IV KRAJOWA KONFERENCJA ELEKTRONIKI
Darłówko Wschodnie, 12 - 15 czerwca 2005
Politechnika Koszalińska, Wydział Elektroniki i Informatyki

Przemyslaw Maziewski
Multimedia Systems Department
Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology, ul. Narutowicza 11/12, 80-952 Gdansk, Poland

Wow defect reduction based on interpolation techniques

Key words: Wow audio defect, restoration, linear interpolation, polynomial interpolation, sinc-based interpolation, non-uniform resampling, incommensurate sampling rate conversion.

ABSTRACT

In this paper the capacity of different interpolation techniques aimed at wow defect reduction is examined. Involved are: linear interpolation, two polynomial based interpolation methods (Hermite and spline) and the windowed sinc-based method. The performance of a synthetic audio signal restored using incommensurate resampling for wow cancellation is evaluated on the basis of standard audio defect measurement criteria and compared for four interpolation techniques.

1. INTRODUCTION

The preservation of mechanically and magnetically recorded sound is an issue of considerable current interest. Extensive recorded sound collections and archives exist world-wide. Normally, in sound recording, it is impossible to obtain absolutely constant speed of the recording medium because of the limited precision of the mechanical drive. This concerns especially old recordings. As a result different forms of distortions arise caused by an irregular motion of the recording medium during the recording, duplicating and reproducing processes. Among them the wow defect appears, which is especially difficult to cancel.

As mentioned above, the main source of the wow defect is the variable velocity of the sound conveyer. The reasons for this can be different. One origin can be found in the excentricity or ellipticity of the audio equipment mechanical parts and the conveyer itself. It is a frequent case regarding vinyl and wax recordings. Motor speed fluctuations can be the reason of wow presence especially in semi-professional cassettes. In case of movies, sound track damages and inappropriate production techniques can trigger wow. Moreover, the distortion can be found in the movie soundtracks, where careless cuts and joins of the separate magnetic audio tape made synchronously with the optical tape, introduce high wow risk [1].

Wow is very difficult to combat with, especially in the analogue domain. Providentially digital signal processing algorithms can be successfully used in the defect extraction [2]. Naturally, in terms of wow evaluation and restoration, it is essential to digitise the analogue audio signals. Then the defect can be compensated for by using, e.g., a resampling method with chosen interpolation techniques as it is done in this paper.

The organization of the paper is the following. Sect. 2 defines the wow defect and presents characteristics useful in the wow reduction. Sect. 3 depicts different interpolation techniques used in experiments. Main results and conclusions are reported in Sect. 4.

2. WOW DEFINITION AND CHARACTERISTICS

The irregular motion of the recording medium introduces an undesired frequency modulation (FM) into the signal. This results in such forms of distortions as: *drift*, *wow*, *flutter* and *FM noise* [3]. Each of them is a FM of the signal characterised by different frequency range leading to a different perception. *Drift* is a FM in the range below approximately 0.5 Hz, resulting in distortion perceived as a slow changing of the average pitch. *Wow* is a FM in the range of approximately 0.5 Hz to 6 Hz perceived as a fluctuation of pitch. *Flutter* is a FM in the range of approximately 6 Hz to 100 Hz causing roughening of the sound quality. *FM noise* is a FM in the range above 100 Hz perceived as a noise added to the signal.
Hereafter we shall focus on the wow defect exclusively. This defect can be described by using a time warping function $f_w(t_{org})$. It characterizes the wow defect as a distortion of the time axis t_{org} of the original signal $x(t_{org})$. Because the sound conveyer playback velocity differs from that used in recording, the function $f_w(t_{org})$ represents the time axis changes relatively to the original recording. Consequently the distorted signal $x(t_{wow})$ can be written as

$$x(t_{wow})=x(f_w(t_{org})) \tag{1}$$

The time warping function is a mapping of the original time axis t_{org} to the distorted time axis t_{wow}, as presented in Fig.1.

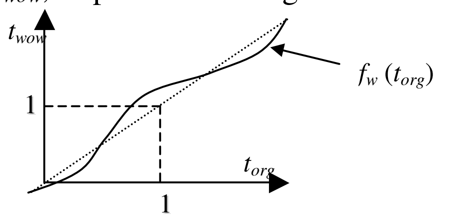

Fig. 1. An example of the time mapping function $f_w(t_{org})$.

The second commonly used wow characteristic is the pitch variation function $p_w(t_{org})$. This function describes the parasite FM caused by the irregular playback. Therefore it is closely connected to the standard wow definition [3]. There exists the following relation between the two functions: f_w and p_w:

$$p_w(t_{org}) = \frac{\mathrm{d}(f_w(t_{org}))}{\mathrm{d}t_{org}} \tag{2}$$

Both these functions are commonly used in the wow defect evaluation. Details can be found in [2]. Based on the distorted signal and the pitch variation curve (or equivalently on the time axis warping function) one can attempt to reduce the wow. Further on the wow reduction is performed by incommensurate sampling rate conversion using different interpolation techniques [4].

3. WOW REDUCTION BASED ON DIFFERENT INTERPOLATION TECHNIQUES

It is possible to recover the signal with the time axis t_{org} using either the pitch variation curve p_w or the time warping function f_w. Eq.3 presents the formula for the restored signal:

$$x(t_{rec}) = x\left(\frac{t_{wow}}{p_w(t_{wow})}\right) \approx x(t_{org}) \tag{3}$$

where the time axis t_{wow} is replaced by the time axis t_{rec}. In case of a discrete-time signal this wow reduction technique can be interpreted as a nonuniform interpolation based on the distorted samples $x(t_{wow})$ to give the restored signal $x(t_{rec}) \cong x(t_{org})$,see Fig.2. In the following a short overview of the interpolation techniques used further in experiments is presented.

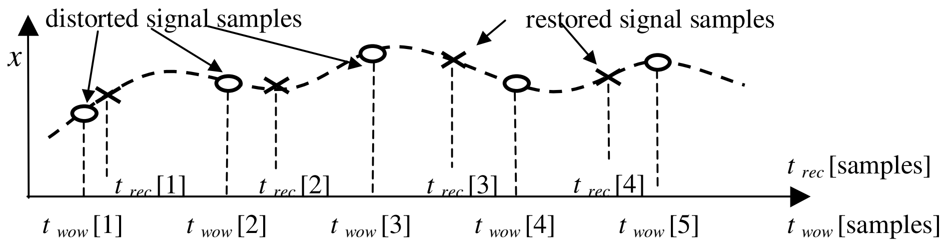

Fig. 2. An illustration of the idea of undistorted signal restoration using nonuniform interpolation.

3.1 Linear interpolation

Linear interpolation can be interpreted as joining two neighbouring signal samples by a straight line and returning the signal value along that line at an appropriate instant of time *t*. It is a very rough type of interpolation. Fig.3 presents a graphical interpretation of this method.

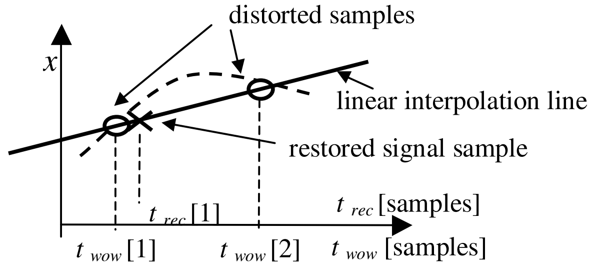

Fig. 3. A restored sample obtained using linear interpolation.

Fig.3 depicts that a significant difference between the true sample of the original signal shown by dashed line and the restored sample (x-mark) can occur. It is the main drawback of this method. However it will be used further in our experiments for the purpose of a comparison because of its simplicity and low computational cost.

3.2 Polynomial interpolation

Linear interpolation means using an interpolating polynomial of degree 1. More generally, an interpolating polynomial of degree $N-1$ for the wow reduction can be created on the basis of given N points: $x(t_{wow}[1])$, $x(t_{wow}[2])$,. . . , $x(t_{wow}[N])$ of a distorted signal. Such polynomial can be written by using the classical Lagrange formula:

$$\begin{aligned}P(t_{wow}) = &\frac{(t_{rec}-t_{wow}[2])\cdot(t_{rec}-t_{wow}[3])...(t_{rec}-t_{wow}[N])}{(t_{wow}[1]-t_{wow}[2])\cdot(t_{wow}[1]-t_{wow}[3])...(t_{wow}[1]-t_{wow}[N])}\cdot x_w(t_{wow}[1]) \\ &+...+\frac{(t_{rec}-t_{wow}[2])\cdot(t_{rec}-t_{wow}[3])...(t_{rec}-t_{wow}[N])}{(t_{wow}[N]-t_{wow}[1])\cdot(t_{wow}[N]-t_{wow}[2])...(t_{wow}[N]-t_{wow}[N-1])}\cdot x_w(t_{wow}[N])\end{aligned} \tag{4}$$

Fig.4 presents a graphical interpretation of this method where the interpolation is done based on a 2nd degree polynomial (i.e. 3 samples are necessary to obtain every one restored sample). Despite the Lagrangian interpolation there exist other, more sophisticated, polynomial-based interpolation methods. One of them is the piecewise cubic Hermite interpolation. The idea of this approach is the following [5].

- On each subinterval $t_{wow}[k] \leq t_{wow} \leq t_{wow}[k+1]$ a cubic Hermitian interpolating polynomial $P(t_{wow}[k]) \cong x(t_{wow}[k]\,)$ is created which interpolates both signal values and the derivatives at a set of nodes.
- $P(t_{wow})$ interpolates the given signal $x(t_{wow})$ in a way that the first derivative $P'(t_{wow})$ is

continuous. However, the second derivative $P''(t_{wow})$ is not necessarily continuous. There may occur jumps at the $t_{wow}[j]$.

- The slopes at the $t_{wow}[j]$ are chosen in such a way that $P(t_{wow})$ preserves the shape of the data and respects monotonicity. This means that on intervals where the data are monotonic, so is the $P(t_{wow})$ and at points where the data have a local extreme, so does the $P(t_{wow})$.

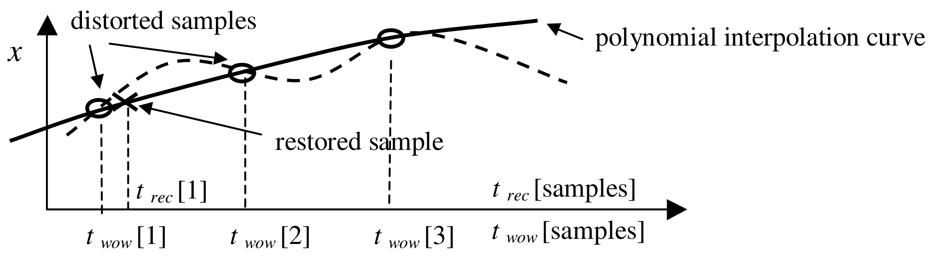

Fig. 4. An illustration of 2nd degree polynomial interpolation.

Another variant is the cubic spline interpolation. Here the interpolation polynomial is constructed in almost the same way as the Hermite one. However, spline chooses the slopes at $t_{wow}[j]$ differently, in order to make the $P''(t_{wow})$ continuous. This has the following effects [6].

- Spline produces a smoother result, i.e. $P''(t_{wow})$ is continuous.
- It produces a more accurate result if the data consist of values of a smooth function.

Both polynomial based methods, Hermitian and spline, are equally numerically expensive. In our experiments the 3rd order polynomials are used. Therefore four distorted samples are involved in the computation of each restored sample.

3.3 Sinc interpolation

The sinc interpolation of irregularly spaced samples is the one most closely connected with the Shannon sampling theorem. In this method the following expression for the sample $x(\tau_n)$ delayed by the amount $-T<\tau_n<T$ relative to $x[n]$ is used [4]:

$$x(\tau_n) = \sum_{m=-M}^{M} x[m]\gamma \,\mathrm{win}\left(\frac{\tau_n}{T} - m\right)\mathrm{sinc}\left(\gamma\left(\frac{\tau_n}{T} - m\right)\right) \tag{5}$$

An illustration is given in Fig.5.

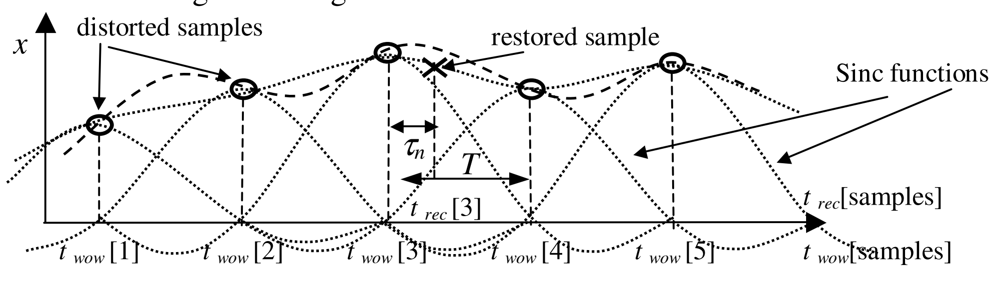

Fig. 5. Restored sample based on sinc interpolation

In Eq. 5 the n-th value $x(\tau_n)$ is a function of the nearest $2M+1$ neighbours. The factor γ is a number given by the minimum of number 1 and the current sampling rate conversion factor. The win function in (5), whose choice is crucial for the accuracy of interpolation, represents the time domain function for sinc windowing. Details can be found in [4].

4. EXPERIMENTS AND CONCLUSIONS

Fig.6 presents a block diagram of the processing used hereafter for the evaluation and comparison of the strength of different resampling routines considered in Sect.3.

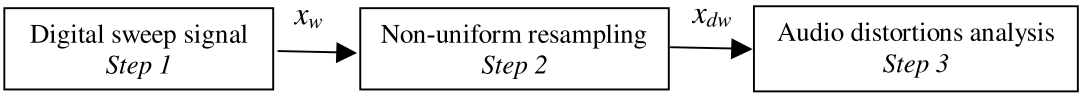

Fig. 6. Block diagram for a comparison of different resampling methods aimed at the wow reduction.

The evaluation method is based on three processing steps. In step 1 a synthetic audio input signal x_w is generated. Digital sweep tone was used for this purpose. Its frequency was varying from 1 kHz to 2 kHz. The signal was build at 8 kHz sampling rate. There are two reasons for choosing these frequency values. Firstly, the relation of the signal frequency to the sampling rate allows the signal spectrum to overlap only in a quarter of the signal bandwidth. Therefore the distortions generated by the investigated methods are clearly seen on the spectrograms. Secondly, the resampling routine generates a tone whose frequency, 1kHz, is well suited to the needs of audio distortion measurements. The second processing step in Fig.6 evolves one of the investigated interpolation techniques. The x_w signal is nonuniformly resampled in step 2 to obtain the output x_{dw} whose wow is reduced. The signal at the output of step 2 should have a constant frequency of 1kHz. It constitutes an input for step 3 - the audio distortion analysis block. Fig.7 presents spectrograms of the x_{dw} signals obtained using the interpolation techniques from Sect.3.

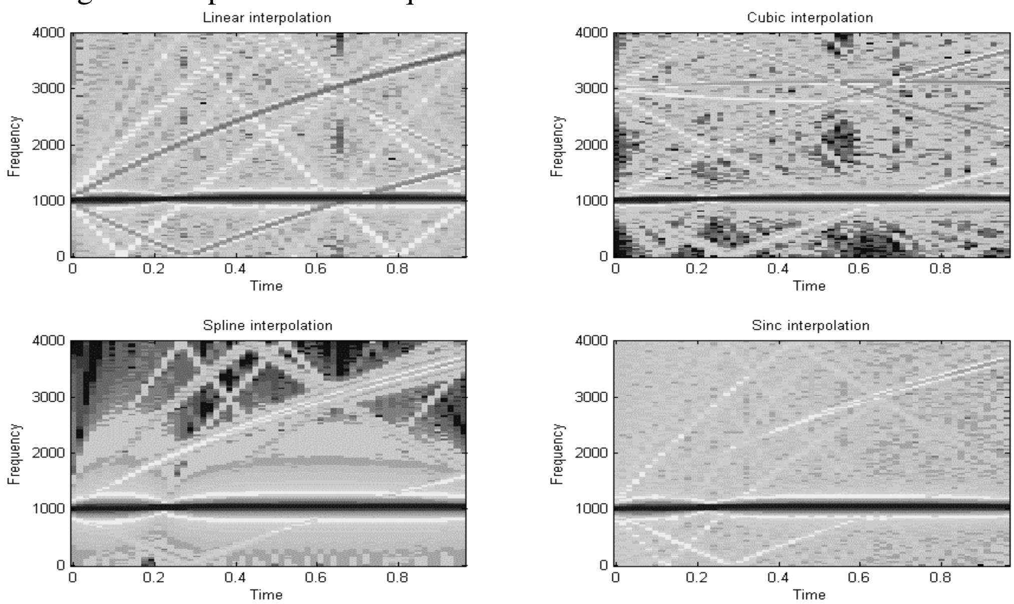

Fig. 7. Spectrograms of the signals at the output of step 2 in Fig. 6.

Within the audio distortions analysis block three standard audio distortion measures were applied. These are the following [7].

- Total Harmonic Distortion (THD) - the ratio of the harmonic power (HP) to the fundamental frequency power (FFP). It is computed by searching over the entire spectrum to find the peak frequency (fundamental) and then calculating the total power in the harmonic frequencies. The THD level is then computed as the ratio of the total HP to the FFP. Residual noise is not included in this calculation.
- Total Harmonic Distortion plus Noise (THD+N) - the ratio of the HP plus noise, to the FFP, computed by searching over the entire spectrum to find the peak frequency (fundamental) and then calculating the total power in the remaining spectrum (harmonics plus noise). The THD level is then computed as the ratio of the total HP plus noise power to the FFP.
- The Signal to Noise Ratio (SNR) - the ratio of the signal peak power level to the total noise level. The SNR is computed by searching over the entire spectrum to find the peak frequency and then calculating the total noise power in the remaining spectrum. The SNR is

then computed as the ratio of the noise power to the peak power and expressed in dB.
Table 1 presents the results. They are divided into two groups. Here the ‘short’ interpolation techniques involved no more than 5 neighbouring samples for one output sample computation. In linear interpolation 2 neighbouring input samples were used for one output sample computation wherein in the cubic and spline techniques 4 samples, and in the ‘short’ sinc interpolation 3 samples (M=1) and 5 samples (M=2) were involved. The ‘long’ interpolation results were obtained based on the windowed sinc technique with more than 20 ($M \geq 10$) samples involved. In experiments the von Hann widow was used.

Table 1. Audio distortions measured for four investigated methods of nonuniform resampling.

Audio distortion measure	Interpolation method						
	‘Short’ interpolation					‘Long’ interpolation	
	Linear	Cubic	Spline	Windowed Sinc M=1	Windowed Sinc M=2	Windowed Sinc M=10	Windowed Sinc M=100
THD [%]	0,67462	2,56364	0,05825	1,22855	0,08022	0,04791	0,04777
THD+N [%]	1,65586	3,05301	0,12263	3,12105	0,24188	0,04793	0,02161
SNR [dB]	35,62	30,305	58,228	30,144	52,353	66,173	73,306

The results gathered in Table 1 indicate that in terms of ‘short’ interpolations the spline technique is the most suitable for wow cancellation. But when the number of the samples involved increases, the distortions generated by the windowed sinc technique decrease and become smallest here. However, it should be mentioned that then the computation time becomes much longer than for the ‘short’ interpolation techniques. To summarize, for the windowed sinc the number of neighbouring samples (M) is very important to achieve the desired high SNR. A relatively high number of window samples must be applied for resampling. Notwithstanding that, because the reduction of the wow defect can be performed offline (when the computation time is not important) the ‘long’ sinc-based interpolation technique seems to be the most appropriate one for wow reduction expressed in terms of low audio noise level among the four interpolation techniques considered here.
It is also important to note a significant difference between the values of distortions measures obtained for the simple linear interpolation and the cubic one. The latter appears worse in the light of measures in Table 1. But the subjective rating of the output signal made by the author indicates an opposite effect. It means the need for more consistent measures directed at the evaluation of wow reduction efficiency. Therefore other new objective methods for restoration rating are currently under investigation. They involve the spectrum shape analysis.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Brock-Nannestad G., *What are the sources of the noises we remove?*, Preservation Tactics, DK-2820 Gentofte, Denmark

[2] Czyzewski A., Maziewski P, Dziubinski M., Kaczmarek A., Kostek B., *Wow detection and compensation employing spectral processing of audio*, 117 Audio Engineering Society Convention, Convention Paper 6212, 28-31 October, San Francisco 2004

[3] *Method for Measurement of Weighted Peak Flutter of Sound Recording and Reproducing Equipment*, AES standard 6-1982 (2003)

[4] Marvasti F., *Nonuniform Sampling Theory and Practice*, Kluwer Acad./Plenum Publishers, New York 2001.

[5] Fritsch, F.N., R.E. Carlson, *Monotone piecewise cubic interpolation*, SIAM J. Num. Analysis, Vol.17, 1980.

[6] de Boor C., *A Practical Guide to Splines*, Springer-Verlag, 1978.

[7] Metzler B., *Audio Measurement Handbook*, Audio Precision, 1973.