Audio Engineering Society

Convention Paper

Presented at the 130th Convention
2011 May 13–16 London, UK

Blind Single-Channel Dereverberation For Music Post-Processing

Alexandros Tsilfidis[1], John Mourjopoulos[1]

[1] *Audio and Acoustic Technology Group, Department of Electrical and Computer Engineering, University of Patras, 26504, Greece*

Correspondence should be addressed to Alexandros Tsilfidis (altsilfidis@upatras.gr)

ABSTRACT
Although dereverberation can be useful in many audio applications, such techniques often introduce artifacts that are unacceptable in audio engineering scenarios. Recently, the authors have proposed a novel dereverberation approach, suitable for both speech and music signals, based on perceptual reverberation modeling. Here, the method is fine-tuned for sound engineering applications and tested for both natural and artificial reverberation. The results show that the proposed technique efficiently suppresses reverberation without introducing significant processing artifacts and the method is appropriate for the post-processing of music recordings.

1. INTRODUCTION

Room reverberation can be regarded as the combination of early reflections and late reverberation. In room acoustics, early reflections are known to result mainly to spectral degradation, perceived as coloration. On the other hand, late reverberation generates the signal's reverberant tails, producing a decaying noise-like effect. Most dereverberation techniques tackle separately the signal distortions produced by the early and late reverberation. In order to deal with the coloration effect produced by the early reflections, methods involving the inverse filtering of the Room Impulse Response (RIR) have been developed based on least-squares [1, 2], frequency warping [3, 4], complex smoothing [5, 6], Kautz filters [7] and spatial clustering [8]. Such methods are common in Room Correction systems, already being incorporated in commercial applications. However, results from subjective tests show that some of these techniques do not always achieve the desired perceptual effect [9, 10, 11]. On the other hand, spectral subtraction has been adapted in the dereverberaiton context in order to compensate for the late reverberation effect. In such cases, late reverberation is considered to be an additive noise and most methods are trying to accurately estimate the late reverberation short-time spectrum and avoid or reduce the processing artifacts [12, 13, 14, 15, 16, 17].

The suppression of late reverberation can be useful in many audio applications. Music is often recorded in venues that are not acoustically treated. In such cases, suppressing the reverberant tails of the recorded signals is often a necessary step before further processing. In the lack of a specialized tool, sound engineers are often obliged to manually suppress the reverberant tails or use tools that are not directly developed to handle such problems (e.g. noise gates). Hence, a specially designed dereverberation tool would be of advantage. In addition, reverberation severely changes the music signal's statistics [18, 19]. Hence, reverberation (and mainly late reverberation) causes a decrease of performance in music signal classification, automatic music transcription, analysis and melody detection, source separation, etc [20, 21, 22, 23, 24].

Most recent dereverberation algorithms have been developed specifically for speech signals. However, such techniques are not always appropriate for processing broadband audio signals as in principle music dereverberation is more challenging. Music signals have broader frequency range and sharper transient structure and typical statistical assumptions made for speech are not always valid for music. Furthermore, music is often reproduced in big auditoria and thus longer RT values are usually involved. In addition, for most speech applications, a deterioration of signal's quality might in principle be acceptable after dereverberation if an increase in ASR performance can be achieved. On the contrary, in realistic sound engineering scenarios, the quality of the produced dereverberated signal should not be compromised.

Recently the authors have proposed a novel method for blind single-channel dereverberation, appropriate for both speech and music signals [25]. The technique is based on perceptual reverberation modeling and the main principle is to employ a time-frequency masking model in order to reduce reverberation relatively to the subjective degradation. Hence, overestimation artifacts that degrade the clean signal components are largely reduced. Note that, psychoacoustic research has pointed that the temporal envelope of a signal should be viewed as a real signal within the auditory system [26] and despite the fact that it is well-known that temporal processing can be used to suppress late reverberation [27], recent late reverberation suppression methods are entirely employed in the spectral domain. On the other hand, the proposed technique suppresses reverberation involving temporal processing in perceptually-significant sub-bands. In this work, the proposed method is fine-tuned for music signals and tested with several types of music but also with measured and artificial impulse responses.

2. LATE REVERBERATION SUPPRESSION BASED ON PERCEPTUAL REVERBERATION MODELING

The blind method utilized here is explained in detail in [25]. However, the main processing steps are briefly described here to facilitate the reading of the remaining paper:

- A spectral preprocessor is used in order to achieve a first rough estimation of the clean signal [12]. For this, a state-of-the art late reverberation suppression method is employed and signal-dependent constraints are applied in order to preserve the signals transients [14]. Furthermore a perceptually-motivated non-linear filtering stage is implemented to reduce the musical noise artifacts[28, 14].
- A time-frequency auditory model, namely the Computational Auditory Masking Model (CAMM) has been incorporated in the dereverberation process [29]. This model quantifies the reverberation distortion throughout the signal's evolution and locates the signal regions where late reverberation is audible, i.e. it is unmasked from the clean signal. Hence, an approximation of the Reverberation Masking Index (RMI) $D_k(n)$ [30] is derived in the sub-band domain [25] (n denotes the discrete time index).
- The proposed method employs a selective signal-processing approach, where only the signal components that are badly contaminated

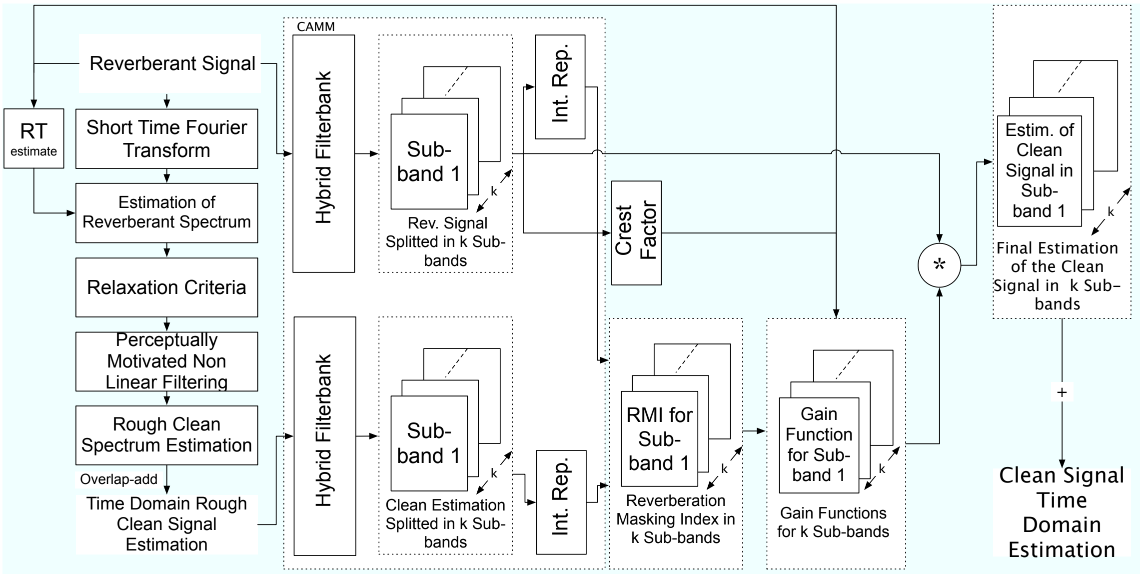


Fig. 1: Signal flow of the proposed dereverberation method method.

from late reverberation are processed. Such signal regions are identified as the time intervals between each local RMI maximum and minimum which typically correspond to signal offsets containing perceptually-detectable late reverberation energy [25]. These signal regions are processed through a hybrid gain function $\mathcal{G}_k(n)$ which is adaptively adjusted based on indicators of the severity of the reverberation degradation (an estimation of the RT [31] and the Crest Factor Cf_k in each sub-band).

To express formally the above steps, let L_k be the total number of local $D_k(n)$ extrema in a sub-band, $M_k(i)$ and $m_k(i)$ the time intervals where each local maximum and the succeeding local minimum occur respectively. By definition $m_k(0) = 0$ and i designates the $\frac{L_k-1}{2}$ consecutive pairs of local extrema. Between each pair of a local maximum and the succeeding minimum (i.e. when $M_k(i) \leq n \leq m_k(i)$) the gain function $\mathcal{G}_k(n)$ in each sub-band decreases exponentially:

$$\mathcal{G}_k(n) = G_k(i)\left(\frac{g_k(i)}{G_k(i)}\right)^{\frac{n-M_k(i)}{m_k(i)-M_k(i)}} \tag{1}$$

In any other case (i.e. when $m_k(i) \leq n \leq M_k(i+1)$), the gain function $\mathcal{G}_k(n)$ takes the value 1:

$$\mathcal{G}_k(n) = H(n - m_k(i)) - H(n - M_k(i+1)). \tag{2}$$

where H is the Heaviside step function:

$$H(n) = \begin{cases} 1 & \text{when } n \geq 0 \\ 0 & otherwise \end{cases} \tag{3}$$

Note that $G_k(i)$ represents the value of the gain function at each local maximum derived as:

$$G_k(i) = \frac{1}{1.8^{RT_{60}}}\left(\frac{1-Cf_k}{1+Cf_k} + \gamma_1\right) \tag{4}$$

and $g_k(i)$ is the value of the gain function at a consecutive local minimum:

$$g_k(i) = G_k(i) - \gamma_2\left(max(D_k(n)) + RMS(D_k(n))\right) \tag{5}$$

where $RMS(D_k(n))$ is the RMS value of the RMI for the signal area of interest $M_k(i) \leq n \leq m_k(i)$:

$$RMS(D_k(n)) = \sqrt{\frac{1}{m_k(i)-M_k(i)} \sum_{j=1}^{m_k(i)-M_k(i)} D_k'^2(n)} \tag{6}$$

A block diagram illustrating the proposed method is presented in Fig. 1. Equations 4 and 5 incorporate two free parameters: γ_1 and γ_2. The values of

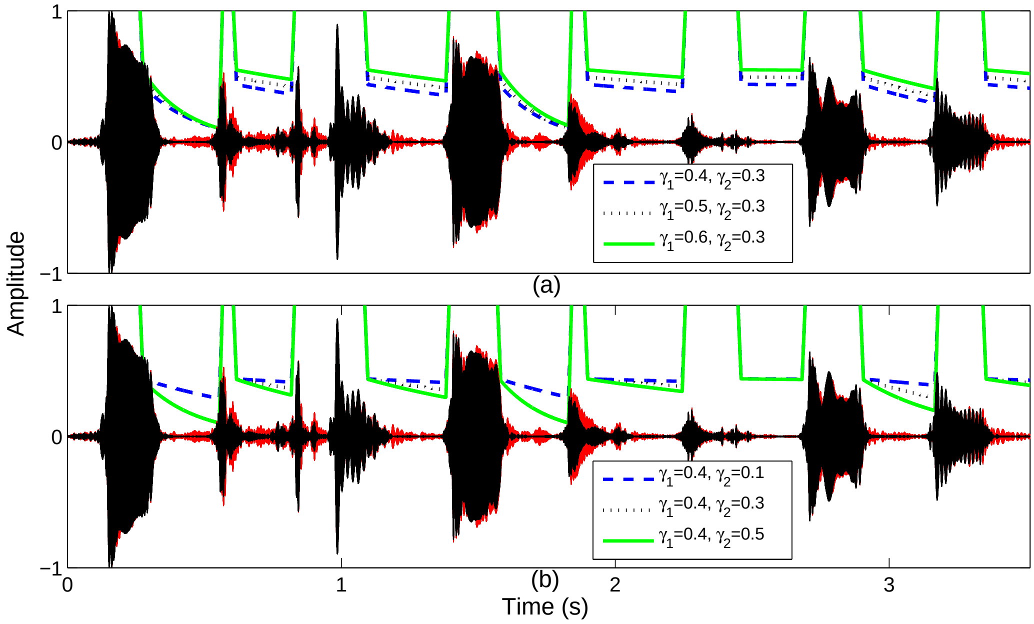


Fig. 2: Gain functions for typical γ_1 and γ_2 values (in a sub-band centered at 2 kHz). The reverberant (red) and anechoic (black) signals corresponding to an acoustic guitar recording are also shown.

these constants are adjusted depending on the dereverberation context to improve the quality of the produced signals. The suppression rate is controlled through γ_1, while the parameter γ_2 controls the decaying rate of the gain function between each pair of local extrema. Hence, smaller γ_1 and bigger γ_2 values result to substantial reverberation reduction in each sub-band. Fig. 2 presents the gain functions for different γ_1 and γ_2 values and the corresponding reverberant and anechoic audio signal (guitar) in a sub-band centered at 2 kHz.

3. TESTS AND RESULTS

An anechoic signal dataset was constructed including samples from the Archimedes database [32] and the TKK anechoic recordings [33]. The dataset contained samples from: (i) acoustic guitar, (ii) timpani, (iii) soprano voice, (iv) male speaker, (v) orchestra and (vi) trumpet. In order to produce the reverberant samples, the anechoic samples were convolved with impulse responses obtained from real room measurements at a source-receiver distance of 1 m and also with impulse responses extracted from standard presets of two commercially available artificial reverberators. The properties of the corresponding impulse responses are presented in Table 1.

Name	*Type*	$RT_{60}(s)$
Room 1	Auditorium	1
Room 2	Music Hall	1.47
Art. Rev. 1	Digital Reverb	3
Art. Rev. 2	Digital Reverb	6

Table 1: Properties of the tested impulse responses

In [25], it is mentioned that γ_1 typically ranges between 0 and 1.5 and γ_2 between 0.1 and 0.5 (see Equations 4 and 5). Here, the authors conducted unofficial listening tests in order to define a suitable range of γ_1 and γ_2 values for music applications. Setting $\gamma_1 = 0.4,\ 0.5,\ 0.6$ and $\gamma_1 = 0.05,\ 0.1,\ 0.2$ was found to produce significant reverberation suppression without introducing processing artifacts for all tested cases and the corresponding nine parameter pairs were used for the tests presented below.

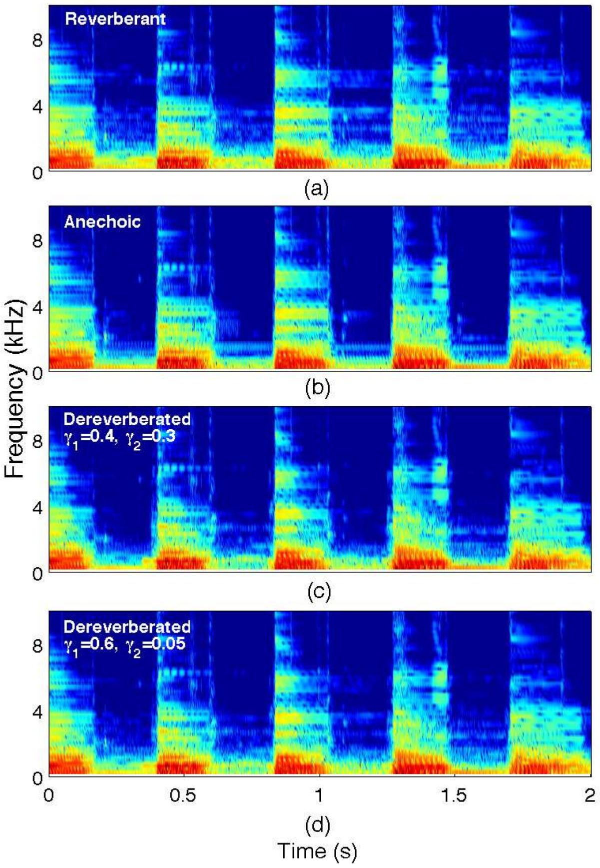


Fig. 3: Typical spectrograms for an acoustic guitar utilizing artificial reverberation Art. Rev. 2: (a) reverberant, (b) anechoic, (c) proposed dereverberation using $\gamma_1 = 0.4$ and $\gamma_2 = 0.3$, (c) proposed dereverberation using $\gamma_1 = 0.6$ and $\gamma_2 = 0.05$

4. EVALUATION IN TERMS OF SPECTROGRAM IMPROVEMENT

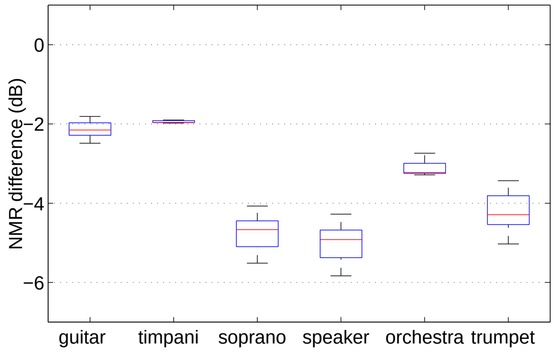
In Fig. 3 typical spectrograms of (a) reverberant guitar sample, (b) anechoic guitar sample, (c) clean signal estimation produced from the proposed method using using $\gamma_1 = 0.4$ and $\gamma_2 = 0.3$ and (d) clean signal estimation produced from the proposed method using using $\gamma_1 = 0.6$ and $\gamma_2 = 0.05$ are shown. By comparing Fig. 3 (b) and (a) it can be seen that reverberation has filled the time-frequency silence regions between the subsequent guitar notes. In Fig. 3 (c) a drastic version of the proposed method has recovered the silence parts of the original anechoic sample. Moreover, the less intrusive approach shown in Fig. 3 (d) has also significantly reduced the reverberation effect. Note that in both dereverberation implementations the anechoic signal has been well preserved and the processing has not introduced any artifacts.

5. EVALUATION IN TERMS OF NMR IMPROVEMENT

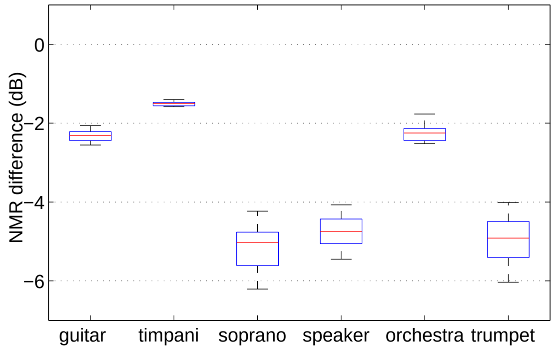
The relative improvement has been also evaluated in terms of (segmental) NMR difference. For this, the NMR over the estimated clean signal and the clean signal and the NMR over the reverberant and the clean signal were calculated. Then their difference was derived as:

$$\Delta NMR = NMR_{estimate} - NMR_{reverberant}. \quad (7)$$

Note that negative values of ΔNMR denote a relative improvement. In Fig. 5 the NMR differences for the tested sample types and for (a) Room 1 and (b) Room 2 are shown. The presented box plots correspond to the results from the nine different implementations of the proposed dereverberation approach. In all tested cases a significant NMR improvement was achieved ranging approximately between -2 and -6 dB. Although the improvement is consistent for all tested audio genres, it seems that sample type affects the overall performance. This was also noticed in the case of the artificial reverberators as shown in Fig. 5 where the achieved NMR differences for (a) Art. Rev. 1 and (b) Art. Rev. 2 are presented. Apart from the case of the orchestra sample reverberated with Art. Rev. 1 it seems that the proposed method achieved a NMR improvement. This improvement reaches the -6 dB for the soprano and speaker samples reverberated with Art. Rev. 1 but the overall dereverberation performance is reduced when compared to the natural reverberation in Fig. 5. In theory, dereverberation of reverberant signals generated with artificial reverbs should be easier since those are much more predictable than the real reverberant environments. However, in this work both natural and artificial reverbs were chosen to represent realistic user cases. Hence, the chosen artificial reverbs present extremely high RT values and low Direct to Reverberation ratios. Although in such conditions dereverberation becomes a really challenging problem, the proposed method seems to substantially enhance the produced signals. Finally,



(a)



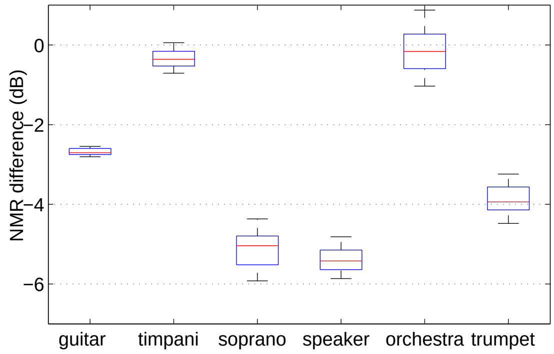
(b)

Fig. 4: NMR difference for (a) Room 1 and (b) Room 2

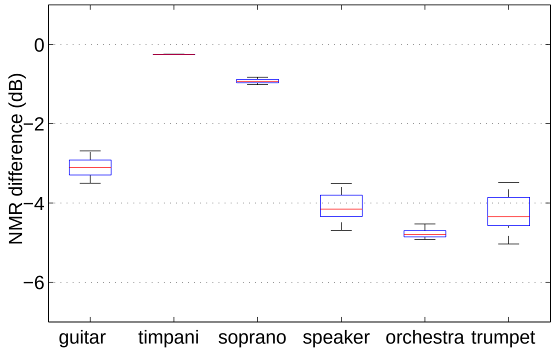
Fig. 6 presents the mean NMR improvement over all tested samples. Here, it is verified that dereverberation performance is signal dependent and as discussed in the Introduction the dereverberation of music is more demanding than speech dereverberation. In addition, it is shown that the proposed technique significantly enhances all tested signal types

6. CONCLUSIONS

A recent single-channel dereverberation method was adjusted for music applications. The proposed approach is based on perceptual reverberation modeling and employs a Computational Auditory Masking Model in order to locate and suppress the signal regions where reverberation is unmasked from the clean signal. The method is tested for both natural and artificial reverberation and the results show significant reverberation reduction for all tested signal types. Hence, the proposed technique can be used



(a)



(b)

Fig. 5: NMR difference for (a) Art. Rev. 1 and (b) Art. Rev. 2

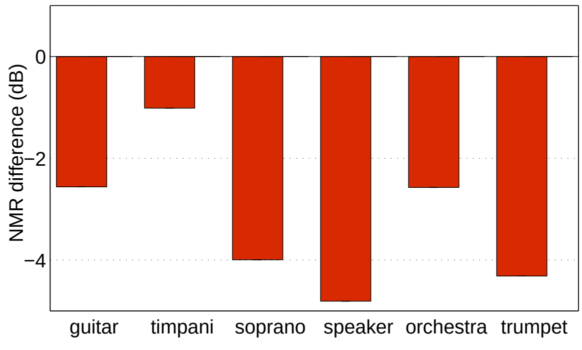


Fig. 6: Overall NMR improvement for the tested sample types

for the post-processing of music recordings.

ACKNOWLEDGEMENTS

The research activities that led to these results, were co-financed by Hellenic Funds and by the European Regional Development Fund (ERDF) under the Hellenic National Strategic Reference Framework (ESPA) 2007-2013, according to Contract no. MICRO2-07/E-II-A

7. REFERENCES

[1] S Neely and J B Allen. Invertibility of room impulse response. *Journal of the Acoustical Society of America*, 66:165–169, 1979.

[2] J Mourjopoulos, P Clarkson, and J Hammond. A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals. In *Proc. of the IEEE ICASSP*, 1982.

[3] A Harma, M Karjalainen, L Savioja, V Valimaki, U Laine, and J Huopaniemi. Frequency-warped signal processing for audio applications. *Journal of the Audio Engineering Society*, 48:1011–1031, 2000.

[4] L Savioja and V Valimaki. Multiwarping for enhancing the frequency accuracy of digital waveguide mesh simulations. *IEEE Signal Processing Letters*, 8(5):134–136, 2001.

[5] P Hatziantoniou and J Mourjopoulos. Generalized fractional-octave smoothing of audio and acoustic responses. *Journal of the Audio Engineering Society*, 48:259–280, 2000.

[6] P Hatziantoniou and J Mourjopoulos. Errors in real-time room acoustics dereverberation. *Journal of the Audio Engineering Society*, 52:883–899, 2004.

[7] M Karjalainen and T Paatero. Equalization of loudspeaker and room responses using Kautz filters: direct least squares design. *EURASIP Journal on Applied Signal Processing*, 2007.

[8] S Bharitkar, P Hilmes, and C Kyriakakis. Robustness of spatial average equalization: A statistical reverberation model approach. *Journal of the Acoustical Society of America*, 116(6):3491 – 3497, 2004.

[9] S Norcross, G Soulodre, and M Lavoie. Subjective investigations of inverse filtering. *Journal of the Audio Engineering Society*, 52:1003–1028, 2004.

[10] P Hatziantoniou, J Mourjopoulos, and J Worley. Subjective assessments of real-time room dereverberation and loudspeaker equalization. In *Proc. of the 118th Convention of the Audio Engineering Society*, 5 2005.

[11] S Olive, J Jackson, A Devantier, and D Hunt. The subjective and objective evaluation of room correction products. In *Proc. of the 127th Convention of the Audio Engineering Society*, 10 2009.

[12] K Lebart and J Boucher. A new method based on spectral subtraction for speech dereverberation. *Acta Acustica united with Acustica*, 87:359–366, 2001.

[13] M Wu and D Wang. A two-stage algorithm for one-microphone reverberant speech enhancement. *IEEE Transactions on Audio, Speech and Language Processing*, 14:774–784, 2006.

[14] A Tsilfidis and J Mourjopoulos. Signal-dependent constraints for perceptually motivated suppression of late reverberation. *Signal Processing*, 90:959–965, 2010.

[15] G A Soulodre. About this dereverberation business: A method for extracting reverberation from audio signals. In *Proc. of the 129th Convention of the Audio Engineering Society*, 11 2010.

[16] A Tsilfidis, E Georganti, and J Mourjopoulos. Binaural extension and performance of single-channel spectral subtraction dereverberation algorithms. In *Proc. of the IEEE ICASSP*, 2011.

[17] A Tsilfidis, E Georganti, E Kokkinis, and J Mourjopoulos. Speech dereverberation based on a recorded handclap. In *Digital Signal Processing Conference (DSP) (submitted)*, Corfu, Greece, 2011.

[18] E Georganti, J Mourjopoulos, and F Jacobsen. Analysis of room transfer function and reverberant signal statistics. In *Proc, of the Acoustics '08 Conference*, Paris, France, 2008.

[19] E Georganti, T Zarouchas, and J Mourjopoulos. Reverberation Analysis via Response and Signal Statistics. In *Proc. of the 128th Convention of the Audio Engineering Society*, London, UK, 2010.

[20] M Barthet and M Sandler. On the effect of reverberation on musical instrument automatic recognition. In *Proc. of the 128th Convention of the Audio Engineering Society*, May 2010.

[21] N Yasuraoka, T Yoshioka, T Nakatani, A Nakamura, and H Okuno. Music dereverberation using harmonic structure source model and Wiener filter. In *Proc. IEEE ICASSP*, 2010.

[22] E Kokkinis and J Mourjopoulos. Unmixing Acoustic Sources in Real Reverberant Environments for Close-Microphone Applications. *Journal of the Audio Engineering Society*, 58(11):1–10, November 2010.

[23] J W. Beauchamp, R C. Maher, and R Brown. Detection of musical pitch from recorded solo performances. In *Proc. of the 94th Convention of the Audio Engineering Society*, 3 1993.

[24] T Wilmering, G Fazekas, and M Sandler. The effects of reverberation on onset detection tasks. *Proc. of the the 128th of the Audio Engineering Society*, 2010.

[25] A. Tsilfidis and J Mourjopoulos. Blind single-channel suppression of late reverberation based on perceptual reverberation modeling. *Journal of the Acoustical Society of America*, 129(2), 2011.

[26] L Wiegrebe and R D Patterson. Quantifying the distortion products generated by amplitude-modulated noise. *Journal of the Acoustical Society of America*, 106(5):2709–2718, 1999.

[27] J Mourjopoulos and J K Hammond. Modelling and enhancement of reverberant speech using an envelope convolution method. In *Proc. of the IEEE ICASSP*, 1983.

[28] D Tsoukalas, J Mourjopoulos, and G Kokkinakis. Speech enhancement based on audible noise suppression. *IEEE Transactions on Speech and Audio Processing*, 5:497–513, 1997.

[29] J Buchholz and J Mourjopoulos. A computational auditory masking model based on signal-dependent compression. I. Model description and performance analysis. *Acta Acustica united with Acustica*, 90:873–886, 2004.

[30] T Zarouchas and J Mourjopoulos. Modeling perceptual effects of reverberation on stereophonic sound reproduction in rooms. *Journal of the Acoustical Society of America*, 126(1):229–242, 2009.

[31] A Tsilfidis, J Mourjopoulos, and D Tsoukalas. Blind estimation and suppression of late reverberation utilizing auditory masking,. In *Proc. of the Hands Free Speech Communication and Microphone Arrays (HSCMA)*, 2008.

[32] V Hansen and G Munch. Making recordings for simulation tests in the Archimedes Project. *Journal of the Audio Engineering Society*, 39(10):768–774, 1991.

[33] J Patynen, V Pulkki, and T Lokki. Anechoic recording system for symphony orchestra. *Acta Acustica United With Acustica*, 94(6):856–865, 2008.