DSP Techniques for Determining "Wow" Distortion*

ANDRZEJ CZYZEWSKI, *AES Fellow,* **ANDRZEJ CIARKOWSKI, ANDRZEJ KACZMAREK,**
(ac@pg.gda.pl) (rabban@sound.eti.pg.gda.pl) (akacz@sound.eti.pg.gda.pl)

JOZEF KOTUS, MACIEJ KULESZA, AND PRZEMEK MAZIEWSKI
(joseph@sound.eti.pg.gda.pl) (maciej_k@sound.eti.pg.gda.pl) (przemas@sound.eti.pg.gda.pl)

Multimedia Systems Department, Gdansk University of Technology, Gdansk, Poland

Algorithms for determining the wow distortion characteristic are proposed. These are the power-line-hum tracking algorithm, the high-frequency-bias tracking algorithm, and the algorithm based on an adaptive analysis of the center of gravity of the spectrum of the distorted signal. All of the algorithms presented allow a hardware- or software-based implementation.

0 INTRODUCTION

Wow is an audio distortion perceived as an undesired frequency modulation (FM) in the range of approximately 0.5 to 6 Hz, which affects analog recordings. The distortion is introduced to a signal by an irregular velocity of the analog medium. The irregularities can originate from various mechanisms, depending on medium type, production technique, random damage of a signal's conveyor, and other factors [1]–[3]. Consequently the resulting parasitic FM [4] can range from periodic to transient, having different instantaneous values.

Regardless of preventative measures taken in the analog domain, wow cannot be avoided [1]–[3]. Therefore the distortion can be found in numerous recordings on wax cylinders, vinyl records, magnetic tapes, compact cassettes, and also on audio tracks in videos and movies. Only digital recording can be safe from wow.

In many respects the problem of wow remained unsolved in the analog domain. However, the digital-signal-processing (DSP) era permitted new approaches [5]–[7]. DSP algorithms can process sampled analog recordings in order to reduce the unwanted FM. Some research on wow reduction has already been reported in the literature [8]–[14]. These papers present only a few DSP algorithms for wow reduction. The algorithms, though interesting, suffer from some drawbacks, doubtful assumptions, and a small number of verified results. In addition, many problems related to wow processing have not been thoroughly studied. A more detailed discussion of these subjects is given in the next section. Thus despite the reported preliminary research, wow reduction in the digital domain remains an open topic in many aspects.

There are no commercially available tools for wow reduction known to the authors of this paper. The lack of DSP applications for wow processing was first identified by the archivists who are facing the challenge of preserving numerous analog archives. Many actions are undertaken to rescue and provide access to historical audiovisual contents. Among others, the European PrestoSpace project (PS) deals with the digital restoration of analog recordings [15]. One of the project's objectives is to provide a DSP application for wow reduction.

Being involved in the PrestoSpace project we faced the problem of researching and developing the wow reduction application. The task was found to be nontrivial because of the complex nature of the distortion and the lack of in-depth research on DSP-based wow reduction methods. Different approaches studied here resulted in several algorithms. As the PrestoSpace project enabled cooperation with the archive community, the algorithms researched were tested on real-life archive recordings. In this paper only the most successful and complementary DSP methods for estimating and reducing the distortion are presented. Other methods were indicated in our previous publications [16]–[21]. The algorithms described in this paper

*Manuscript received 2006 May 30; revised 2006 December 3 and 29.

were implemented in a personal computer (PC) application, which is the first step toward the missing DSP tool for wow reduction.

A general scheme for wow evaluation and reduction in the DSP domain is presented in Fig. 1. As can be seen from the diagram, wow processing is preceded by an evaluation of distortion annoyance. The evaluation stage is necessary because wow is not always audible; thus it is not always necessary to reduce it. Nonetheless, the distorted sounds with the perceptibly relevant wow should be restored. As the wow perceptibility and annoyance evaluation issues were discussed extensively in the past [22]–[24], the preprocessing stage is not studied here. In other words, an a priori assumption is made that the processed signal is distorted by perceptibly relevant wow.

In this paper we propose several algorithms for determining the wow characteristic. Consequently the determination procedure can be iterative, involving different algorithms as well as additional semimanual routines. The iterative approach is indicated in Fig. 1 by the decision block. After satisfying results are obtained, the distortion can be reduced using an appropriate resampling technique. A comparison of different resampling methods used for wow reduction was reported in our previous paper [20]. The current paper, however, presents only the algorithms for determining the wow characteristic.

The outline is as follows. In Section 2 the study of art is given, presenting the open issues of wow processing and explaining more deeply the research motivation. In Section 3 new algorithms for determining the distortion characteristic are presented. As the algorithms were designed and verified using some real-life archival recordings, the descriptions are accompanied by an analysis of archival sound examples. In Section 4 the results are discussed and a proposal is presented for the methodology of choosing the right algorithm for the determination of the wow characteristic.

1 REVIEW OF THE ART

Parasitic wow modulation is one of the audible effects of time-based distortions (TBD) caused by variations in the speed of the recording medium. The other effects are drift, flutter, and modulation noise. The effects differ in their modulation frequency, resulting in a different subjective perception. The wow modulation having its range as stated in the previous section is perceived as a fluctuation of pitch [4], [23], [25].

Time-based distortion is introduced into the signal because of speed irregularities of the recording medium. The irregularities can occur not only during the recording or duplicating process but also when reproducing [4], [23], [25]. Aging brings additional risk of time-based distortion [1], [2]. As a detailed discussion of the causes and effects of speed variations can be found in the literature [1]–[3], [10], [12], [14], [26], it is not addressed here. However, we want to point out that the various factors that come into play lead to the very complex nature of time-based distortion [23]. As a result the wow distortion, being a special case of time-based distortion, can have various characteristics, ranging from periodic to transient.

Various approaches to time-based distortion were introduced in analog machinery [1]–[3], [27]. A particularly interesting technique was the pilot tone recording of the magnetic tapes. In such a case the additional tone, namely, the pilot tone, with the FM depending on the tape speed, was recorded simultaneously with the sound track. Thus the tone modulation was depicting the speed variations of the tapes and could be used (involving demodulation) in the reproduction to steer the tape speed [1], [27]. All of the analog methods aimed at reducing the amount of speed variation. Nonetheless time-based distortions could not be avoided totally, and they are present in many archival recordings.

DSP allows for a different approach to the problem of wow reduction. The first ideas concerning algorithms for wow determination and reduction were introduced by Gerzon [5]. He proposed using the high-frequency bias found in magnetic tape recordings as the pilot tone, which depicts the wow characteristic. Thus the idea was similar to the analog pilot tone recordings. It was also Gerzon who first proposed to use the nonuniform resampling technique, controlled by the determined wow characteristic, for reducing the distortion.

Recently an implementation of Gerzon's ideas was reported by Howarth and Wolfe [14]. They presented a system that was a combination of the hardware part for capturing the bias and the software part for wow reduction. Although interesting, the report lacks a clear description of the DSP algorithms used for determining the wow characteristic. Thus it is impossible to analyze it or implement a similar one. Therefore in this paper the algorithm for bias analysis is presented. As other high-frequency pilot tones can be used for wow determination, such as the National Television Standards Committee (NTSC) video disturbances, the algorithm presented can be tuned easily for such an analysis. In addition a novel algorithm for analyzing power-line hum, which could also be viewed as a specific "pilot tone," is presented. As the hum is positioned very low on the frequency scale (50 or 60 Hz), the

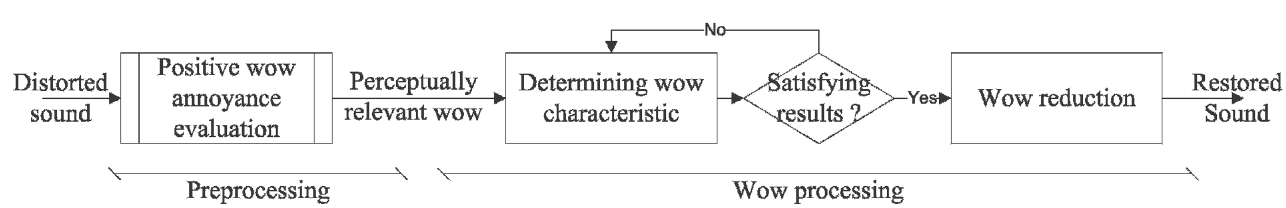

Fig. 1. General diagram for wow determination and reduction.

autoregression-based spectrum estimation algorithm was utilized. Both algorithms for pilot tone tracking presented here allow for software and hardware implementations.

Despite Gerzon's ideas to track the pilot tones, the first algorithms for wow evaluation and further reduction were built on different assumptions. Since wow leads to undesirable changes in all sound frequency components, algorithms that utilize methods adopted from sinusoidal modeling [28] were found very useful in the determination of the wow characteristic. Such an approach was first studied by Godsill [6]–[8] and reported thereafter [9], [10]. It was also explored by Nichols [11], [12].

Godsill algorithms were dedicated to the automatic correction of smooth pitch variations over long time scales. He proposed a three-step scheme. First the preprocessing stage, being the standard sinusoidal modeling analysis, provided information on the frequency partials, that is, trajectories made of spectral peaks, which are believed to be tonal. Second the partials were analyzed in order to determine the wow characteristic. Third nonuniform resampling, controlled by the characteristic evaluated, was used for the distortion reduction. The main emphasis in Godsill's work was on the determination of the wow characteristic. Here the Bayes' estimation scheme was used. Different a priori models were proposed to describe the distortion characteristic, including the simple sinusoidal as well as more general auto regressive processes.

Although Godsill's method offers some interesting properties, foremost the attempt to determine and reduce wow automatically, it suffers from some drawbacks. Most of all it is complicated, that is, it requires choosing and tuning the distortion model, which can be problematic for nonspecialists. Moreover it addresses only smooth and repetitive pitch variations over long time scales whereas the real-life wow distortion is frequently unique and short in time. Furthermore, some open questions regarding the method remain, such as the preprocessing influence of sinusoidal modeling on the results as well as the vibrato versus wow discrimination. Thus it appears that Godsill's fully automatic approach to wow reduction, although interesting, is not the most convenient in real-life situations.

A similar approach to wow processing was studied by Nichols [11], [12]. He presented the iterative application J_PITCH with two algorithms for wow determination (and an additional one for drift processing) and the nonuniform resampler for the final restoration. The first of Nichols' algorithms for wow determination was similar to Godsill's concept of partial processing. Nichols, however, introduced the iterative procedure for a better estimation of the spectral peaks' frequency as well as some additional features, such as spectrum warping. The second algorithm was based on a novel concept of graphical processing of the spectrogram. The time–frequency spectrogram was treated as a two dimensional graphical object. It was then searched for peaks, which, after the preprocessing necessary to exclude some false elements, were joined to form partials. Then the partials were postprocessed in the same manner as those from the first algorithm.

Although Nichols' approach refines the accuracy of sinusoidal modeling compared with Godsill's work, his procedure for determining the wow characteristic seems to be a step back. Nichols utilizes only one distortion model—sinusoidal. As the distortion is assumed to be periodic within the processing frame, the Nichols algorithms discern only the gross and periodic pitch variations. Thus even if the algorithms are sufficient to restore wow from the wax cylinder recordings, they are insufficient for the other recording mediums. This limitation is at least true for most of the examples.

Methods similar to Godsill's and Nichols' were studied in [19]. Two main problems were encountered. First the sinusoidal modeling involved in the preprocessing stage had a great impact on further processing. Choosing and tuning the sinusoidal modeling components, such as the short-time Fourier transform (STFT) parameters, peak selection, or partial building, was very important and highly signal-dependent. The second problem encountered was the model selection for the wow characteristic. The simple sinusoidal models were found insufficient whereas the more sophisticated autoregressive models were difficult in terms of choosing the right parameter set. It was especially noticeable when dealing with short and chaotic wow distortions. Consequently the method requires a skilled and experienced human operator to properly set the parameters for different recordings but the results are not always satisfying.

To overcome these problems we studied a different approach. In the new algorithm only the most prominent frequency component is being analyzed. The component can be chosen by the operator or picked automatically from the long-term Fourier spectrum. As it is only one component, a simple spectral characteristic, namely, the center of gravity, was chosen for its analysis. The analysis is performed adaptively in successive frames, providing information on the component's frequency variations. The variations are then processed to obtain the wow distortion characteristic. As there is no a priori model assumed, different types of distortions can be analyzed. In addition as the algorithm employs a straightforward spectral analysis, there is no need for the peak selection and partial building procedures. Therefore the method proposed herein overcomes typical constraints of the sinusoidal modeling analysis and a priori model selection.

The final wow reduction using the appropriate resampling technique is another issue, which was not studied thoroughly so far. In the literature different interpolation techniques were used for resampling. Godsill employed the truncated sinc interpolator [6]–[10]. Nichols used the bicubic interpolator [12]. Howarth and Wolfe, although they presented a description of the truncated sinc interpolator [13], in their paper describing the wow correction system itself [14] claim that the polynomial-based interpolators, namely, Hermitian or splines, are preferable to the sinc. As there is no clear comparison of these methods it was unclear which one is best for the wow reduction. The question of proper wow reduction was addressed in our previous paper, where we compared different resampling methods in terms of their influence on the audio quality [20]. Thus the most appropriate resampling technique for wow reduction can be chosen based on the results reported therein.

To sum up, this paper presents methods for determining the wow distortion characteristic. The determining algorithms are given in the next section. They are the algorithm for the bias analysis, a novel algorithm for the power-line-hum analysis, and an algorithm for adaptive analysis of the spectral center of gravity. The final section presents a discussion of the results.

2 ALGORITHMS FOR DETERMINING THE WOW CHARACTERISTIC

2.1 Wow Distortion Characteristic

As wow is a time-based distortion, it can be characterized using a function that depicts timing errors, that is, the time warping function $f_w(t_{org})$. This function characterizes wow as a distortion of the original time axis t_{org}. 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. 2.

The second commonly used wow characteristic is the pitch-variation curve $p_w(t_{org})$. The pitch-variation curve describes the parasitic FM caused by the irregular playback. Therefore it is closely related to the standard wow definition [4]. The pitch-variation curve can be defined as

$$p_w(t_{org}) = \frac{d[f_w(t_{org})]}{dt_{org}}. \tag{2}$$

If the pitch is constant, that is, there is no wow; $p_w(t_{org}) = 1$. Deviations from unity illustrate the wow modulation. In most real-life recordings the pitch-variation curve is close to a constant unity function with seldom varying parts.

Based on the distorted signal and the pitch-variation curve (or equivalently the time-warping function) one can attempt to reduce wow. Algorithms presented in the following subsections are aimed at the determination of the pitch-variation curve.

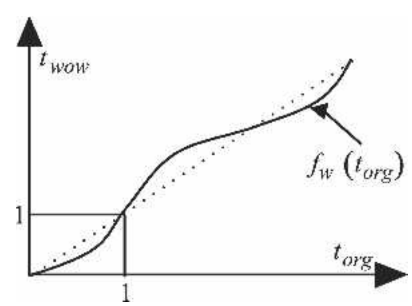

Fig. 2. Example of time-mapping function f_w (t_{org}).

2.2 Power-Line-Hum Analysis

2.2.1 Introduction—Characteristics of Hum Signal

Contemporary high-end studio equipment allows recording audio signals with virtually no hum. However, that was not always the case. In fact, many of the archival sound samples subjected to the authors' analysis were contaminated with the power-line hum. Its typical causes are poor power-supply stabilization of the vacuum-tube-era recording equipment and improper shielding of sensitive microphone cables. The parasitic hum signal was difficult to eliminate by analog filtering because of its rich harmonic structure and the relatively low frequency that requires large, expensive, and bulky high-inductance coils. This, and the fact that the hum frequency is typically lower than the frequencies of the useful audio components, makes it a particularly good carrier of information on the parasitic frequency modulation.

Some assumptions about the characteristics of the power-line-hum signal are as follows.

- It is assumed that hum was introduced in the audio path during the recording phase and was subjected to the same wow distortion as the useful signal.
- The hum frequency was stable throughout the process of recording.
- The sound material being restored was played back and digitized with modern high-class audio equipment, which introduces only a negligible amount of hum.

These assumptions are met in most real-life situations.

In all European countries the power-line frequency equals 50 Hz, in North America it is typically 60 Hz, and it is always one of these two values in all other countries. In the majority of archival recordings the hum frequency is slightly below or at the edge of the useful audio band. If caused by poor power-supply stabilization, the hum-related artifacts include also lots of harmonics (the rectification effect), which make for a particularly unpleasant listening experience. Sometimes sub harmonics are also included. Both the base-frequency hum signal and its harmonics may be used for tracking wow modulation. However, typically harmonics in excess of 100 Hz are hardly visible in the spectra because they are masked by the overlapping useful components of a higher power, as illustrated in Fig. 3. The figure shows the power spectrum magnitude plot of an exemplary archival recording with the clearly visible 50-Hz hum signal at −20 dB, its second harmonic at −25 dB, and the 25-Hz subharmonic at −45 dB. Higher harmonics are masked and therefore unsuitable for analysis.

The main advantage of using the power-line hum for the purpose of determining the wow characteristic is that the hum is relatively stable with regard to its frequency and level, and in a typical recording there is no or very little musical content in the frequency band it occupies. The main drawback is that the properties of such a low-frequency signal are extremely difficult to analyze with satisfactory precision. Moreover, the low power of the hum signal makes the analysis highly susceptible to accidental disturbances.

2.2.2. Hum Analysis through Autoregressive Modeling

It is a well-known property that achieving an arbitrarily high resolution simultaneously in both the time and the frequency domains is impossible. This stems from the inequality commonly known as the uncertainty principle [29]. In particular, in order to overcome the DFT constraints and fulfill the resolution requirements of the hum-based pitch-variation curve determination a method based on the autoregressive model was proposed. Autoregressive modeling is a method of so-called parametric spectrum estimation, based on the assumption that the observed signal is 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 pth-order autoregressive model has the form

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

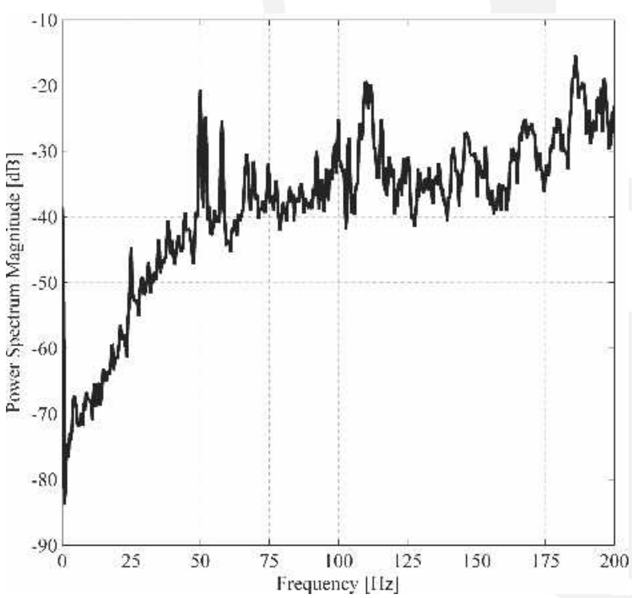

Fig. 3. Power-spectrum magnitude of archival recording with significant hum.

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

There are many methods that will allow estimating the parameters of the autoregressive model. The most common are autocovariance, modified autocovariance, and Burg and Yule–Walker methods [30]. They all share the common quality of producing highly accurate results based on extremely short observations, thus yielding a high resolution in both time and frequency.

2.2.3 Hum Tracking Algorithm

The method described in this section utilizes autoregressive modeling to determine the wow distortion characteristic through tracking of the hum frequency. A high-level overview is presented in Fig. 4. Since the algorithm was designed to extract the data from archival recordings, sometimes very noisy and of poor quality, a very important step is the preprocessing of the input data. Therefore it has been emphasized in the following description along with some hints based on the empirical knowledge gained throughout the process of algorithm development.

The original full-band signal analyzed apart from the power-line hum contains also some other tonal components, which typically have a much higher power than the hum. The analysis of such a signal would require the use of autoregressive models of a very high order. The higher order analysis would make for a very high computational overhead and could lead to a high number of false results due to the inclusion of so-called spurious peaks, which appear in the autoregressive model when its order is too high. Therefore in order to facilitate the solution of these problems, the input signal is downsampled prior to being split into frames. The downsampling allowed eliminating most of the non-hum-related tonal components from the signal's spectrum, thus making it possible to use a very low-order autoregressive model and consequently short frames. To further reduce the presence of noise and other non-hum-related components, the signal is bandpass filtered. The passband boundaries correspond strictly to the minimum and maximum relative detune of the hum trace. They are input as method parameters since they may be read from the spectrogram. The filter is designed as an

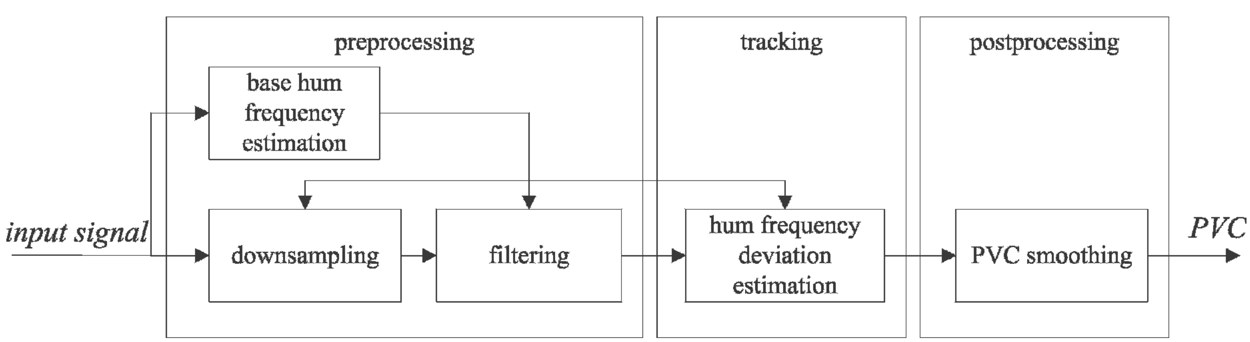

Fig. 4. High-level overview of hum-tracking method.

optimal equiripple finite impulse response (FIR) bandpass filter using the Parks–McClellan algorithm [31].

The next step is to track the hum properly with autoregressive modeling. For this purpose the downsampled and filtered signal is split into frames. Both frame length and overlap are the parameters of the method. It is important to note that the autoregressive model estimation described later performs poorly when the level of the hum signal changes rapidly within the analysis frame. Several first and last frames include transient states caused by filtering, so they are omitted. Omitting is done because the analysis of each subsequent frame depends on the results obtained for previous frames, and the inclusion of frames with transient states may cause the error to accumulate. The exact number of omitted frames is determined with regard to the length of the bandpass filter and frame as well as the frame overlap. It is derived from the equation

$$t_{\mathrm{f}} = \left\lceil \frac{\lfloor \text{filter length}/2 \rfloor}{f_{\mathrm{l}} - f_{\mathrm{o}}} \right\rceil \tag{4}$$

where t_{f} is the number of omitted frames, f_{l} is the frame length in samples, and f_{o} is the frame overlap, also in samples.

During the analysis each input frame is assigned a single value representing its estimated relative detune. These values ordered in a vector represent a rough estimate of the curve. The ignored starting (or ending) frames are assigned the pitch-variation curve value of the first (or last) nonignored frame in order to maintain pitch-variation-curve continuity.

Autoregressive modeling is used in the process of estimating the pitch-variation curve for each frame. For a given frame, the coefficients of the corresponding IIR filter are calculated using the modified autocovariance method. As a result the coordinates of transmittance poles are obtained. Subsequently one pole is selected, and the frequency corresponding to its angle normalized with the base hum frequency is chosen as the pitch-variation curve value. The pole selection algorithm is presented in Fig. 5.

The pole for each frame is selected depending on the mean frequency of M previously selected poles. A frequency memory vector is used, which initially is filled in with some start frequencies, defaulting to the base hum frequency. For a given model order, filter poles are calculated, and their frequencies are tested against the proximity constraint with the mean of the memory frequency. If none of the poles meets the constraint, the model order is incremented and the tests are performed until the model order reaches half the length of the frame. If during these iterations the pole that meets the proximity constraint is found, its frequency is stored in the frequency memory and the iteration stops. Otherwise the pole that was closest to meeting the constraint is chosen. The length M of the frequency memory is a method parameter, and it controls how susceptible the algorithm is to rapid changes. The memory length allows to limit the possibility of the algorithm "tuning into" high noise or accidental tonal components with frequencies close to hum, if there are any. The proximity constraint predicate is described by

$$|f_{\mathrm{mean}} - f_{\mathrm{curr}}| < f_{\mathrm{d}} \left(\frac{f_{\mathrm{l}} - f_{\mathrm{o}}}{f_{\mathrm{l}}} \right) \tag{5}$$

where f_{mean} is the mean of the memory frequency, f_{curr} is the frequency of the current pole, and f_{d} is the frequency deviation parameter.

Typically an f_{d} parameter in the range of 0.001 to 0.005 yields satisfactory results for most wow distortions. The dynamic matching of the model order improves the accuracy of frequency selection significantly, but at a slightly increased numerical cost.

During the experiments it was tested whether the algorithm can benefit from taking into account the magnitude of the pole in the selection process, but no noticeable increase in the algorithm accuracy was discovered. Therefore the use of magnitude was abandoned. Nevertheless it

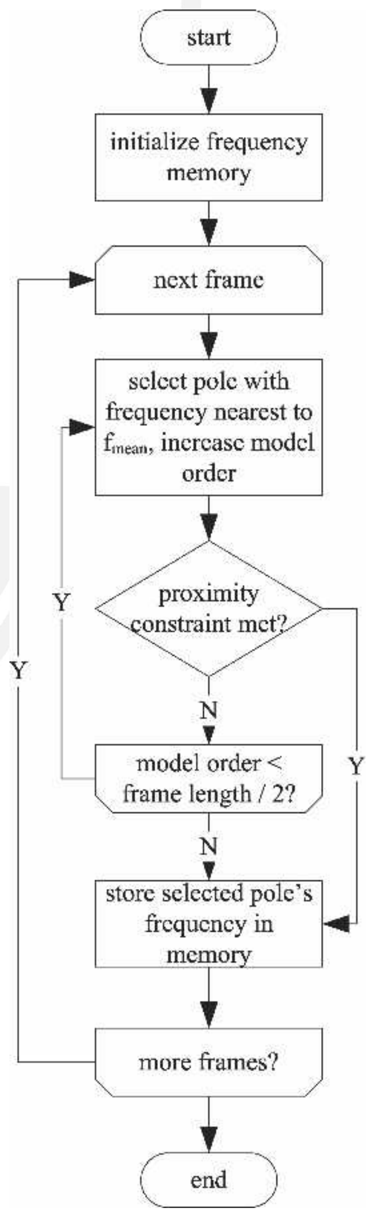

Fig. 5. Pole-selection algorithm.

was observed that major changes (reflecting several orders of magnitude) in the flow of pole magnitudes might indicate regions where the analysis is flawed. The observation may lead to the development of some kind of certainty measure based on the pole magnitude changes selected.

As a result of the pole selection algorithm described, a vector representing the estimated pitch-variation flow is obtained. It is further subjected to smoothing, which is achieved in a two-stage procedure. First median filtering is used, which rejects possible extreme values, and then a moving-average filter that smoothes the remaining "steps" is applied.

2.2.4 Experiments

A number of tests were performed in order to evaluate the usefulness of the algorithm described. First we used an excerpt taken from an archival magnetic recording. The spectrogram of the fragment, presented in Fig. 6, has highly compressed dynamics in order to better visualize the hum trace. As visible in the spectrogram, the fragment presents an uncommon situation where the hum trace is detuned at the beginning of the fragment. The initial detune puts a requirement to initiate the frequency memory with the correct starting conditions. Failure to do so will result in a highly erroneous pitch-variation curve estimate (an example is presented in Fig. 7), an effect of the difficulties the algorithm may encounter when tuning into the hum trace. To overcome the "tuning into" problem, an initial condition was specified, and the frequency memory was initiated to the value of 54 Hz read from the spectrogram in Fig. 6. The estimate thus obtained, shown in Fig. 8, proved to be accurate.

The next fragment for processing is also taken from an archival recording. Its spectrogram is presented in Fig. 9. As seen, the hum trace fades out momentarily and reappears between 5 and 6 seconds, possibly due to some kind of a dynamic processing (such as a noise gate). Fading causes transient states to appear. As discussed earlier, the method described performs poorly in such conditions where the number of fluctuations increases, as is visible in Fig. 10. Nevertheless the hum trace estimate is quite accurate and proves that the implemented algorithm performs well under disadvantageous conditions, allowing to track even rapid detunes, like the one visible in Fig. 9 between 9 and 10 seconds.

The method described produces good results when used appropriately. Its important advantage is that the method is numerically cheap, that is, even a Matlab implementation not optimized numerically typically runs several times faster than the real-time process. Though the algorithm has many "tunable" parameters, it was found empirically that for most signals, some predetermined, default values yield correct results. In fact, only a minimal amount of user interaction is required, so the restoration process may be highly automated.

2.3 High-Frequency Bias Analysis

2.3.1 Introduction

During the process of recording, a high-frequency bias signal of the analog recording device is used. Such a tech-

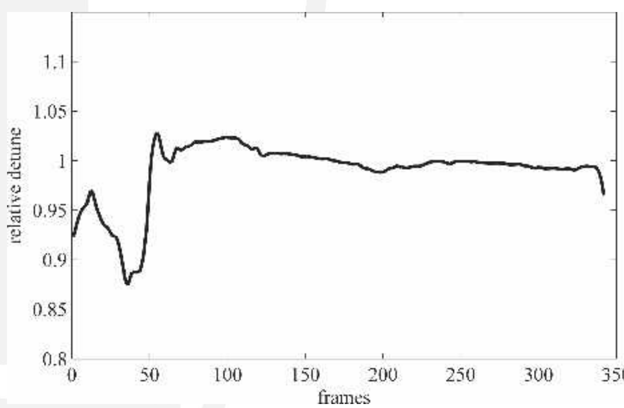

Fig. 7. Plot of pitch-variation curve extracted from an archival recording (incorrect initial conditions). Length of x axis—≈18.3 s; positions of markers—0, 2.6155 s, 5.2310 s, 7.8465 s, 10.4620 s, 13.0775 s, 15.6930 s, and 18.3085 s.

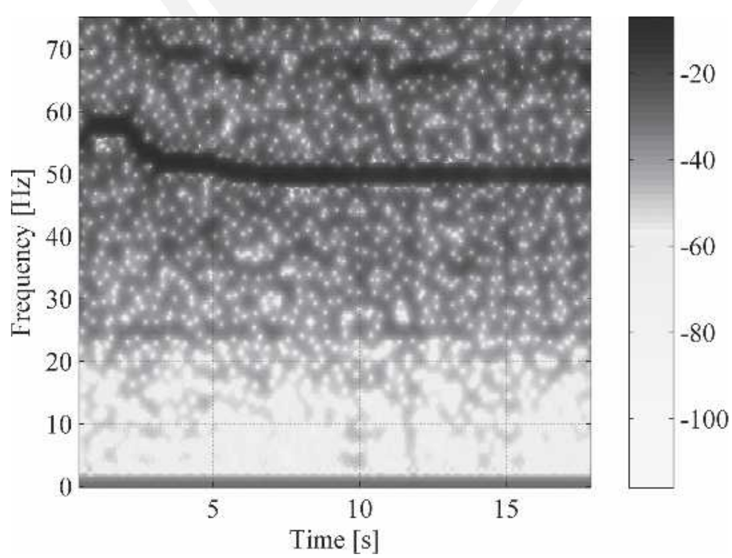

Fig. 6. Spectrogram of archival recording sample.

nique permits us to avoid distortions related to the magnetization process of the tape. Since the bias frequency is initially constant, wow distortions may be reflected by the deviations of the bias signal which is then used easily to retrieve the distortion properties (that is, the pitch-variation curve). Further it can be utilized in the restoration process. The frequency of the bias signal is related to the type of material of which the tape is made. The bias frequency may vary from 30 to 435 kHz [14]. Moreover, it is almost impossible to predict the nominal frequency of the bias signal for a particular recording and the range of its frequency variations. Due to the limitations of the tape player frequency characteristic (too narrow bandwidth of the player head to encompass the bias) and the restrictions described, a direct digitization of the audio material containing the bias is almost impossible [18]. Thus an appropriate method, which can be viewed as a preprocessing procedure for the tracking algorithm, must be employed when digitizing the bias signal.

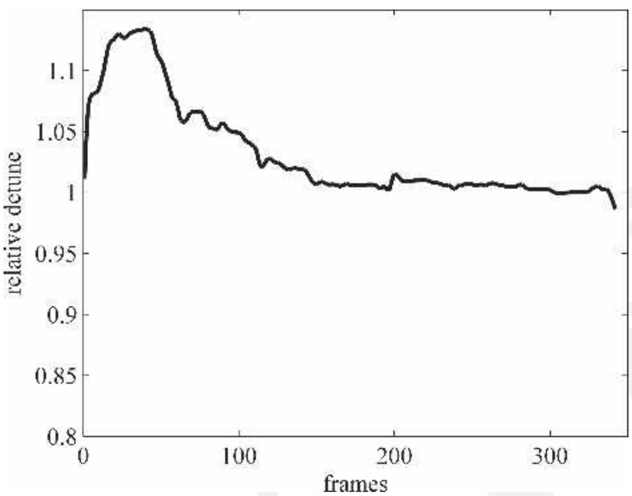

Fig. 8. Plot of pitch-variation curve extracted from an archival recording with correct initial conditions (initial hum frequency set to 54 Hz according to spectrogram). Length of x axis—approximately 18.3 s; positions of markers—0, 5.2310 s, 10.4620 s, and 15.6930 s.

In order to capture the bias signal, techniques usually utilized in analog or digital radio receivers may be applied. One straightforward method is to mix the bias signal with a locally generated constant-frequency sinusoidal tone and then apply the low-pass filtering procedure. Shifting the bias signal down to the frequency suitable for the analog-to-digital conversion may also be achieved by employing the undersampling technique [18]. Of course, for these two methods a dedicated interface must be constructed and added. Furthermore it is usually required that the built-in tape player head be replaced to ensure that the bandwidth of the reproduced signal is wide enough to include the bias component.

The method used during the practical experiments presented in this paper does not require the use of an additional hardware interface or rebuilding a tape recorder. It could be applied when the nominal frequency of the bias

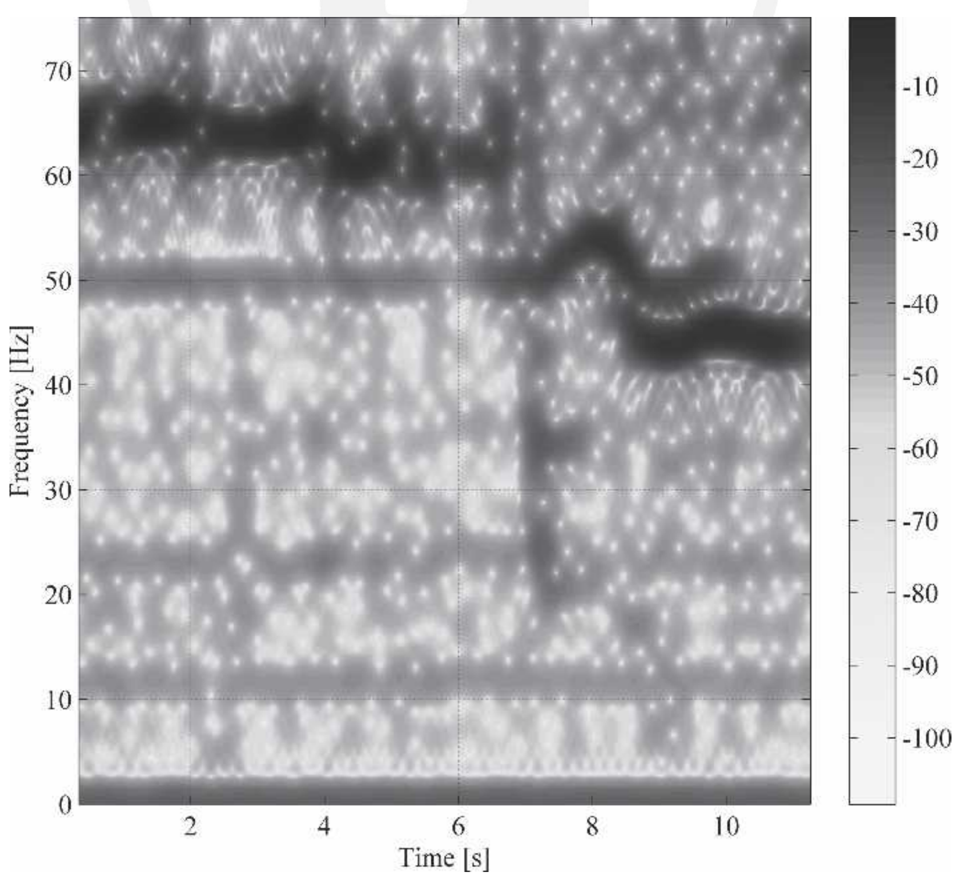

Fig. 9. Spectrogram of archival recording fragment.

is not higher than about 100 kHz. The idea is to reduce the tape speed during playback (digitization). As a result the audio signal bandwidth is compressed and the frequency of the bias signal is shifted downward. If the tape speed reduction rate is high enough, the tape player is able to reproduce the transposed bias signal, which enables capturing the time-stretched audio containing the bias signal [17]. After the analog-to-digital conversion process is completed, the original time relations of the audio content must be restored. There is no need to involve any DSP technique to obtain the original time relations. In fact, a new value of the sampling frequency must be set in the wave file header. The idea is to increase the sampling frequency to the reciprocal of the rate of the tape speed reduction. The new sampling frequency must be set according to the formula

$$f_{s_{\mathrm{new}}} = \frac{f_s}{R} \tag{6}$$

where f_s stands for the sampling frequency during the digitization process and R is the tape speed reduction rate.

The tape speed reduction method for tracking the bias may introduce some additional distortions to the audio content. The audio signal is filtered according to the frequency characteristic of the entire signal chain of the tape recorder, during playback at the reduced speed. The filtration is mostly perceived for low-frequency components of audio signals, which are usually transposed beyond the useful bandwidth of the electronic circuits of tape recorders. It can be viewed as high-pass filtering.

During the experiments with bias tracking the tape was played back using the Kudelski Nagra III reporter tape recorder and also the Studer Revox B77 studio master tape recorder. The bias nominal frequencies for tapes played using these two tape recorders were 73.7 kHz and 80 kHz, respectively. Archival audio signals (originating from the 1970s) stored on ¼-inch magnetic tapes were utilized. It is possible to switch the tape playback rate from 3.75 through 7.5 to 15 in/s with the Nagra III recorder. The audio material contaminated with the wow distortion, originally recorded at the rate of 15 in/s, was played back at 3.75 in/s. As a result the signal bandwidth was compressed, and the frequency of the bias signal was transposed down to a frequency reproducible by the player head. It was detected that the bandwidth of the Nagra III tape player head does not exceed 25 kHz. Therefore the 48-kHz sampling frequency was found high enough to digitize the stretched audio signal. Sampling four times the stretched audio material at a frequency of 48 kHz can be treated as equivalent to sampling the same signal at 192 kHz if played at the original speed. The initial time relations of the digitized audio content were restored by increasing the value of the sampling frequency to 192 kHz in the wave file header. In further experiments the Revox B77 recorder was used to play the audio material, recorded with the tape running at 7.5 in/s. In this case the sampling frequency was 96 kHz and the tape was running at 3.75 in/s during the digitization. Consequently in the last step the initial time relations of the sampled audio material were restored.

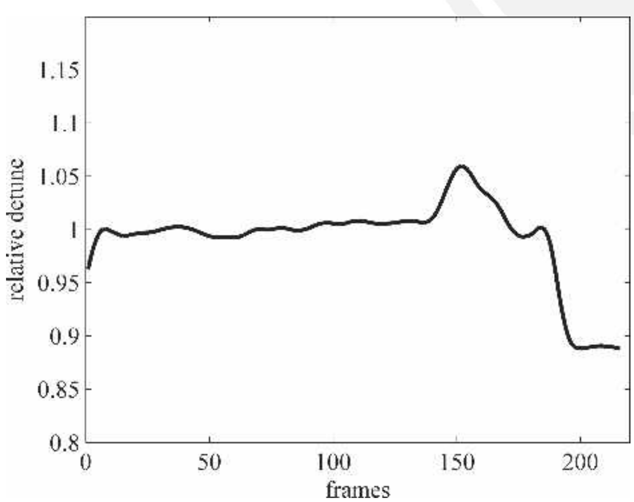

Fig. 10. Plot of pitch-variation curve extracted from an archival recording. Length of x axis—≈11.26 s; positions of markers—0, 2.6065 s, 5.2130 s, 7.8194 s, and 10.4259 s.

Obviously it is sensible to apply the bias tracking method only to those archival recordings that still contain the bias signal. Although there is no clear rule to predict the presence of bias in a particular magnetic recording, it can be noted that aging degrades the bias signal, causing irregularities in its amplitude, and leads to fading. The experiments with archival recordings revealed that it is possible to discover the bias signal on magnetic tapes recorded more than 30 years ago. According to the high energy level of the bias signal detected in those recordings, it could be assumed that even older archival recordings may contain it. Thus tracking the bias seems to be a practical method for determining the wow characteristic.

2.3.2 Algorithm for High-Frequency Bias Tracking

The algorithm developed here makes the determination of the pitch-variation curve possible, which is done based on the high-frequency bias properties. It has to be assumed, however, that the disturbance was introduced after the initially constant-frequency bias signal was recorded, and that the recorded material was digitized using a method that allows capturing both the audio content and the bias signal [17].

Since the bias signal is of high frequency, STFT analysis seems to be a suitable tool for detecting its time–frequency variations. A diagram illustrating the engineered algorithm for determining the bias-based pitch-variation curve is presented in Fig. 11.

In the first stage the input signal is divided into short segments (in conformity with the STFT analysis concept). The operator, in accordance with the bias signal properties, may adjust the lengths of the blocks as well as the overlap. The von Hann window weights every block of the segmented signal. Further, the Fourier spectrum for each block is calculated. As a result of the spectrum calculations the spectrogram matrix, representing the time–frequency properties of the signal, is obtained. Spectral components representing frequencies below 25 kHz are set to zero in order to remove the high-energy components that are related to the audio signal and may obscure the bias. Setting to zero can be viewed as high-pass filtering. In addition each amplitude spectrum (each column of the spectrogram matrix) is weighted by an appropriate preemphasis curve, allowing enhancement of the high-frequency components. The curve properties (slope) are also user adjustable according to the formula

$$y[n] = \left(\frac{n}{N-1}\right)^s, \qquad n = 0, \ldots, N-1 \tag{7}$$

where n is the sample number as before, $y[n]$ is the preemphasis curve, N is the block size, and s is the slope ratio.

The characteristics of the preemphasis for $s = 1, 2, 3, 4$, and $N = 512$ are presented in Fig. 12.

The preemphasis leads to a modified time–frequency representation with enhanced components representing the bias signal. Since aging degrades the bias signal, causing significant irregularities in its amplitude, detection of the spectral peak representing the bias may be difficult, or even impossible for some of the STFT segments. Thus an appropriate smoothing is necessary. In practical experiments a moving average filter operating in the frequency and time domains, smoothing, respectively, each row and column of the spectrogram matrix, allows to limit the amplitude irregularities, and to determine the bias frequency peak in the spectrum of each processing block. The equation of the moving average filter is

$$y[n] = \frac{1}{M_{\mathrm{Ma}}} \sum_{i=0}^{M_{\mathrm{Ma}}-1} x[n+i] \tag{8}$$

where M_{MA} is the filter order.

The spectrograms of an unaltered signal containing the bias and processed according to the procedure described are presented in Fig. 13. In this example the preemphasis slope ratio was set to 2 and the order of the spectrogram-smoothing filter operating in both the frequency and time domains was adjusted to 3. One can notice that after smoothing, the components representing the bias signal are enhanced and also the bias amplitude irregularities are reduced significantly. Consequently the tracking procedure may be applied.

It is assumed that the highest energy peaks are the bias peaks. In addition to enhance the accuracy of frequency detection, each peak is interpolated with a parabolic curve. The correction value of the index representing the frequency of the bias for each spectrum is given by

$$i_{\mathrm{corr}} = -\frac{b}{2a} \tag{9}$$

where

$$b = \frac{A[i_{\max}+1] - A[i_{\max}-1]}{2} \tag{10a}$$

$$a = A[i_{\max}-1] - A[i_{\max}+1] + b \tag{10b}$$

and where $A[i]$ is the value of the amplitude spectrum and $i_{\max}$ is the index of the maximum amplitude peak.

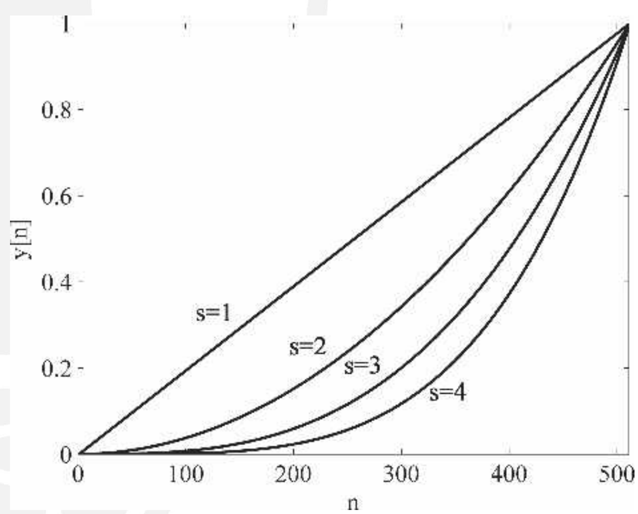

Fig. 12. Preemphasis curves for slope ratio $s = 1, 2, 3, 4$.

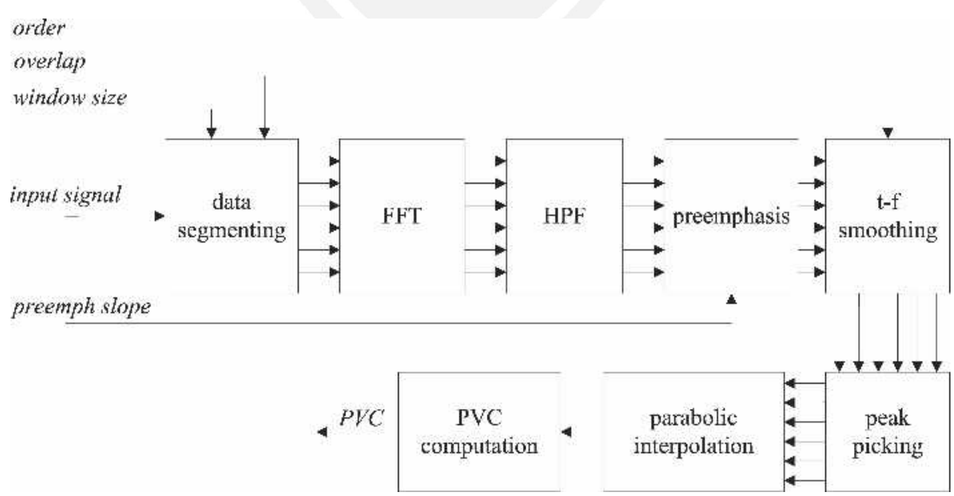

Fig. 11. Algorithm for determination of pitch-variation curve. FFT—fast Fourier transform; HPF—high-pass filtering, t-f—time–frequency.

The vector containing the estimated frequencies of the bias for each data segment can then be represented by

$$\mathrm{freq}_{\mathrm{bias}}[j] = \frac{f_s}{2} \frac{i_{\max}[j] + i_{\mathrm{corr}}[j]}{L/2 + 1} \tag{11}$$

where j is the number of the signal segment (block of samples) and L is the length of the block.

Fig. 14 shows the spectrogram of a sound sample containing the bias signal (black line) along with the detected bias track (white line) corresponding to the wow disturbance. Although the engineered algorithm requires setting the analysis parameters individually for every sound record, it can be seen from Fig. 14 that it is effective in terms of producing the bias track without noticeable artifacts.

In order to estimate the pitch-variation curve, the vector containing the bias frequencies for all blocks must be normalized by the value of the nominal bias frequency. Although the median value of detected bias frequencies would be an efficient way to estimate the nominal bias frequency, it would also require the off-line mode of operation. Therefore to allow on-line implementation of the bias-tracking algorithm, the first detected bias frequency is assumed as its nominal value and is used to calculate the pitch-variation curve (PVC). Thus the PVC value for each signal segment can be calculated according to the formula

$$\mathrm{PVC}[j] = \frac{\mathrm{freq}_{\mathrm{bias}}[j]}{\mathrm{freq}_{\mathrm{bias}}[0]} \tag{12}$$

where j is the number of the signal block, as before, and $\mathrm{freq}_{\mathrm{bias}}[j]$ is the estimated frequency of the bias.

The experiments confirm that such simplicity does not influence the reliability of the pitch-variation curve and leads to an accurate estimate.

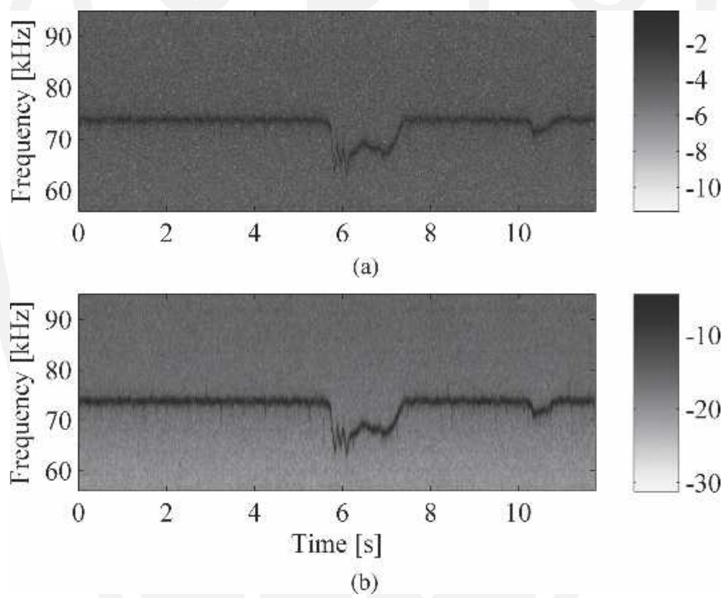

Fig. 13. Spectrograms. (a) Unaltered bias signal. (b) Processed according to procedure described.

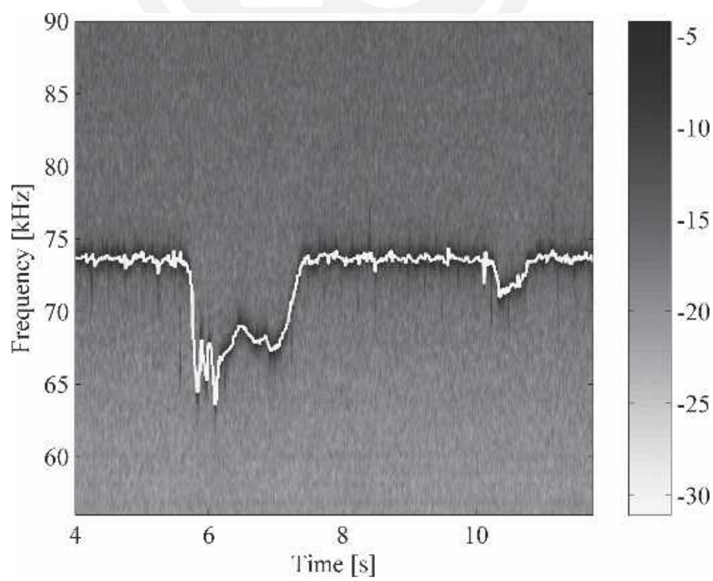

Fig. 14. Part of a spectrogram of a signal containing bias (black) and showing also detected bias track (white).

2.3.3 Experiments

Since the algorithm proposed for the PVC determination using the bias is not completely automated, the calibration of the processing parameters for particular sound samples must be performed. The user can adjust the values of five processing parameters:

- Block size for DFT analysis
- Overlap size
- Preemphasis slope ratio
- Filter order for frequency-domain smoothing
- Filter order for time-domain smoothing.

Obviously the time and frequency resolution of the bias analysis is determined by the block size and overlap size parameters. It was found experimentally that using a block length ranging from 2.66 to 21.33 ms (512 to 4096 samples for the 192-kHz sampling frequency) yields satisfactory results in most cases. A shorter block size should be used in order to obtain proper estimates of the pitch-variation curve if the wow distortion changes rapidly with time. Rapid pitch disturbances may be triggered, for example, by inaccurate tape splitting. On the other hand, using greater block sizes and nonzero overlap allows a more precise determination of the pitch-variation curve in the case of slow and shallow pitch changes.

As mentioned previously the preemphasis operation can be perceived as a specific high-pass filtering. The higher the nominal frequency of the bias, the higher the value of the preemphasis slope ratio should be. The preemphasis rule is absolutely true if the bias frequency rises above its nominal value. Special attention must be given to the range of frequency changes when wow introduces a downward pitch modulation. If a wide pitch variation occurs it is suggested to set the preemphasis slope ratio parameter initially to 1.

Since the preemphasis operation exposed the bias component in the signal spectrum, the frequency-domain smoothing prevents the influence of noise on the accuracy of PVC determination. It was found experimentally that averaging the frequency over the range from 375 to 2250 Hz is suitable for most cases. In consequence, the filter order parameter for time-domain smoothing should be chosen in accordance with the level of the bias signal amplitude irregularities, the block length, and the value of the overlap parameter. Usually it is necessary to adjust higher values of the time-domain smoothing order parameter when restoring the contaminated archival recordings, and if high time-resolution analysis is required.

According to the aforementioned rules the bias frequency determination procedure was applied to sound samples captured using the Kudelski Nagra III (sound samples 1 and 2) and the Studer Revox B77 (sound sample 3) tape recorders. Table 1 presents the processing parameters chosen for these particular sound samples. The bias characteristics obtained are plotted in Figs. 15–17.

As can be seen in Figs. 15–17 the engineered algorithm is effective in producing the bias track without any artifacts. Note that for the first sound sample, a relatively high preemphasis slope ratio was chosen with regard to upward frequency changes of the bias. A high value of the time-domain smoothing parameter is crucial for the second sound sample because of frequent bias signal discontinuities. As the bias frequency was deeply modulated by the

Table 1. Processing parameters chosen for particular sound samples.

	Sound Sample		
Parameter	1	2	3
Block length	512	1024	1024
Overlap [%]	25	50	75
Preemphasis slope ratio	3	1	2
Frequency-domain smoothing order	3	3	2
Time-domain smoothing order	3	4	2

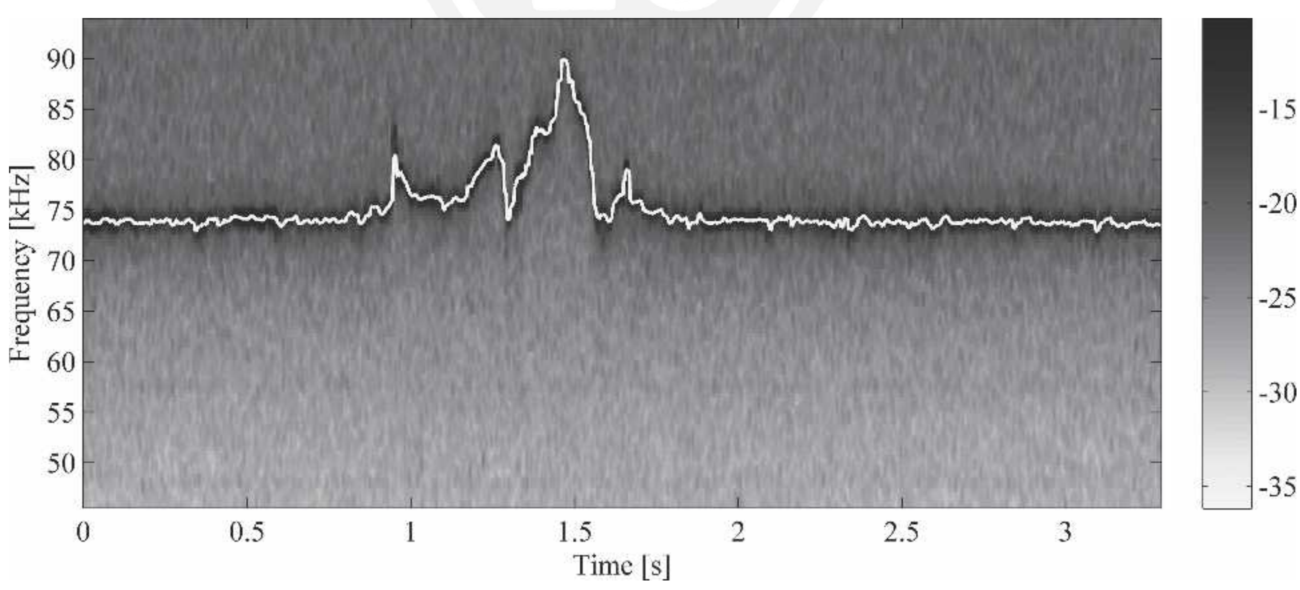

Fig. 15. Track of bias pilot tone (sound sample 1).

wow distortion a low value of the preemphasis slope parameter had to be adjusted. The sound sample reproduced with the Revox B77 tape recorder is noiseless, and thus only a negligible smoothing in the time and frequency domains was applied. In order to improve the accuracy of the bias tracking, a greater overlap size was chosen for this particular sound sample.

Finally, normalization of the bias track by the bias nominal frequency yielded the pitch-variation curve estimate, which can be utilized easily to reduce the presence of wow in the original sound sample. The listening tests revealed that sound samples were restored nearly perfectly in terms of pitch disturbance. In addition in restored sound samples the bias frequency remains constant over time, which proves the tracking accuracy. It is also worth mentioning that the proper selection of the processing parameter values is fairly intuitive and does not require much experience. Notwithstanding the difficulty of capturing both the audio content and the bias signal, the procedure proposed herein is a robust method for estimating the pitch-variation curve.

2.4 Spectrum Center of Gravity Analysis

2.4.1 Introduction

Wow introduces FM to all spectral components of the distorted audio signal. The modulation is well visible when analyzing pilot tones such as hum or bias, which are separated from the useful audio signal. Regrettably the pilot tones are not always available or reliable. In such situations the spectrum of the useful audio signal, which is also affected by the wow modulations, can provide the information necessary to determine the distortion characteristic. In the past some approaches to determining the pitch-variation curve using a simultaneous analysis of numerous audio-related spectral peaks were reported [6]–[12]. The algorithms described, though interesting, suffered from some drawbacks (see Section 1). The

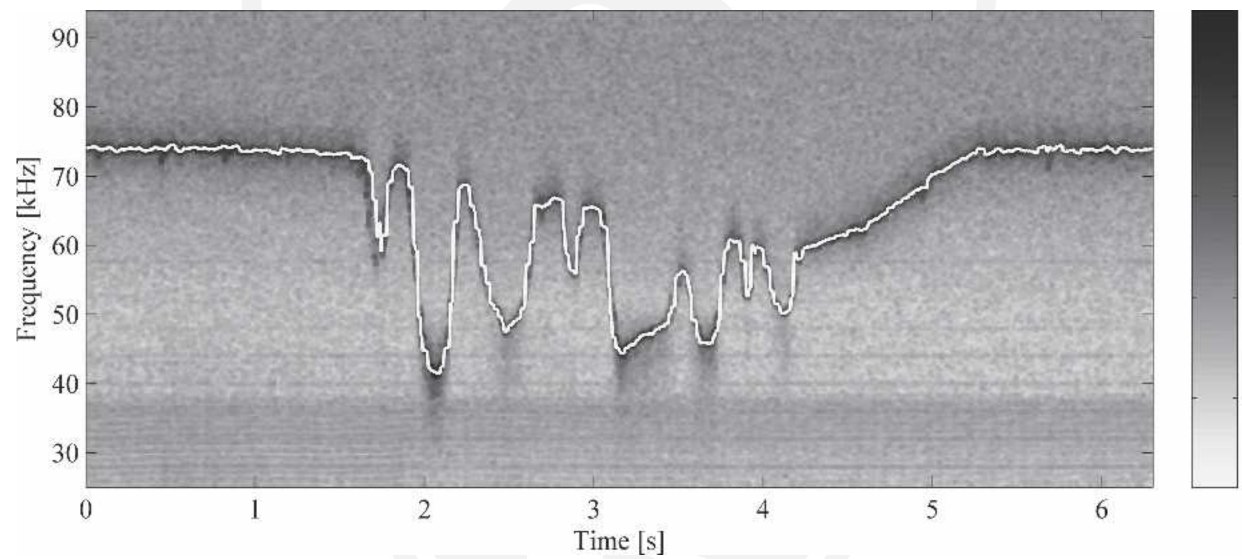

Fig. 16. Track of bias pilot tone (sound sample 2).

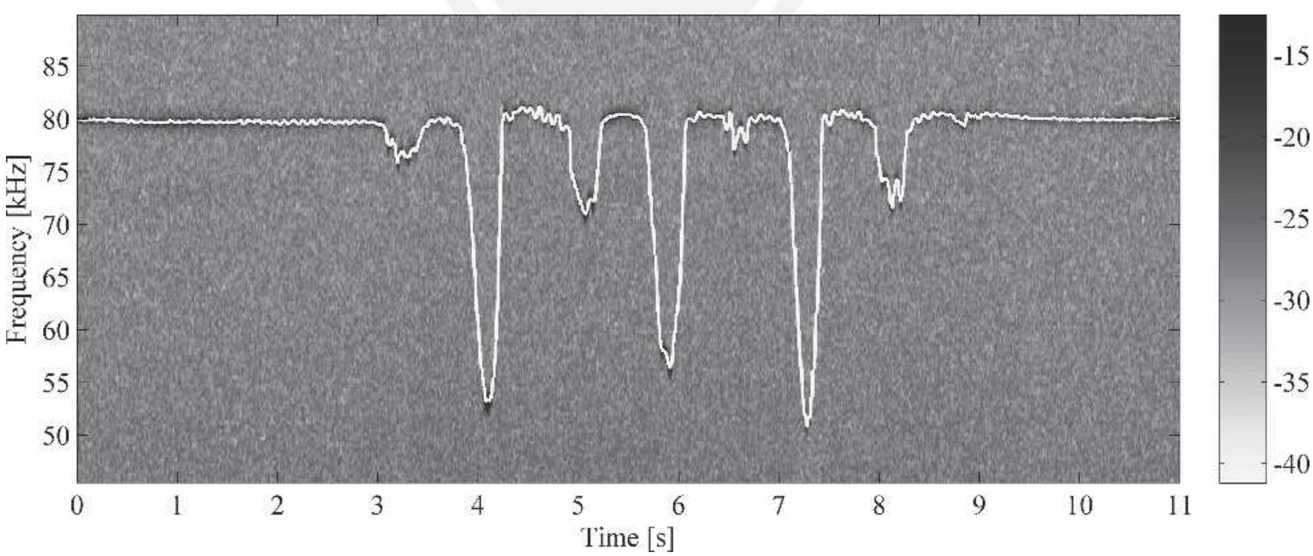

Fig. 17. Track of bias pilot tone (sound sample 3).

method presented in the following paragraphs overcomes some of the disadvantages of the previously published algorithms. It takes advantage of the observation that wow modulation is noticeable when only one particular spectral component is analyzed. Thus an analysis performed by calculating values of a simple and intuitive spectral descriptor, which depicts the component's position, allows the determination of the wow characteristic. The spectral center of gravity (SC) was chosen to describe the properties of the spectral component. The following paragraphs present the algorithm for adaptive center-of-gravity analysis used for determining the pitch-variation curve.

2.4.2 Algorithm for Adaptive SC Analysis

The spectral center of gravity (or the audio spectrum centroid [32]) is a well-known spectral descriptor. It is defined on the basis of the signal power spectrum [32]. Although there are other definitions with the amplitude spectrum involved [33], the one given by Eq. (13) seems to be the most convenient. It is because of the simple and intuitive interpretation of the denominator-scaling factor, that is, the sum of the spectral signal components equals the total signal power,

$$\mathrm{SC}_{\mathrm{Hz}} = \frac{\sum_{l=0}^{N_\mathrm{b}} S_x[l]\, f_l}{\sum_{l=0}^{N_\mathrm{b}} S_x[l]} \tag{13}$$

where l is the frequency bin index, N_b is the total number of frequency bins, f_l is the lth bin frequency given in hertz, and $S_\mathrm{x}[l]$ is the discrete power spectrum of the lth bin.

When dealing with a simple spectrum having one prominent peak, the center of gravity points out the peak's frequency. In case of a more complex spectrum with numerous peaks the center of gravity indicates the spectrum's weighted mean. In order to analyze only the prominent peak chosen from the complex spectrum, the center of gravity must be defined only for the frequency band closely enclosing the peak. In addition, as the FM caused by wow has a multiplicatory nature, that is, frequency deviations are greater for higher carrier frequencies, the center-of-gravity computation must be performed on a logarithmic scale. Given this, the expression for the band-pass octave-scale center of gravity can be set as

$$\mathrm{SC}_{\mathrm{oct}} = \frac{\sum_{l=N_\mathrm{L}}^{N_\mathrm{U}} S_x[l] \log_2 f_l}{\sum_{l=N_\mathrm{L}}^{N_\mathrm{U}} S_x[l]} \tag{14}$$

where l, f_l, and $S_\mathrm{x}[l]$ are as in Eq. (13), N_L is the number of the frequency bin closest to the lower frequency border f_L, and N_U is the number of the frequency bin closest to the higher frequency border f_U.

In order to track the peak's position variations in time the analysis of $\mathrm{SC}_{\mathrm{oct}}$ must be performed in consecutive time frames. In such a case the boundary frequencies must be adaptively modified in a way to track the component variations. f_L and f_U for the consecutive time frames can be evaluated according to the formulas

$$f_\mathrm{L}[i+1] = 2^{(\mathrm{SC}_{\mathrm{oct}}[i] - \Delta f_\mathrm{L})} \tag{15a}$$

$$f_\mathrm{U}[i+1] = 2^{(\mathrm{SC}_{\mathrm{oct}}[i] + \Delta f_\mathrm{U})} \tag{15b}$$

where Δf_L and Δf_U are half-widths of the frequency band analyzed and are constant in all time frames,

$$\Delta f_\mathrm{L} = \mathrm{SC}_{\mathrm{oct}}[1] - \log_2 f_\mathrm{L} \tag{16a}$$

$$\Delta f_\mathrm{U} = \log_2 f_\mathrm{U} - \mathrm{SC}_{\mathrm{oct}}[1]. \tag{16b}$$

When performing the progressive center-and-gravity analysis, additional difficulties arise. The edges of the frequency band analyzed can be biased by some random spectral components migrating in and out of the analyzed band in the successive frames. In order to prevent such migration, spectrum windowing can be used. In the experiments reported the von Hann window was involved [34]. Similar to Eq. (14) the octave scale was used for the frequency windowing,

$$W_{\mathrm{oct}}(f) = 0.5 - 0.5\cos\left(2\pi \frac{\log_2 f - \log_2 f_\mathrm{L}}{\log_2 f_\mathrm{U} - \log_2 f_\mathrm{L}}\right) \tag{17}$$

where f is the frequency in the $\langle f_\mathrm{L}, f_\mathrm{U}\rangle$ band.

Frequency windowing eliminates the outermost frequencies from the analysis, leaving the most prominent center part. In case of the octave scale the window's maximum values are placed near the geometrical mean of the analyzed band. Owing to the frequency windowing, the band-pass octave-scale windowed center of gravity is given as

$$\mathrm{SC}''_{\mathrm{oct}} = \frac{\sum_{l=N_\mathrm{L}}^{N_\mathrm{U}} S_x[l] W_{\mathrm{oct}}(f_l) \log_2 f_l}{\sum_{l=N_\mathrm{L}}^{N_\mathrm{U}} S_x[l]} \tag{18}$$

whereas the final value for the analyzed frame given in the linear (hertz) scale is

$$\mathrm{SC}'_{\mathrm{Hz}} = 2^{\mathrm{SC}''_{\mathrm{oct}}}. \tag{19}$$

The center of gravity depicts the peak position in the current time frame. Analysis in the consecutive frames yields the information on the peak variations in time. Thus as the analysis reveals the peak FM, it can be used for the determination of the wow characteristic. The block diagram for determining the pitch-variation curve using adaptive center-of-gravity analysis is given in Fig. 18.

According to the scheme presented the distorted signal is divided into short time frames (the i index). Then for each ith frame the power spectrum S_x is calculated. In the proposed algorithm the periodogram was used to estimate the power spectra [35],

$$\hat{S}_x[k] = \frac{1}{K} |X_\mathrm{w}[k]|^2 \tag{20}$$

where k is the frequency bin index, $X_\mathrm{w}[k]$ is the DFT of the signal $x_\mathrm{w}[n]$, and K is the number of samples in the time frame.

Based on the $\hat{S}_x$ estimate the center of gravity of the chosen frequency band is calculated [Eqs. (18) and (19)]. Selection of the initial boundary frequencies f_L and f_U can be done manually (as in the following experiments) or using a simple algorithm for choosing the valleys around the most prominent peak in the signal's long-term DFT. In the algorithm the boundary frequencies are updated according to Eqs. (15a)–(16b).

Fig. 19 shows the spectrogram of a certain recording depicting the wow-modulated tonal components. The track selected for the analysis is placed around 1400 Hz. The solid white line presents the computed center of gravity, whereas the dashed lines present the frequency band borders. It can be seen that the center of gravity as well as the boundary frequencies adaptively follow the tonal component variations.

Having the center of gravity values for the successive time frames, the distortion characteristic can be calculated. The pitch-variation curve is computed in two steps. First the relative frequency deviations are calculated. Second, in order to eliminate the threat of a global pitch disturbance introduced by the pitch-variation curve with a nonzero mean value, the characteristic's mean is subtracted.

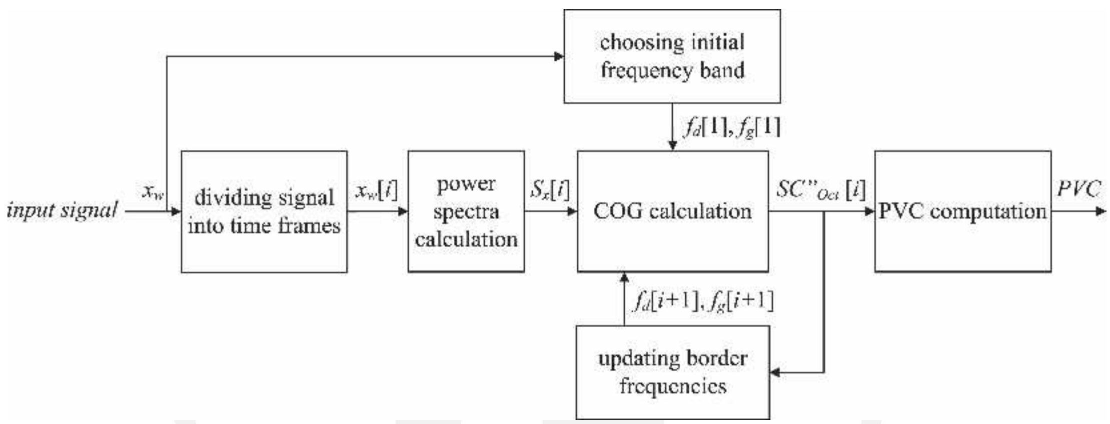

Fig. 18. Block diagram of adaptive center-of-gravity analysis for determining pitch-variation curve.

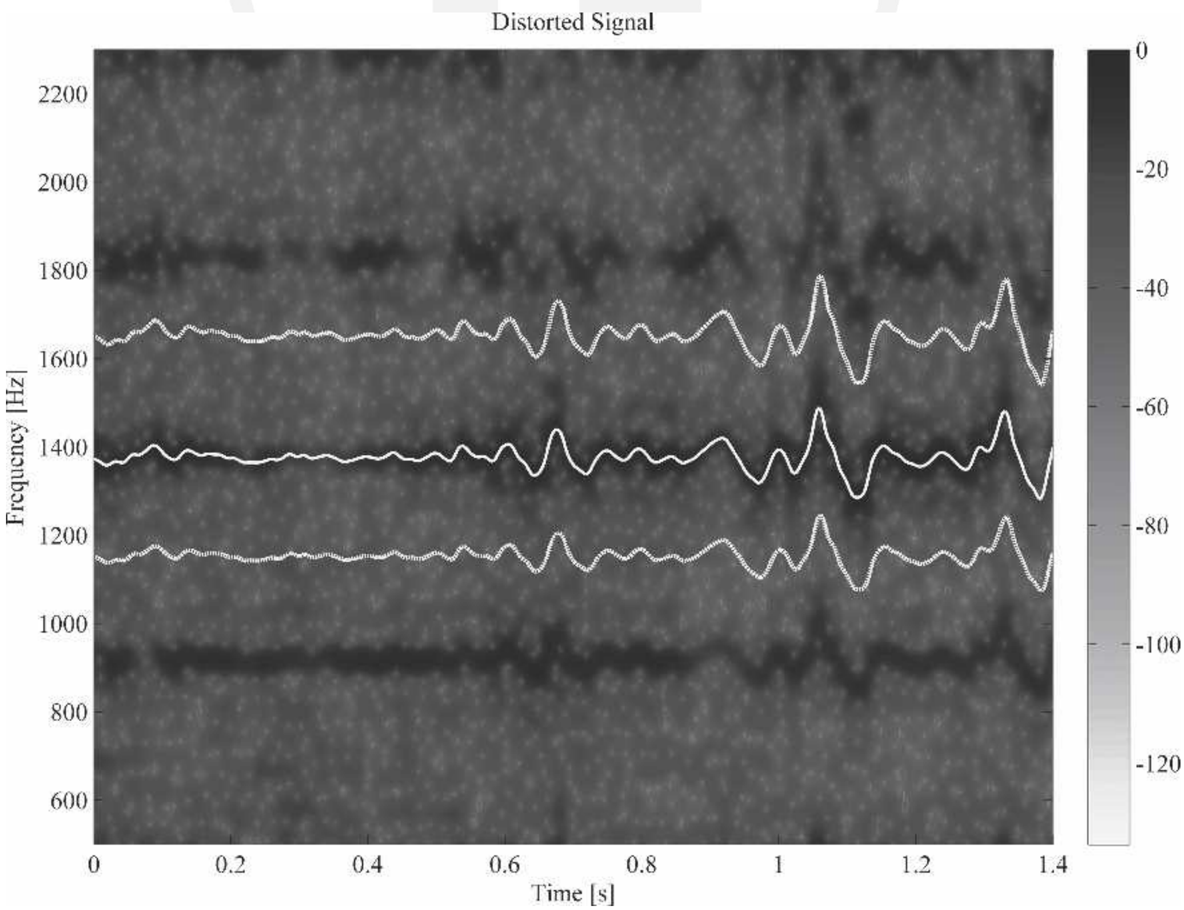

Fig. 19. Distorted signal (excerpt 1) with computed center of gravity (thin white line) and boundary frequencies (dashed white lines).

2.4.3 Experiments

Fig. 19 presents an example of a successfully determined pitch-variation curve using adaptive center-of-gravity analysis. The sound recording utilized (excerpt 1) was taken from the soundtrack of a Polish film recorded in the late 1930s. The second example (excerpt 2), taken from the same soundtrack, is given in Fig. 20. The figure presents the spectrogram of the wow-distorted part together with the calculated center of gravity, which depicts the wow modulation. It can be noticed that the wow modulation is well tracked.

The third example (excerpt 3) is taken from another archive movie, also recorded in the late 1930s (see Fig. 21). Here the spectral structure was more complex and dense. Nonetheless, carefully set boundary frequencies allowed for a successful center-of-gravity analysis.

In all of the examples presented the center-of-gravity analysis yielded the pitch-variation curves that allowed for almost perfect restorations. The method was easy to use, requiring only the starting boundary frequencies to be set by the operator. Such a selection is straightforward when using a spectrogram, that presents the modulated components. In addition it can be further facilitated by a simple analysis of the long-term Fourier spectrum aimed at choosing the valleys around the most prominent peak. As the method is easy to use it overcomes one of the main problems of the algorithms reported earlier [6–12]. Also, there is no a priori pitch-variation-curve model involved. Thus the center-of-gravity analysis can track various modulations.

3 DISCUSSION AND CONCLUSIONS

Pilot tones are available in the archival sound recordings. The presence of the pilot tones was the first observation gained during the analysis of the archival sounds available. The power-line hum was found in most of the magnetic tapes analyzed. The bias signal, though needing a special approach for capturing the high-frequency content, was detected in many magnetic tapes recorded more than 30 years ago. In addition other pilot tones were encountered, such as the standard NTSC video disturbance at 15.734 kHz. Only in the optical soundtrack recordings no pilot tones were found. Nonetheless, because of the presence of the tones the pilot tone analysis seems to be a good approach to the determination of the wow characteristic.

Two algorithms for pilot-tone tracking, which were presented here, were found very useful in determining the wow characteristic. The first, the bias tracker, yielded more precise results. The increased precision is due to the obvious observation that the wow modulation is depicted more precisely by high-frequency components, that is, the modulations have a logarithmic frequency scale. Thus the bias analysis allowed for more precise pitch-variation curves. Furthermore the bias algorithm can be tuned easily for analyzing different pilot tones placed in the high-frequency band, such as the above-mentioned 15.734-kHz tone encountered in the NTSC videos. The other advantage of the high-frequency tones is that they are well separated from the genuine audio content. Thus the analysis is simple and can be almost automatic. However, it can be

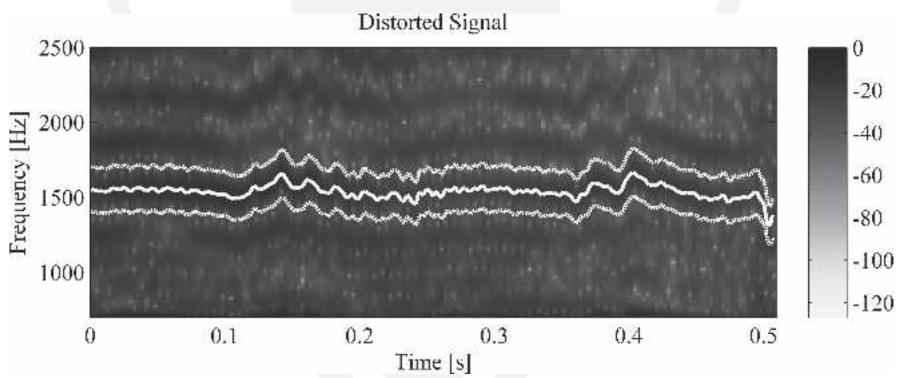

Fig. 20. Center-of-gravity vector for wow modulation from excerpt 2.

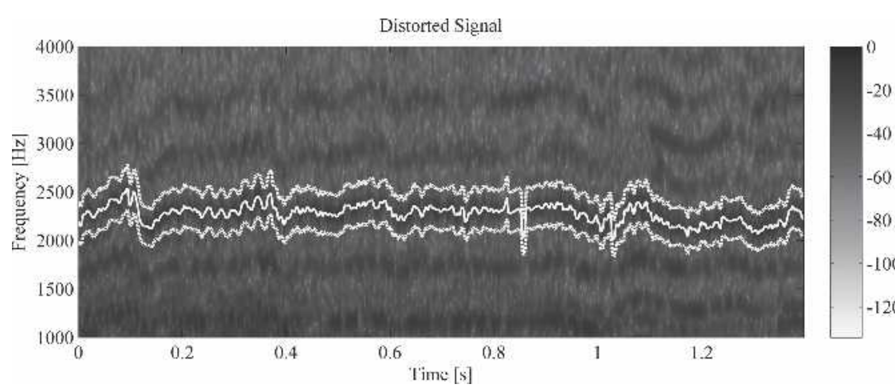

Fig. 21. Center-of-gravity vector for wow modulation from excerpt 3.

inferred that a human operator is necessary to judge and correct the automatically determined characteristics.

The power-line-hum tracker is the second algorithm for the pilot tone analysis. It was found that the results obtained when using it are rougher. The pitch-variation curve generated depicts only the general trends of the parasitic pitch variations. This is due to the nature of wow distortions, namely, as the hum lies very low on the frequency scale it is less affected by the distortion. In many cases the hum is not unique, that is, there can be two or even three tones placed around 50–60 Hz. This is probably due to copying or other processing of the original source materials. In addition, the hum can be affected by the subharmonics of the genuine audio content. In consequence the hum tracker requires more human assistance to assess and correct the automatically generated pitch-variation curves.

No high-frequency pilot tones were found in optical soundtracks. In such circumstances the audio itself must provide information for determining the wow characteristic. Adaptive center-of-gravity analysis was used in such situations. It was found that this approach yields good results when analyzing short audio parts with a distinct tonal structure of the signal spectrum. Thus a human operator is needed for choosing the right audio part. Nonetheless, when the fragment was found, the use of center of gravity was simple and intuitive, providing accurate pitch-variation curves.

Finally it was found beneficial to use the algorithms for determining the pitch-variation curves iteratively. It is practical to correct some false results of the bias tracker by using the center-of-gravity analyzer. In the case of an iterative approach to the determination of the pitch-variation curve, fast resampling is needed for monitoring the intermediate restoration results. Based on the research reported in [20] the authors suggest using the spline interpolation technique for processing the intermediate results. Nonetheless, when the final pitch-variation curve is determined, long sinc resampling (involving more then 20 neighboring samples) should be used for the final restoration. Sinc resampling requires more computational overhead, but it introduces a smaller amount of distortion.

The presented observations and conclusions allowed for general clues on the wow processing application. First, as most of the determining algorithms require a human operator, the application should be interactive. A fully automatic analysis, though it can be imagined for some of the materials with distinct and unique high-frequency pilot tones, is not likely to be possible for all of the distorted recordings. Second, as the distorted samples can originate from different analog recording mediums, where it is unknown whether pilot tones will be available, different algorithms allowing for various approaches to the determination of wow should be provided. Finally, the application should afford at least two types of resampling techniques, the first for fast monitoring of the intermediate results and the second, more precise one, for the final wow reduction.

4 ACKNOWLEDGMENT

This research was funded by the Commission of the European Communities, Directorate-General of the Information Society, within Integrated Project FP6-507336, entitled "PRESTOSPACE—Preservation towards Storage and Access. Standardized Practices for Audiovisual Contents Archiving in Europe." The authors would like to thank the anonymous reviewers for helpful comments and suggestions, which improved the presentation of this paper considerably.

5 REFERENCES

[1] E. D. Daniel, C. D. Mee, and M. H. Clark, *Magnetic Recording: The First 100 Years* (IEEE Press, New York, 1999).

[2] P. Read and M. P. Meyer, *Restoration of Motion Picture Film* (Butterworth Heinemann, Oxford, UK, 2000).

[3] E. S. Busby, "Analog Tape Recording," in *Standard Handbook of Audio and Radio Engineering,* J. Whitaker and B. K. Benson, Eds. (McGraw-Hill, New York, 2003), pp. 6-21–6-42.

[4] AES Std. 6-1982, "Method for Measurement of Weighted Peak Flutter of Sound Recording and Reproducing Equipment," Audio Engineering Society, New York (2003).

[5] M. Gerzon, "Don't Destroy the Archives!," unpublished (1992); available at www.waves.com/german/htmls/service/faq/dont.html (accessed 2006 Oct.).

[6] S. J. Godsill and P. Rayner, "The Restoration of Pitch Variation Defects in Gramophone Recordings," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (Mohonk, NY, 1993 Oct. 17–20), pp. 148–151.

[7] S. J. Godsill, "The Restoration of Degraded Audio Signals," Ph.D. dissertation, Cambridge University, Cambridge, UK (1993).

[8] S. J. Godsill, "Recursive Restoration of Pitch Variation Defects in Musical Recordings," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing 2,* (Adelaide, Australia (1994 Apr. 19–22), pp. 233–236.

[9] S. J. Godsill and P. J. W. Rayner, "Restoration of Pitch Variation Defects," in *Digital Audio Restoration—A Statistical Model-Based Approach* (Springer, London, 1998), pp. 171–190.

[10] S. J. Godsill, G. Reid, and C. Hicks, "Restoration of Smooth Pitch Variations over Long Timescales," presented at the Joint Technical Symposium, Toronto, Canada (2004 June 24–26); available at http://www.jts2004.org/english/proceedings/Reid.htm (accessed 2006 Oct.).

[11] J. Nichols, "An Interactive Pitch Defect Correction System for Archival Audio," in *Proc. AES 20th Int. Conf.* (Budapest, Hungary, 2001 Oct. 5–7), paper 1941.

[12] J. Nichols, "High Performance, Low-Cost Wax Cylinder Transcription System," in *Proc. AES 20th Int. Conf.* (Budapest, Hungary, 2001 Oct. 5–7), paper 1937.

[13] P. Wolfe and J. Howarth, "Nonuniform Sampling Theory in Audio Signal Processing," presented at the 116 Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 52, p. 813 (2004 July/Aug.), convention paper 6123.

[14] J. Howarth and P. Wolfe, "Correction of Wow and Flutter Effects in Analog Tape Transfers," presented at the 117 Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 53, p. 91 (2005 Jan./Feb.), convention paper 6213.

[15] D. Terrugi, "Can We Save Our Audio-Visual Heritage?," *Ariadne,* vol. 39 (2004 Apr.); available at http://www.ariadne.ac.uk/issue39/teruggi/ (accessed 2006 Oct.).

[16] A. Czyzewski, P. Maziewski, M. Dziubinski, A. Kaczmarek, and B. Kostek, "Wow Detection and Compensation Employing Spectral Processing of Audio", presented at the 117 Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 53, p. 90 (2005 Jan./Feb.), convention paper 6212.

[17] A. Czyzewski, M. Dziubinski, A. Ciarkowski, M. Kulesza, P. Maziewski, and J. Kotus, "New Algorithms for Wow and Flutter Detection and Compensation in Audio," presented at the 118 Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 53, p. 669 (2005 July/Aug.), convention paper 6353.

[18] 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," in *Proc. AES 26th Int. Conf.* (Denver, CO, 2005 July 7–9), paper 3-3.

[19] L. Litwic and P. Maziewski, "Evaluation of Wow Defects Based on Tonal Components Detection and Tracking," in *Proc. XI Int. AES Symp. of Sound Engineering and Tonmeistering* (Polish Section, Krakow, Poland, 2005 June 23–25), pp. 145–150.

[20] P. Maziewski, "Wow Defect Reduction Based on Interpolation Techniques," in *Proc. 4th Nat. Electronics Conf.,* vol. 1/2 (Darlowko Wschodnie, Poland, 2005 June 12–15), pp. 481–486.

[21] P. Maziewski, L. Litwic, and A. Czyzewski. "Accidental Wow Evaluation Based on Sinusoidal Modeling and Neural Nets Prediction," presented at the 120 Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 54, p. 709 (2006 July/Aug.), convention paper 6769.

[22] F. A. Comerci, "Perceptibility of Flutter in Speech and Music," *IRE Trans. Audio,* pp. 62–70 (1955 May–June).

[23] J. G. McKnight, "Time Base Distortion in Continuous Recording Systems. Its Terminology, Measurement, Causes and Effects," *J. Audio Eng. Soc.,* vol. 10, pp. 44–48 (1962 Jan.).

[24] H. Sakai, "Perceptibility of Wow and Flutter," *J. Audio Eng. Soc.,* vol. 18, pp. 290–298 (1970 June).

[25] W. Wolf, "Electromechanical Analogs of the Filter Systems Used in Sound Recording Transports," *IEEE Trans. Audio Electroacoust.,* vol. AU-14, pp. 66–85 (1966 June).

[26] Y. Fujimoto, M. Suzuki, K. Fujio, K. Sasamoto, and Y. Satoh, "A New Method of Reducing Direct-Drive Motor Vibration in Turntables," *J. Audio Eng. Soc.* (*Engineering Reports*), vol. 31, pp. 246–252 (1983 Apr.).

[27] L. Ryder, "Synchronous Sound for Motion Pictures," *J. Audio Eng. Soc.,* vol. 16, pp. 291–295 (1968 July).

[28] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation." *IEEE Trans. Acoust. Speech, Signal Process.,* vol. 34, pp. 744–754 (1986 Aug.).

[29] J. Smith, *Mathematics of the Discrete Fourier Transform (DFT)* (W3K Publ., Stanford, CA, 2003), also available at http://www-ccrma.stanford.edu/~jos/mdft/.

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

[31] J. S. Lim and A. V. Oppenheim, *Advanced Topics in Signal Processing* (Prentice-Hall, New York, 1988).

[32] "Information Technology—Multimedia Content Description Interface—Part 4: Audio," ISO/IEC JTC 1/SC 29/WG 11, International Organization for Standardization, Geneva, Switzerland (2001 June).

[33] D. Arfib, F. Keiler, and U. Zolzer, "Source-Filter Processing," in *DAFX: Digital Audio Effects,* U. Zolzer, Ed. (Willey, Chichester, UK, 2002), p. 363.

[34] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing,* 2nd ed. (Prentice-Hall International, Upper Saddle River, NJ, 1999), p. 468.

[35] S. J. Orfanidis, *Introduction to Signal Processing* (Prentice-Hall International, Englewood Cliffs, NJ, 1996), p. 731.

THE AUTHORS

A. Czyzewski

A. Ciarkowski

A. Kaczmarek


J. Kotus


M. Kulesza


P. Maziewski

Andrzej Czyzewski was born in Gdansk, Poland, in 1956. He received an M.Sc. degree in sound engineering from the Gdansk University of Technology in 1982, a Ph.D. degree in 1987, and a D.Sc. degree in 1992.

He was granted the title of professor in 1999 and he has held the position of full professor at Gdansk University since 2002. His main interest is in research and education in the domain of multimedia and studio recording as well as in modern information and communication technologies.

Dr. Czyzewski is a fellow of the AES.

Andrzej Ciarkowski was born in Gdansk, Poland, in 1979. He received an M.Sc. degree in sound engineering from the Gdansk University of Technology in 2003. His research interests are mainly in VoIP telephony and multimedia streaming applications. Currently he is a Ph.D. candidate in the Multimedia Systems Department of the Gdansk University of Technology.

Andrzej Kaczmarek received an M.Sc. degree in automatics and computer science from the Technical University of Gdansk in 1972. He joined the staff of the Sound Engineering Department in 1985. His research interest is focused on computer analysis of musical sounds. He received a Ph.D. degree in 1995 in this subject. His interests are also focused on applications of audio engineering to speech recognition and to sound restoration. In his research he uses intelligent computation methods such as neural networks, fuzzy logic, and rough sets.

Dr. Kaczmarek has presented more than 30 scientific papers in journals and at conferences.

Jozef Kotus was born in Gdansk, Poland, in 1976. He received an M.Sc. degree from the Faculty of Electronics, Telecommunications and Informatics of the Gdansk University of Technology in 2001. His research interest is principally in multimedia techniques, especially in audio recording, noise control, multimedia programming, and digital signal processing in real time. Since 2001 he has been a Ph.D. candidate in the Multimedia Systems Department.

Maciej Kulesza was born in Gdansk, Poland, in 1979. He received an M.Sc. degree in sound engineering from the Gdansk University of Technology in 2003. He is interested in digital audio processing techniques and DSP-based electronic systems engineering. Currently he is a Ph.D. candidate in the Multimedia Systems Department. His principal subject is related to introducing psychoacoustic modeling of hearing to speech coding.

Przemyslaw Maziewski was born in Suwalki, Poland, in 1978. He received an M.Sc. degree from the Faculty of Electronics, Telecommunications and Informatics of the Gdansk University of Technology, Poland. His thesis was related to the problem of sound virtualization techniques implemented in stereo headphones. His fields of interests are related to audio recording, digital signal processing, and sound archiving. Currently he continues his studies at the Gdansk University of Technology as a Ph.D. candidate.