MUSIC DEREVERBERATION BY SPECTRAL LINEAR PREDICTION IN LIVE RECORDINGS

Katariina Mahkonen[1], Antti Eronen[2], Tuomas Virtanen[1], Elina Helander[1], Victor Popa[1], Jussi Leppänen[2], Igor D.D. Curcio[2]

[1] Department of Signal Processing Tampere University of Technology, Tampere, Finland
firstname.lastname@tut.fi

[2] Nokia Research Center
Tampere, Finland
firstname.lastname@nokia.com

ABSTRACT

In this paper, we present our evaluations in using blind single channel dereverberation on music signals. The target material is heavily reverberated and dynamic range compressed polyphonic music from several genres. The applied dereverberation method is based on spectral subtraction regulated by a time-frequency domain linear predictive model. We present our results on enhancing music signal quality and automatic beat tracking accuracy with the proposed dereverberation method. Signal quality enhancement, measured by improvement in signal to distortion ratio, is achieved for both reverberant and dynamic range compressed signals. Moreover, the algorithm shows potential as a preprocessing method for music beat tracking.

1. INTRODUCTION

Reverberation is a phenomenon of sound energy persisting within a space due to a multitude of echoes from surrounding surfaces. Reverberation impacts the coloring of sounds. Usually the early reflections of the sound due to walls and other reflectors around are considered comfortable for human perception. Therefore, concert halls are designed to have some amount of reverberation and artificial reverberation is used as an artistic effect in music production. However, under heavy reverberation, the intelligibility of speech [1] and pleasantness of music decreases. Furthermore, the accuracy of automatic audio analysis algorithms decreases [2, 3].

The process of suppressing reverberation within audio signals is called dereverberation or acoustic channel equalization. When there is no information about the acoustic impulse response (AIR), dereverberartion is named blind. There are both time-domain [4,5] and frequency-domain [4,6,7,8,9] techniques for this task available. Time-domain techniques aim at estimating an AIR, and suppressing the echoes using signal deconvolution. There are however several problems in this approach although it is theoretically appealing. Estimating the AIR and its frequency domain representation, the acoustic transfer function (ATF), is fairly difficult as ATF is generally not minimum phase. Also it is very sensitive to even small deviations to recording geometry, thus in all practical cases it must be considered time-varying, and ideally infinite AIR must be estimated as a finite sequence. However, many techniques, for example [11], use this approach to remove few early reflections and use a spectral technique for suppressing late reverberation part. The methods operating purely in frequency domain have adopted the idea of spectral subtraction and attenuate amplitudes within spectral bands according to some criterion.

Most of the dereverberation studies conducted so far have considered speech signals, aiming at increasing both the intelligibility of speech for humans and automatic speech recognition (ASR) performance [4]. Some studies have included music signals in evaluations of applied dereverberation algorithms [5, 6], and few experiments have focused primarily on music signals [7, 8]. Wilmering & al. have explored using dereverberation as a preprocessing step for music onset detection and musical instrument recognition [8]. For evaluation they used single-instrument recordings generated from MIDI. In [7], Yasuraoka & al. have performed music dereverberation on monophonic musical recordings and evaluated the results with the log spectral distance improvement (LSDI). However, such a measure, which uses only the spectral magnitude does not take into account the phase disturbances, which are a very common source of artifacts in framewise audio processing.

Our goal in this work is to discover whether the chosen dereverberation method is effective in processing polyphonic musical signals which are deteriorated by dereverberation and subjected to dynamic range compression (DRC). Many dereverberation algorithms, including the proposed one, are based on linear prediction (LP), which assumes that reverberation is a linear process. However, DRC is often applied to audio recordings, and as it is a nonlinear operation, it potentially disturbs LP-based dereverberation algorithms.

We analyze how well the applied dereverberation method can suppress the reverberation in terms of improvement in the signal to distortion ratio (SDR) [10]. We compare the achieved SDR improvement (SDRI) for the signals initially deteriorated by reverberation and the same signals after subjecting them to DRC to see how the DRC affects the dereverberation performance. We also test how the dereverberation affects the accuracy of automatic beat tracking when used as a preprocessing step.

This paper is organized as follows. In Section 2 the used dereverberation method is explained. Section 3 describes the performed evaluations, whereas the results are presented in Section 4. Discussion and conclusions complete the paper in Section 5.

2. METHOD

In this section, the proposed dereverberation method, which is adopted mainly from [11], is introduced. The mathematical model for the reverberation, model parameter estimation and the methods used for dereverberating the observed signal are de-

scribed. This is followed by a brief description of the beat tracking method used.

2.1. Spectral Subtraction via Linear-predictive model

The music signal is processed in successive frames which are partially overlapping and smoothed with a Hanning window. The spectrum of each frame is computed with the discrete Fourier transform (DFT). The model we use to describe the spectral magnitude $|X(n,f)|$ of the reverberant signal in each frame n and each frequency f is formulated as

$$|X(n,f)| = |S(n,f)| + |R(n,f)|.$$

$|S(n,f)|$ and $|R(n,f)|$ are respectively the spectral magnitudes of the clean source signal and the reverberation noise part in the observed signal in the same time-frequency-bin. The reverberation part of the signal within each frequency f is modeled by a linear predictive system

$$\left|\hat{R}(n,f)\right| = \sum_{p\epsilon P} a_f(p)\, |X(n-p,f)|, \qquad (1)$$

where P defines a set of frame indices anterior to index n, which are considered to be involved with the late reverberation within frame n. Generally $P = \{1...p_{\max}\}$ if the LP-model order is $p_{\max}$. However, some frames can be omitted from the full set $P = \{1...p_{\max}\}$ to prevent subtraction of early reflections or to prevent the LP-solution from being affected by the regular beat of music. As an alternative to frequencies f given by DFT for the model, we may prefer to model the average of the spectral magnitude within some frequency bands, such as frequency bands with center frequencies spaced evenly on the perceptually motivated mel-scale. In this case we replace $|X(n-p,f)|$ in (1) with a mean spectral magnitude within a frequency band k for $f = \{ f_{\min}{}^k \ldots f_{\max}{}^k \}$. Then the magnitude of the reverberation noise $\left|\hat{R}(n,f)\right|$ is considered equal for all the frequencies f within the frequency band k.

The parameters $\boldsymbol{a}_f = [a_f(1), a_f(2), \ldots, a_f(p_{\max})]^{\mathrm{T}}$ of the reverberation model are estimated separately for each recording and frequency f or band k. We determined the weight vectors $\boldsymbol{a}_f$ by the standard Least Squares solution

$$\boldsymbol{a}_f = \left(V_f{}^T V_f\right)^{-1} V_f{}^T \boldsymbol{v}(p_{\max}+1, f)$$

where $V_f = [\boldsymbol{v}(p_{\max}, f),\ \boldsymbol{v}(p_{\max}-1, f), \ldots, \boldsymbol{v}(1, f)\,]$ and $\boldsymbol{v}(i,f)$ is defined as $\boldsymbol{v}(i,f) = [|X(i,f)|,\ |X(i+1,f)|, \ldots,\ |X(i+N-p_{\max}-1,\ f)|\,]^{\mathrm{T}}$. N is the number of frames in one recording. Additionally, we calculated vectors $\boldsymbol{a}_f$ with an algorithm from [15] forcing all the values to be non-negative, i.e. $\boldsymbol{a}_f(p) \geq 0$ for all p. This constraint was chosen heuristically, as a physical nature of sound energy is to decay in time throughout the spectrum almost invariably in natural environments.

In order to prevent undesirable processing effects, a frequency dependent parameter $\beta(f)$ is used in the dereverberation stage to limit the amount of dereverberation as follows

$$\left|\hat{S}\,(n,f)\right| = |X(n,f)| - \beta(f)|\hat{R}(n,f)|.$$

The complex spectrum of the dereverberated signal is generated from the dereverberated magnitude spectrum $\left|\hat{S}\,(n,f)\right|$ using the phase information from the originally observed signal spectrum as

$$\hat{S}(n,f) = \left|\hat{S}(n,f)\right| e^{-i\angle X(n,f)}\,.$$

Each dereverberated signal frame is produced via the inverse discrete Fourier transform (IDFT) and the frames $n = \{1 \ldots N\}$ are combined after IDFT by summing their contributions together with the overlap-add method.

2.2. Beat tracking

The proposed dereverberation method is also evaluated as a preprocessing method for a music beat tracker. The beat tracker combines the elements from the methods presented in [12] and [13] and is only briefly described here, highlighting the essential novel parts. Beat tracking starts by obtaining an estimate of the average tempo of the signal with the tempo estimation method of [12]. The method computes a pitch-chroma based accent signal to measure the degree of spectral change and music accentuation over time. The accent signal is processed by a generalized autocorrelation function to compute periodicity vectors, and then k–nearest neighbor regression is applied on the periodicity vectors to obtain an estimate of the signal tempo. The beat tracking step takes the tempo as an input and estimates the most likely sequence of beat times from the signal, using the effective dynamic programming routine described in [13]. Compared to the beat tracking system described in [13], this beat tracker provides superior accuracy which is attributed to the inclusion of the robust k-nearest-neighbor based tempo estimation step described in detail in [12].

An obvious way to implement the dereverberation as a preprocessing for beat tracking is to input the dereverberated signal to the beat tracker. However, this was not found to give any improvement in beat tracking accuracy on the used dataset. On the contrary, often a decrease in the accuracy was observed. Instead, it was found better to input the dereverberated signal to the tempo estimation step, and to perform the beat tracking step on an accent signal calculated from the original, non-dereverberated signal. That is, we perform the accent signal analysis described in [12] on both the original signal and the dereverberated signal. The accent signal computed from the dereverberated signal is used in tempo estimation. Then, the tempo estimate and the accent signal computed from the original signal are input to the beat tracking step.

The accent signal measures the change in the spectrum of the signal and exhibits peaks at onset locations. The goal of the beat tracking step is to find the most likely sequence of beat times, given the tempo estimate and the accent signal. Beat tracking is performed with the method described in [13]. The dynamic programming step takes as inputs the accent signal and the beat period, performs smoothing of the accent signal with a Gaussian window, and then finds the optimal sequence of beat times through the smoothed accent signal.

3. EVALUATION AND RESULTS OF SIGNAL QUALITY ENHANCEMENT

Both artificially reverberated and real-world reverberant signals were used in testing the algorithm. For objective signal quality evaluation purposes, two sets of non-echoic polyphonic musical signals were generated from tracks stored in MIDI format using

the Timidity synthesis software. The tracks used were from classical, pop and jazz genres. The evaluation set consists of 13 sound segments of length from 7s to 29s, and these sounds were used for system parameter estimation. The test set consists of 31 segments of 20s to 9min in duration to be used for evaluation of the dereverberation performance. To resemble a reasonable real-life concert situation, a room impulse response (RIR) from the AIR database [14] with a reverberation time $T_{60} \approx 3$s was used for reverberating the dry music signals. To evaluate the effect of dynamic range compression, DRC with a compression ratio 3:1 above the threshold -20dB was applied. The timespans for root mean square (RMS) signal power level estimation for increasing and decreasing the DRC-gain were τ_{attack}= 5ms and $\tau_{release}$= 200ms respectively.

As an evaluation metric for these artificially distorted signals we used the signal to distortion ratio (SDR) [10]. SDR is more suitable for evaluation of dereverberation performance than the measures operating purely in the frequency domain, such as the log spectral distance improvement used in [7]. This is due to the fact that the SDR-calculation segregates the source+early reflections part s_{clean} from the evaluated signal s in the time domain by considering s's projections to the known clean source signal s_{dry} and its slightly delayed versions. Then SDR in dB is calculated as the logarithmic energy ratio of s_{clean} and the remaining noise part as

$$\mathrm{SDR} = 10\log_{10}\frac{||s_{\mathrm{clean}}||^2}{||s-s_{\mathrm{clean}}||^2} \ ,$$

where s is either the distorted or the dereverberated signal. The amount of signal quality enhancement was calculated as the SDR-improvement (SDRI) as

$$\mathrm{SDRI} = \mathrm{SDR}_{\mathrm{dereverberated}} - \mathrm{SDR}_{\mathrm{reverberant}} \ .$$

The optimal values for most of the system parameters were selected according to SDRIs given by the evaluation sound set and kept unchanged for producing the results with the test sound set. According to the SDRIs on the evaluation sound set, forcing all the linear prediction weights $\boldsymbol{a}$ to be non-negative was found beneficial. The set P={1,2,3} for the LP-model was selected as sufficient. Only with very short processing frames, say 20ms, increasing p_{max} was found beneficial. Reducing the full set from P = {1…p_{max}} was found to decrease the performance. Empha sizing dereverberation on certain spectral area with frequency dependent $\beta(f)$ was tested with ascending, descending and smooth window-function –like $\beta(f)$:s. None of these was found out to have strong positive effect on dereverberation result, the weight for the lowest frequencies dominated the result in every case, thus constant β –value was used in the test phase.

The results with the test set, introduced in Figure 1, show that dereverberation can be done successfully with this method. Comparisons of SDRIs when the processing frame length and the amount of dereverberation β are varied are shown in Figure 1 (a) and (b) respectively. Nonlinear DRC was found not to deteriorate the dereverberation performance. The average SDR-values prior to dereverberating are 6.1 for only reverberated and 5.2 for the reverberated and DRC-processed signals. Thus the achieved higher SDRIs for DRC-modified signals do not imply higher final SDR.

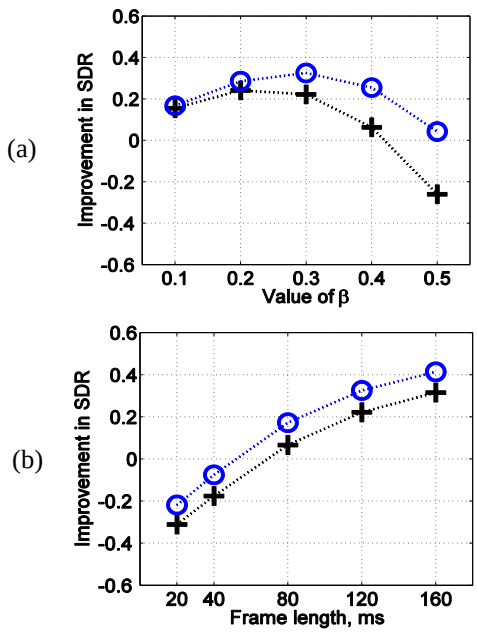

Figure 1: *Improvement in signal to distortion ratio due to dereverberating the signal, when (a) the amount of dereverberation, i.e. value of β or (b) the length of a processing frame are varied, and the rest of the system parameters are kept constant. The crosses correspond to results with signals suffering only from reverberation. The circles correspond to results with signals subjected also to DRC.*

4. EVALUATION AND RESULTS IN MUSIC BEAT TRACKING

The dataset for testing the effectiveness of the method as a preprocessing step for beat tracking comprises 113 musical excerpts captured with mobile devices in live situations. The music material is mainly from mainstream pop, rock, and dance genres, with a few salsa and progressive tracks. The amount of reverberation in the recordings varies from highly reverberant to not reverberant and also the amount of DRC varies. It is desired that the method should not decrease the beat tracking accuracy even if reverberation is not present, and therefore also non-reverberant examples are included. Some of the signals contain distortion and noises from the audience, presenting a very challenging scenario for beat tracking.

The ground truth beat annotations were input by human experts by tapping along to the pieces. The beat tracking accuracy is measured with the performance criteria described in more detail in [15]. "Correct" denotes the percentage of pulse estimates where both the period and phase are correct within a 17.5 % tolerance. "Accept d/h" allows consistent tempo halving and doubling whereas "Period correct" ignores the phase and considers only the period, i.e., the tempo.

The results for beat tracking are depicted in Table 1. The baseline denotes the results of the beat tracking method when no dereverberation is applied, and the results in the row "Dereverberated" denote the results when the tempo estimation is performed on the dereverberated signal. The results are shown for the best parameter combination, where the length of the processing frame was 120ms, P = 1, the number of mel-frequency

bands K used for estimating the reverberation $|R(n,k)|$ for k=1...K is 128, and β = 0.2. This parameter combination was obtained by varying the dereverberation method parameters and running the system on the complete dataset, using the beat tracking accuracy as the parameter selection criterion. From the results, a small improvement is observed, indicating potential of the method as a preprocessing stage for tempo estimation and beat tracking in reverberant conditions.

Table 1: *Results of beat tracking.*

	Correct	Accept d/h	Period correct
Baseline	58%	65%	81%
Dereverberated	60%	67%	84%

5. CONCLUSIONS

The goal of this paper was to verify whether the proposed dereverberation method, which is based on spectral subtraction regulated by a linear predictive model, is effective in enhancing reverberant polyphonic music signals. The performance of the method was evaluated using a signal to distortion ratio improvement measure. Using SDRI, we also investigated the effect of dynamic range compression on dereverberation performance. The performance of the method on automatic beat tracking, when the dereverberation was done in a preprocessing step, was measured too.

The results show that music dereverberation can be achieved by this method and the presence of dynamic range compression does not deteriorate the performance. Even better, the signal to distortion ratio improvement turned out to be higher when the music had been subjected to DRC. Generally the quiet signal parts are relatively harder to dereverberate than the loud signal parts. Thus the more even dynamics of DRC-processed music is beneficial for dereverberation. Also, as the louder signal parts dominate the value given by SDR-measure, the more stationary and even dynamics of DRC-processed signals may be an asset. Anyhow, the absolute SDR-values both before and after dereverberation were lower for the signals distorted by DRC than for the signals containing only reverberation. Altogether, this is very interesting result and it gives us verification that this kind of nonlinearity is not a problem for the linear dereverberation method used.

The method shows promise for improving beat tracking accuracy in highly reverberant conditions but the improvement is too small for strong conclusions to be made. Somewhat unexpectedly, no improvement in beat tracking accuracy was observed when beat tracking was performed on the dereverberated signal, although one could have anticipated such behavior based on the reported increase in sound onset detection accuracy in [8]. A small increase in overall beat tracking accuracy was observed only when tempo estimation was performed on the dereverberated signal while performing the beat tracking on the original signal. A possible explanation for this behavior is that the dereverberation is successful in enhancing the pulse sensation in music. Indeed, informal listening experiments indicate that the beat pulse is slightly better audible in the highly reverberant signals after dereverberation, which may explain why it helps the tempo estimation. However, since the beat tracking accuracy is not improved if performed on the dereverberated signal, the dereverberation may be causing too much artifacts on the temporal accent signal shape for the beat positioning accuracy to improve.

6. REFERENCES

[1] M. Klatte, T. Lachman and M. Meis, “Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting”, *in Noise and Health*, Vol. 12, Issue 49, 2010, pp. 270-282.

[2] T. Wilmering, G. Fazekas and M. B. Sandler, “The effects of Reverberation on Onset Detection Tasks”, *Audio Engineering Society Convention 128*, London, UK, May 2010.

[3] T. Virtanen, R. Singh and B, Raj, *Techniques for Noise Robustness in Automatic Speech Recognition*, pp. 42-43, Wiley, 2012

[4] P.A. Naylor, N.D. Gaupitch, “Speech Dereverberation*”, Signals and Communication Technology*, Springer, London 2010.

[5] T. Okamoto, Y. Iwaya and Y. Suzuki, “Wide-band dereverberation method based on multichannel linear prediction using prewhitening filter”, *in Applied Acoustics*, Vol.73, Nr. 1, 2012, pp. 50-55.

[6] A. Tsilfidis and J. Mourjopoulos, “Blind single-channel suppression of late reverberation based on perceptual reverberation modelling”, *Journal of the Acoustical Society of America*, Vol. 129, Issue 3, March 2011.

[7] N. Yasuraoka, T, Yoshioka, T. Nakatani, A. Nakamura, H. G. Okuno, ”Music dereverberation using harmonic structure source model and Wiener filter”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 53-56.

[8] T. Wilmering, M. Barthet and M. B. Sandler, “Dereverberation of Musical Instrument Recordings for Improved Note Onset Detection and Instrument Recognition”, *Audio Engineering Convention 131*, New York USA, October 2011.

[9] K. Lebart, J.M.Boucher, P.N.Denbigh, “A new nethod Based on Spectral Subtraction for speech Dereverberation”, *Acta Acustica*, Vol 87, 2001, pp. 359-366.

[10] E. Vincent, R. Gribonval and C. Févotte, “Performance measurement in Blind Audio Source Separation”, *IEEE Transactions on Audio, Speech and Language*, Vol. 14, Nr. 4, 2006.

[11] K. Furuya and A.Kataoka, “Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction”, *IEEE Transactions on Audio, Speech and Language Processing*, Vol.15, No. 5, July 2007.

[12] A. Eronen, A. Klapuri,“Music tempo estimation with *k*-NN regression” *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, Nr. 1, 2010, pp. 50-57.

[13] D.P.W. Ellis, ”Beat tracking by dynamic programming” *in Journal of New Music Research*, Vol. 36, Nr. 1, 2007, pp. 51-60.

[14] M. Jeub, M. Schäfer and P.Vary, “A binaural room impulse response database for the evaluation of dereverberation algorithms” *in proceedings of the 16th international conference on Digital Signal Processing*, DSP’09, Greece, 2009, pp. 550-554.

[15] A. Eronen, A. Klapuri, J. Astola,” Analysis of the meter of acoustic musical signals”, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, Nr. 1, 2006, pp. 342-355.

[16] C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems*, Prentice Hall, 1974, Chapter 23, p.161.