# Evaluation of wow defects based on tonal components detection and tracking

## Wyznaczenie przebiegu zniekształceń drżenia na podstawie detekcji i śledzenia zmian tonalnych składowych dźwięku

**Łukasz Litwic**
**Przemysław Maziewski**

**Multimedia Systems Department**
**Gdańsk University of Technology**

### SUMMARY

This paper presents a method for wow defect evaluation. The method is based on tonal components detection in STFT frames. Extracted harmonic elements are joined together to form tracks. The pitch variation curve *pvc* is created from the selected tracks. *Pvc* course is considered to depict the wow defect, therefore it can be used for parasite modulation removal.

## 1. INTRODUCTION

The wow and flutter effects (wow) are very common among analogue audio materials. They could be found in old gramophone recordings, wax cylinders and also on the magnetic and optical sound tapes. The origin of those parasite modulations can be found in audio equipment mechanical parts excentricity or ellipticity. Motor speed fluctuations, tape damages and inappropriate production techniques (e.g. cutting and joining tracks) can also trigger mentioned defects. As wow leads to undesirable changes of all of the sound frequency components, sinusoidal sound analysis originally proposed by McAulay and Quatieri [1] was found to be very useful in the defects evaluation. In such approach tracks depicting tonal components changes are processed to obtain precise wow characteristics [2-5]. Different method employing graphical processing of spectrogram was presented by Nichols [6]. Also medium features (e.g. bias) were utilized in wow correction [7]. Notwithstanding all of the cited proposals there is still a need for further algorithmic approach to the wow restoration as it can be very complex sharing periodic or accidental nature. Therefore this paper addresses the problem of wow extraction.

Wow can be characterised by means of the time warping function $f_w(t)$ or equivalently by the pitch variation function $p_w(t)$. First expression describes wow as a distortion of the time domain of the original signal *x(t)*. As the sound carrier playback velocity differs from the reference the $f_w(t)$ function depicts time axis changes relatively to the original recording.

$$x_w(t)=x(f_w(t)) \tag{1}$$

where $x_w$ denotes distorted signal.

Second characteristic i.e. $p_w(t)$ depicts parasite frequency modulation caused by irregular playback. There is a simple connection between the two descriptors.

$$p_w(t)=\frac{d(f_w(t))}{dt} \tag{2}$$

Based on the foregoing characteristics wow restoration in digital time domain can be achieved by means of the incommensurate-ratio sampling rate conversion [8]. The nonuniform resampling routines were presented in details by Marvasti [9]. Additional details are also demonstrated by Laakso et all [10]. As this paper address only the problem of wow evaluation, reconstruction is beyond its scope.

## 2. PROPOSED ALGORITHM

The proposed algorithm for wow defect evaluation employs several processing methods to estimate the pitch variation function $p_w(n)$ from the contaminated input signal $x_w(n)$. The algorithm consists of four, frame-based processing steps presented in the Fig.1. The results of every stage form a data matrix for the input of the successive stage.

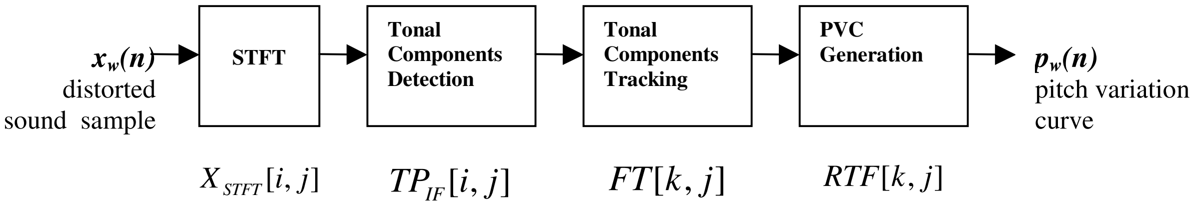

Fig.1 Block diagram of the wow evaluation process

The first stage of the algorithm, depicted as STFT in the Fig.1, results in a time-frequency representation of the input signal. The distorted input signal is initially divided on time-frames (analysis frames) with an appropriate overlapping. Sizes of the analysis frame and the overlapping are essential for time and frequency resolution. The framed signal is windowed with Hamming window for better side-lobe suppression. For the purposes of further processing two additional steps are performed. To gain frequency resolution the windowed signal is zero padded and then it is packed into the buffer for a zero phase spectrum [11]. Finally, Discrete Fourier Transform (DFT) is evaluated for every time-frame buffer to obtain the time-frequency representation $X_{STFT}[i,j]$.

Once the time-frequency representation is computed, the next stage of the algorithm is to detect tonal components of the signal. The detection process is performed in three consecutive steps presented in Fig 2.

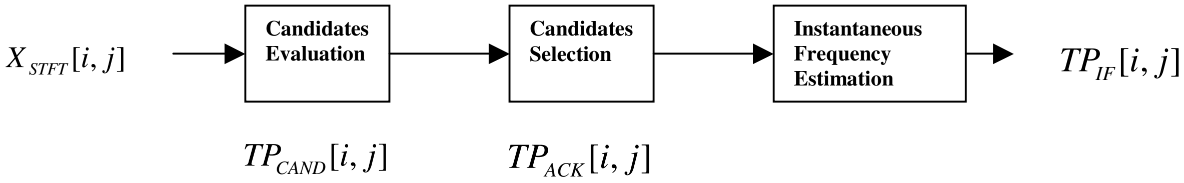

Fig.2 Structure of the Tonal Components Detection Block

In the first step candidates for tonal components $TP_{CAND}[i,j]$ are detected as local maxima (peaks) of a magnitude spectrum stored in $X_{STFT}[i,j]$. These peaks result either from the main-lobes or the side-lobes of the spectral components. Thus it is necessary to exclude the latter ones. As the proposed algorithm's approach is based on tonal components it also essential to reject the peaks which are valid spectral components but result from localised noise [14].

This task has to be performed according to an appropriate criterion. The most intuitive and frequently used one is an amplitude threshold. A candidate is recognised as a tonal peak when its amplitude is above the threshold. As this criterion suffers from many drawbacks and is insufficient when harmonic structure of the signal varies, other criterions for peak validation have been proposed [12-14]. Nevertheless, none is enough sufficient by oneself for distorted and noisy musical signals which are of interest of this paper.

We propose a heuristic algorithm, depicted as Candidates Selection block in Fig.2, containing three independent criterions. Those are as follow:

- Sinusoidal Likeness Measure (SLM) – the criterion assumes that the tonal peaks in analysis spectrum are the result of multiplication of the tonal components by the window function. Thus the maxima of cross-correlation function of the main-lobe's spectrum of the candidate and the analysis window would indicate the presence of a sinusoidal component [12].

$$\Gamma(\omega) = \left| \sum_{k,|\omega-\omega_k|<B} X(\omega_k) W(\omega-\omega_k) \right| \tag{3}$$

  $X()$ and $W()$ are spectra of analysed signal and window respectively, $B$ is the low-pass bandwidth within the cross-correlation is evaluated.

  The main drawback of this preliminary criterion is that it may acknowledge a side-lobe component as a tonal part.

- Phase measurement – this criterion assumes that the phase of the tonal component's main-lobe varies significantly less than those resulted from noise. To gain the measurement reliability as well as remove the linear trend in the phase spectrum the zero padding and buffering are introduced in STFT stage (as mentioned earlier). This criterion validates the result of the SLM criterion selection.
- Relative Amplitude Threshold [14] – this criterion can discard some side-lobes as well as components of minor significance.

$$h(k_p) = \left|X(k_p)\right|_{dB} - \frac{\left|X(k_{v+})\right|_{dB} + \left|X(k_{v-})\right|_{dB}}{2} \tag{4}$$

where, $X()$ depicts the DFT array, $k$ - the frequency bin index, subscript $p$ represents the peak, and subscripts $v+, v-$ represent the adjacent local minima.

Having determined the tonal components $TP_{ACK}[i,j]$ the next step is to estimate their true (instantaneous) frequency values as it is relevant because of a time-frequency resolution trade-off in the STFT representation. As mentioned earlier the zero padding was applied to gain frequency resolution but still this resolution is constant over the signal bandwidth. Moreover, for signals that are non-stationary within the analysis frame inter-frame modulations which effect in peaks smearing may occur [13].

To overcome all this, the spectral reassignment method, depicted as Instantaneous Frequency Estimation block (Fig.2), is employed. This method, originally proposed by Auger and Flandrin [15] as time-frequency reassignment, assigns the tonal component value to the STFT frequency bin's center of gravity. Eq. 5 presents the appropriate formula:

$$\hat{\omega}(x;,t,\omega) = \omega + \Im\left\{\frac{STFT_{dw}(x;,t,\omega)}{STFT_{w}(x;,t,\omega)}\right\} \tag{5}$$

where, $STFT_{w}()$ is STFT using analysis window ***w***, $STFT_{dw}()$ is STFT using the first derivative of a window function.

After the tonal components detection the tracking stage, in which the peaks are linked to create trajectories is launched. Fig. 3 presents the block diagram of this step.

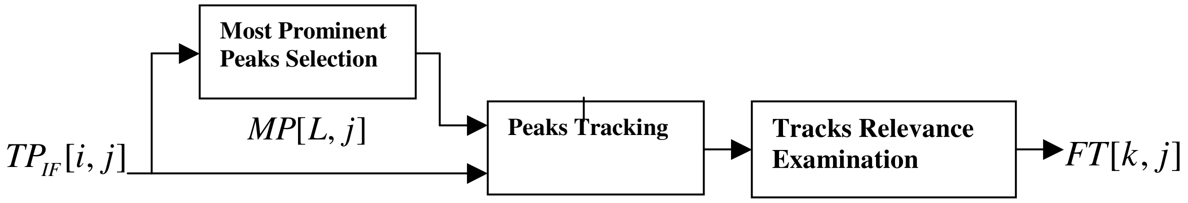

Fig.3 Structure of the Tonal Component Tracking Block

Since the pitch variation curve $p_w(n)$ is evaluated from the trajectories it is proposed to take into account only the relevant tracks, that can depict the wow defect well. Therefore it is reasonable to form only these trajectories which are based on the most prominent tonal peaks from the magnitude spectrum. This approach differs somewhat from the original MQ approach. Nevertheless, the idea for joining tonal components is the same as it applies the following frequency criterion:

The *K-th* tonal component $TP_{IF}[K,j]$ is joined to the *P-th* track $FT[P,j]$ when:

$$|TP_{IF}[K,j] - FT[P,j-1]| == \min(|TP_{IF}[:,j] - FT[P,j-1]|) \tag{6}$$

and

$$|TP_{IF}[K,j] - FT[P,j-1]| < f_{Dev}$$

where, $\min(\;)$ denotes minimal value and $f_{Dev}$ is the maximum frequency deviation.

The most prominent tonal components, stored in *MP* matrix, respond to the peaks of the greatest magnitude hence are considered to be the most perceptive ones. A new track is "born" only from the components stored in *MP* matrix that were not fitted to any existing tracks. A track is "dead" when there is no continuation according to the frequency criterion (Eq. 6). However before the track is labelled as "dead" a "zombie" state is introduced [11]. The "zombie" state can handle situation when a peak is erroneously rejected during the detection stage or when the criterion (Eq. 6) is not met due to a fast change in the harmonic structure (e.g. when accidental wow occurs). If there are some tonal components in the successive frames that can continue the track then the "zombie" peaks are interpolated.

Afterwards trajectories are examined for the relevance. The tracks which are too short or contain too many zombie states are discarded. Then the remaining tracks are sorted within the *FT* matrix accordingly to their relevance.

The last stage of the presented algorithm, according to the diagram in Fig.1, is the *PVC* generation stage in which the pitch variation function $p_w(n)$ is computed. Firstly the relative frequencies are calculated for all of the tracks stored in the *FT* matrix. Next, for each track its frequency values are divided by the preceding point values. Thereby *RFT* (Relative Frequency Tracks) matrix is obtained. Secondly median is calculated in *RTF* columns (i.e. discrete time moments). Also mean as well as weighted mean or median values can be utilized in this step (see [5] for details). Finally, $p_w(n)$ is obtained as a cumulative product of the mean values computed for each discrete time moment.

## 3. EXPERIMENTS

The proposed algorithm was tested on several archival sound samples recorded in the Polish National Film Library and The Documentary and Feature Film Studio. Presented example epitomise obtained results. Fig. 4 depicts the spectrogram of the sample simultaneously with the detected tracks and the evaluated pitch variation curve (*PVC*). As can be noticed from the spectrogram only the most relevant tracks are taken into account for the *PVC* extraction.

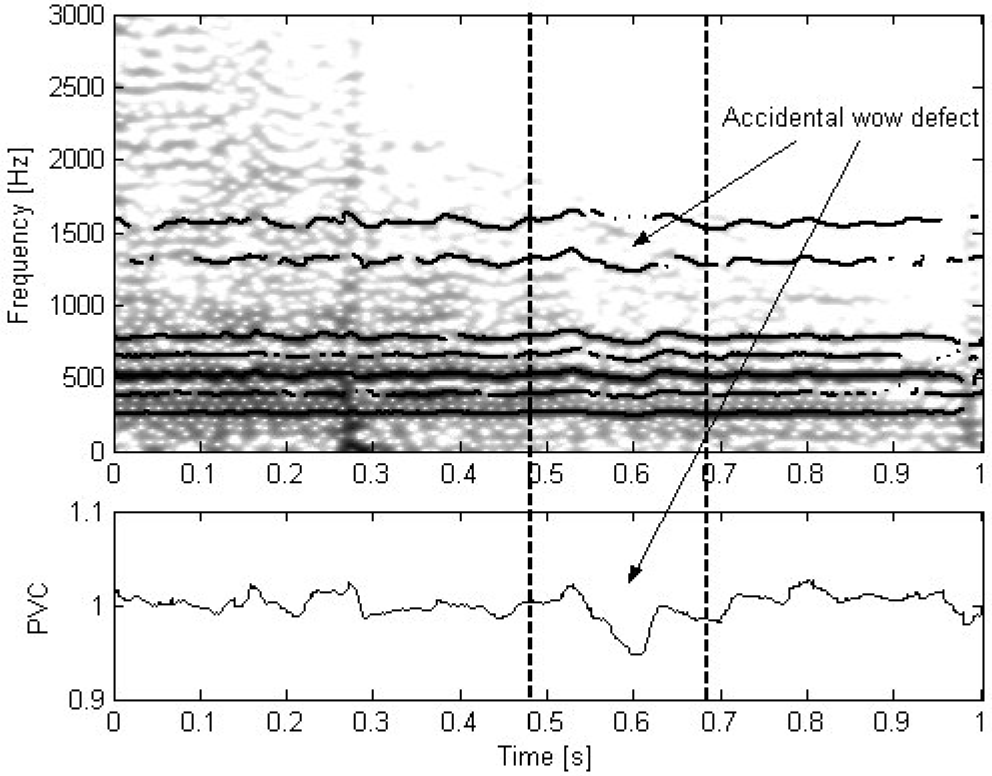

Fig.4 Tracks detected by the algorithm plotted on the spectrogram with the simultaneously plotted pitch variation curve

This example clearly demonstrates computed *PVC* and wow defect convergence. It is also worth to mention that the obtained *PVC* characteristics were successfully utilized in restoration process.

# 4. CONCLUSIONS

As the preliminary experiments result in satisfactory wow defect evaluation, the presented approach appears to be valid. The presented algorithm sufficiently manages to detect mild changes of the pitch variation function $p_w(t)$. However, in case of strong variations of $p_w(t)$ (e.g. accidental wow defect), it still needs to be more robust, especially with regard to the tracking stage. Therefore, a further development of this algorithm is underway.

# 5. REFERENCES

1. McAulay .J., Quatieri T.F., *Speech analysis/synthesis based on a sinusoidal representation, IEEE Trans. Acoustics*, Speech, and Signal Processing, 34(4) pp. 744-754, August 1986.

2. Godsill J. S., Rayner J. W., *The restoration of pitch variation defects in gramophone recordings*, Applications of Signal Processing to Audio and Acoustics, IEEE, 1993.

3. Godsill J. S., *Recursive restoration of pitch variation defects in musical recordin*gs, Proc. International Conference on Acoustics, Speech, and Signal Processing, vol. 2 pp 233-236, Adelaide, April 1994.

4. Walmsley P. J., Godsill S. J., Rayner P. J. W., *Polyphonic pitch tracking using joint Bayesian estimation of multiple frame parameters*, Proc. 1994 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York 1999.

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

6. Nichols J., *An interactive pitch defect correction system for archival audio*, AES 20th International Conference, Budapest, October 2001

7. Howarth J., Wolfe P., *Correction of Wow and Flutter Effects in Analog Tape Transfers*, 117 Audio Engineering Society Convention, Convention Paper 6213, 28-31 October, San Francisco 2004

8. Wolfe P., Howarth J., *Nonuniform Sampling Theory in Audio Signal Process*sing, 116 Audio Engineering Society Convention, Convention Paper 6123, 8-11 May, Berlin 2004

9. Marvasti F., *Nonuniform Sampling Theory and Practice*, Kluwer Academic/Plenum Publishers, New York 2001

10. Laakso T., Valimaki V., Karjalainen M., Laine U., *Splitting the Unit Delay*, IEEE Signal Processing Magazine, pp 30-60, January 1996

11. Serra X., Musical *Sound Modeling with Sinusoids plus Noise*, published in. Pope S., Picalli A., De Poli G., Roads C. Ed., "*Musical Signal Processing*", Swets & Zeitlinger Publishers, 1997

12. Rodet X., *Musical Sound Signal Analysis/Synthesis: Sinusoidal + Residual and Elementary Waveform Models*, Proc. IEEE Symp. *Time-Frequency and Time-Scale Analysis*, 1997

13. Lagrange M., Marchand S., Rault J.B., *Sinusoidal parameter extraction and component selection in a non-stationary model*, Proc. of the 5th Int. Conference on Digital Audio Effects, Hamburg, September 2002

14. Masri P., Computer *Modeling of Sound for Transformation and Synthesis of Musical Signals*, PhD thesis, University of Bristol, 1996

15. Auger F., Flandrin P., *Improving the readability of time-frequency and time-scale representations by the reassignment method*, IEEE Transactions on Signal Processing, vol.43, no. 5, pp. 1068-1089, May 1995

# ACNOWLEDGMENTS

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