Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

Accidental Wow Evaluation Based on Sinusoidal Modeling and Neural Nets Prediction

Przemyslaw Maziewski, Lukasz Litwic and Andrzej Czyzewski

Gdansk University of Technology, Multimedia Systems Department

Narutowicza 11/12, 80-952 Gdansk, Poland
{przemas, llitwic, andcz}@sound.eti.pg.gda.pl

ABSTRACT

In this paper an algorithmic approach to the wow defect characteristic evaluation is presented. The approach is based on a sinusoidal analysis comprising both amplitude and phase spectra processing. The frequency trajectories depicting the distortion are built on a basis of amplitude, frequency and phase dependencies and are further used for wow characteristic evaluation. Additionally the experiments concerning the neural-network-based prediction applied to the characteristic are performed. The obtained results are compared to linear-prediction.

1. INTRODUCTION

Wow defect is a distortion defined as parasitic frequency modulation and is perceived as pitch fluctuations of audio program. It is introduced into audio signal by motor speed fluctuations, tape damages and inappropriate editing techniques [1].

The perceived pitch variation comes from the fact that wow modulates the spectral components of audio signal. Thus the estimation of parasitic modulation waveform can be performed by the analysis of spectral components variations. It was found that both genuine and artifact components can be utilized in wow defect determination process. The first approach concerns the analysis of tonal components. The latter, concerns the analysis of high-frequency-bias [2-3] and hum [3]. The analysis of artifact components is alluring since their variations are not correlated with genuine pitch variations. Thus the wow estimation can be performed over long passages independently from the melodic line. However, the analysis employing artifact components may be often merely feasible. The high-frequency bias can be found in digitized archives only when special precautions are performed during digitization. The hum component can originate from various devices in recording/reproducing chain, thus its variations may not correspond to parasitic frequency modulation. For these reasons, there is a strong motivation for a wow determination algorithm based on tonal components analysis. Such an approach was proposed by Godsill [4-5]employing the statistical analysis of tonal components in a distorted sound. Tonal components are determined by means of sinusoidal modeling (McAulay&Quatrieri analysis)[6]. The statistical model, proposed by Godsill, favors the smooth periodic variations which are common to most components assuming these variations originate in wow defect. Other variations, non-smooth or not common to all tones, are assumed to originate from noise or genuine variations and are rejected.

We also studied these problems and the results of our findings are presented in this paper which is divided into two parts. The first part presents the algorithm for wow determination based on sinusoidal modeling. The algorithm is dedicated to accidental wow defect determination, which is characterized by a short duration and a strong modulation. The short duration of the defect allows analyzing it independently of genuine changes of the audio program.

The second part concerns the possible employment of a neural-network-based prediction for enhancing the partial tracking process. The neural network (NN) prediction is also compared to linear prediction (LP) presented in other papers.

2. SINUSOIDAL MODELING

2.1. Basic Sinusoidal Modeling Algorithm

Sinusoidal modeling, which is based on Fourier Theorem, expresses the audio signal as a sum of sinusoidal components having slowly-varying frequencies and amplitudes. This relation can be shown:

$$x(t) = \sum_{p=1}^{P} a_p(t)\cos(\phi_p(t)) \quad (1)$$

$$\phi_p(t) = \phi_p(0) + 2\pi\int_0^t f_p(u)du \quad (2)$$

where P corresponds to a number of sinusoidal components (partials) in analyzed sound. The vectors f_p, a_p, ϕ_p contain the partial's frequency, amplitude and phase values respectively. The successive values of f_p create the frequency track.

The consecutive values of f_p, a_p and ϕ_p are determined in a frame-based algorithm. The block-diagram showing the operation blocks of the algorithm is presented in Fig. 1.

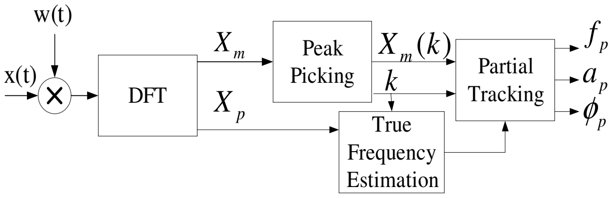

Figure 1: Block-diagram of sinusoidal modeling.

An input signal $x(t)$ is divided into analysis (time) frames using windowing. Hamming window is often used in sinusoidal modeling since it ensures a good main-lobe to side-lobe rejection ratio. The zerophase windowing is performed to remove the linear trend from a phase spectrum [7]. The windowed frame is also zeropadded to gain a frequency resolution. Next a complex spectrum X is obtained via DFT. Magnitude and phase spectra which are used in further processing are obtained according to the following formulas:

$$X_m = 20\log_{10}|X| \quad [dB] \quad (3)$$

$$X_p = \tan^{-1}\frac{\Im\{X\}}{\Re\{X\}} \quad [rad] \quad (4)$$

The candidates for tonal components in every time-frame are determined as meaningful peaks of a magnitude spectrum according to the following expression:

$$X_m(k \pm 1) < X_m(k) \quad (5)$$

where k denotes a spectrum bin.

The presented formula indicates all local maxima in the magnitude spectrum. In order to reject the peaks resulting from side-lobe components or localized noise, the peaks are validated. Different algorithms for tonal component validation have been introduced such as in

MPEG-1or SLM (Sinusoidal Likeness Measure) [8]. This group of tonality validation algorithms is frame based and adjacent frame information is not considered. There is also another approach which assumes that the non-tonal components are rejected during partial tracking stage.

The resolution of time-frequency analysis is limited according to Gabor-Heisenberg Uncertainty Principle [9]. It results in biased estimation of spectral components. The determination of *true* (or *instantaneous*) frequencies is applied to overcome the limited resolution. The true frequencies can be evaluated by means of several estimators, from which the most popular is *parabolic estimation.* A survey on true frequencies estimation can be found in literature [10]. This step is significant in wow determination algorithm, since finite frequency resolution can mask the wow variations. An adequate example is shown in Fig. 2.

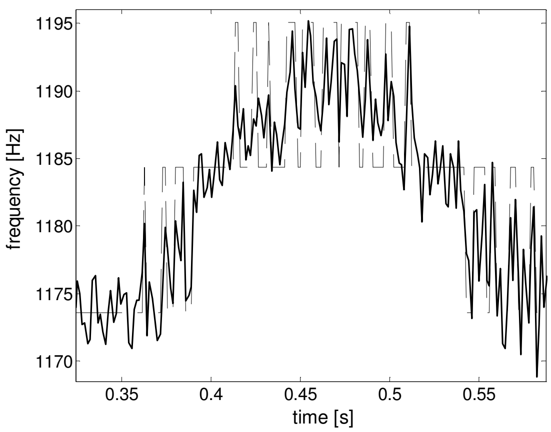

Figure 2: The estimation of true frequency of the component selected for parasitic frequency modulation estimation.

The last step of sinusoidal modeling is partial tracking. The determined tonal components either are linked to the existing tracks or discarded. The decision is made upon a matching criterion. The most popular criterion was proposed by McAulay and Quatieri in [6] employing the frequency values in a frame-to-frame processing:

$$\left| f_k^{i-1} - f_l^{i} \right| < \Delta_f \qquad (6)$$

where f_k^{i-1} is a frequency of the processed track in frame $i-1$ and f_l^{i} is a frequency of a candidate peak in the frame i. Frequency deviation parameter Δ_f is a maximum frequency distance between the track and its continuation. The component which is the closest to the processed track in terms of frequency and which frequency distance is smaller than Δ_f, is selected for the track continuation. Other components which are not matched to existing tracks create new tracks. Tracking of partials, for which the matching criterion (Eq. 6) is not fulfilled, is terminated.

2.2. Partial Tracking of Polyphonic Signals

Sinusoidal modeling was initially employed in additive synthesis systems, but was also found useful in vast number of applications including audio restoration [11][12]. However, sounds from archival recordings merely fit to the expression (1) considering that sounds are likely to be contaminated by several distortion e.g., noise, clicks, which are represented by non-sinusoidal components. Moreover, the wow distortion itself deforms the tonal structure of a contaminated sound. Figure 3 shows tonal component extracted from a clarinet sound while Fig. 4 shows extracted tonal components from a wow-distorted signal.

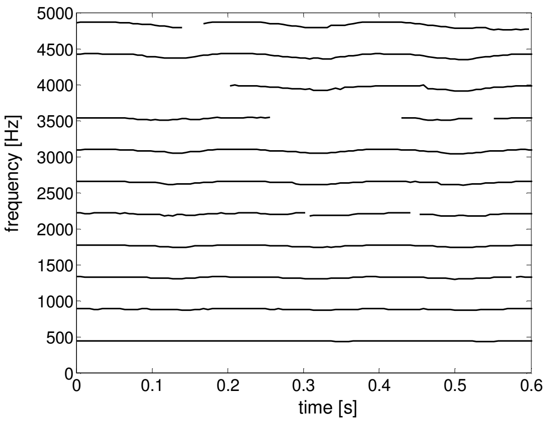

Figure 3: Extracted tonal component from a clarinet sound.

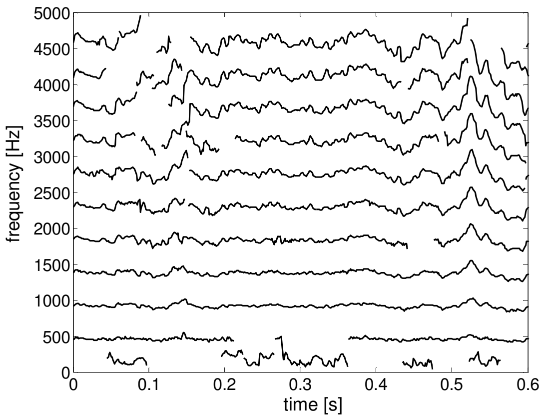

Figure 4: Extracted tonal components from a wow distorted signal (sound from Polish archives). Noticeable distortions of tonal components correspond to accidental wow distortion.

The distortions of tonal components in wow contaminated signals force to use more robust matching criterion than (6). The frequency matching criterion can be enhanced with a magnitude matching criterion:

$$\left|a_k^{i-1} - a_l^i\right| < \Delta_a \quad (7)$$

where a_k^{i-1} is a magnitude of a processed track, a_k^i is a magnitude of a candidate peak in frame i and Δ_a is the maximum magnitude deviation of a track in two adjacent frames. Matching criterion based on Eq. 6 and Eq. 7 expresses the assumption that both frequency and magnitude changes over time should be smooth ones.

It is important to notice that both Eq. 6 and Eq. 7 operate on magnitude spectrum. However, phase spectrum may be also employed for trajectory matching process. For analyzed partial having the phase value ϕ_k^{i-1} in the frame $i-1$, the predicted phase value in frame i is equal to:

$$\hat{\phi}_k^i = \phi_k^{i-1} + \frac{R}{N_{DFT}} \quad (8)$$

where N_{DFT} is the length of zeropadded DFT analysis and R is the frame hop distance. The phase prediction error can be evaluated for every candidate component:

$$\phi_{err} = \left|\hat{\phi}_k^i - \phi_k^i\right| \quad (9)$$

where ϕ_k^i is the actual phase value of a candidate.

The values of phase prediction error ϕ_{err} are in the range $[0;\pi)$. The value near 0 suggests phase coherence in two adjacent frames. Otherwise, if the value is near π, it may indicate phase incoherence.

2.3. Partial tracking based on trajectories exploration

The different approach to partial tracking was proposed by Depalle, Garcia and Rodet in their work [13]. This approach utilized more than two frames information to estimate the trajectory. The algorithm employed the analysis of peaks frequencies and magnitudes in order to indicate the smoothest sequence of peaks in consecutive frames by means of Hidden Markov Models.

The latest approach in this field was proposed by Lagrange, Marchand and Rault and reported in the literature [14][15]. Authors used a linear prediction is used to forecast the successive frequency and amplitude values of analyzed components. Reported results show that the order of LP analysis and number of coefficients must be in a range which conforms to the signal properties e.g. vibrato, portamento. The Burg method is used to minimize the prediction error. The matching criterion favors the peaks for which the prediction error is the smallest. The method is reported to enhance the partial tracking and was used for interpolation of gaps in audio signal. However, it must be noticed that the core assumption for LP employment is that the components variations are sinusoidal in case of vibrato or tremolo or exponentially increasing or decreasing in case of portamento.

The partial tracking enhancement by means of prediction appears to be encouraging idea nevertheless the signal properties must be taken into considerations. However in case of accidental wow a neural-network-based prediction can probably be more effective than the LP-based one.

3. ARTIFICIAL NEURAL NETWORKS

Artificial neural networks have been used in a variety of DSP applications involving classification tasks, pattern recognition as well as decision making and others (see [16] for an overview). However, time series forecasting is by far one of the main neural nets' application area. Although pioneering prediction applications date back to 1960s, it was the introduction of the backpropagation training algorithm in 1980s that made the turning point in the forecasting application development (see [17] for more details on the historical background). From that moment NNs were commonly used in different scientific researches and practical appliances [16 - 23].

There are many reasons why NNs provide an attractive alternative for well known prediction techniques such as the linear prediction (LP). First of all artificial NNs are nonparametric methods. Thus only few priori assumptions are needed about the processed data, whereas, in many traditional forecasting techniques detailed information about the data generation process is necessary to achieve acceptable results. Furthermore, NNs offer useful generalization abilities and can easily adopt to new data only by analyzing them. Hence NNs are called data driven methods. Additionally NNs can model linear [18] as well as nonlinear processes [19-21]. The later is especially important to this research as the frequency trajectories depicting wow defects are highly nonlinear and chaotic, which can be noticed e.g. in Fig 4.

Different NNs structures have been used in forecasting applications. Among many the feedforward (FF) networks and especially the multi-layer perceptrons (MLPs) seem to be the most popular ones [17 - 23]. A typical MLP structure is presented in Fig. 5.

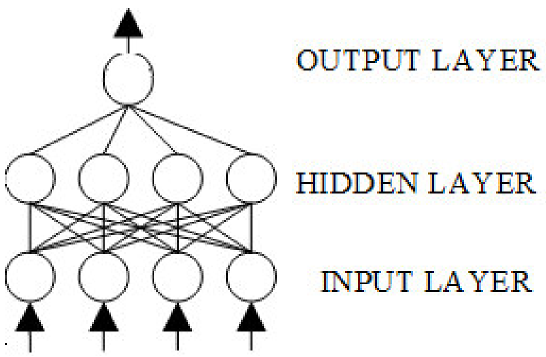

Figure 5: Typical multi-layer perceptron (MLP).

The key factors defining the MLP structure are the numbers of: inputs, outputs and hidden layers. Furthermore the nodes interconnection and neurons' activation functions play a important role in achieving the desired forecasting abilities. The latter determinant is crucial when MLP is used for modeling the nonlinear processes. Other important factors are: the training procedure as well as input data normalization and splitting. Since the MLP was used in the performed experiments, thus detailed reasoning behind the selected structure settings is given in Section 3.3.

4. THE ALGORITHM FOR WOW EVALUATION

4.1. The core algorithm

The algorithm for wow determination takes as an input a contaminated signal $x(t)$ and outputs the Pitch Variation Curve (PVC), which controls the non-uniform resampler during the reconstruction process [3].

The core algorithm for wow determination comprises two main parts. The first is a sinusoidal modeling block which outputs the frequency tracks. The sinusoidal modeling stage is applied according to the algorithm described in Section 2.1. The partial tracking is developed since it utilizes joint magnitude and phase spectra information, which is expressed in the matching criterions presented in Eq. 6, Eq. 7 and Eq. 9. The second is a PVC estimation block which computes the PVC from determined trajectories. Figure 6 presents the basic idea of the wow determination algorithm.

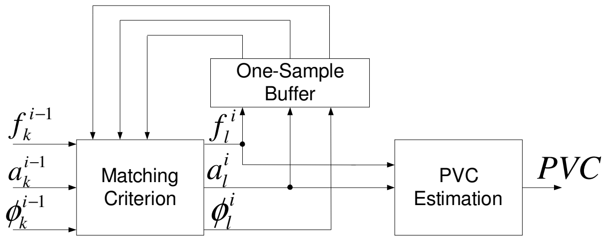

Figure 6: Wow determination algorithm.

The matching criterion block showed in Fig. 6 consists of two steps. Firstly, the candidate for continuation for a given partial is determined using Eq. 6. Next, Eq. 7 and Eq. 9 are applied to validate the candidate. If the positive validation is not achieved the track terminated Otherwise the continuation is acknowledged. Figures 7 and 8 show the track validation according to phase prediction error (Eq. 9).

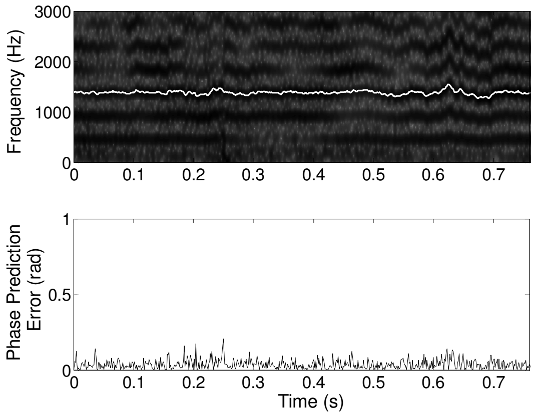

Figure 7: The valid frequency track and corresponding values of phase prediction error

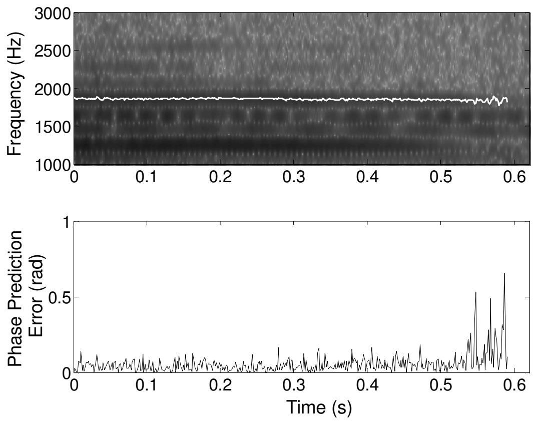

Figure 8: Terminated frequency track due to joint frequency-magnitude matching criterion. Notice high values of phase prediction error for last track values

The obtained vectors $F=[f_1 \quad f_2 \quad f_3 ... \quad f_p]$ and $A=[a_1 \quad a_2 \quad a_3 ... \quad a_p]$ are processed to obtain the PVC. It is assumed that the parasitic modulation is most reliably depicted by the most significant components. Thus for every estimated frequency track a confidence value parameter is computed:

$$CV(k)=L_k \cdot mean(a_k) \qquad (10)$$

where L_k is the length of kth track and a_k is the magnitude level of this track.

Next, the maximum values of CV are determined in every time-frame to indicate the most significant components. The selected tracks are then normalized according to the following formula:

$$RF_k(i)=\frac{f_k^i}{f_N} \qquad (11)$$

where f_N is the normalization value. PVC is assembled from the selected tracks after normalization. The normalization value f_N is set to ensure the value of 1 in the beginning frame and smooth changes of PVC between the parts originating in different frequency tracks.

4.2. Neural-Network-based predictor

This paragraph presents the proposed architecture of a Neural-Network based predictor. The aim of the predictor is to enhance the partial tracking process. Fig. 9 presents the possible employment of such predictor.

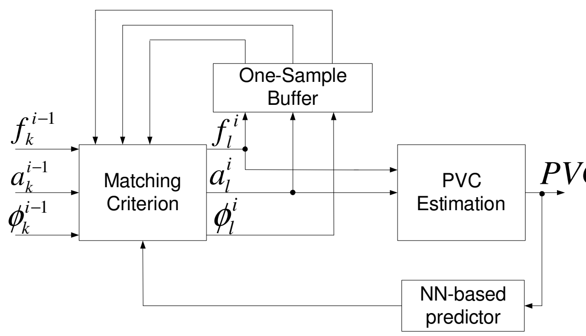

Figure 9: Wow determination algorithm enhanced by NN-based predictor.

The FF MLPs are one of the most popular NN structures used for time series prediction (for examples see [17 - 23]). Typical MLP diagram is given in Fig. 4. Even this simple presentation indicates that several factors can compose the overall MLP structure. Firstly the number of layers and corresponding number of layer's nodes must be set correctly. Nodes in neighboring layers must be properly connected and appropriate functions must be used for activating neurons in different layers. Also other issues such as: choosing and tuning the right learning algorithm as well as prepossessing the input data are important, because they can influence obtain forecasting abilities. Following paragraphs give detailed reasoning behind the chosen values of the aforementioned MLP parameters.

Several studies give some practical guidelines for preparing the input data and choosing the proper NN architecture. An comprehensive overview on setting the MLPs parameters is given by Zhang [25]. However, commonly-accepted standard procedures do not exist.

First of all the input data must be correctly prepared. The size of the input set is important and long time series are preferable. Thus, in the performed experiments long PVC with more then 800 samples was used. Further, the data must be spitted into two separate subsets. The in-sample set is used for NN training, whereas, the out-of-sample set is employed in the prediction abilities evaluation. Most common splitting ratios are 70%:30%, 80%:20%, and 90%:10% [25]. In the following experiments 20% of the PVC samples constituted the evaluation subset. Data preprocessing is another recommendation made by majority of authors. Azoff gives an overview on the most common normalization procedures [26]. In the performed experiments external normalization was applied. The data were rescaled to the range from –0.9 to 0.9 according to the formula given in Eq. 12:

$$x_n = 1.8(x_n - x_{\min})/(x_{\max} - x_{\min}) - 0.9 \qquad (12\ )$$

where :
x_n – the input data sample;
$x_{\min}$ – the minimal value in the input data set;
$x_{\max}$ – the maximal value in the input data set;

After the linear normalization data were statistically processed on the basis of Eq. 13:

$$x_n = (x_n - \overline{x}) / s \qquad (13\ )$$

where :
x_n –the input data sample;
$\overline{x}$ – the mean value of the input data set;
s – the standard deviation of the input data set;

Following the normalization the data were smoothed using a 3-rd order moving average zero-phase filter.

Choosing the MLP architecture is another necessity. In the reported experiments only one-sample-ahead forecasting was investigated. Hence, only one node was selected for the output layer. Another important issue is the hidden nodes number. The one-hidden-layer perceptrons are most commonly used MLPs in the forecasting applications [25]. It is probably due to some theoretical findings showing that a single hidden layer is sufficient to approximate any nonlinear function [27-28]. Thus, structures with one hidden layer were chosen for the presented experiments. Other meaningful issues are the numbers of input and hidden nodes. In the investigated MLPs the input layer sizes were set to: 2, 4, 8, 16 and 32 samples. This allowed for results comparison with the examined LP method. Considering the hidden nodes' number, results reported by Zhang, Patuwo and Hu [19] indicate that in case of the MLP-based one-sample-ahead forecasting greater attention should be put on the input's number. Therefore, in the experiments the number of hidden-layer nodes was equal to the inputs number. Thus, all of the examined networks had similar architectures, i.e., U-like structure. The neuron-activation functions were chosen as follows: the logistic functions were used for hidden nodes and the linear function for the output node. This popular convention, represented by many authors [17], allows for nonlinear data processing.

Among the most popular training algorithms are basic backpropagation procedures as well as more sophisticated second-order approaches [25] [29]. In the experiments both the standard backpropagation and the Levenberg-Marquardt procedure were investigated.

Various error indicators can be used for the forecasting performance evaluation (see [17] for an overview). There is no common agreement which indicators are appropriate in most cases, nevertheless some recommendations can be found in the literature [19] [30]. Since in the performed experiments the forecasting methods were examined using the same input data, the mean square error (MSE, defined by Eq. 14) was computed. It allowed for direct comparison of algorithms' results:

$$\text{MSE} = \frac{\sum (e_t)^2}{N} \qquad (14\ )$$

where :
e_t – the individual forecast error;
N – number of error terms;

Furthermore, the median absolute percentage error (MdAPE, defined by Eq. 15) allowing for assessing the relative forecasting performance, was also computed:

$$\mathrm{MdAPE} = median(|e_t|) \tag{15}$$

where :
e_t – the individual forecast error;

5. EXPERIMENTS AND RESULTS

5.1. Wow characteristic evaluation

The experiments were conducted on archival sounds obtained in Polish archives. The spectrogram of a distorted sound is shown in Fig. 10. The presented results were obtained using the algorithm presented in Section 4.1. The length of DFT was 512 samples and was zeropadded with the factor of 8. Hop distance was set to 64 samples.

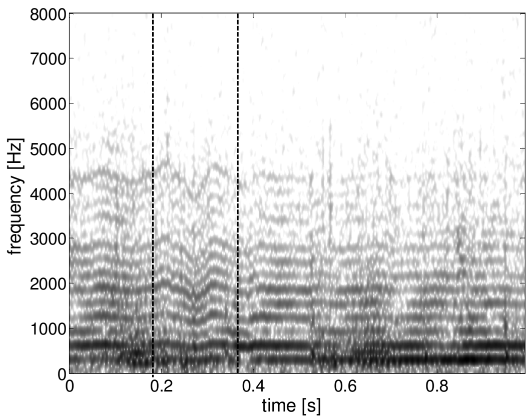

Figure 10: The spectrogram of a processed sound. The accidental wow distortion can be noticed between 0.2-0.3 s.

The evaluated frequency tracks are shown in Fig. 11. The track depicted by a solid line was selected for PVC computation due to the highest confidence value.

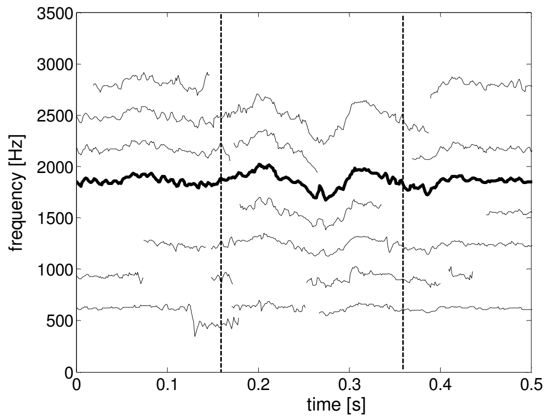

Figure 11: Evaluated frequency tracks. The solid line indicates the track used for PVC evaluation.

The evaluated PVC was scaled and is shown on the spectrogram in Fig. 12.

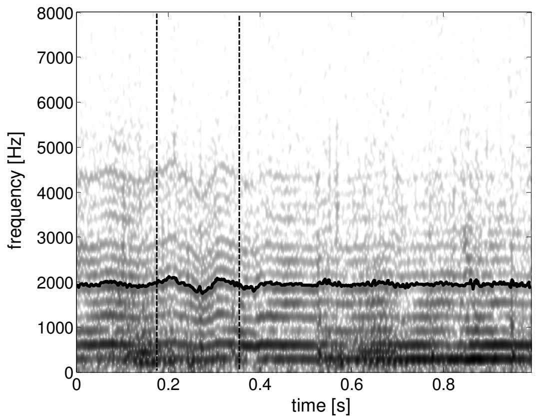

Figure 12 Evaluated PVC scaled and presented on spectrogram of the analyzed sound.

The restoration of sound was performed by means of non-uniform resampling, described in literature []. The evaluated PVC was used for restoration. The spectrogram of restored sound is shown in Fig. 13.

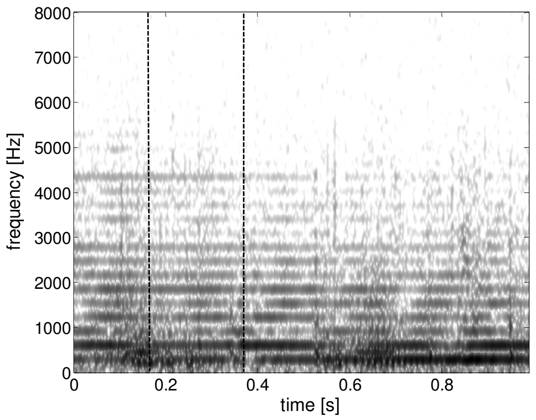

Figure 13: The restored sound on basis of evaluated PVC.

The result of restoration shows that the presented algorithm can effectively determine the accidental wow characteristic.

5.2. Prediction experiments

Two prediction methods were studied. The first involved different settings of the linear prediction algorithm. The second used various neural-network structures. In both cases the PVC from the described wow evaluation algorithm (see previous paragraph) was used. The characteristic, however, was preprocessed according to Eqs. 12 and 13. The resulting PVC is given in Fig. 14.

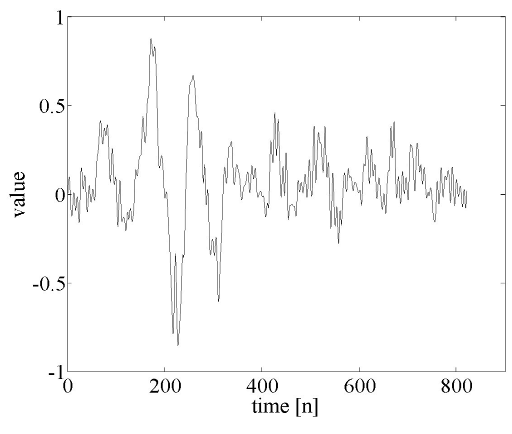

Figure 14: Preprocessed PVC.

5.2.1. LP-based prediction

The LP experiment was performed according to the following outline: firstly the input data was divided into small subsets, then the prediction filters were build and the forecasting performance was evaluated. A sliding-time-window technique was used to divide the input data into subsets. In each subset a one-sample-prediction filter was computed using the autocorrelation method of autoregressive modeling (AR). The prediction procedure can be called adaptive, because the adaptation filter coefficients were calculated in each time window.

The experiment was repeated with different combination of algorithm's settings in order to examine LP parameters influence on forecasting results. The following variables were chosen: the LP order which defines the number of prediction filter coefficients, and the LP length representing the number of past samples used for calculating the coefficients. The obtained results are given in Table 1.

LP length	LP order				Error Measure
	2	4	8	16	
4	0,003947				MSE
	0,557844				MdAPE
8	0,002605	0,002817			MSE
	0,448259	0,426178			MdAPE
16	0,001830	0,001824	0,002058		MSE
	0,339048	0,336988	0,320679		MdAPE
32	0,001604	0,001463	0,001599	0,001660	MSE
	0,320402	0,227850	0,258056	0,279518	MdAPE

Table 1 LP-based prediction errors.

The smallest prediction error was achieved using the prediction filter with 4 coefficients calculated on 32 past samples.

The results indicate also that the main factor which affects the algorithm's performance is the LP length. The prediction error decreases with greater lengths, which is mostly noticeable in the table's second column. The smallest prediction errors were obtained for the longest LP. Increasing the prediction filter order also influences the error level, however, its influence is smaller and less obvious as higher prediction orders can introduce greater errors. This can be noticed in the table's last column where the prediction error decreases at first and then rises up again. The probable reason behind the error variation is the chaotic and accidental nature of the utilized PVC (see Fig. 14), whereas, the LP tries to model it as a linear process.

5.2.2. NN-based prediction

The multi-layer preceptors (MLP) together with the sliding time-window technique were used in the forecasting experiment. The MLPs input layer sizes were set to 2, 4, 8, 16 and 32. The number of hidden-layer nodes was equal to inputs number. The logistic functions were used for hidden nodes and linear function for the output node.

The analysis was carried out according to the following outline. Firstly the input data were divided into training (80%) and validating sets (20%). Subsequently, the networks learning procedures were applied and followed by the evaluation stage. The Levenberg-Marquardt was used as the training algorithm, because it converged faster on the processed data then the standard gradient descent with momentum and adaptive learning rate back-propagation. Both the goal MSE value and the minimum performance gradient were set to 0.001. The maximal number of epochs was set to 1000.

The experiments on each MLP structure were repeated 200 times with randomly initiated weights and biases. However, during each evaluation cycle the NN was constantly updating its weights allowing for adaptation. The mean MSE and the median MdAPE from all the repetition were taken as the prediction performance indicators. The obtained error measurement results are given in Table 2.

Error Measure	NN structures				
	2-2-1	4-4-1	8-8-1	16-16-1	32-32-1
MSE	0,000980	0,000511	0,000448	0,000387	0,000442
MdAPE	0,301662	0,184748	0,170394	0,161080	0,163078

Table 2 NN-based prediction errors.

The results show that among the examined NNs the most suitable is build of both 16 input and 16 hidden nodes. The findings indicate also that the input size plays important role in the prediction performance. Increasing the number of nodes can lead to better forecasting. It can be noticed also in decreasing error values for NN structures from 2-2-1 to 16-16-1. However, after reaching a certain point the error level goes up again as in the 32-32-1 structure. This is probably caused by the under-fitting in the learning stage as the last NN structure is quite large. However, the most evident and important observation, is that the NN-based forecasting outperforms the linear prediction (compare Tables 1 and 2). The MSEs from all the NN structures are smaller then the smallest LP error and only the simplest NNs have a greater MdAPE error then longer LPs.

6. CONCLUSIONS

The sinusoidal analysis algorithm for wow defect evaluation was presented in the paper. The algorithm uses both magnitude and phase spectra for precise evaluation of frequency trajectories. The trajectories are utilized in PVC computation. Performed experiments showed that wow characteristic determination based on confidence value analysis leads to satisfactory results. Long tracks parts were employed besides single values for PVC estimation. Additionally the tracking procedure can be enhanced using some predictive methods. The results of experiments performed show also that neuralnetworks can be utilized in PVC forecasting, effectively. The reported results also showed that NN-based prediction outperforms the previously proposed linear prediction methods.

7. ACKNOWLEDGEMENTS

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

8. REFERENCES

[1] A. Czyzewski, P. Maziewski, M. Dziubinski, A. Kaczmarek and B. Kostek, "Wow detection and compensation employing spectral processing of audio," presented at 117th AES Convention San Francisco 2004 USA, October 28-31.

[2] J. Horwarth and P. Wolfe, "Correction of Wow and Flutter Effects in Analog Tape Transfers," presented at the 117th AES Convention San Francisco USA October 28-31.

[3] 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 118th AES Convention Barcelona 2005 Spain, May 28-31.

[4] J. S. Godsill and J. W. Rayner, “The restoration of pitch variation defects in gramophone recordings,” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (1993 October).

[5] J. S. Godsill, “Recursive restoration of pitch variation defects in musical recordings,”. in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 233-236 (1994, April).

[6] R. J. McAulay, and T. F. Quatieri, “Speech Analysis/Synthesis Based on a Sinusoidal Representation,” in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744-754 (1986 August).

[7] X. Serra, “Musical Sound Modeling with Sinusoids plus Noise” in *Musical Signal Processing*, ed. S. Pope, A. Picalli, G. De Poli, and C. Roads, (Swets & Zeitlinger Publishers, 1997).

[8] M. Lagrange, S. Marchand and J. B. Rault, “Sinusoidal parameter extraction and component selection in a non-stationary model,” in. *Proceedings of the 5th Int. Conference on Digital Audio Effects, Hamburg* (2002 September).

[9] F. Auger and P. Flandrin, “Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method” in *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068-1089 (1995 May).

[10] F. Keiler and S. Marchand, “Survey on Extraction of Sinusoids in Stationary Sounds,” in *Proceedings of the 5th International Conference on Digital Audio Effects, Hamburg* (2002 September).

[11] R. C. Maher, “A Method for Extrapolation of Missing Audio Data”, in *J. Audio Eng. Soc*., vol. 42, no.5, pp. 350-357 (1994).

[12] A. Czyzewski, ”Learning Algorithms for Audio Signal Enhancement, Part 2: Implementation of the Rough-Set Method for the Removal of Hiss,” in *J. Audio Eng. Soc*., vol. 45, no. 11, pp. 931-943 (1997 November).

[13] P. Depalle, G. Garcia and X. Rodet, “Analysis of Sound for Additive Synthesis: Tracking of Partials Using Hidden Markov Models,” in *Proceedings of the IEEE International Conference on* Acoustics, *Speech, and Signal Processing*, vol. 1, pp. 225 - 228 (1993 April).

[14] M. Lagrange, S. Marchand and J. B. Rault, ”Tracking Partials for Sinusoidal Modeling of Polyphonic Sounds,” in *Proceedings of IEEE International Conference on Speech and Signal Processing (ICASSP'05)* (2004 March).

[15] M. Lagrange and S. Marchand, ”Long Interpolation of Audio Signals Using Linear Prediction in Sinusoidal Modeling,” in *J. Audio Eng. Soc*., vol. 53, no. 10, pp. 891-905 (2005).

[16] B. Widrow, D. Rumelhart and M. A. Lehr, “Neural networks: Applications in industry, business and science,” in *Communications of the ACM*, vol. 37, no. 3, pp. 93-105 (1994).

[17] Guoqiang P. Zhang, Eddy B. Patuwo and Michael Hu, “Forecasting with artificial neural networks: The State of the art”, in *International Journal of Forecasting,* vol 14, pp. 35-62 (1998 March).

[18] Guoqiang P. Zhang, “An investigation of neural networks for linear time-series forecasting,” in *Computers & Operations Research*, vol. 28, no. 12, pp. 1183-1202 (2001 October).

[19] Guoqiang P. Zhang, Eddy B. Patuwo and Michael Hu, “A Simulation study of artificial neural networks for nonlinear time-series forecasting”, in *Computers & Operations Research*, vol. 28, pp. 381-396 (2001 April).

[20] Yoshio Kajitani, Keith W. Hilpel and A. Ian McLeod “Forecasting Nonlinear Time Series with Feed-Forward Neural Networks: A Case Study of Canadian Lynx Data,” in *Journal of Forecasting*, vol. 24, pp. 105–117 (2005)

[21] Stefania Troncia, Massimiliano Gionab and Roberto Barattia, "Reconstruction of chaotic time series by neural models: a case study," in *Neurocomputing*, vol. 55, pp. 581 – 591 (2003).

[22] R. J. Frank, N. Davey and S. P. Hunt, "Time series prediction and neural networks," in *Journal of Intelligent Robotic Systems*, vol. 31, no. 1-3, pp. 91-103 (2001 May).

[23] Yuansong Liao, John Moody and Lizhong Wu, "Applications of Artificial Neural Networks to Time Series Prediction," in *Handbook of Neural Network Signal Processing*, ed. Yu Hen Hu, Jenq-Neng Hwang (London: CRC Press, 2002).

[24] Andrzej Czyzewski, "Learning Algorithms for Audio Signal Enhancement Part 1: Neural Network Implementation for the Removal of Impulse Distortions," in *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 815-831 (1997 October).

[25] Guoqiang P. Zhang, "Business Forecasting with Artificial Neural Networks Neural Networks: An Overview" in *Business Forecasting*, ed. Guoqiang P. Zhang, pp. 1-22 (London: Idea Group Publishing, 2004).

[26] E. M. Azoff, *Neural Network Time Series Forecasting of Financial*, (Chichester: John Wiley & Sons, 1994).

[27] G. Cybenko, "Approximation by superpositions of a sigmoidal function," in *Mathematical Control Signals Systems*, vol. 2, pp. 303–314 (1989).

[28] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators." in *Neural Networks*, vol. 2, pp. 359–366 (1989).

[29] M. Bishop, *Neural Networks for Pattern Recognition*, (Oxford: Oxford University Press, 1995).

[30] E. S. Gardner and D. G. Dannenbring, "Forecasting with exponential smoothing: some guidelines for model selection," in *Decision Sciences*, vol. 11, pp. 370-383 (1980).