained a standing position as a video delivery platform in HAMK. Microsoft Stream is designed to be a video collaboration platform inside the organization. In the case of University, this would mean a platform to interact with video between faculty and students, faculty or students themselves or any such combination. Key fraise here being "inside the organization". Both Kaltura VPaaS and YouTube for education are services developed to delivering

video to widespread audiences. YouTube especially being one of the world's leading video on demand (VOD) services. HAMK uses both YouTube and Kaltura services with an emphasis being that educational video material for self-study and background materials produced at HAMK, are delivered using Kaltura system.

YouTube remains the only system where HAMK has widespread access and where live streams can be published. This makes choice of delivery platform easy when the goal is to deliver purely video based live event capture. At chapter 5.2, this thesis presents an open source alternative for a live streaming platform with adaptive streaming capabilities and compatibility with multiple widely used end-user devices and players.

2.6   End-user players and environments

Using different video delivery systems allows for multiple different kinds of end user devices to be used. For example, YouTube video player will adaptively scale itself to screen size and the system will try to dynamically deliver material most suitable for end system and used bandwidth. For this to happen, original video material has to be with as large resolution and with as high bitrate as the highest viewing system is using. This is usually not a problem for mobile devices. Mobile devices such as Android-based phones or tables are using screens that are in multiple sizes and shapes. To be able to adjust to these sizes, Android developer guide gives out and comparison chart displayed in Figure 11. This chart is a collection of data collected from Google Play store visited devices. (Android.com 2017) It shows that most Android users are using devices with the normal sized screen in extra high dpi (dot per inch). Android developer emulator translates this to resolutions of 480 x 800 pixels or 480 x 854 pixels depending on screen symmetry. Other emulator supported screen sizes can be found from Figure 12. (Android.com 2017.) This corresponds to educational video in demand for maximum resolution and size for video. Most devices still try to offer crisper image by using higher resolutions and

higher bitrates, but these image sizes should be enough to fill full-screen sized viewport on most Android devices.

	ldpi	mdpi	tvdpi	hdpi	xhdpi	xxhdpi	Total
Small	0.7%						0.7%
Normal		1.5%	0.2%	30.5%	36.7%	21.3%	90.2%
Large	0.1%	2.9%	1.6%	0.5%	0.9%	0.1%	6.1%
Xlarge		2.0%		0.5%	0.5%		3.0%
Total	0.8%	6.4%	1.8%	31.5%	38.1%	21.4%	
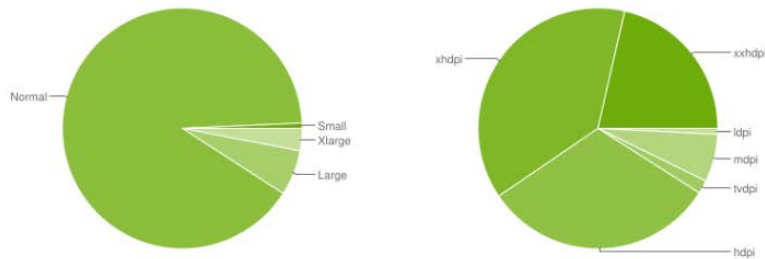
Data collected during a 7-day period ending on October 2, 2017.
Any screen configurations with less than 0.1% distribution are not shown.

FIGURE 11. Android screen sizes and densities (Android.com 2017)

	Low density (120), *ldpi*	Medium density (160), *mdpi*	High density (240), *hdpi*	Extra-high-density (320), *xhdpi*
Small screen	QVGA (240x320)		480x640	
Normal screen	WQVGA400 (240x400) WQVGA432 (240x432)	HVGA (320x480)	WVGA800 (480x800) WVGA854 (480x854) 600x1024	640x960
Large screen	WVGA800 (480x800) WVGA854 (480x854)	WVGA800 *(480x800)* WVGA854 *(480x854)* 600x1024		
Extra-Large screen	1024x600	WXGA (1280x800)† 1024x768 1280x768	1536x1152 1920x1152 1920x1200	2048x1536 2560x1536 2560x1600

To emulate this configuration, specify a custom density of 160 when creating an AVD that uses a WVGA800 or WVGA854 skin.
* To emulate this configuration, specify a custom density of 120 when creating an AVD that uses a WVGA800 or WVGA854 skin.
† This skin is available with the Android 3.0 platform

FIGURE 12. Android SDK emulator representative resolutions (Android.com 2017)

For PC screens, the situation is more flexible. Website statcounter.com keeps records from over 2 million sites globally. (StatCounter 2017.) Company then publishes these statistics in multiple categories. One of

these is screen resolutions. Latest of these collections are from October 2017. Figure 13 shows that mostly used screen resolution is 1366 x 768 pixels, second most used is full HD resolution of 1920 x 1080. The corresponding conclusion is that at the most, full HD size is appropriate for most usages, and that going over that is still, at the time of writing this, just curiosities. This will eventually change, but for the current generation of end users, full HD image represents commonly used format.
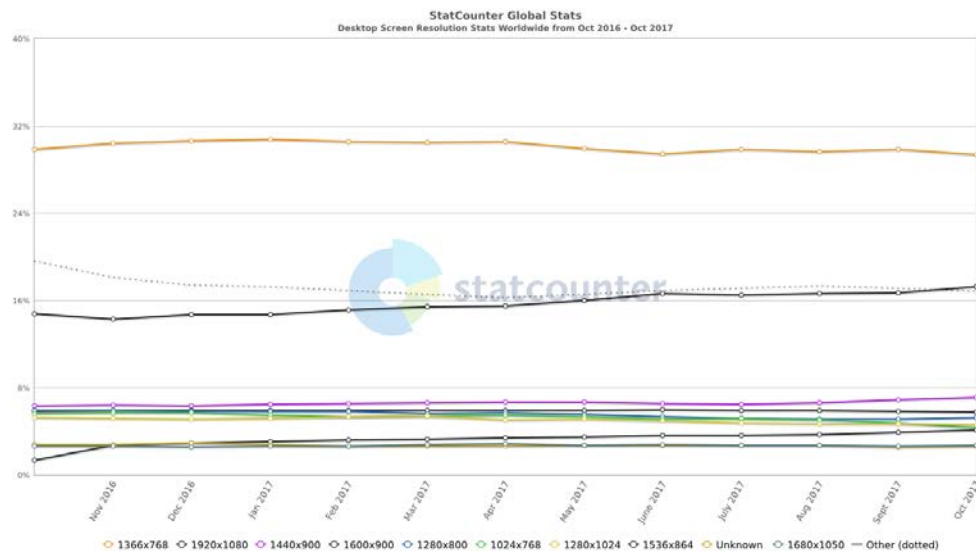
FIGURE 13. Desktop Screen Resolution Stats Worldwide (StatCounter 2017)

Most of the live event delivery will be done on different kinds of video conferencing systems at HAMK. This actually does not change the end user environments or demands drastically. Users are still using that same device, that they would use in their everyday lives, only the player component will in case of collaboration system or conferencing system be an app. For mobile users, this app will still commandeer a part of the usable screen, and video presentation inside of workspace is still smaller part of these systems. Same kind of thing happens with most of the conferencing systems on pc platform as well. For purposes of this theses, deliverable screen resolution for video shall be at the most, full HD. Anything over would be a waste of resources and under would cut the best user experience from almost quarter of the audiences. Full HD size image

is also the standard starting point for post-production editing and enhancement. These considerations in mind, today's video production in general, is mostly limited to resolutions of 1920 x 1080 pixels and up to 60 frames per second. Even this much is stretching the capabilities of most end users' devices.

3   LECTURE CAPTURE

With new technology, lecture capture systems and practices have experienced a significant turn to everyday, affordable and easy to use technology. Lectures are primarily done with educator speaking. This provides a audio track, over the subject matter and simultaneously presents slides or other material to augment presentation.

Lecture capture practices can be divided into the ways of production. At HAMK Riihimäki there are couple basic methods to produce lecture captures. These include screencast, classroom capture and live event capture by way of the live video stream. Screencast systems usually depend on few production tools to capture video and audio. Also, systems used to create content usually use some system to utilize stylus or another kind of drawing aid. To use a stylus, there are two different methods used at HAMK Riihimäki. The primary method is to use drawing screen, a touch display, to input drawings and pictograms digitally. The secondary method is to use a drawing tablet, but this method requires some familiarization to get used to disjointed hand-eye coordination. Image 1 displays an usual interactive screen and conferencing microphone combination. With this kind of setup combined with electrically adjustable table height, basic notation making during lecture is easy.

IMAGE 14. Interactive screen and conferencing microphone

To use any kind of drawing or writing tool, lecturer needs to have a digital canvas or paper to use. At HAMK Riihimäki, this is done with Microsoft OneNote 2016. OneNote is a software tool originally thought to be a note taking a tool, that in our case is been repurposed to be a lecturing aid.

OneNote functionality includes handwriting support, as well as compatibility with almost any platform in the market today. Together withdrawing aid of pen display, the lecturer can interact with screen capture, and use signaling that is important in creating compelling lectures.

3.1   Screen Capture

To make a video of screen presentation, there are two ways to do so. Almost any computer today has a way to output monitor signal to multiple monitors. To capture screen presentation as a video, along with lecturers' audio input, there is an option to just connect a digital video recorder to computers video and audio outputs, duplicate screen to this output and route microphone to recorder. This method would create an identical video out of presentation including the OS interface and pointer cursor movements. Another method to capture screen is to use capture software, that records screen content out of the graphics buffer by interacting with screen API. By using capture software, the user can select specific area, software being used or whole screen to capture. At HAMK Riihimäki, there are few different kinds of software in use for this purpose.

To capture the specific physical screen of a computer, the lecturer can use Kaltura CaptureSpace lite desktop recorder or Open Broadcasters Software. To capture virtual conference, the lecturer can use conference tool's recording function. At HAMK Riihimäki, there are two different conferencing tools in use, Microsoft Skype Professional 2016 or Cisco WebEx platform. Both conferencing tools have an option of recording conference, although both platforms have limitations on quality and recording method.

Lecture capture needs to have an audio track, that matches modality with the presented visual material. This is achieved mostly by using computers audio input or attached headphone-microphone device. To have a quality audio input, the audio device needs to have a full range microphone and a recording capability to make a recording of sufficient quality. These are described in Figure 14.
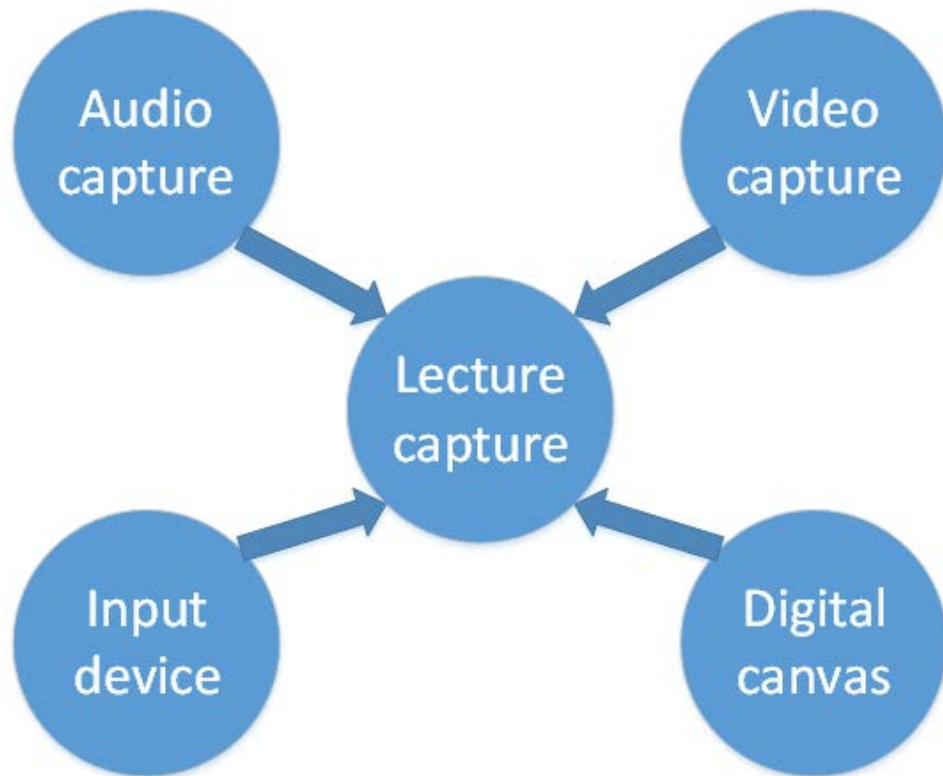
FIGURE 154. Components of capturing lectures

Lecture capture should always result in the video file. To ensure files combability with selected delivery systems at HAMK Riihimäki, the file should always be wrapped in MP4 container, using MPEG-4 AVC as video compression method and AAC as the audio compression method. The resolution should be selected with delivery channel kept in mind, but resolution should be at least 720p and it should not go over 2160p. If the video is intended to be post-processed, then the original material should always be at least one logical step higher than its future delivery channel resolution will be. For example, if the delivery channel is meant for 720p material, best quality will be reached using original video material of 1080p or higher and downscaling it to 720p for delivery purposes.

3.2   Screen and lecture capture

When adding a lecturer as part of the video, there are few elements to consider. Adding actual video, once it is captured, is a matter of selecting appropriate software option. CaptureSpace Lite software includes the option of using a webcam as a video source, and when it is selected, places image from a webcam in front of desktop capture image. This is an easy application of picture in picture function. The picture in picture or PiP is presented in figure 15.
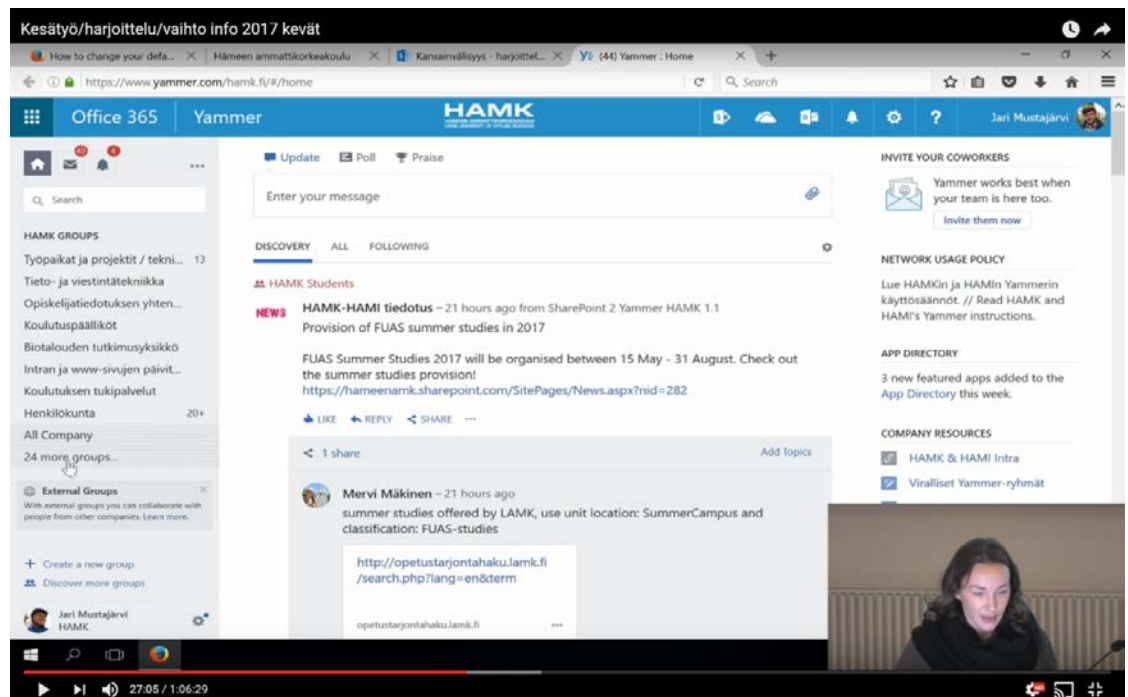
FIGURE 15. Picture In Picture example

This application of PiP is the very simple usage of video channels that are situated on top of each other. To get an effect of placing the lecturer on top of lecture materials, some application of video keying should be used. As discussed earlier, chroma or luma keying can be done in real time, if certain key factors are being met. To have a successful keying video, video background should be with chroma keying, a mono-color surface, usually green or blue color, and with luma keying a black surface. This is easy to reach if the video is being produced at the studio, but usually

harder or impossible to do, if the lecture is captured as is, in classroom or auditorium.

With lecture capture covering part of the screen, the best result would come out if lecture materials were available to preview before adjusting the lecturer capture video location on the screen. With keying same consideration is valid, but the placement of lecturer on the screen is more subject to notes usage of colors and illumination of presentation.

Another element to consider is lighting the video of presenter or educator. This requirement is more prevalent when using webcams or cameras integrated with laptop i.e. Cameras that have a smaller sensitivity to light and smaller aperture size to use with lensing. Solution to lighting a lecturer can be solved with either or both of following. First, and easier option, is to use a high-quality broadcast video camera, with the very high-quality image sensor. The second option would be to increase indirect or direct lighting of the presenter. This can be accomplished in multiple ways, but will always bring visible lights and light stands near presenter. This is not a problem if the presenter is alone but can be a hindrance when capture is done in the classroom with a live audience present.

Video signal from the camera, if it is not an integrated or USB webcam variety, will have to be captured with some device. As previously discussed, these devices range from one channel USB connect capture cards to multichannel video capture cards, which can process simultaneously multiple non-synced video signals. Using any of the capture cards always requires a support of the recording system.

3.3   Screen, notes, lecturer and event place capture

Adding more inputs to an educational video, will probably increase information load of the viewer and make presentation or subject matter discussed harder to follow. With exception of event place or live audience capture and when the lecture is held in an interactive manner. Technically this requires either camera that is capable of panning, i.e. manned camera or remotely controlled camera or two or more cameras and a system capable of switching between video signals from cameras.
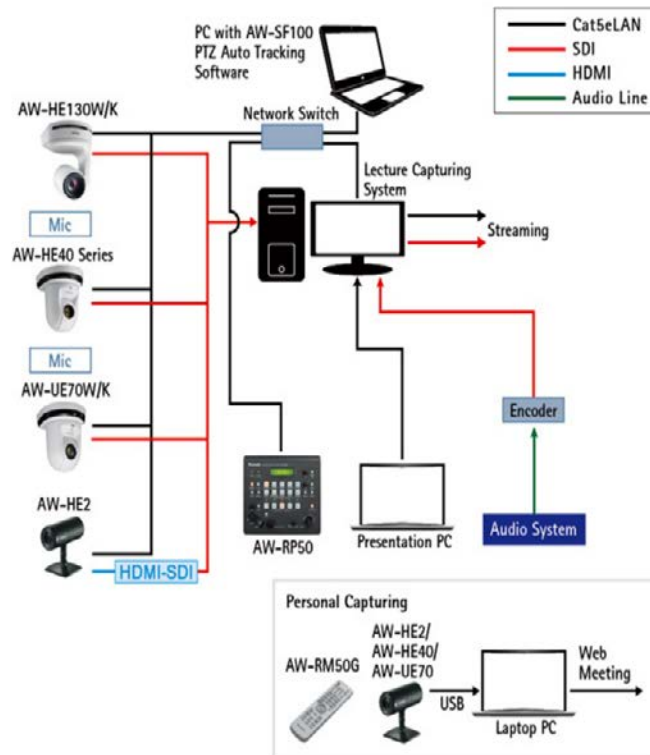
FIGURE 16. Panasonic flexible lecture capture system (Panasonic Corporation 2017)

Figure 16 presents an approach to lecture capture system that relies on multiple camera systems, that are controlled by the auto tracking system and feed image signal to lecture capture system PC. This method would require a multi-input capable Serial Digital Image (SDI) capture device along with a method of capturing presentation PC screen. Panasonic

system presented in Figure 16, does not take any stand on the operation of the sound system or audio capture.

Multiple source system described here would still not be able to deliver notes from the lecturer or from the audience at the same time with video. This should, therefore, be done at the delivery system. Using video delivery system to hand out lecture notes frees the production system from the need to expressly holding one of the video channels for the lecture notes capture. There is also the consideration of where on the screen can the notes be places, without making an image of presentation or presenter needlessly small on the subjective screen of the viewer. As pointed out previously, annotated slides can increase student's retention rate, but making annotations appear farther away from the main presentation will decrease information retention and impact of the video (Mayer 2001).

4 LIVE EVENT CAPTURE

Capturing and delivering live event makes video production more challenging on multiple levels. Live production usually means that errors made in production cannot be corrected before consumption. Also, artificially enhancing captured audio or video must be done in real-time and this will make live event capture intensive production method.

4.1 Goals and limitations of events and delivery systems

Capture of live events at HAMK Riihimäki usually involves some or all of the three elements: digital material, usually Powerpoint slides or other presentation material, presenter or lecturer and audience. These elements define minimum capabilities needed to be present at the capture equipment. Live event capture can happen with just one or two of the elements previously discussed, but usually live-event consists of all of the three basic elements.

These elements are deciding elements on capture location and devices. Figure 17 presents a decision matrix for live event capture preparation and location selection. This matrix is produced as a starting point, when planning a live event and consequent capture.
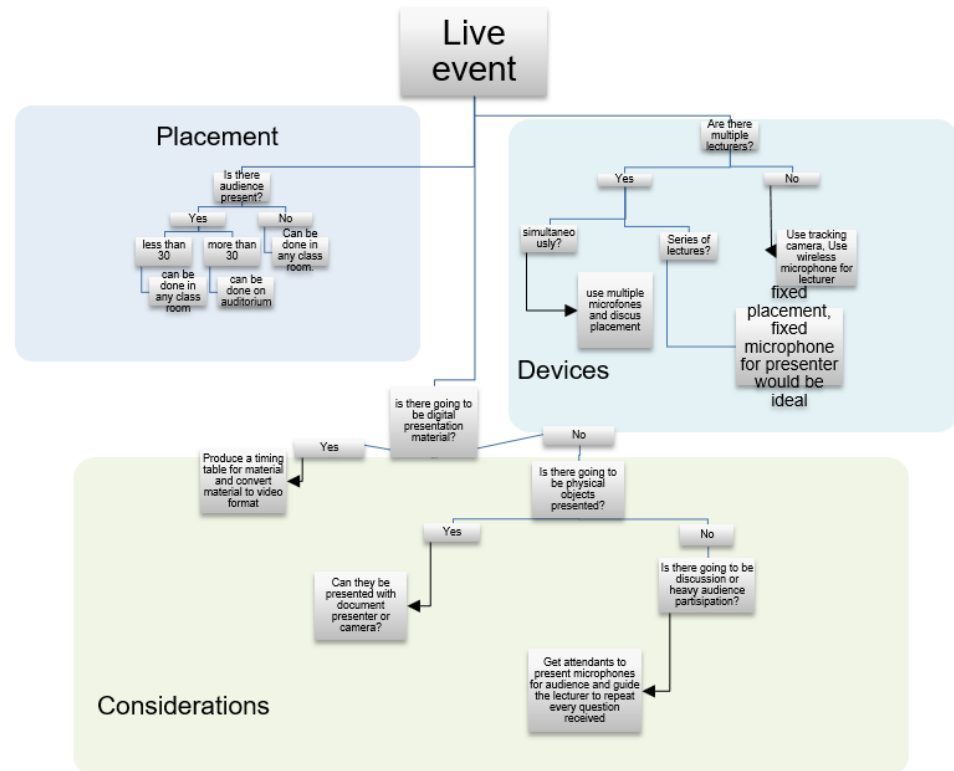
FIGURE 17. Decision matrix for live events

The goal of live-event capture is seen as two-fold. Firstly to convey live-event to an audience not present at the event site. And secondly to get a recording of an event to be consumed at a later time or to be a material for self-study. These goals are main drivers behind live-event capture and as such, give out a basic needs for capture and delivery system.

Live event also needs an delivery medium, where access control and a return channel for an audience can be implemented. For purely video standpoint, live events can be broadcasted using Youtube for education service, that incorporates access control mechanism in itself. Limiting factor for using Youtube as a delivery platform, comes with limitations Youtube has placed on itself and those that Youtube enforces automatically. These include copyright material usage such as known brands like Cisco or media content such as music. This is done through a system Youtube calls Content ID. (Google 2017.)

Content ID system scans uploaded material, live or otherwise, against a database of files, that are submitted by content copyright holders. Content ID system will recognize content similarities against the database and when found, stop the delivery of the media. For live event stream, this might mean audio mute or video blank being inserted into stream content during the event. Youtube does not, at the time of writing, provide a system to declare a paid license to use copyrighted material for live streams. Copyright claimant can be done after the fact, by submitting a counter notification, but these are processed by Google legal team and take at least 10 days to process. (Google 2017.)

4.2 Systems and facilities present at HAMK Riihimäki Campus

There are four types of spaces available at the campus, to record an educational video. Basic classroom, lecturing hall or community area such as the cafeteria, would be the most common space to make lecture capture or live lecture capture. Basic places are all either equipped with conference microphones, computers, and basic software in for of Capture space live and thus these spaces can be used to make a rudimentary educational video. Almost always using these spaces, the recording is from classroom lecture, with digital material presentation and voice capture. These videos mostly function as reminders and background materials for courses. They are most likely published in the format of "as is", without any editing besides pasting of topic slide and outro slide with copyright information in it. Usual medium for delivering this kind of video is thru video.hamk.fi site, using Kaltura media system.

Second most common space for producing educational video is the lecturer offices and workstations. All of the faculty members have been equipped with wireless headsets, computers with screen capture software and at least an office desk or private office. Videos produced at lecturers office is usually information clip, with the length of 5 to 10 minutes, explaining one phenomenon or function. These videos usually follow a script and contain easy explanation with some kind of pointer used to

highlight the subject of the video. These videos sometimes are enhanced with annotated slides delivered with video, using Kaltura system. Usually, these videos are collected in playlists and delivered using Moodle learning environment as self-study materials for given topic. This kind of production method usually involves Kaltura CaptureSpace lite screen capture software. Using such a software presents a challenge for produced material, given the limitation of the software to only record video with 15 frames per second.

With frame rate being so slow, every movement, be it computer cursor or video from webcam, will look choppy. The motion will start to look seamless from 25 fps and upwards. This does not present a problem when presenting in large parts static content such as computer software interface, but it has to play a role in designing the video. Presentation, or video, in this case, cannot contain information in small rapidly moving windows or slides. This can be compensated by the presenter using the time to explain each and every movement and screen part before moving forward with the presentation.

Third space is dedicated distance learning studio or educational video producers space. These spaces were designed and furnished during this thesis and are partially influenced by research done in this theses. These studios have approximate furnishing and devices shown in the Figure 18. The figure describes a generic idea for space and does not make any disclaimers about the actual devices used. In actuality, monitor show on the table is replaced with Wacom interactive pen and touch display. This gives the video producer a digital surface to write and draw upon. The figure does not actually show any lighting or sound dampening on the room. Actual rooms have optional lighting done with led panels if the lecturer needs or wants to use green screen technology or present some physical object as a study object.
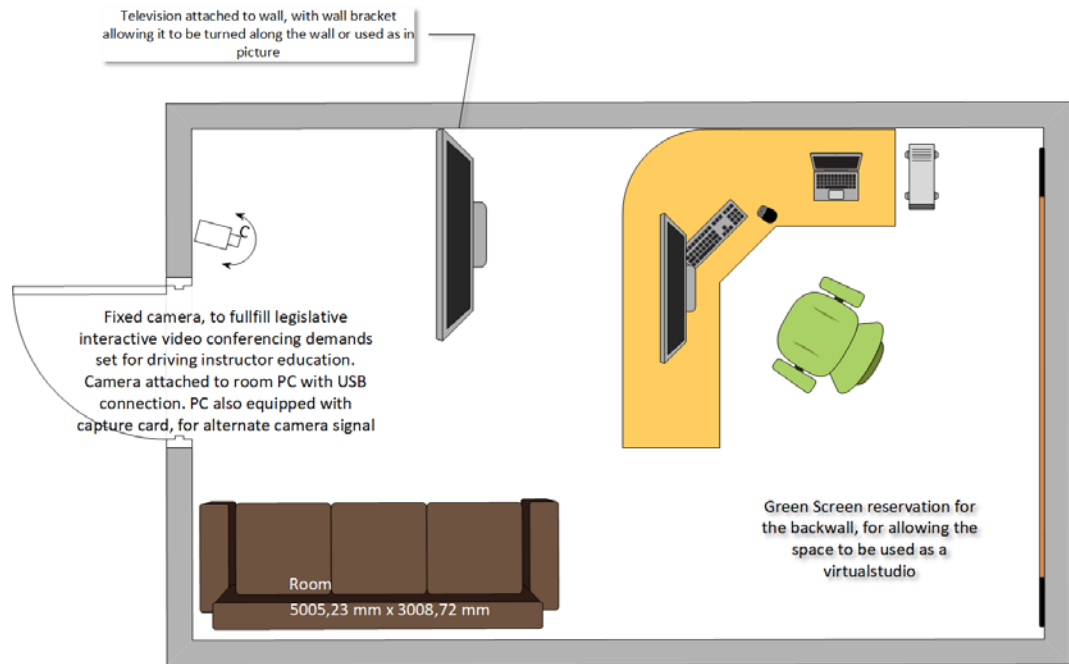
FIGURE 18. Distance learning studio

Studios also have typical Finnish school building construction from the '70s, so the walls of the rooms are brick or another hard surface. This means that the rooms have very loud and distinct echo. To cancel some of the echoes, without over dampening the room. Some acoustic materials were placed on the walls. Idea was to dampen speech frequencies (250 – 4000 Hz) so that resulting echo delay would be near 0,6 seconds. This according to Finish standards association standard SFS 5907 is below the recommended classroom reverberation time. Image 2 shows typical equipment used at studio. On the background there is sound dampening material applied to walls. Picture was taken during the construction of the studio spaces and does not display a television or larger display on the wall.

IMAGE 2. Distance learning studio equipment

Studios have also been furnished with USB connected dynamic microphones with internal mixer. Microphones have been installed using a spring-loaded desk stand and pop filters to ensure optimal placement for the microphone and to reduce noises such as pops or keyboard noise from the table. Ideally, this should ensure good sound quality on video and allow the lecturer to concentrate on presentation without worry on sound quality.

Both of the studios also have a video camera mounted on the frame of the external display. This is done with mounting brackets so that camera viewpoint is directly focused on the lecturer. Cameras used here are Panasonic AW-HE2 multi-purpose HD cameras. This model was chosen because cameras offer both USB and HDMI connectivity and it has an option of using the camera as part of remote-controlled IP cameras. These connectivity options combined with compatibility with all of the conferencing software's used at HAMK made this camera as the ideal choice for education video purposes.

Workstations on the studios have also been equipped with capture cards so that multiple cameras or video sources can be attached to the production environment. This allows for multi-aspect video production or attaching multiple computers for multiple lecturer presentations. It also allows demonstrations between two different systems presented simultaneously.

Software used here is either Telestreams Wirecast or Open Broadcaster Software studio application. This allows for higher frame rates and more complex video configuration to be used. Both software's are capable of handling multiple input streams and making a layered presentation of them. This allows, for example, overlaying a keyboard shortcut information on top of screen presentation simultaneously with the recording of the screen capture. This also allows for presenting multiple live video feeds on top of presentation material on the screen.

Last, and the largest physical space where educational video can be produced is campus's auditorium. Space is furnished for approximately 350 participants and has fixed camera and streaming equipment. The system here is based on Panasonic system of remotely operated integrated cameras with live switching and object tracking. The system comprises of two Panasonic AW-HE40, one AW-UE70 and one AW-HEA10 camera, compact live switcher AW-HS50N and compact remote camera controller AW-RP50. Image 3 displays functioning camera system with video switcher device. Lower left corner displays Blackmagic design recording device.
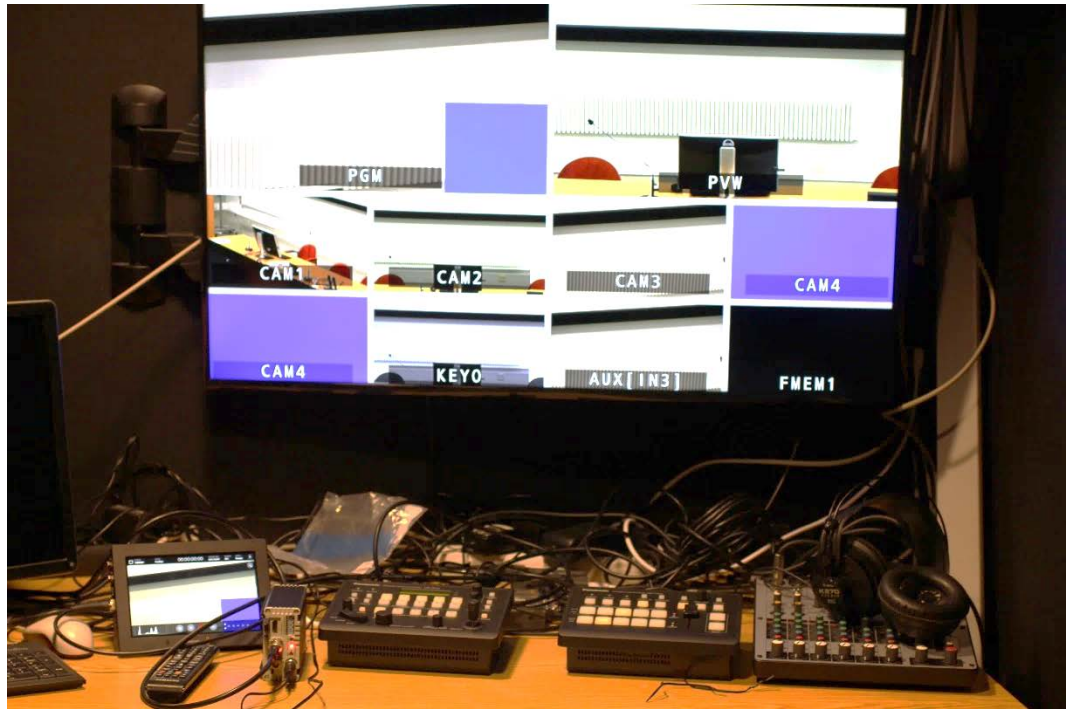
IMAGE 3. Video switching and controlling system

One of the cameras can be placed in an auto-tracking mode where the camera will be remotely controlled by auto-tracking software. The software recognizes face image with the cranial object and is capable of tracking the face image with the camera. This functionality requires a sample of tracked persons face, either captured with systems camera or imported as an image. This allows for automatic operation of the camera without the need to have a camera operator present. The system still needs an operator to start and stop the tracking and recording or streaming functions but otherwise can be operated without an operator. The end result for this system was presented in figure 15 from page 32 in form of frame capture.

5   DELIVERY OF CAPTURED MATERIALS

5.1   Delivery systems presently in use at HAMK

HAMK has multiple video delivery systems at the present. These systems are all selected for use, by their availability and suitability. Before current strategy was published, educational video usage at HAMK was limited to occasional YouTube or YLE Archive video presented at class using presentation computer or DVD player. This practice was then evolved by using file shares on local LAN network. These early videos were often recorded using specialized equipment, such as high speed or heat-sensing cameras. Videos produced with these types of cameras were often in small resolutions and difficult video formats requiring a purpose-built video players or codecs for access. Videos were there for not easily accessed and required special knowledge just to get them playing.

Delivery of videos went through a revolution when HAMK signed a contract with YouTube Education system and started using YouTube as primary video delivery platform. This was a direct result of HAMK strategy for reaching 2020 being published at the end of 2013. The strategy states that one of the main parts of strategy is to use digitalization as one of the main components of success. Furthermore, the strategy states that for educational purposes, one of the main components is a usage of virtual campus where using educational video is a natural component. (Häme University of applied sciences Ltd. 2015.)

During 2016 HAMK deployed a dedicated video delivery platform, using Kaltura Video platform as a basis. This was done in collaboration with CSC – IT center for Science LTD. and NORDUnet, a collaboration service between national research and education networks of the five Nordic countries. This video platform later called video.hamk.fi or just Kaltura is an embodiment of CSC service called Funet Etuubi Kaltura. Our platform is, in fact, a branch of Etuubi service modified to HAMK. During 2015 and

the official launch in 2016, Kaltura service was in its pilot stages and HAMK with two other Finnish universities of applied sciences, Lahti and Laurea, was its first pilot users. Since the pilot has reached its conclusion, and Etuubi is being deployed to all Funet participating universities. (Leiponen 2016.)

From these three different main systems in use, Kaltura system has been developed to deliver an educational video alongside with lecture notes and can even be used as a platform for questionnaires and video-based learning environments.

Kaltura system is, at the moment, licensed just for video on demand applications at HAMK. This severely limits its usage for the event and live session delivery channel. YouTube as a service allows recognized individuals to live stream content on their channel, within the boundaries of copyright. This is an educational environment, where teacher or lecturer holds all the copyrights related to teaching material is not a problem. With material licensed to be used in the classroom, such as Cisco Network Academy materials, republishing them in video format at public service such as YouTube for Education would require a different kind of licensing deal with Cisco Corporation. Also, YouTube copyright governing system is automated on most of the live channels, and it will perceive the show material as a copyrighted content. This means that content displayed is automatically discarded, and video is stopped. Same functionality applies to Video on Demand content.

YouTube system that governs and detects the copyrighted material is also able to recognize the audio input of copyrighted material. This affects live streaming on events where there is background music or live performance of music. The system allows for corrections but only after the fact, and as such is not applicable to live, once in a lifetime, performances.

5.2   Demo system for video delivery with next-generation video platform

Nginx or engine X, as the hypertext transfer protocol (HTTP) server is called, offers an open source solution for streaming Real-Time Messaging Protocol (RTMP) -based streams. This solution allows streaming in http live streaming (HLS) and dynamic adaptive streaming over http (DASH) formats with backup solutions to file-based streaming. Both HLS and DASH formats make use of multiple compression formats, including H.264 AVC and H.265. This flexibility in compression usage makes both of these streaming formats available to use on multiple different platforms simultaneously.

Engine X system for streaming live multimedia streams is to use RTMP receptor, forwarding streams to third-party encoder or transcoder and the delivering resulting MPEG fragments over HTTP protocol to the client. Same kind of methodology is in use with multiple common streaming platforms such as YouTube and Vimeo.

Modularized functionality is what makes Engine X system powerful. Using modules to offer different functions allow these functions to be spread out to multiple systems thus delivering computational power from multiple different systems. This also allows scaling in the horizontal dimension, by dividing for example transcoding load to multiple same kinds of systems. This division can be made possible with the use of task scheduler service such as RabbitMQ. Depending on the basic system technology, even one FFMPEG transcoder is usually powerful enough to transcode multiple different bitrate streams simultaneously. This makes possible to use this kind of system as a basis for dynamic stream production.

Dynamic streams are streams that incorporate multiple different bitrates to serve different clients capable of using different maximum bitrate streams. The change between bitstreams can be made to be automatic or manual depending on what kind of service is been offered. The basic requirement is that bitstreams, which make up published AV stream, are synchronized for their keyframes. This is usually accomplished with simultaneous

encodings of bitstreams, with identical settings for encoding apart from bitrate and resolution. The important part is to make sure streams have keyframe structure that matches. This allows streams to be interchangeable at the keyframe locations. For stream structure, this means balancing between the complexity of the bitstream and an increase of needed bitrate to reach same subjective quality for each component streams.

The system described here is put together on purpose to explain and demonstrate streaming service components and functionality.

Demonstration system uses a third-party module, made by Roman Arutyunyan. (Arutyunyan, 2017) This module introduces an RTMP protocol listener, Http Live Streaming, and MPEG-Dynamic Adaptive Streaming over Http segmenters and external programs integration system. This system also uses FFMPEG implementation of H.264 codec. FFMPEG is the leading multimedia framework for decode, encode, transcode, mux, demux, stream, filter and play multiple audiovisual formats. (FFmpeg.org 2017.)

The system uses latest engine X HTTP server over latest Ubuntu server system build with RTMP (Real Time Multimedia Protocol) module support and FFMPEG. As this system is meant for demonstrating adaptive media delivery using open source systems, there is no focus on system streamlining or system security. System build for this demonstration uses a Dell PowerEdge R720 server system with Intel processors as hardware. To make processing load easier on the server system, the stream is compressed with external encoder before delivering it to the delivery system.

Delivery system then translates received media stream to multiple transcoding processes that are necessary to deliver adaptive system stream to the player. Demo system uses an FFMPEG transcoder to make elementary streams for both audio and video. Adaptively requires the streams to be key frame synced between the elementary video and audio

streams. For RTMP module to be able to make fragments for HLS or DASH delivery purposes, each individual elementary video stream needs to have a complimentary audio stream with each segment delivered. This is done using FFMPEG command of "-force_key_frames "expr:get(t,n_forced*3)" for each elementary stream. Keyframe interval stays the same for each if the framerate of the streams is identical. (ffmpeg.org 2017) Command will insert a keyframe in each resulting stream at the interval of three seconds. Figure 19 explains resulting video structure, assuming video is encoded at two frames per second.

Frame	Time in Seconds (t)	Expression, based on gte(t,n_forced*3)	Evaluates as	Keyframes forced so far (n_forced)
1	0	If 0 >= 0*3	TRUE – force keyframe	1
2	0	If 0 >= 1*3	FALSE	1
3	1	If 1 >= 1*3	FALSE	1
4	1	If 1 >= 1*3	FALSE	1
5	2	If 2 >= 1*3	FALSE	1
6	2	If 2 >= 1*3	FALSE	1
7	3	If 3 >= 1*3	TRUE – force keyframe	2
8	3	If 3 >= 2*3	FALSE	2
9	4	If 4 >= 2*3	FALSE	2
10	4	If 4 >= 2*3	FALSE	2
11	5	If 5 >= 2*3	FALSE	2
12	5	If 5 > 2*3	FALSE	2
13	6	If 6 >= 2*3	TRUE – force keyframe	3
14	6	If 6 >= 3*3	FALSE	3
15	7	If 7 >= 3*3	FALSE	3
16	7	If 7 >= 3*3	FALSE	3
17	8	If 8 >= 3*3	FALSE	3
18	8	If 8 >= 3*3	FALSE	3
19	9	If 9 >= 3*3	TRUE – force keyframe	4

FIGURE 19.16 FFMPEG expression evaluation model (Herron 2017)

Selection of adaptation layers composition, in terms of resolution, frame rate, and bitrate, should be a considered process in regards to original material and projected audience. Encoding settings are usually selected

so that adjacent bit rates for encoded streams are factored to be 1,5 to 2 times higher or lower than currently playing the stream. This approach is adapted originally from Apple recommended best practices guide for HTTP Live Streaming Media. (Apple Inc. 2016.) Figure 20 represents recommendations from Apple Technote tn2224 regarding HLS streaming. These recommendations are done in regard to Apple manufactured devices and their capabilities but are usually used widely as a basis for any adaptation process regardless of the application or audience.

Clients			Dimensions for 16:9 aspect ratio	Dimensions for 4:3 aspect ratio	Frame rate	Video bit rate (average)	Video bit rate (peak)	Audio bit rate	Total bit rate
	CELL		416 x 234	400 x 300	12	145	200	64	264
	CELL	ATV	480 x 270	480 x 360	15	365	400	64	464
WiFi	CELL	ATV	640 x 360	640 x 480	29.97	730	800	64	864
WiFi	CELL	ATV	768 x 432	640 x 480	29.97	1100	1200	96	1296
WiFi		ATV	960 x 540	960 x 720	29.97 or source	2000	2200	96	2296
WiFi		ATV	1280 x 720	960 x 720	29.97 or source	3000	3300	96	3396
WiFi		ATV	1280 x 720 or source	1280 x 960 or source	29.97 or source	4500	5000	128	5128
WiFi		ATV	1280 x 720 or source	1280 x 960 or source	29.97 or source	6000	6500	128	6628
WiFi		ATV	1920 x 1080	1920 x 1440	29.97 or source	7800	8600	128	8728

Source frame rate may be as high as 60 fps
Audio sample rate: 48 khz
Keyframe: Every 2 seconds (i.e., frame rate × 2)
Segment Size: 6 seconds

Bit Rate Variability - Should not exceed 10% of target bit rate

FIGURE 20. Recommended Encoding Settings for HTTP Live Streaming Media (Apple Inc. 2016)

After video stream has been split into its adaptation layers, layers are then transported back to RTMP module for fragmentation and playlist and MPD creation. For HLS streams, RTMP module will create an m3u8 type of playlist with references to each adaptation layer and subsequent segment information. For the purely live streaming system, the playlist is an

updating file, that only contains up to date information of the media stream. For recording system, it will also contain information about all the previous segments of the media.

For DASH stream, RTMP module will create an MPD file, that will also point to presentation segment of each of the adaptation layers. Both MPD and M3U8 type of file will automatically update at a fixed interval according to hls_fragment setting at the RTMP module config. The default for this is set at 5 seconds. (Arutyunyan 2017.)

For demo system, last missing part is a player. To be able to view the streams build by NGINX-RTMP module and FFmpeg transcoder, a player is needed.

Player for the live stream can be embedded in some HTML page and delivered to the viewer over HTTP. There are multiple players in the market with HLS and DASH support. To mimic the functionality of Kaltura system used by HAMK, a similar HTML 5 based video player was used. JWPlayer is an html5 based video player with support to HLS and DASH-based delivery (Longtail Ad Solutions Inc. 2017). Streams can also be viewed with a desktop application such as VideoLAN media player, VLC. For complete configuration reference look for appendices

6 CONCLUSION AND FUTURE DEVELOPMENT TOPICS

This thesis has been about finding a common workflow for producing educational video, using equipment and facilities present at HAMK Riihimäki campus. Thesis research and knowledge base have been directed on aspects of the video and production methodology that are both obtainable and directly influential on the product, an educational video. For conclusion, the educational video takes multiple different kinds of manifestations, from the basic recording of a lecture to live multiple view performance with interaction. These different videos have common functionalities and techniques presented in this thesis and they can be utilized for any kind of educational video, independent if it is a short clip presenting a specific subject matter or longer lecture capture dealing with multiple topics. Figure 21 presents a decision matrix for both live and video on demand material. This matrix is meant for a starting point dealing with different video formats and matrix only contains basic steps to ensure handling of few production elements. The idea behind this type of matrix is to ask video producer to consider aspects of the video and what kind of material is been produced.
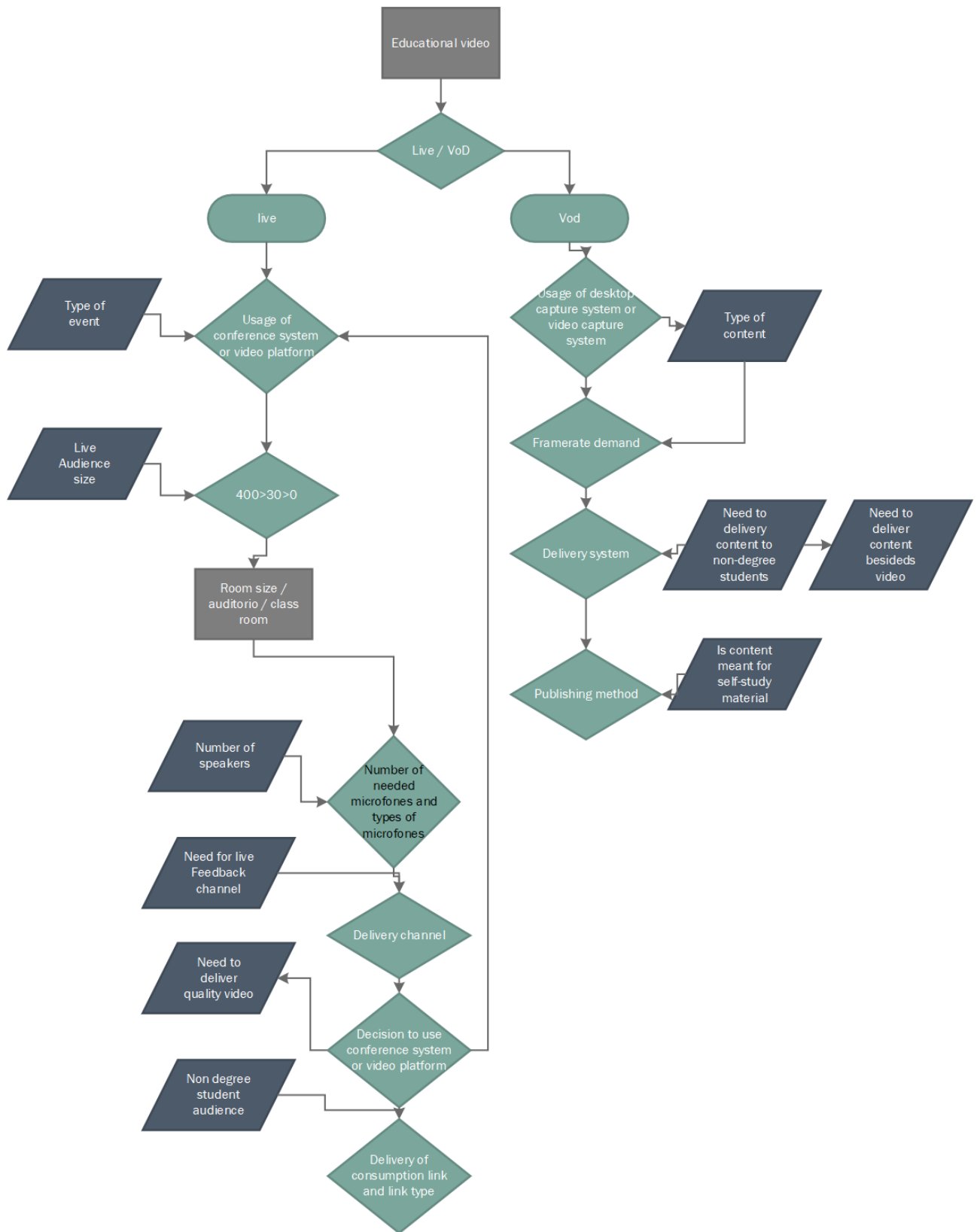
FIGURE 21. Video production decision matrix

For basic lecture capture, with minimal movement or live video feed being captured, the workflow is simple. At campus, all of the lecturing halls are equipped with conference microphone, and room computer has connections to both video projector and loudspeakers. Figure 22 gives out basic steps to capture lecture using equipment at HAMK Riihimäki campus. This workflow is a skeleton structure, that does not contain anything else than just capture of classroom held lecture, with or without camera equipment.

Ensure computer has Kaltura CaptureSpace light software installed by lauching 'add new' fuction on video.hamk.fi site

check capture parameters for CaptureSpace lite use attached microphone and capture is been done at 1080p resolution.

start capture by selecting Screen capture option or screen and webcam combination

Give out lectura as usual. Remember to speak out anything not on computer screen. Use pointer fuction on the desktop recorder software to point out key elements on the lecture notes.

Upload video when lecture is concluded and use kaltura video site to publish video to preferred delivery site.
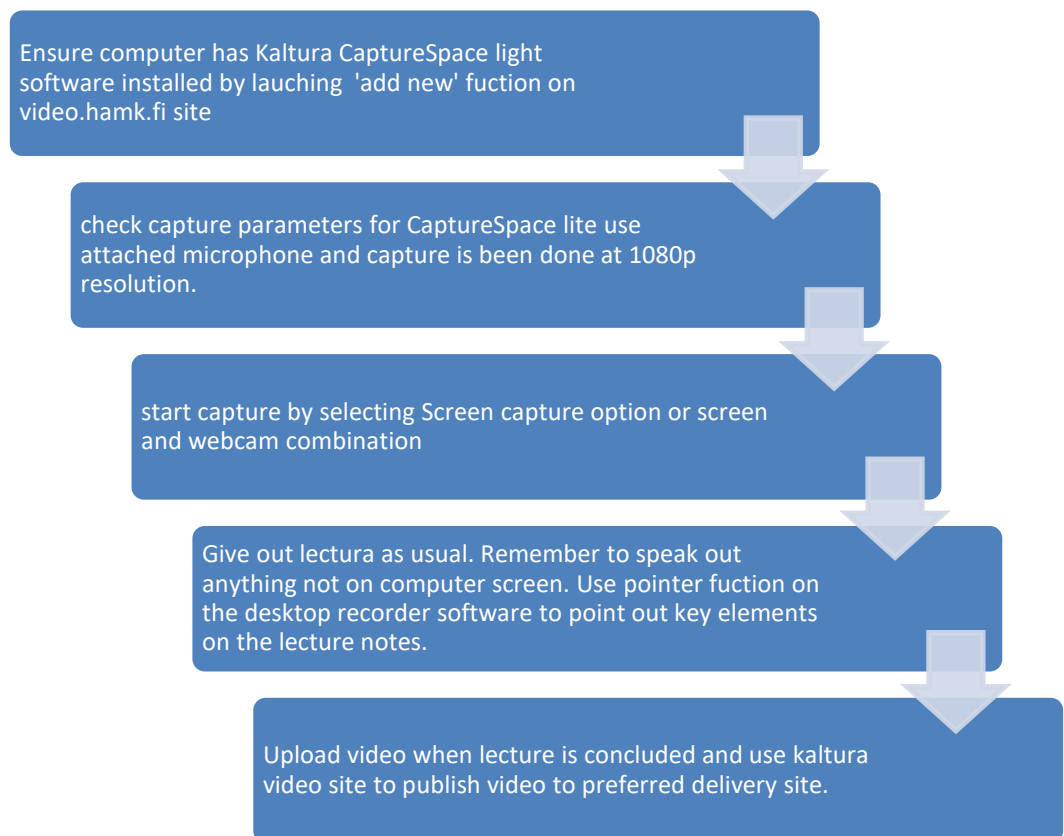
FIGURE 172. Lecture capture

When lecture capture is needed to be broadcasted live, then capture should be done using the conferencing software. As discussed previously, the conferencing software can have its own delivery channel for recorded material. For the workflow, this means adaptation is needed for publishing method. At the time of writing this, recommended publishing method is to use Kaltura video site. For conferencing software recordings, this means

transcoding the recorded file format to the more common H.264 codec and uploading resulting video to the site.

To progress further, the workflow for high movement content or multiple aspect view contents, capture system needs to change. The basic conclusion here is to use widely available OBS software in conjunction with capture hardware such as Magewell capture card. This combination gives out a flexible platform for capture, with the option to insert different medias to the captured video feed. Basic tutorial type of educational content usually needs at least a capture of educator, a video feed of subject of the tutorial and notices or written pointers alongside verbal lecture. These are easily combined with OBS layered approach to video production.

Open Broadcasters Software can be also used for live streaming and as such is entry level approach for HAMK Riihimäki live video production. Alternative software available for this is Telestream Wirecast software that typically can achieve similar or same end results than OBS software. The limiting factor for use of this software is a number of licenses available.

The live production of event capture is next step up in complexity requirements and production equipment needed. For HAMK Riihimäki, this level of production is usually met with the usage of auditorium space and PTZ camera equipment installed there. The same functionality can be transferred elsewhere, on or off the campus, but typically conferences and such events are held at largest free space on the campus, ergo auditorium. This approach utilizes four remotely controlled pan-tilt-zoom cameras, that produce video feeds that are mixed with computer screen video feed. Figure 16 demonstrated this on the static level. Mixed video feeds are then processed with audio embedder, that embeds video with audio from auditoriums audio system. Audio capture is done from direct feeds of auditorium microphones. Audio is mixed at its own mixer and the combination is embedded in video feed as its left and right audio channel. This video is then recorded at Blackmagic Video Assist device that simultaneously, using passthrough connections give video out to the

streaming system. For the live streaming system, HAMK Riihimäki campus uses Digital Rapids StreamZHD multi-format ingest system. This system is capable of delivering H.264 coded full high definition video to multiple live streaming delivery sites. The main endpoint for HAMK live videos is, at the time of writing this, YouTube For Education channels.

When choosing delivery channels, presenter have to consider availability and accessibility of the videos. Most of the video on demand type of material, will be delivered using Kaltura platform and embedded to Moodle learning environment. This is done with inbuilt video tool inside Moodle, that uses Kaltura video platform as a content source. This also gives out an added benefit of user authentication to both educational contents, video and more static content types at the same time. Using Kaltura platform also delivers a new tool to educator's arsenal. A video-based quiz tool, that uses educators own video material and gives an opportunity to follow students learning and needs for material. Differences shows up in video quiz results when learners are at the approximately same level before attempting to complete quizzes. This type of material does not fully replace the interaction of face to face learning but can be a fully autonomous resource for both teachers and students.

In conclusion of this thesis, there are presently multiple viable methods of creating educational videos and systems in place from where to present them to students. This thesis has presented possible ways to produce videos and techniques to improve or enhance the viewer participation and retention of knowledge. Thesis has also touched upon the production environments, tools and limitations present on today's systems. Demo system described has also showcased possible solutions on truly independent delivery system where limitations of production are not inherent to platform or practice.

6.1   Future development paths

Today's video systems are sufficient for today's needs. When considering tomorrows probable needs, decision points are many and not all of them

are readily available or easily predicted. Most likely future will see even more emphasis on distance learning or at least learning curriculum will be enhanced with distance learning methodology. To respond to this, video production should be more autonomous, non-lecturer participation besides teaching needed. This particularly can be partly answered by using automatically tracking camera systems with automated video switching and ingestion for streaming.

Video for future educational needs, will most likely need to develop to contain more than just two-dimensional images. This could mean either going on the way of augmented reality and incorporating virtual environments as back drop for educator or by augmenting video with more static content like PowerPoint presentations or portable document files. This in some form is reality even today, but to incorporate these more fully to video by making them interactive parts of the video feed, should make these more integral for learning experience.

During the writing process of this thesis, research for background information revealed clearly two pronged approach to educational video production. Compared to HAMK approach to use digital surfaces and software instead of traditional black board, Jyväskylä University of applied sciences uses cameras to capture same information. JAMK approach is to use carbon free glass as a surface from where cameras can capture chalk felt tip drawings. Recording is done through clear glass panel, and image is digitally flipped during recording process. (Järvinen 2017.)

JAMK approach is there for to develop traditional board teaching method to digital age. This will allow lectures to be captured, without having the lecturer to change their teaching methods where as HAMK method uses totally different platform and requires teacher to adapt to the new environment. Upside of HAMK method is that the material produced can be reused as written material and is available to low bandwidth situations as purely digital document.

Methods used today, are more or less requiring a direct connection between cameras, capture equipment and production platforms. Company called NewTek recently published their approach to using computer network between active devices. Their technology is called Network Device Interface (NDI), this allows devices to directly influence other devices inside interconnected production environment. (NewTek inc. 2017.)

For educational video, this would mean a live production environment where common local area network would allow lecture capture devices to communicate with other lecturers and content to be produced as a collaboration between video platforms. This would allow for example mobile devices such as phones to be used as source devices for video, and educators to input their materials over the available sources. Live demonstrations of different techniques without massively expensive equipment to record the demonstration, should allow for engaging content creation with multiple possible applications.

7   REFERENCES

Android.com, 2017. Developer dashboards. [referenced 5.11.2017].
Available at: https://developer.android.com/about/dashboards/index.html

Android.com, 2017. Supporting Multiple Screens. [referenced 5.11.2017].
Available at:
https://developer.android.com/guide/practices/screens_support.html

Apple Inc., 2016. Best Practices for Creating and Deploying HTTP Live
Streaming Media for Apple Devices. [referenced 1.11.2017]. Available at:
https://developer.apple.com/library/content/technotes/tn2224/_index.html

Arutyunyan, R., 2017. Github. [referenced 20 August 2017]. Available at:
https://github.com/arut/nginx-rtmp-module

Arutyunyan, R., 2017. nginx-RTMP module wiki. [referenced 15 August
2017]. Available at: https://github.com/arut/nginx-rtmp-
module/wiki/Directives#hls_fragment

Bing, B., 2015. Next-Generation video coding and streaming. Hoboken,
New Jersey: Wiley.

Bitmovin inc, 2017. Adaptive Streaming. [referenced 19.11.2017].
Available at: https://bitmovin.com/adaptive-streaming/

BlackMagic Design, 2017. Video Assist product page. [referenced
7.7.2017]. Available at:
https://www.blackmagicdesign.com/fi/products/blackmagicvideoassist

Brandenburg, K., 1999. MP3 and AAC explained. Erlangen, Fraunhofer
Institute for Integrated Circuits FhG-IIS A.

Cisco, 2016. VNI Complete Forecast Highlights Tool. Available at:
http://www.cisco.com/c/m/en_us/solutions/service-provider/vni-forecast-
highlights.html

Dongsong, Z., Zhou, L., Briggs, R. O. & Nunamaker Jr., J. F., 2006. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. Information & Management, January, p. 15-27.

FFmpeg.org, 2017. FFmpeg.org. [referenced 27.8.2017]. Available at: http://ffmpeg.org/ffmpeg.html

FFmpeg.org, 2017. FFmpeg.org. [referenced 5.10.2017]. Available at: http://ffmpeg.org/about.html

Google, 2017. Counter Notification Basics. [referenced 29.8.2017]. Available at: https://support.google.com/youtube/answer/2807684?hl=en-GB

Google, 2017. Youtube Help. [referenced 8.10.2017]. Available at: https://support.google.com/youtube/answer/2797370?hl=en-GB

Greenberg, A. D. & Zanetis, J., 2012. The Impact of Broadcast and Streaming Video in Education. [referenced 10.4.2017]. Available at: http://www.cisco.com/c/dam/en_us/solutions/industries/docs/education/cis covideowp.pdf

Guo, P. J., Kim, J. & Rubin, R., 2014. How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. Atlanta, ACM Digital Library, p. 41-50.

Hansch, A. et al., 2015. Video and Online Learning: Critical Reflections and Findings From the Field. Available at: http://dx.doi.org/10.2139/ssrn.2577882

Herron, P., 2017. Forcing keyframes with ffmpeg. [referenced 19.11.2017]. Available at: http://paulherron.com/blog/forcing_keyframes_with_ffmpeg

Holman, T. & Baum, A., 2013. Sound for Digital Video. Burlington: Focal Press.

Häme University of applied sciences Ltd., 2015. HAMK Strategy 2020. Available at: https://spark.adobe.com/page/ZNikZ/

Ibrahim, M., Antonenko, P., Greenwood, C. M. & Wheeler, D., 2012. Effects of segmenting, signaling, and weeding on learning from educational video. London, Routledge Informa Ltd, p. 220-235.

Jagetiya, A. & Sharma, S. P., 2015. Seminar on MPEG 1 and 2. Chandigarh, Panjab University.

Järvinen, J., 2017. JAMK Blog. [referenced 11.11.2017]. Available at: http://blogit.jamk.fi/aavee/2017/09/11/tuberoom-ja-lightboard-opettajille-avuksi-videotuotantoihin/

Kaltura, 2016. State of video in Education 2015 Report. [referenced 10.12.2016]. Available at: http://site.kaltura.com/Kaltura_Survey_State_of_Video_in_Education_201 5.html

Korpinen, P., 2005. Äänipää. [referenced 5.6.2017]. Available at: http://www.aanipaa.tamk.fi/aanit_1.htm

Leiponen, T., 2016. One service, many names – a shared video platform for the Federation of Universities of Applied Sciences. [referenced 5.11.2017]. Available at: https://www.csc.fi/web/atcsc/-/sama-palvelu-monta-nimea-ammattikorkeakoulujen-fuas-liittouman-yhteinen-videoalusta

Leis, J. W., 2011. Digital Signal Processing Using MATLAB for Students and researchers. 1. Edition ed. John Wiley and Sons Inc.

Longtail Ad Solutions, Inc, 2017. HTML5 VIDEO PLAYER. [referenced 13.9.2017]. Available at: https://www.jwplayer.com/video-solutions/html5-video-player/

Magewell, 2017. Magewell Pro Capture AIO 4K Plus product page. [referenced 7.7.2017]. Available at: http://www.magewell.com/pro-capture-aio-4k-plus

Matrox, 2017. Matrox vs4 product page. [referenced 7.7.2017]. Available at: http://www.matrox.com/video/en/products/vs4/

Mayer, R. E., 2001. Multimedia Learning. Cambridge: Cambridge University Press 2001.

Mayer, R. E., 2005. The Cambridge handbook of multimedia learning. Cambridge: Cambridge University Press.

Mayer, Richard E. & Johanson, Cheryl I., 2008. Revising the Redundancy Principle in Multimedia Learning. Journal of Educational Psychology 2008, vol 100, no. 2, p. 380-386

Mayer, R. E. & Moreno, R., 2003. Nine Ways to Reduce Cognitive Load in Multimedia Learning. EDUCATIONAL PSYCHOLOGIST, p. 43-52.

Monteiro de Castro Silva, I. & Feitosa, M., High-frequency audiometry in young and older adults when conventional audiometry is normal. Brazilian journal of otorhinolaryngology, 72(5), p. 665–72.

NewTek inc., 2017. NewTek. [referenced 11.11.2017]. Available at: https://www.newtek.com/ndi/

Panasonic Corporation, 2017. Broadcast and Professional AV global web site. Available at: http://pro-av.panasonic.net/en/sales_o/camera/app/index.html

Pandit, K. & Speranza, R., 2014. Cornell University. [referenced 19.11.2017]. Available at: http://people.ece.cornell.edu/land/courses/ece5030/FinalProjects/s2014/kkp37_rjs483/kkp37_rjs483/AudioGram.html

Siemens, G., Gašević, D. & Dawson, S., 2015. Preparing for the digital university: a review of the history and current state of distance, blended, and online learning. p. 234.

StatCounter, 2017. Frequently Asked Questions. [referenced 5.11.2017]. Available at: http://gs.statcounter.com/faq#methodology

Vaishnavi, V. & Kuechler, W., 2015. www.desrist.org. [referenced 19.11.2017]. Available at: http://www.desrist.org/design-research-in-information-systems/

Youtube, 2016. Statistics. [referenced 11.11.2016]. Available at: https://www.youtube.com/yt/press/statistics.html

APPENDICES

nginx configuration with rtmp plugin

```
#user  nginx;
worker_processes  1;

#error_log  logs/error.log;
#error_log  logs/error.log  notice;
#error_log  logs/error.log  info;

#pid        logs/nginx.pid;


events {
    worker_connections  1024;
}

rtmp {
        server {
                listen 1935;
                allow play all;
                application live {
                        live on;
                        exec ffmpeg -re -i
rtmp://localhost/live/$name -c:a aac -strict -2 -b:a 64k -c:v
libx264 -b:v 512k -force_key_frames "expr:gte(t,n_forced*3)" -f
flv rtmp://localhost/hls/$name_low -c:a aac -strict -2 -b:a 128k -
c:v libx264 -b:v 1024k -force_key_frames "expr:gte(t,n_forced*3)"
-f flv rtmp://localhost/hls/$name_mid -c:a aac -strict -2 -b:a
128k -c:v libx264 -b:v 2048k -force_key_frames
"expr:gte(t,n_forced*3)" -f flv rtmp://localhost/hls/$name_hi;
                }
                application hls {
                        live on;
                        hls on;
                        hls_path /tmp/hls;
                        hls_nested on;

                        hls_variant _low BANDWIDTH=576000;
                        hls_variant _mid BANDWIDTH=1152000;
                        hls_variant _hi BANDWIDTH=2176000;
                }
        }
}
http {
    include        mime.types;
    default_type  application/octet-stream;

    #log_format  main  '$remote_addr - $remote_user [$time_local]
"$request" '
    #                  '$status $body_bytes_sent "$http_referer" '
    #                  '"$http_user_agent"
"$http_x_forwarded_for"';

    #access_log  logs/access.log  main;
```

```
    sendfile        on;
    #tcp_nopush     on;

    #keepalive_timeout  0;
    keepalive_timeout  65;

    #gzip  on;

    server {
        listen       80;
        server_name  localhost;

        #charset koi8-r;

        #access_log  logs/host.access.log  main;

        location / {
            root   html;
            index  index.html index.htm;
        }

        #error_page  404              /404.html;

        # redirect server error pages to the static page /50x.html
        #
        error_page   500 502 503 504  /50x.html;
        location = /50x.html {
            root   html;
        }
         # rtmp stat
        location /stat {
                rtmp_stat all;
                 rtmp_stat_stylesheet stat.xsl;
            }
        location /stat.xsl {
                # you can move stat.xsl to a different location
                root /usr/local/nginx/html/rtmp-stat/;
            }

       location /hls {
            # Serve HLS fragments
            types {
                application/vnd.apple.mpegurl m3u8;
                video/mp2t ts;
            }
            root /tmp;
            add_header Cache-Control no-cache;
            add_header 'Access-Control-Allow-Origin' '*';
        }

        # proxy the PHP scripts to Apache listening on
127.0.0.1:80
        #
        #location ~ \.php$ {
        #    proxy_pass   http://127.0.0.1;
        #}

        # pass the PHP scripts to FastCGI server listening on
127.0.0.1:9000
        #
        #location ~ \.php$ {
```

```
        #    root             html;
        #    fastcgi_pass    127.0.0.1:9000;
        #    fastcgi_index   index.php;
        #    fastcgi_param   SCRIPT_FILENAME
/scripts$fastcgi_script_name;
        #    include         fastcgi_params;
        #}

        # deny access to .htaccess files, if Apache's document
root
        # concurs with nginx's one
        #
        #location ~ /\.ht {
        #    deny  all;
        #}
    }


}
```