SOME TECHNIQUES FOR WOW EFFECT REDUCTION

Andrzej Czyzewski, Przemyslaw Maziewski

Multimedia Systems Department,
Gdansk University of Technology
ul. Narutowicza 11/12, 80-952 Gdansk, Poland

ABSTRACT

Wow distortion reduction has not attracted an adequate scientific attention so far. Only few papers on the subject are available, concerning mostly archive gramophone records, wax cylinders, and magnetic tapes affected by wow. This paper outlines researched wow reduction algorithms concerning archive movie soundtracks, or more generally audio recordings accompanying archival visual contents. The methods presented here are based on the pilot tone tracking, on the spectral analysis of genuine audio components, and on non-uniform resampling. The paper provides only a short overview of the concepts founding those methods; other studied approaches to the wow processing, as well as a more detailed description of the presented ones, can be found in referenced papers.

Index Terms— Audio recording, FM distortion, signal restoration

1. INTRODUCTION

Wow is an audio distortion perceived as an undesired frequency modulation (FM) in the range of approximately 0.5 Hz to 6 Hz [1], which affects original analogue recordings. The distortion is introduced to a signal by an irregular velocity of the analogue medium. As the irregularities can originate from various mechanisms, depending on: a) medium types, b) production techniques, c) random damages of a signal's conveyor, d) other factors, the resulting parasitic FM distortions can range from periodic to accidental, having different instantaneous values.

The problem of reducing wow remains unsolved in many respects in the analogue domain., However, digital signal processing (DSP) era enabled new approaches. First ideas on how to reduce wow in digitized analogue recordings were introduced by Gerzon [2]. Next, they were employed in a more practical implementation [3]. Some research on wow reduction was also reported by other authors [4-5]. All the reported methods, though interesting, suffer from some drawbacks, e.g. missing descriptions of the pilot-tones tracking algorithms, complicated (involving model selection) or too simple algorithms for genuine audio analysis addressing only smooth and repetitive wow. Thus, despite the available research, the wow reduction in the digital domain remains an open topic in many aspects.

Being involved in the PrestoSpace (PS) European project, the authors of this paper faced the problem of researching and developing the wow reduction algorithms for movie soundtracks. The task was found to be nontrivial because of the complex nature of the distortion and the lack of an in-depth research on the DSP-based wow reduction methods. We studied various approaches, some inspired by available literature, which resulted in several algorithms. Owing to the PS project, which enabled cooperation with the archive community, the researched algorithms were tested on real-life archive recordings. In this short paper only an overview of the most successful and complementary DSP methods for determining and reducing the distortion can be given. A more in-depth description is available in our previous papers [6-11].

Wow reduction consists of two processing steps. First, the distortions characteristic must be determined. Next, the distortion can be reduced. In the following parts an overview of algorithms for determining the wow characteristic are given first. Then the wow reduction methods are presented..

2. ALGORITHMS FOR WOW CHARACTERISTIC DETERMINATION

The wow distortion can be characterized by the pitch variation curve (PVC) [4]. This function describes the parasitic frequency modulation (FM) caused by the irregular velocity $V(t)$ of the recording medium:

$$PVC(t) = \frac{V(t)}{V_{nom}} \tag{1}$$

where V_{nom} represents the nominal, constant speed value. If the pitch (speed) is constant, i.e. there is no wow, the PVC equals one. PVC deviations from unity illustrating pitch variations, indicate the wow depth. In most real-life recordings the PVC is close to a constant unity function with occasionally varying fragments indicating wow. The algorithms presented next aims at the PVC determination.

2.1. Pilot tone analysis

Additional tones, besides the genuine audio, can be found in archival recordings. Some of the magnetic recordings contain a high-frequency bias (bias), which is a leftover from the magnetic recording head. Another pilot tone can be found in the NTSC stereo soundtracks which were recorded together with a 15.734 kHz tone. Moreover, older recordings likely contain a low-frequency hum (hum). Tracking the additional tones' frequency variations was found useful in the PVC determination.

Tracking the bias allows determining the PVC. Since the bias is of high frequency, Short Time Fourier Transform (STFT) is a suitable tool for detecting its time-frequency variations. Thus, in our algorithms the input signal is divided into STFT frames. The Hann window weights each frame and then the Fourier spectra are calculated. As a result of the spectrum calculations, the 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 bias. Additionally, each amplitude spectrum (each column of the spectrogram matrix) is weighted by an appropriate preemphasis curve, allowing the bias enhancement:

$$y[n] = \left(\frac{n}{N-1}\right)^{s}, n = 0, \ldots, N-1 \tag{2}$$

where n is the sample number, $y[n]$ is the preemphasis curve, N is the block size and s is the slope ratio. Then each column is searched for a maximal peaks, which, after correcting their amplitude and frequency estimation accuracy, are processed to obtain the PVC [7-8, 11]. An example of such a PVC, rescaled in order to plot it together with the original bias track - the carrier frequency of 73.7 kHz and strong modulations - is given in the left part of Fig. 1. We proved only recently that the same algorithm can be used for tracking the NTSC-related tone.

Another approach was used for tracking the hum. In all European countries the power-line frequency equals to 50 Hz. In Americas it is typically 60 Hz, and always one of these two values in all other countries. Thus, the hum is placed in a frequency region where achieving a desired high STFT resolution is impossible (due to the "uncertainty principle"). Therefore, the method proposed by the authors of this paper utilizes the auto-regressive (AR) modeling to track the hum frequency.

The AR modeling is a method of so called "parametric" spectrum estimation which bases on an assumption that the observed signal represents a response of some infinite impulse response (IIR) all-pole filter to an excitation being a realization of white noise. The difference equation of a p-order AR model has the form:

$$x[n] = -\sum_{m=1}^{p} a_m x[n-m] + e[n] \tag{3}$$

where $x[n]$ is the known observation, a_m with $m = 1, 2, \ldots, p$ are the AR model coefficients (parameters) and $e[n]$ is an excitation - the realization of white Gaussian noise with a mean value of 0 and unknown power.

There are many methods allowing estimating the parameters of the AR model. The most common are: so-called "autocovariance", "modified-autocovariance", Burg and Yule-Walker methods. As it is known, they all share a common quality to produce highly accurate results based on extremely short observations, thus yielding a high resolution in both time and frequency [13].

In our approach the input signal is downsampled prior to being split into frames. The downsampling allows eliminating most of non-hum related tonal components from the signal's spectrum. To reduce the presence of noise further, and other non-hum related components, the signal is bandpass-filtered. The next step is tracking properly the hum employing AR-modeling. The downsampled and filtered signal is split into frames for this purpose.. The coefficients of the corresponding AR filter (see Eq. 2) are calculated using the modified-autocovariance method for a given frame. The coordinates of transmittance poles are obtained as a result. Subsequently, one pole is selected and the frequency corresponding to its angle, normalized with the base hum frequency, is chosen as the PVC value [7-8, 11]. The right part of Fig 1 serves as an example of such a PVC, rescaled in order to plot it together with the original hum track. In the example the hum's carrier frequency equals 50 Hz and is modulated at beginning of the recording.

2.2. Genuine Audio Content Analysis

Wow introduces FM to all spectral components of the distorted audio signal. The modulation is well visible when analyzing the pilot tones like hum or bias, being separated from the useful audio. Regrettably, the pilot tones are not always available or reliable. In such situations the spectrum of the useful audio, which is also affected by the wow modulations, can provide information necessary to determine the distortion characteristic.

Analyzing separated tonal audio components allows a PVC determination. Wow is noticeable mainly for sounds with a distinct tonal structure, thus silent or noisy parts do not have to be processed. Tonal components analysis can be performed for several components simultaneously, by using algorithms adapted from the sinusoidal modeling [4-6, 10-11]. We found out that such an analysis is more reliable provided that one tonal component depicting the wow most noticeably, is used. Additionally, in movie soundtracks wow is mostly accidental and short in time. Therefore our approach involves an human operator who selects

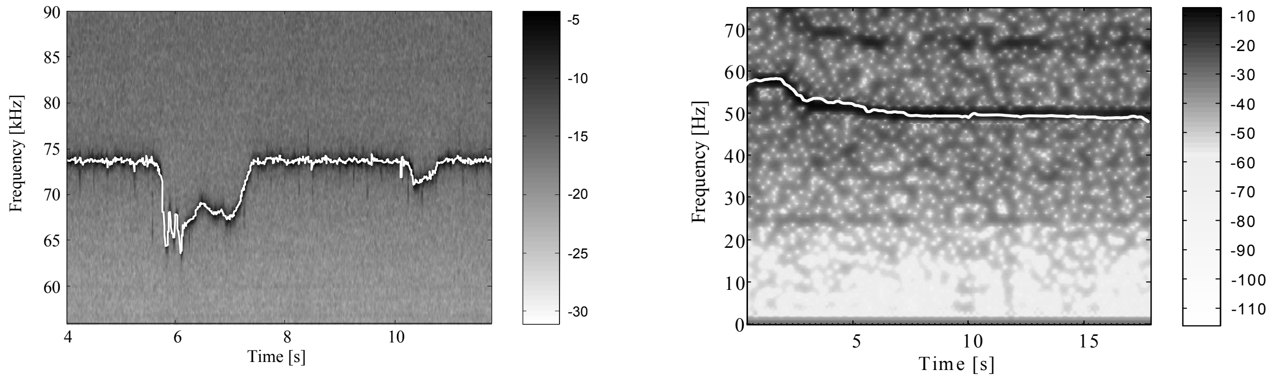

Fig. 1 Rescaled PVCs (white lines) plotted together with bias (left) and hum (right).

the time-frequency region with the varying tonal component. Based on his selection, the centre-of-gravity (COG) is calculated:

$$SC = \left(\sum_{l=N_L}^{N_U} S_x[l] W(f_l) \log_2 f_l\right) / \left(\sum_{l=N_L}^{N_U} S_x[l]\right) \tag{4}$$

where l stands for the frequency bin index, $< f_{N_L}, f_{N_U} >$ defines the frequency band enclosing the tonal component, $S_x[l]$ is the discrete power spectrum of the l-th bin, and $W(f_l)$ is the appropriate time window given in the logarithmic scale, i.e. the von Hann window:

$$W(f) = 0.5 - 0.5\cos\left(2\pi \frac{\log_2 f - \log_2 f_{N_L}}{\log_2 f_{N_U} - \log_2 f_{N_L}}\right) \tag{5}$$

As the FM caused by wow has a multiplicatory nature, i.e. frequency deviations are greater for higher carrier frequencies, the COG computation must be performed on a logarithmic scale

Next the COG values from successive time frames are used to compute the PVC [8,11]. An example of such a PVC, rescaled in order to plot it together with the original component track around 1.4 kHz, is illustrated in Fig 2.

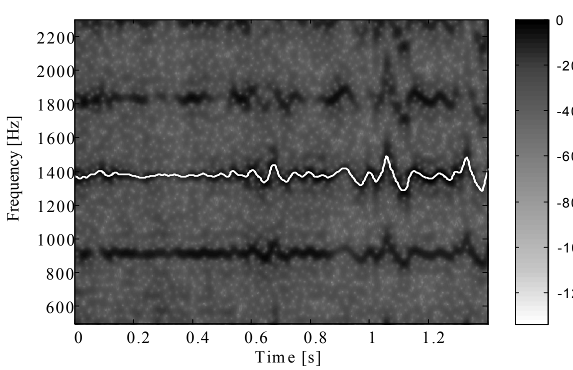

Fig. 2 Rescaled PVC plotted together with original tonal components.

3. WOW REDUCTION ALGORITHMS

The final wow reduction using the appropriate resampling is an issue which has not been studied thoroughly so far. Different interpolation techniques were used in the literature for resampling [3-5]. As there is no clear comparison of these methods, it was unclear which one is the best for the wow reduction. The authors of this paper addressed this question by comparing different resampling methods in terms of their influence on the audio quality [9].

The resampling can be done by building an interpolation function which preserves the signal values at the sampling points t_{wow} and then calculating intermediate values for the t_{rec} points resulting from the PVC processing:

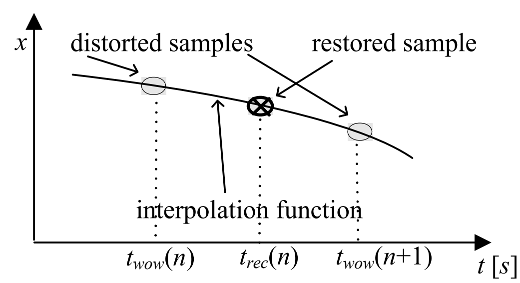

Fig. 3 Illustration of signal resampling using interpolation.

It was found out by the authors of this paper that the sinc and spline-based resampling are most appropriate for the wow reduction. The sinc-based resampling of irregularly spaced samples is the one most closely connected with the Shannon sampling theorem. In the sinc method, the equation used to obtain the restored sample $x(t_{rec})$, delayed by the amount T_{dw}>0 relative to the closest distorted sample $x(t_{wow})$, is given by following Eqs. :

$$x(t_{rec}) = \sum_{m=-N/2}^{N/2} x(t_{wow}(m)) \cdot h\left(\frac{T_{dw}}{T} - m\right) \tag{6}$$

$$h(t) = \gamma \cdot \mathrm{win}(t)\mathrm{sinc}(\gamma \cdot t) \quad (7)$$

In Eq. 6 the $x(t_{rec})$ value represents a sum of the N nearest distorted samples (the block size). The factor γ in Eq. 7 is a number given by the minimum of 1 and the current sampling rate conversion factor T/T_{dw}, where T represents the current sampling period. The $\mathrm{win}(t)$ function, choice of which is crucial for the accuracy of the interpolation, represents the time domain windowing.

Spline-based resampling, on the other hand, utilizes one of the widely known osculatory interpolation technique. The basic concept of the osculatory interpolation states that the interpolation function, based on which the reconstructed sample can be calculated, is build from several low-order polynomials in intermediate interpolation regions, i.e. the time periods between the neighboring distorted sound samples. In the spline-based resampling, the interpolating function is constructed in a way that makes its second derivative continuous at the transition points, i.e. successive sound samples. Therefore, the spline produces smooth and accurate results provided the data consist of values of a smooth function, which is the case in low-pass anti-aliasing filtered digital recordings. More on the spline method can be found in the literature [12].

The experiments conducted by the authors of this paper showed that longer sinc-based resamplers, e.g. $N = 21$ (see Eq. 6), allows for a better quality, whereas the splines are the best choice in terms of quality, especially when the computational cost is an issue. Experiments involved: generating a clean, only wow distorted audio signal; reducing wow using various resamplers; and measuring resultant audio distortions, i.e. signal-to-nose ratio, total harmonic distortions, and total harmonic distortions plus noise [9].

4. CONCLUSIONS

The knowledge gained through the research and development of algorithms briefly described above, as well as results of testing them using real-life archive recordings, brought the following conclusions. Firstly, majority of designed algorithms requires a human operator. A fully automatic analysis, though imaginable for some of the materials with distinct and unique high frequency pilot tones, is not practically applicable for most recordings. Secondly, as wow-distorted samples can originate from different analogue recording media, it might happen that some pilot tones will not be available, thus algorithms based on different approaches to the wow determination should be made available. Finally, the wow reduction may demand making a choice from among at least two types of resampling techniques.

5. ACKNOWLEDGMENTS

Research funded by the Commission of the European Communities, Directorate-General of the Information Society within the Integrated Project No. FP6-507336 entitled: “PRESTOSPACE – Preservation towards storage and access. Standardized Practices for Audiovisual Contents Archiving in Europe”.

6. REFERENCES

[1] Audio Engineering Society, *Method for Measurement of Weighted Peak Flutter of Sound Recording and Reproducing Equipment*, AES Standard, No. 6-1982, 2003.

[2] M. Gerzon, “Don't Destroy the Archives!”, unpublished manuscript (1992).
Available at www.waves.com/german/htmls/service/faq/dont.html

[3] J. Howarth, P. Wolfe, “Correction of Wow and Flutter Effects in Analog Tape Transfers”, *Proc. 117 AES Convention*, paper No. 6213, San Francisco, USA, 28-31 October 2004.

[4] S. J. Godsill, P. Rayner, “The restoration of pitch variation defects in gramophone recordings”, *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, NY State, USA, pp. 148-151, 17-20 October 1993.

[5] J. Nichols, “An interactive pitch defect correction system for archival audio”, *Proc. AES 20th International Conference*, paper no. 1941, Budapest, Hungary, 5-7 October 2001.

[6] A. Czyzewski, P. Maziewski, M. Dziubinski, A. Kaczmarek, B. Kostek, “Wow Detection and Compensation Employing Spectral Processing of Audio”, *Proc. of the 117 AES Convention*, paper No. 6212, San Francisco, USA, 28-31 October 2004.

[7] A. Czyzewski, M. Dziubinski, A. Ciarkowski, M. Kulesza, P. Maziewski, J. Kotus, “New Algorithms for Wow and Flutter Detection and Compensation in Audio”, *Proc. of the 118 AES Convention*, paper No. 6353, Barcelona, Spain, 28-31 May 2005.

[8] A. Czyzewski, P. Maziewski, M. Dziubinski, A. Kaczmarek, M. Kulesza and A. Ciarkowski, ”Methods for detection and removal of parasitic frequency modulation in audio recordings”, *Proc. of the AES 26th Int. Conf.*, paper no. 3-3, Denver, USA, 7-9 July 2005.

[9] P. Maziewski, “Wow Defect Reduction Based on Interpolation Techniques”, *Bulletin of the Polish Academy of Science: Technical Sciences*, vol. 54, no. 4, pp. 469-477, (2006).

[10] P. Maziewski, L. Litwic, A. Czyzewski. “Accidental Wow Evaluation Based on Sinusoidal Modeling and Neural Nets Prediction”, *Proc. of the 120 AES Convention*, paper No. 6796, Paris, France, 20-23 May 2006.

[11] A. Czyzewski, A. Ciarkowski, A. Kaczmarek, J. Kotus, M. Kulesza, P. Maziewski, “DSP Techniques for Determining “Wow” Distortion”. *Journal of the Audio Engineering Society*, vol. 55, no.4, pp.266-284, (April 2007).

[12] C. de Boor, *A Practical Guide to Splines*, New York, USA: Springer, 2001.

[13] V. K. Madisetti, D. B. Williams, *Digital Signal Processing Handbook*, New York, CRC Press, 1998.