Spontaneity and Delay Considerations in Distributed TV Productions

Susanne Naegele-Jackson, Michael Graeve, and Peter Holleczek

Regional Computing Center Erlangen (RRZE) of the University of Erlangen-Nuremberg

1 Preface

During the online edition of video material for television over data networks a major concern is the accumulation of delay over the production chain. Before a video signal of a camera arrives at a remote studio to be edited live, it travels over video adapters and analog/digital converters to codecs which in turn compress the signal to comply with small bandwidth constraints. This technical process adds a significant amount of delay that makes it difficult for the camera crew and director to interact spontaneously. An additional amount of delay is added through their human reaction times as well. Meaningful interaction requires however, that the overall delay times stay within tight limits.

The study evaluates the chain of events during the distributed online production of television shows in the project "Uni-TV" and investigates where delay times can be reduced significantly during the process. The use of low-latency codecs and cameras that are remote controlled over the network is proposed to reduce the interaction times between a camera crew on location and the director at the distant studio. Additional equipment such as analog/digital converters only add very insignificant amounts of delay to the process and are therefore considered fixed parameters.

2 Basic Concepts of the Project "Uni-TV"

The main focus of the project "Uni-TV" is to produce university lectures online for television. This distributed production requires high bandwidths to ensure studio quality and low delay times to allow for a speedy interaction between the camera crew on location and the director at the remote studio. During the production process delay is introduced mainly through the compression and decompression of the video signals to reduce the data volume for transmission. Additional delay is added through the human interactions during the editing process. Whereas so far the camera crew and studio personnel have been separated at distinct locations, it is now being investigated how delay times could be minimized by remotely controlling the cameras from the studio over the data network.

Both the University of Erlangen-Nuremberg (FAU) and the Technical University of Munich (TUM) provide lectures and technical staff for the television productions. The video signals are transferred from especially equipped auditoriums to the studio in Munich at the Institut fuer Rundfunktechnik (Institute of Broadcasting Technology) (IRT) where a director of the Bayerischer Rundfunk (Bavarian Broadcasting Channel) (br) is standing by to edit the arriving video signals online. A script for the production is developed by the Hochschule fuer Fernsehen und Film (University for Television and Filming) (HFF). The material is broadcast on a regular basis on the educational channel br-Alpha of the br. The distributed production takes place via the Gigabit Testbed South/Berlin with its nodes at the Regional Computing Center Erlangen (RRZE) and the Leibniz Computing Center (LRZ) in Munich, and also via the campus networks of the universities and the citywide network Mnet of Munich.

3 Introduction of Delay During the Production Process

During the production process three cameras are used to film the presentation at the auditorium. Two of these cameras are remote controlled from within the classroom (Figure 1). The analog video signals travel from the camera to their remote control units and are then converted into valid digital signals with 270 Mbit/s each. Although the Gigabit Testbed South/Berlin offers three channels of 2.5 Gbit/s ATM bandwidth per channel the data volume is reduced by compressing the signal into MPEG-2 (4:2:2) format. Once the signal has been transmitted over the network, the process is reversed and the signal decoded into a digital video stream again. Afterwards the video signal is synchronized and displayed on studio equipment for the director to see. The director then produces an online edition using the three different camera signals and the electronic version of the presentation material.

Figure 1 - Remote Controlled Camera

Since the director also instructs the camera personnel on additional focusing, zooming or changes in camera positions, there is a constant interaction between camera crew and studio on an audio transmission line where the added delay must be kept to an absolute minimum if spontaneous reactions or surprising effects are to be captured and the live element of the event is to be preserved. Before the camera crews are able to follow the director's instructions, however, the digital audio signal must travel from the microphone of the studio to the encoder and over the network to the decoder, before it can finally be converted back to an analog signal and pass through an audio mixer onto the headsets of the camera crew at the auditorium. After the initial reaction of the cameras it takes another cycle of conversion, encoding and decoding of the signal, until the new camera angle can be displayed and the director can actually tell that the instruction has been carried out. Due to the large amounts of delay involved the process requires a lot of discipline and patience on both sides.

4 Fixed and Variable Intervals of Delay

The delay of the production process can be described as follows (Figure 2):

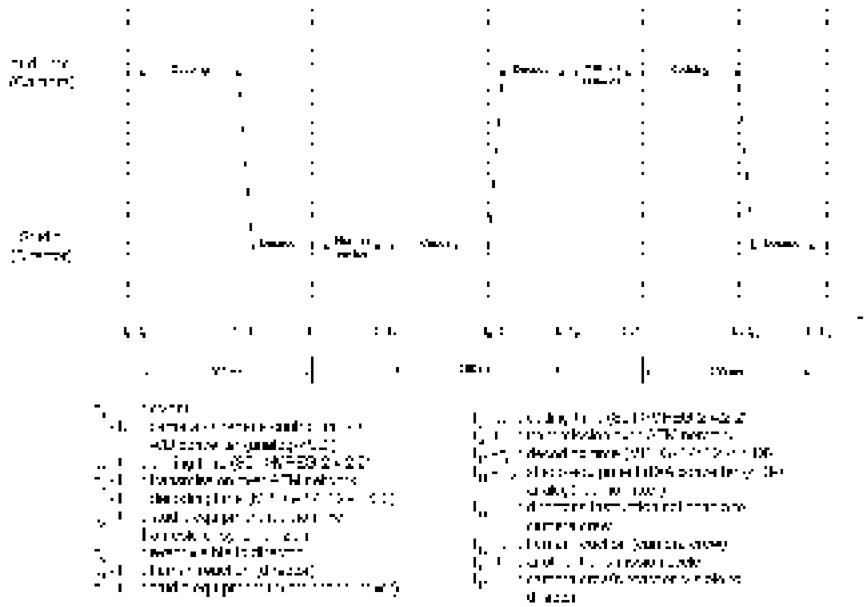

Figure 2 - Delay Intervals

At time t_0 the actual event takes place. During the time frame t_1-t_0 the camera signal is transported to the remote control unit of the camera and sent to an analog-digital converter from there. The A/D converter translates the analog video signal into a valid SDI signal which in turn can now be compressed into MPEG-2 (4:2:2) format and packed into ATM cells from time t_2-t_1. The ATM cells are transmitted during the time frame t_3-t_2 and then decoded during the time period t_4-t_3. After decompression the video signal is available at the studio for further processing. An additional delay t_5-t_4 is introduced for the frame synchronization before the video signal is displayed on a digital monitor and digital video editing system. At timestamp t_5 the event is visible to the director.

By then the director in the studio takes time t_6-t_5 to react to the pictures and to give instructions to the editing personnel. The audio signal carrying the instructions passes through the studio equipment during time frame t_7-t_6. Afterwards these audio signals are compressed by the MPEG-2 codec during the time slot (t_8-t_7) and then are retransmitted to the auditorium (t_9-t_8). The signal is decoded during time frame t_{10}-t_9 and travels through the studio equipment at the auditorium before the director's instruction is finally noticeable to the camera crew at time t_{11}. The camera personnel reacts to the director's instructions during time period t_{12}-t_{11}. The changes made by the camera crew will then need another transmission cycle until the signal arrives at the studio for the director to see the result at time t_{17}.

During the whole chain of events the longest delay is caused by the encoding and decoding processes of the codecs. The delay times are dependent on the codecs being used and their compression settings. For the project "Uni-TV" the video signals are compressed into MPEG-2 (4:2:2) I-frame only with 625 lines per frame and 720 pixels per line. The ATM PVCs (Permanent Virtual Circuits) configured for the transmission of the camera signals over the ATM network are implemented as Cbr (Continuous Bit Rate) traffic streams with 50 Mbit/s bandwidth to include embedded audio.

Codecs that compress in MPEG-2 (4:2:2) format usually take between 200ms and 400ms for both encoding and decoding (one way) depending on the hardware. The transmission time for the ATM cells from Erlangen to Munich and back to Erlangen again across the Gigabit Testbed has been measured to range from 2.2870ms (in the case of a 0.67% workload on a STM-1 interface (1 Mbit/s)) to 2.2981ms (in the case of a 99.99% workload on a STM-1 interface (149.745Mbit/s)). The cell delay variations ranged from 5.4μs to 10.9μs. The mere optical signal transmission time takes about 2ms over a round trip distance of approximately 400km and can be calculated as (Distance/Speed of optical signal over fiber) or (400km / 200*10^3km/s = 2ms).

The delay times t_3-t_2 and t_9-t_8 of the ATM transmissions are therefore considerably shorter than the total latency of 200ms to 400ms introduced by the encoding and decoding of the video signals into MPEG-2 format. The time periods of analog and digital conversions or frame synchronizations can also be neglected in comparison to the codec compression times. Alteration of these delays does not pay off and the intervals can therefore be considered fixed variables.

5 Minimizing Delays

During the end-to-end process there are two main areas where a lot of delay time is spent: The human reaction time and the time interval needed for encoding and decoding (Figure 3).

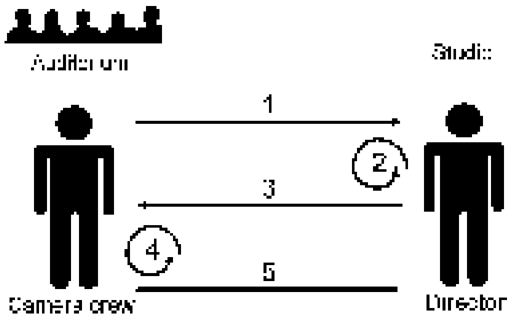

Figure 3 - Main Delay Categories

1. Initial encoding/decoding
2. Human reaction (director)
3. Encoding/decoding of instruction (for a new camera setting)
4. Human reaction (camera crew)
5. Encoding/decoding of new camera settings

Although a delay around one second may seem small (Figure 2), it can be very annoying during the interactions between the camera personnel and the director, since additional reaction time of the camera crew is introduced before the director is finally able to see his instructions take effect over the distance. The ITU-T recommendation G.114 indicates a limit of only 150 ms as appropriate for a one way transmission time (ITU-96) for bidirectional traffic.

The human reaction of the director and the camera crew cannot be accelerated with technical equipment. Using a stop watch and taking the amount of time it takes a person to stop the clock a reaction period of 100ms to 120ms can be measured. In this experiment the person pressing the button of the stop watch is prepared for the event and will press the button fairly quickly. In traffic situations, for instance, where events are not expected, reaction times are considerably longer and last from 300ms to 1000ms depending on the alertness of the person involved. In the studio environment the situation is similar, since the crew behind the cameras has to react quickly to a sudden request for change by the director. Even with a lot of experience and practice the time span can only be optimized to stay within a minimal limit due to the human factor.

One way to minimize delay is to place the camera personnel next to the director at the studio (Figure 4) and use remote controlled cameras.

This way the time it takes to convey the instructions of the director to the camera crew is almost minimized to zero (3a), since the camera personnel only needs to hear the spoken word in the studio and no encoding, decoding and transmission of the director’s new instructions are necessary. This reduces the regular encoding and decoding period by 200ms to 400ms depending on the hardware of the codecs. However, there is still some delay involved (3b) for transmitting control information from the control panel to the camera to

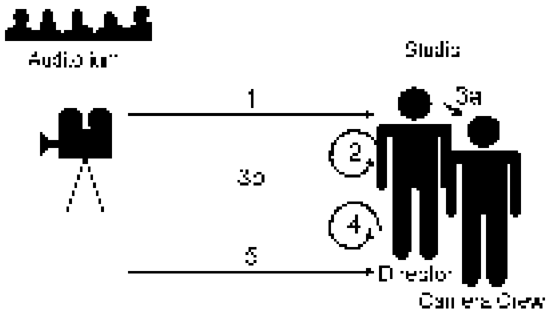

Figure 4 - Camera Crew Close to the Director

actually change the settings. Even by handling the cameras via remote control from the studio two encoding and decoding cycles remain (1 + 5). To further speed up the process the time intervals consumed by the codecs must be shortened.

Encoding and decoding delay times are very much determined by the equipment that is used and as such have - next to the underlying quality of the network - a strong impact on the overall quality of a transmission. A pair of codecs can be fine-tuned for optimal picture quality, latency or limited use of bandwidth. All of these parameters cannot be optimized at the same time, however, since their realizations are conflicting goals. In order to be able to process the SDI video signals of 270 Mbit/s in studio quality the broadcasting environment insists on the compression format MPEG-2 (4:2:2) with 40 Mbit/s bandwidth and I-frame only encoding. Whereas the latter favors short latencies, since complicated prediction patterns such as IBBP are avoided, the bandwidth demands are high in order to preserve contrasts, color and luminance information.

Another way to avoid the delay caused by the encoding and decoding cycles of the codecs is to replace the devices with ATM adapters that capture the digital video signal and map it directly into ATM cells with a lossless algorithm. Such ATM adapters have just recently been developed at the IRT in connection with the Fraunhofer Institute and are capable of mapping a 270 Mbit/s video stream onto an ATM PVC with a delay of only 350 μs.

Therefore, the coding and decoding intervals become negligible. Figure 5 shows the new chain of events using an ATM adapter (again with camera personnel acting on location and the director at the studio). With such small latencies the spontaneity of the human interaction should not be noticeably impaired.

6 Summary and Future Outlook

The end-to-end delay that accumulates during the production of video material over a network is mainly caused by human reaction times and by the encoding and decoding process of the codecs. While human reaction times can only be reduced to a certain level through experience and practice, the technical delays can be shortened by using cameras that are remote controlled over the data network and by providing low latency codecs.

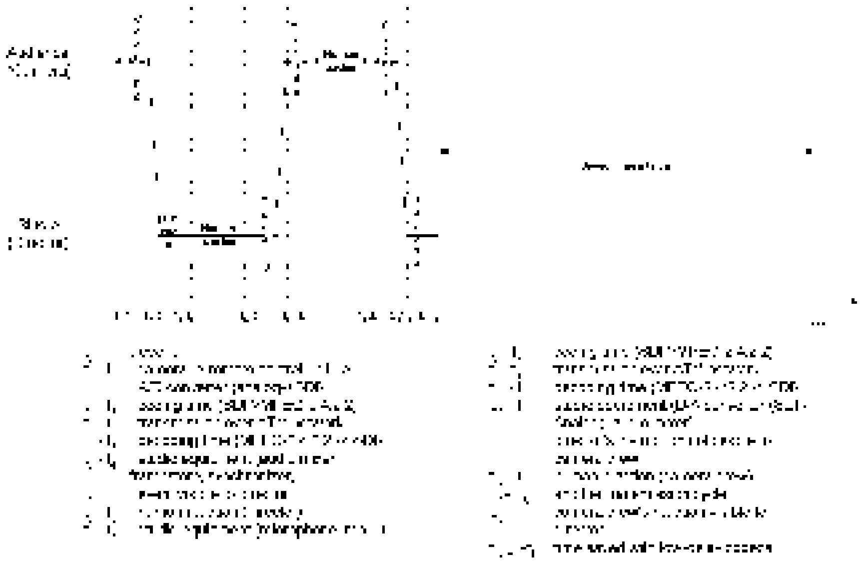

Figure 5 - Usage of Low Latency Equipment

The continued focus of the project "Uni-TV" will be to convert the current standardized production into an improved online editing process. Delays will be minimized with the use of ATM adapters and remote controlled equipment.

Bibliography

ITU-96 ITU-T International Telcommunication Union ITU-T Recommendation G.114: Transmission Systems and Media. General Characteristics of International Telephone Connections and International Telephone Circuits ITU 1996