ARCHIVES OF ACOUSTICS
33, 1, 125–131 (2008)

MODULATION FREQUENCY CONSTRAINTS ON WOW AND FLUTTER DETERMINATION

Przemysław MAZIEWSKI

Gdańsk University of Technology
Multimedia Systems Department
Narutowicza 11/12, 80-952 Gdańsk, Poland
e-mail: przemas@sound.eti.pgd.pl

(*received June 15, 2007; accepted November 30, 2007*)

Hitherto developed algorithms for wow and flutter characteristic determination were found useful in some real-life restoration procedures, however, there was no study of their capabilities in terms of the maximal parasite modulation frequency that can be determined with their application. This paper presents the study for three algorithms, i.e. the power-line hum tracker, high frequency bias tracker, and the center of gravity tracker. The short description of the algorithms is followed by the theoretical study on the maximal modulation frequency determined by the algorithms. Next, the experimental part is given, with simulations which confirm the findings from the theoretical part, and conclusions summarizing obtained results.

Keywords: wow and flutter determination, maximal modulation frequency.

1. Introduction

Wow and flutter are audio distortions perceived as an undesired frequency modulation (FM) in the range of: 0.5 Hz to 6 Hz for wow, and of 6 Hz to 100 Hz for flutter [1]. The distortions are introduced into signals by the irregular velocity of analogue media. As the irregularities can originate from various mechanisms, the resulting parasitic FMs can range from periodic to accidental, having different instantaneous values.

The problem of reducing wow and flutter remains unsolved in many aspects in the analogue domain, however, the era of digital signal processing (DSP) enabled new approaches. First, ideas on how to reduce wow and flutter 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]. Despite the available research, the wow reduction in the digital domain remains an open topic in many aspects.

The author of this paper took part in researching and developing new algorithms for wow and flutter reduction [6–13]. Different approaches were studied, some inspired by available literature, which resulted in several algorithms. The most successful algorithms are briefly presented in the next section. A more in-depth description is available in the previous papers of the author [11, 12]. The third section gives a theoretical study of the maximal modulation frequency ($FM_{\max}$) determined by the developed algorithms, whereas the fourth section presents simulations' results confirming the theoretical findings. The paper ends with conclusions.

2. Algorithms Overview

The wow distortion can be characterized by the pitch variation curve (PVC) [4, 11]. This function describes the parasitic FM caused by the irregular velocity $V(t)$ of the recording medium:

$$\mathrm{PVC}(t) = V(t)/V_{\mathrm{nom}}, \tag{1}$$

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. The PVC deviations from unity illustrating pitch variations, indicate the wow depth.

2.1. *Genuine audio content analysis*

Wow introduces FM to all spectral components of the distorted audio signal. Thus the spectrum of useful audio, which is also affected by wow modulations, can provide information necessary to determine the characteristic of the distortion.

Analyzing separated tonal audio components allows the PVC determination. Several tonal components can be analyzed simultaneously using the algorithms adapted from sinusoidal modeling [4, 5, 10]. However, such analysis is more reliable provided that the tonal component depicting wow most noticeably is used. Additionally, as wow in most cases is accidental and short in time, the presented approach involves a human operator who selects the time-frequency region with the varying tonal component. Based on this selection, the spectral center-of-gravity (COG) is calculated. The COG values from successive time frames are used to compute the final PVC [11].

2.2. *Pilot tone analysis*

Besides the genuine audio, additional tones can be found in archival recordings. Some magnetic recordings contain a high-frequency bias (bias). Another pilot tone (15.734 kHz) can be found in the NTSC (National Television Systems Committee) stereo soundtracks. 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 to determine the PVC. Since the bias is of high frequency, the Short Time Fourier Transform (STFT) is a suitable tool for detecting its time-frequency variations. Thus, in the researched algorithm 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 these 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 tracking. Additionally, each amplitude spectrum (each column of the spectrogram matrix) is weighted by an appropriate preemphasis curve, allowing the bias enhancement. Then all columns are searched for maximal peaks, which, after correcting their amplitude and frequency estimation accuracy, are processed to obtain the PVC [11]. It was proven that a similar algorithm can be used for tracking the NTSC-related tone [13].

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 the frequency region where achieving the desired high STFT resolution is impossible (due to the "uncertainty principle"). Therefore, the proposed algorithm utilizes the auto-regressive (AR) modeling to track the hum frequency.

In the presented approach the input signal is downsampled prior to being split into frames. Downsampling allows to eliminate most of the non-hum related spectral components from the signal's spectrum. To further reduce the presence of noise, and other non-hum related components, the signal is bandpass-filtered. The next step is tracking properly the hum which employs the AR-modeling. The downsampled and filtered signal is split into frames for this purpose. The coefficients of the corresponding AR filter 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 [11].

3. Maximal modulation frequency

The algorithms presented in the previous section aim at the PVC determination, however, the maximal modulation frequency determined by those algorithms is unknown. In order to establish the $FM_{\max}$, two factors must be analyzed: a) the Nyquist sampling theorem for the discrete PVC, b) the quasi-stationarity criteria for the infra-frame modulations.

The discrete version of wow and flutter characteristic, i.e. PVC$(nT_{s\mathrm{PVC}})$, is defined by the PVC sampling period $T_{s\mathrm{PVC}}$ (or the PVC sampling frequency $F_{s\mathrm{PVC}}$), and represents its continuous counterpart until the maximal frequency $FM_{\max} < 0.5F_{s\mathrm{PVC}}$. The PVC sampling period can be expressed in a number of distorted signal's samples as:

$$R = T_{s\mathrm{PVC}}/T_s = F_s/F_{s\mathrm{PVC}}, \tag{2}$$

where T_s (F_s) represents the sampling period (frequency) of the distorted signal. R can be interpreted as the hop size of the time-windowing used in the algorithms presented in the previous section. Knowing that in the presented algorithms the hop size is always a quarter of the frame size (M), the $FM_{\max}$ limitation, being a consequence of the sampling theorem, can be expressed as:

$$FM_{\max} < F_s/0.5M. \tag{3}$$

The second factor which limits the $FM_{\max}$ comes from the quasi-stationarity criteria. The criteria eliminates time aliasing between the adjacent frames, and can be expressed as the *a priori* assumption that the spectral components must be constant within the analysis frame. For a specific frame size, it can be formulated as:

$$M \ll T_{FM_{\max}}/T_s = F_s/FM_{\max}. \tag{4}$$

Minimal M value which meets these criteria must be at least two times smaller then the FM period, assuming small frequency deviations. Thus the $FM_{\max}$ limitation coming from the quasi-stationarity criteria can be expressed as:

$$FM_{\max} < F_s/2M. \tag{5}$$

Assuming that R is always smaller then M, which is true for the algorithms presented in this paper, Eq. (5) fulfills also the Nyquist theorem for the discrete PVC given by Eq. (3). Thus the final $FM_{\max}$ limitation is given by the quasi-stationarity criteria. Based on the determined limitation the equation for the frame length in the wow processing algorithms can be formulated:

$$M = \lfloor F_s/2FM_{\max} \rfloor. \tag{6}$$

It allows a more intuitive handling of the algorithms where the user specifies only the $FM_{\max}$ value.

Knowing typical F_s values of the distorted signals and typical frame lengths of the presented algorithms, the $FM_{\max}$ can be calculated – see Table 1.

Table 1. $FM_{\max}$ values for the presented wow and flutter tracking algorithms.

Algorithms and their settings	$FM_{\max}$
Power line hum tracker; M=8 Sa; F_s=200 Hz;	$FM_{\max} < 12.5\,\mathrm{Hz}$
COG tracker; M=1024 Sa; F_s=44100 Hz;	$FM_{\max} < 21.5\,\mathrm{Hz}$
High-frequency bias tracker; M=1024 Sa; F_s=192000 Hz	$FM_{\max} < 93.75\,\mathrm{Hz}$

The simulations presented in the following section aim at confirming the given $FM_{\max}$ values.

4. Experiments

The experiments were performed as follows. For each algorithm, three test signals were prepared. Each signal – a pure sinusoidal tone with a constant carrier frequency – was modulated with a simple FM pattern equaling: a half, a whole, and a double

$FM_{\max}$ (see Table 1). The carrier frequencies were set to: a) 50 Hz for the hum tracker, b) 3150 Hz for the COG tracer [1], c) 30 kHz for the bias tracker. Next, the modulated signals were processed by the tracing algorithms and the resulting PVCs were compared against known FM patterns. The results are as follows.

The results of the COG tracking, are given in Fig. 1. The left chart shows, that the algorithm performs well for the slowly modulated signal. The 20 Hz modulation illustrated in the center chart is also well determined, however, there is a minor underestimation of frequency deviations due to the time aliasing. The right chart shows even a greater underestimation caused by the unfulfilled quasi-stationarity criteria.

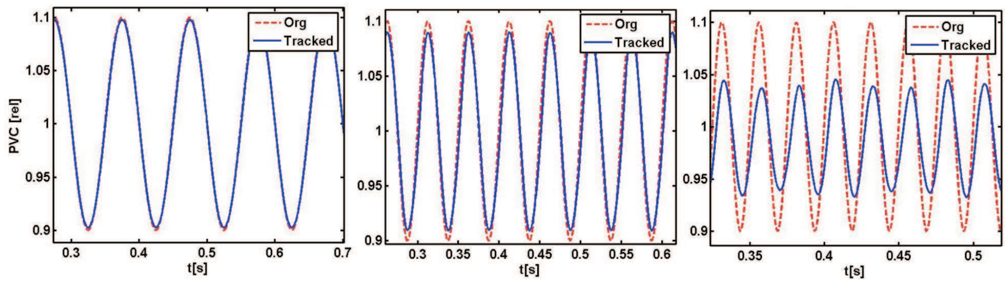

Fig. 1. Simulation results of the COG tracking. The FM values from the left side are: 10, 20 and 40 Hz.

The results of the hum and bias tracking, given in Figs. 2 and 3, present similar situations. The simulated tones are well tracked for the FMs below the $FM_{\max}$.

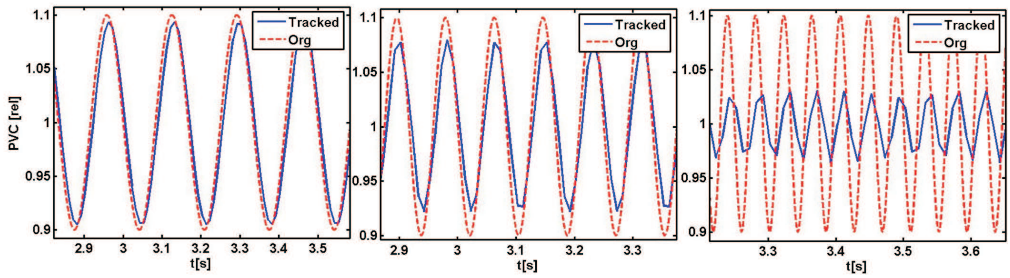

Fig. 2. Simulation results of the artificial hum tracking. The FM values from the left side are: 6, 12 and 24 Hz.

The third simulation for the bias tracker differs from the COG and hum tracking results. It is because for the first two algorithms the examples with the doubled $FM_{\max}$ still hold the Nyquist criteria (Eq. (3)), thus the modulation frequency is preserved and only the deviation is underestimated. The 180 Hz modulation in the bias tracker exceeds the $FM_{\max}$ set by the quasi-stationarity criteria as well as by the Nyquist criteria. As a result, the modulation frequency in this example is not preserved.

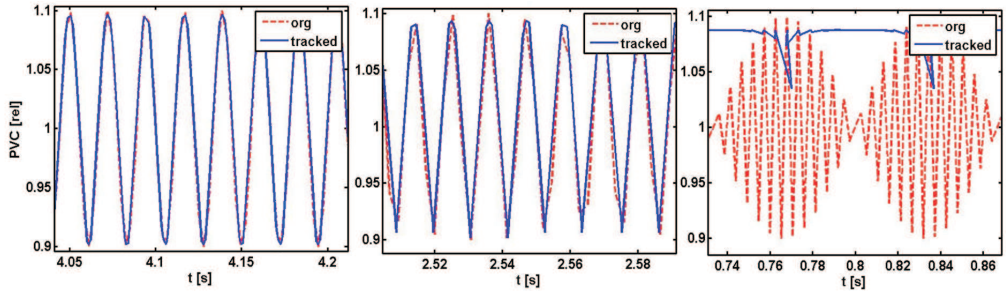

Fig. 3. Simulation results of the artificial bias tracking. The FM values from the left side are: 45, 90 and 180 Hz.

5. Conclusions

The following conclusions can be derived both from the presented theoretical study and the experiments. First of all, $FM_{\max}$ is determined mostly by the quasi-stationarity criteria. Based on the criteria the $FM_{\max}$ for each of the presented algorithms can be calculated. Consequently, it can be stated that all of the algorithms can track wow modulations. Further, the hum tracker can determine flutter modulations up to 12 Hz approximately; the COG tracker can determine flutter up to 20 Hz approximately; and the bias tracker can determine the full range of flutter modulations. Additionally, the presented theoretical study allows formulating the equation for the frame length calculation (Eq. (6)) leading to a more intuitive handling of the algorithms.

Acknowledgments

The research was 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"

References

[1] Audio Engineering Society, *Method for measurement of weighted peak flutter of sound recording and reproducing equipment*, Standard No. 6-1982, 2003.

[2] GERZON M., *Don't Destroy the Archives!*, unpublished manuscript, 1992.

[3] HOWARTH J., WOLFE P., *Correction of wow and flutter effects in analog tape transfers*, Proceedings of the 117th Audio Engineering Society Convention, paper no. 6213, San Francisco, USA, 28–31 October 2004.

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

[5] NICHOLS J., *An interactive pitch defect correction system for archival audio*, Proceedings of the Audio Engineering Society 20th International Conference, paper no. 1941, Budapest, Hungary, 5–7 October 2001.

[6] CZYŻEWSKI A., DZIUBIŃSKI M., LITWIC L., MAZIEWSKI P, *Intelligent algorithms for optical track audio restoration*, Lecture Notes Computer Science: Rough Sets, Data Mining, and Granular Computing, **3642**/2005, 283–293 (2005).

[7] CZYŻEWSKI A., KOSTEK B., MAZIEWSKI P., LITWIC L., *Accidental wow defect evaluation using sinusoidal analysis enhanced by artificial neural networks*, Lecture Notes in Computer Science: Rough Set and Knowledge Technology, **4062**/2006, 389–395 (2006).

[8] CZYŻEWSKI A., DZIUBIŃSKI M., LITWIC L., MAZIEWSKI P., *Intelligent algorithms for movie sound tracks restoration*, Lecture Notes in Computer Science: Transaction on Rough Set V, **4100**/2006, 123–145 (2006).

[9] MAZIEWSKI P., *Wow defect reduction based on interpolation techniques*, Bulletin of the Polish Academy of Science: Technical Sciences, **54**, 4, 469–477 (2006).

[10] MAZIEWSKI P., LITWIC L., CZYŻEWSKI A., *Accidental wow evaluation based on sinusoidal modeling and neural nets prediction*, Proceedings of the 120th Audio Engineering Society Convention, paper no. 6796, Paris, France, 20–23 May 2006.

[11] CZYŻEWSKI A., CIARKOWSKI A., KACZMAREK A., KOTUS J., KULESZA M., MAZIEWSKI P., *DSP techniques for determining "wow" distortion*, Journal of the Audio Engineering Society, **55**, 4, 266–284 (2007).

[12] CZYŻEWSKI A., MAZIEWSKI P., *Some techniques for wow effect reduction*, Presented at the IEEE International Conference Image Processing 2007, San Antonio, Texas, USA, 16–19 September 2007.

[13] PASTUSZAK P., *MTS pilote tone tracking algorithms for "Wow" distortion estimations in NTSC audio recordings*, Presented at the 12th International Symposium on Sound and Vision Engineering and Mastering, Gdańsk, Poland, 15–16 June, 2007.