AUTOMATIC SOUND RESTORATION SYSTEM
Concepts and Design

Andrzej Czyzewski, Bozena Kostek and Adam Kupryjanow
Multimedia Systems Department, Gdansk University of Techonlogy, Narutowicza 11/12, Gdańsk, Poland

Keywords: Audio restoration, Web service, Tape shrinkage, Noise and wow and flutter reduction.

Abstract: A concept of a system for automatic audio recording reconstruction is described. It is supported by the video image reconstruction algorithm, focused on the video instability analysis. Sound restoration is performed focusing on noise and wow and flutter analysis. Presented algorithms are designed to be automatic and to reduce the human effort during the restoration process. A web service designed especially for automatic restoration process is envisioned as an integration platform for these algorithms and for repository of recordings.

1 INTRODUCTION

Audio recordings, especially those from previous epochs are one of the main valuables of our times. These archival recordings should be made available for the society, in order to help people learn and understand the changing reality. In the archives of the Polish Radio and Television there are countless hours of unique audio and video, movie recordings stored on various types of storage media: magnetic tapes, film tapes with optical and magnetic soundtracks, that should be digitized, reconstructed and stored. In order to facilitate the reconstruction procedure, it is necessary to design algorithms for automatic audio-video content quality assessment and restoration.

First the objectives of this study are to be presented. The main causes of film degradation and resulted distortions are pointed out. Next methodology proposed is shortly described along with some conceptual use scenarios of the system designed. The last part of this paper deals with the technology and reviews shortly the main algorithms that are used in the system. Future plans are outlined in Conclusion.

2 STUDY OBJECTIVES

The objectives of this work are to design a system and to integrate it with the archive repository. Further goals are to evaluate algorithms for automatic audio archive material reconstruction. The algorithms are designed for supporting analogue media digitalization by fully automatic recordings distortions assessment. Nowadays the process of analogue recording digitalization is highly related to a human subjective assessment of the recording quality. This process is time consuming and arduous because it has to be performed during the real-time reproduction of the archive materials. The developed methods enable the digitalization and reconstruction of movie soundtracks without human control. Time of these processes should be shorter than the real-time track duration. Moreover, methods that are created could be used for automatic quality assessment and restoration of the digital multimedia libraries.

Many of the distortions in the video recordings may be introduced to the original signal during the recording, producing, monitoring or duplicating stages. However some of them occur because of tape shrinkage (Brun, 2007, Maziewski, 2008). It is caused by loss of water, solvent, and plasticizer in movie tapes. Both nitro-based and acetate-based films are exposed to this process. Tape shrinkage results in that perforations do not match the distances between the cogs of the sprocket roller, which in turn causes movie tape displacements. This is schematically shown in Fig. 1.

Among the most common distortions noise and/or wow and flutter may be pointed out. An example of noise that is introduced to the original

Czyzewski A., Kostek B. and Kupryjanow A..
AUTOMATIC SOUND RESTORATION SYSTEM - Concepts and Design.
DOI: 10.5220/0003527702070211
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2011), pages 207-211
ISBN: 978-989-8425-72-0

signal is the low-frequency power-line hum. This low-frequency signal may be however useful in determining undesired frequency modulation (FM) (Czyzewski, 2007, Wolfe, 2004). Wow is an audio distortion perceived as an undesired frequency modulation in the range of approximately 0.5 to 6 Hz, which affects analog recordings. In case of flutter this range is between 6 and 100Hz. The distortion is introduced to a signal by an irregular velocity of the analog medium. As mentioned before, irregularities can originate from various mechanisms, depending on medium type, production technique and other factors (Czyzewski, 2007, Godsill, 1994, Malecki, 1953, Nichols, 2001, Ryder, 1968).

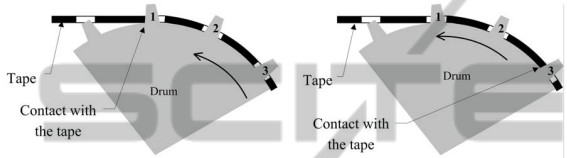

Figure 1: Tape shrinkage effects on the film tape transportation system.

3 METHODOLOGY

Distortion of the audio-video materials may be analyzed in two separate areas. In the beginning quality of a video frame is measured, then audio quality is assessed. The following audio defects are measured: level of the noise and wow and flutter. Distortion of the video frame, that is assessed, is interpreted as the image instability produced by the shrinkage of the movie tape. Those distortions often occur in the archival materials and require a long time to be described and removed. Algorithms for the quality assessment are integrated with algorithms for automatic audio-video material reconstruction.

The proposed algorithms are a part of the digital online library that is to be engineered at the Multimedia Systems Department of the Gdansk University of Technology (GUT). This library is designed to be open for everyone, on the condition that person will register and upload an example of the multimedia content (movie or audio recording).

Automatic recording restoration could be performed within three possible scenarios. In the first scenario, the recordings are first digitized. The digitized material is quality-assessed basing on the distortion types specified during the process of digitization. Then this material is saved in the library without any modifications and is reconstructed in accordance with the quality assessment. The next scenario concerns recordings added by the Internet service. In this case, the quality control should be performed as in Scenario 1. In the third scenario, recordings are saved in the library and an automatic restoration is performed for all database content that is not reconstructed.

4 TECHNOLOGY DESCRIPTION

Video frame irregularity analysis algorithm is conducted on separate film frames digitalized using a dedicated movie tape scanner. This algorithm is combined with an algorithm for wow and flutter reduction (wow and flutter reduction algorithm is described in the further part of this paper). Instability of the video frame is calculated according to the image content using information about the original size of the movie tape, i.e. frame height and perforation hole height. In Fig. 2 images of the digitalized movie frames with the optical soundtrack are presented.

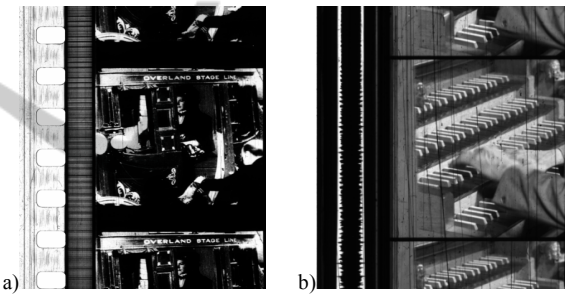

Figure 2: a) Image of the 35 mm movie frame with the optical variable density soundtrack and perforation holes image, b) image of the 16 mm movie frame with optical variable area soundtrack.

The defined size of the movie frame and perforation holes are invariant for every type of film (35 mm, 16 mm etc.), these values could be compared with the actual size of the specific element in the analyzed movie frame. To determine height of the movie frame, the mean intensity of the frame is analyzed in the horizontal direction. For all movie frames a ratio between actual and original height is calculated. A set of the ratio values calculated for the whole film creates the so-called pitch variation curve (PVC) (Czyzewski, 2007). If PVC equals 1, the video frame is free of mechanical distortion otherwise irregularity of the video frame is detected. PVC is also used to describe changes in the soundtrack of the movie.

Reduction of this distortion is obtained by performing 2-dimensional non-uniform resampling

according to the PVC. Algorithms used for the reconstruction and analysis of this distortion have been thoroughly researched at the Multimedia Systems Department (GUT) and described in earlier publications (Czyzewski, 2007; Czyzewski, 2008; Czyzewski, 2010) however Section 4.1 provides some details concerning the wow determination algorithm.

4.1 Video Frame Irregularity Assessment and Reconstruction

Wow introduces FM to the whole signal, this means that all spectral components are affected by this distortion. Thus audio spectrum can provide information necessary to determine the characteristics of the distortion. Several tonal components can be analyzed simultaneously using algorithms adapted from sinusoidal modeling (Godsill, 1994). The wow distortion can be characterized by the pitch variation curve (PVC) (Czyzewski, 2007, Kupryjanow, 2007). This function describes the parasitic FM caused by irregular velocity $V(t)$ of the recording medium:

$$PVC(t) = V(t) / V_{nom} \quad (1)$$

where V_{nom} represents the nominal, constant speed value. As mentioned before, if the pitch (speed) is constant, i.e. there is no wow, the PVC equals one. The PVC deviations from unity illustrating pitch variations indicate the wow depth.

Besides the genuine audio, additional tones can be found in archival recordings. This concerns magnetic recordings which contains a high-frequency bias (HFB). Another pilot tone at 15.734 kHz can be found in the NTSC stereo soundtracks. This is the so-called Multichannel Television Sound (MTS) pilot (Maziewski, 2008, Pastuszak, 2008).

Tracking the pilot tone allows for determining the PVC. For this purpose the Short Time Fourier Transform (STFT) is used for detecting its time-frequency variations. Thus, in the algorithm proposed, the input signal is divided into STFT frames. Some earlier study indicated that Hann window should be applied to weight each frame. Fig. 3 shows block diagram of the pilot tone tracking (Czyzewski, 2007).

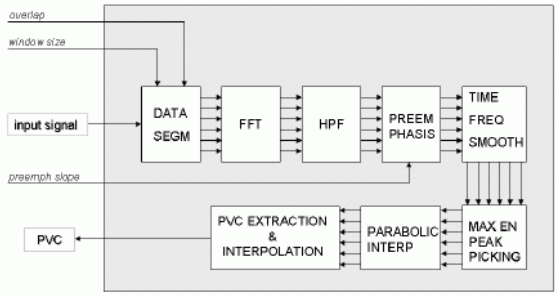

Figure 3: Block diagram of the HFB tracking algorithm.

As a result of STFT calculation, a spectrogram matrix, representing the time-frequency properties of the signal, is obtained. Low-frequency spectral components are set to zero in order to remove the high-energy genuine audio content which may obscure the pilot tone (HFB). In addition, each column of the spectrogram matrix is weighted by an appropriate preemphasis curve, allowing the bias enhancement. Also, the time- and frequency-domain smoothing (both 3rd order) are applied. This has an effect of blurring the whole spectrogram which is helpful to reduce noise and improve frequency estimation. Spectral expansion algorithms (e.g. spectrum raised to 4th power) may also be applied at this time. Then all columns are searched for maximal peaks, which, after correcting their amplitude and frequency estimation accuracy, are processed to obtain the PVC. Parabolic interpolation helps to find the fractional index of the bias frequency bin.

The MTS tracking algorithm operates in two phases, the first phase is similar to the HFP. However different (lower) cut-off frequency is applied, as well as parabolic interpolation is used to estimate the pilot tone frequency more precisely. During the second phase the center of gravity (CoG) within the neighborhood of the nominal pilot tone frequency (15734 ± 250 Hz) is sought. It is used later to eliminate the background pilot tone signal and to correct any potential inexactness of a signal pitch. Calculating CoG instead of spectral maximum reduces accidental jumps to nearby strong interfering signals that are making pilot tone harder to find in the next steps. The algorithm is iterative. Fig. 4 shows spectrogram analysis of wow-distorted signal using the MTS tracking algorithm (Pastuszak, 2008). There are three fragments with oscillating FM distortion (one with deep aperiodical modulation). A background MTS pilot tone is noticeable in the spectrogram.

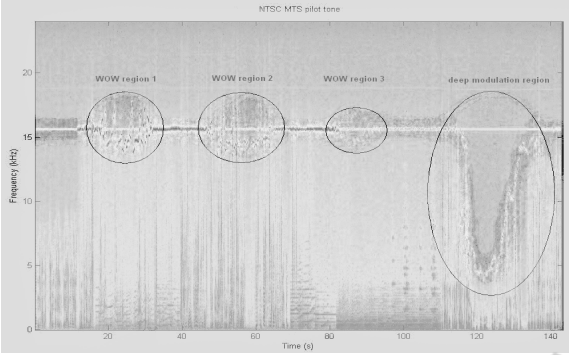

Figure 4: Wow-distorted audio sample spectrogram (Pastuszak, 2008).

After the wow and flutter distortion is detected in the video frame irregularity, then accordingly to the obtained PVC, distortion reduction is performed by the non-uniform resampling of the movie soundtrack.

4.2 Movie Frame Processing

After the reconstruction process of distortions connected with frame images, image processing of frames is performed. In the first step, to present uploaded movie frames as one movie, an algorithm for an automatic film frames cutting is used. This algorithm cuts the images of the movie from the image of the whole movie and connect them into the avi file. Then if images of the movie contain the image of the optical soundtrack, its image is converted to the digital sound.

4.3 Automatic Noise Reduction

Noise reduction is performed only for audio files or for the soundtrack of the video file. Therefore if the video file is assed regarding the level of the noise in the soundtrack, audio file is extracted from the input avi file. An algorithm for an automatic noise reduction is based on whitening and spectral subtraction. This method assumes that noise in the archival recording is the added noise. The algorithm used for noise reduction was designed during the European Union Project named PrestoSpace and is described in details in the paper by one of the Authors and his colleagues (Czyzewski, 2007).

4.4 Web Service Description

To facilitate automatic material reconstruction special web service is designed. During the upload process quality assessment and reconstruction process of the recordings are to be performed according to the block diagram presented in Fig. 5.

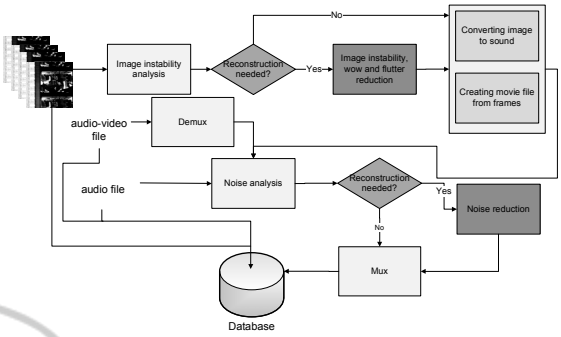

Figure 5: Block schema of the automatic reconstruction web service.

Two scenarios of the data analysis are possible. In both scenarios original files are saved to the database. These files could be used in the future e.g. when more algorithms for quality assessment will be added or if the process of the distortion reduction will not give the sufficient quality of reconstruction.

4.4.1 Scenario 1 – Movie Frame Upload

In the first step the user uploads film frames to the content directory. Then assessment of the image instability is performed. According to the obtained the PVC function a decision if the reconstruction is required is taken. The algorithm of the quality assessment is based on the comparison of the mean value and standard deviation of the PVC function with the thresholds set by the web service user. This method provides an opportunity to set the level of distortions that is acceptable for the user. If the level of the accepted distortions is set to a very low value, the reconstruction algorithm will work slower than for the higher value, because the reconstruction process is performed only for parts of the movie where the PVC function has a value higher than the threshold.

After the reduction of the image instability and wow and flutter defects all images are processed in order to get the avi file with the movie and wav file with the soundtrack. In the next step only soundtrack of the movie is analyzed to detect the level of the distortion. The algorithm noise reduction detects the noise level automatically. A similar quality assessment rule as for the image instability and wow and flutter algorithms is used. The user should define a threshold of the noise level that is satisfactory for him/her and the noise reduction algorithm is performed only when the level of the

noise in the soundtrack is higher than the defined noise level.

In the last step the reconstructed soundtrack and image are connected into one avi file. The last step is to save the avi file in the database, then it could be viewed by other users. The owner of the file has also an opportunity to perform the reconstruction process on the original file as many times as he/she wants setting various values of the threshold. In the database however only the newest version of the reconstructed file is saved, therefore only this version of the file is present in the web service.

4.4.2 Scenario 2 – avi/wav File Upload

The second scenario describes the situation when a movie is saved in the avi file or only the wav file with audio signal is uploaded to the service. In this situation only audio distortions are assessed. The reconstruction is conducted in the same way as in the Scenario 1. The reconstructed recordings are saved in the database and the owner of the file has the same possibilities of its processing as in the Scenario 1.

5 CONCLUSIONS

As a result of the carried out research, a solution for automatic audio-video reconstruction and archiving system has been proposed and is now under development. The proposed quality control mechanism ensures that high quality recordings are stored in the database. The system, currently being implemented as a web application, enables users to access the audio-video restoration services. The Internet service allows users for uploading their own recordings and, in exchange, the system reconstructs them and allows downloading the restored recordings.

ACKNOWLEDGEMENTS

Research funded within the project No. SP/I/1/77065/10 entitled: "Creation of universal, open, repository platform for hosting and communication of networked resources of knowledge for science, education and open society of knowledge", being a part of Strategic Research Programme "Interdisciplinary system of interactive scientific and technical information" funded by the The National Centre for Research and Development (NCBiR, Poland).

REFERENCES

Brun, E., Hassaine, A., Besserer, B., Decenciere, E., 2007. Restoration Of Variable Area Soundtracks. *ICIP IEEE*.

Czyzewski, A., Ciarkowski, A., Kaczmarek, A., Kotus, J., Kulesza, M., Maziewski, P., 2007. DSP Techniques for Determining "Wow" Distortions. *J. Audio Eng. Soc*., *55, 4, 266-284*.

Czyzewski, A., Maziewski, P., Kupryjanow, A., Papaj, M., Wroclaw, 9-11.04.2008. Methods for audio and video distortion. *Proc. KKRRiT, 465-468, (in Polish).*

Czyzewski, A., Maziewski, P., Kupryjanow, A., 2010. Reduction of Parasitic Pitch Variations in Archival Musical Recordings. *Signal Processing, Special Issue of Signal Processing: Ethnic Music Restoration, 981 – 990.*

Godsill, J., Adelaide, April, 1994. Recursive restoration of pitch variation defects in musical recordings. *Proc. International Conference on Acoustics, Speech, and Signal Processing*, vol. 2 pp. 233-236.

Kupryjanow, A., Gdańsk 2007. Converting digital picture of optical soundtrack to digital sound. *ISSVEM 2007, 12th International Symposium on Sound and Vision Engineering and Mastering*.

Malecki, I., Warszawa, 1993. Sound Registration and Reproduction, *PWT* (*in Polish*).

Nichols, J., Budapest, October 2001. An interactive pitch defect correction system for archival audio. 20*th Audio Eng. Soc. International Conference*.

Maziewski, P., Kupryjanow, A., Czyzewski, A., Amsterdam, NL, 2008. Drift, Wow and Flutter Measurement and Reduction in Shrunken Movie Soundtracks. *124th Audio Eng. Soc. Convention, Preprint No. 7392*.

Maziewski, P., 2008. Modulation frequency constrains on wow and flutter determination. *Archives of Acoustics*, *33, 125-131.*

Pastuszak P., 2008. MTS pilot tone tracking for "wow" distortion determination. *Archives of Acoustics, 33, 1, 117-123, 2008.*

Ryder L., July 1968. Synchronous Sound for Motion Pictures. *J. Audio Eng. Soc., 16, 291–295.*

Wolfe P., Howarth J., July/Aug. 2004. Nonuniform Sampling Theory in Audio Signal Processing. *116 Audio Engineering Society Convention, Preprint 6123, J. Audio Eng. Soc. (Abstracts), 52, 813*.