Forum Acusticum 2011 - Aalborg - Denmark

# A generalized framework for binaural spectral subtraction dereverberation

Alexandros Tsilfidis, Eleftheria Georganti, John Mourjopoulos
Audio and Acoustic Technology Group, Department of Electrical and Computer Engineering, University of Patras, 26504, Greece

**Summary**
Adapting single channel dereveverberation techniques to binaural processing is not trivial. Apart from the challenging task of reducing reverberation without introducing audible artifacts, binaural dereverberation methods should also at least preserve the Interaural Time Difference (ITD) and Interaural Level Difference (ILD) cues, since bilateral signal processing affects the source localization. Given that single-channel spectral subtraction is commonly used to suppress late reverberation, here a generalized binaural extension of such methods is presented that utilizes three alternative bilateral gain adaptation schemes. Each algorithm is implemented independently for the left and right ear channel signals resulting to corresponding weighting gains. Then, these gains are combined and different adaptation strategies are investigated. The proposed extensions are tested with various measured Room Impulse Responses and the results reveal the most appropriate binaural extension.

**PACS no. 43.60.Dh, 43.66.Pn**

## 1. Introduction

There are many applications where reverberation is considered as an unwanted distortion deteriorating the quality of acoustic signals. Reducing or completely removing reverberation from audio and speech signals has been a challenging research issue for at least four decades (e.g. [1, 2]). Most recent dereverberation techniques have been developed mainly for speech signals since reverberation (and essentially late reverberation) is known to reduce speech quality and intelligibility and deteriorate the performance of Automatic Speech Recognition (ASR) systems (e.g. [3]). Dereverberation is also important for binaural applications in the context of digital hearing aids, binaural telephony, hands free devices and immersive audio applications (e.g. [4, 5, 6, 7]). However, adapting single or multichannel techniques for binaural processing is not trivial. Apart from the challenging task of reducing reverberation without introducing audible artifacts, binaural dereverberation methods should also at least preserve the Interaural Time Difference (ITD) and Interaural Level Difference (ILD) cues as it has been shown that bilateral signal processing affects the source localization [6].Note that despite the great importance of binaural dereverberation, few studies have been published in the existing literature (e.g [8, 5, 9]).



This work presents a generalized framework for binaural spectral subtraction dereverberation. The proposed approach is based on the binaural extension of state-of-the-art single-channel late reverberation suppression techniques [10, 11, 12, 13] and relies on bilateral gain adaptation [14], a technique which efficiently reduces reverberation and also preserves the binaural localization cues. Significantly, the performance of the proposed framework is investigated here for broadband signals sampled at 44.1 kHz. The objective results show significant reverberation reduction while a subjective test investigates the perceived quality of the dereverberated signals.

## 2. A framework for binaural spectral subtraction dereverberation

### 2.1. Bilateral gain adaptation

Generally speaking, reverberation is a convolutive distortion; however, late reverberation can be considered as an additive degradation with noise-like characteristics [13]. Hence, in the dereverberation context spectral subtraction has been adapted for the suppression of late reverberation. The basic principle of single-channel spectral subtraction dereverberation [10, 11, 12] is to estimate the short time spectrum of the clean signal $S_e(\omega, j)$ by subtracting an estimation of the short time spectrum of late reverberation

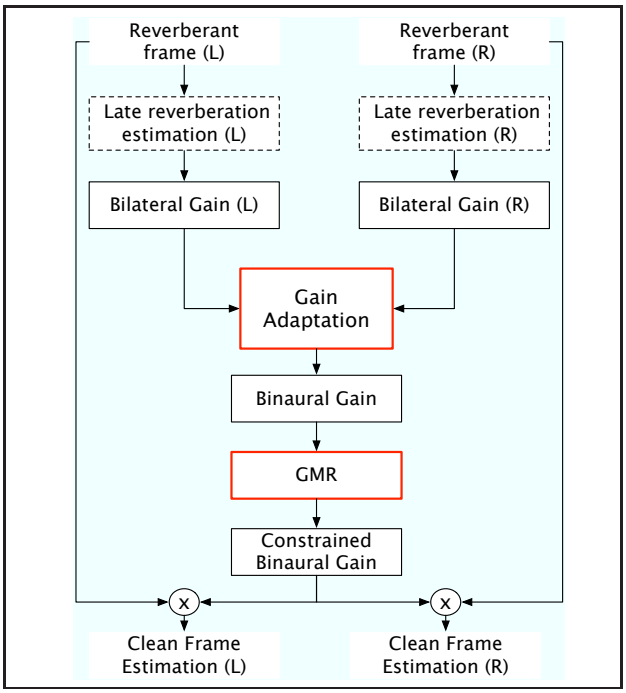


Figure 1. Block diagram of the proposed binaural framework illustrating the gain manipulation steps within a spectral frame (red boxes)

$R(\omega, j)$ from the short time spectrum of the reverberant signal $Y(\omega, j)$:

$$S_e(\omega, j) = Y(\omega, j) - R(\omega, j) \tag{1}$$

where $\omega$ and $j$ are the frequency bin and time index respectively. Following an alternative formulation, the estimation of the short time spectrum of the clean signal can be derived by applying appropriate weighting gains $G(\omega, j)$ in the short time spectrum of the reverberant signal i.e.:

$$S_e(\omega, j) = G(\omega, j)Y(\omega, j) \tag{2}$$

where

$$G(\omega, j) = \frac{Y(\omega, j) - R(\omega, j)}{Y(\omega, j)} \tag{3}$$

Therefore, the dereverberation problem is deduced in an estimation of the late reverberation short time spectrum.

When adapting the above principle in the binaural context, bilateral processing must be avoided in order to preserve the binaural cues [14]. For this, in [15] the authors propose the implementation of a Delay and Sum Beamformer (DSB). However, in binaural applications, the time delay between the left and right channels of the reverberant signal is limited by the width of the human head. Therefore, for the proposed framework the above time delay is assumed shorter than the length of a typical analysis window and the DSB stage is omitted. Moreover, here the binaural processing is realized through bilateral gain adaptation. The late reverberation spectrum has been estimated independently for the left and right ear channel signals resulting to the corresponding weighting gains $G_l(\omega, j)$ and $G_r(\omega, j)$ (as explained in Eq. 3). These gains may be combined into the gain adaptation block shown in Fig. 1 following three alternative adaptation strategies:

(i) The final gain is derived as the maximum of the left and right channel weighting gains:

$$G(\omega, j) = max(G_l(\omega, j), G_r(\omega, j)) \tag{4}$$

This approach (maxGain) achieves moderate late reverberation suppression, but it is also less likely to produce overestimation artifacts.

(ii) The final gain is derived as the average of the left and right channel weighting gains:

$$G(\omega, j) = \frac{[G_l(\omega, j) + G_r(\omega, j)]}{2} \tag{5}$$

This gain adaptation strategy (avgGain) compensates equally for the contribution of the left and right channels.

(iii) The final gain is derived as the minimum of the left and right channel weighting gains:

$$G(\omega, j) = min(G_l(\omega, j), G_r(\omega, j)) \tag{6}$$

The above adaptation technique (minGain) results to maximum reverberation attenuation but the final estimation may be susceptible to overestimation artifacts.

### 2.2. Gain Magnitude Regularization

After the derivation of the adapted gain, a Gain Magnitude Regularization (GMR) technique is applied (see Fig. 1) and the purpose of this step is twofold. Firstly, the GMR has been proved to be a low-complexity approach reducing annoying musical noise artifacts [16, 17]. Furthermore, the GMR is utilized in order to constrain the suppression and prevent from overestimation errors. An overestimation of the late reverberation is less likely to happen in high SRR spectral regions such as signal steady states [18] contrary to low SRR regions. Therefore a low SRR detector is employed [15] and GMR is applied only on the lower gain parts. Hence, the new constrained gain $G'(\omega, j)$ is derived as:

$$G'(\omega, j) = \begin{cases} \dfrac{G(\omega, j) - \theta}{r} + \theta & \text{when } \zeta < \zeta_{th} \\ & \text{and } G(\omega, j) < \theta \\ G(\omega, j) & \text{otherwise} \end{cases}$$

and

$$\zeta = \frac{\sum_{\omega=1}^{\Omega} G(\omega, j)|Y(\omega, j)|^2}{\sum_{\omega=1}^{\Omega} |Y(\omega, j)|^2} \tag{7}$$

Table I. Properties of the BRIRs

| Room | Dist.(m) | Azim. | RT(s) | IC skew |
|---|---|---|---|---|
| Meeting Room | 2.25 | 0° | 0.24 | 1.167 |
| Lecture Hall | 5.56 | 0° | 0.79 | 1.302 |
| Cafeteria | 1.29 | 40° | 1.25 | 1.210 |

Table II. Analysis Parameters

| Meth. | Frame Length | Zero pad. | Frame Overlapp |
|---|---|---|---|
| LB | 2048 | 1024 | 0.5 |
| WW | 8192 | 4096 | 0.25 |
| FK | 8192 | 4096 | 0.25 |

where $\theta$ is the threshold for applying the gain constraints, $r$ is the regularization ratio, $\zeta$ is the power ratio between the enhanced and the reference signal, $\zeta_{th}$ the threshold of the low SRR detector and $\Omega$ is the frame size.

## 3. Tests and results

In [14] the proposed approach has been verified for speech signals sampled at 16 kHz. Here, the method is applied in broadband signals sampled at 44.1 kHz, where longer analysis windows are involved. The proposed framework has been implemented in three single-channel state-of-the-art spectral subtraction algorithms originally proposed by Lebart et al. (LB), Wu and Wang (WW) and Furuya and Kataoka (FK) [10, 11, 12] (see also the Appendix).

A database consisting of 18 anechoic speech samples uttered from 11 male and female speakers has been employed. The reverberant samples were produced by convolving the anechoic signals with measured Binaural Room Impulse Responses (BRIRs), taken from the Aachen and the Oldenburg databases ([15, 19]) and their properties are shown in Table I. In the last column of the table, the Interaural Coherence Skewness (IC skew) of the BRIRs is also given. This measure has been proposed in [20] as a measure of the diffuseness of the reverberant field. Note that the original dereverberation methods (LB, WW and FK) were optimized for lower signal resolutions and here the authors conducted unofficial experiments to choose the optimal values for the analysis parameters. The STFT analysis parameters (total frame length, zero padding and frame overlap) for each tested method are detailed in Table II, the $\theta$ and $\zeta_{th}$ values of the GMR step were set at 0.15 and 0.8 respectively while the regularization ratio $r$ was 4. All parameter values that are not detailed here were set according to the values proposed by the authors of the original works. In addition, for the FK and LB techniques, two additional relaxation criteria were imposed [18] as they were previously found by the authors to have advantageous effects on the performance.

The relative improvement achieved by the tested methods has been evaluated in terms of segmental Signal to Reverberation Ratio (SRR) and Bark Spectral Distortion (BSD). The SRR measure is the equivalent to the well-known Signal to Noise Ratio (SNR) when reverberation is considered as an additive noise [14] and quantifies the reverberation reduction. Hence, the SRR over the estimated clean signal and the clean signal and the SRR over the reverberant and the clean signal were calculated and their difference was derived as:

$$\Delta SRR = SRR_{estimate} - SRR_{reverberant}. \quad (8)$$

Furthermore, the BSD is a perceptually motivated measure of spectral distortion [15] and evaluates the overall distortion by calculating the distance between loudness vectors of the reference and the processed speech. Again, the BSD difference between the BSD over the estimated clean signal and the clean signal and the BSD over the reverberant and the clean signal was calculated, noting that in this case negative BSD values denote the relative improvement. In Figures 2, 3 and 4 the SRR and BSD results obtained from the LB, WW and FK methods are presented. In Fig. 2 (a) the SRR improvement achieved by the LB method is presented an it is more pronounced in the Lecture Hall where a longer source-receiver distance is employed. The same applies for the WW and FK methods as seen in Fig. 3 (a) and Fig. 4 (a) respectively. Moreover, in most cases the minGain adaptation scheme seems to suppress more reverberation, noticing also that all tested gain adaptation techniques achieve significant reverberation reduction. On the other hand, when looking at the BSD results (see Figures 2 (b), 3 (b) and 4 (b)) it seems that all methods reduce the BSD when compared to the reverberant signal in the Meeting Room and the Lecture Hall but fail in the case of the Cafeteria. In this case, no relative improvement was noticed probably due to the longer RT and to the shorter source-receiver distance. The use of the DSB seems to produce slightly better BSD results in the Meeting Room while the minGain adaptation technique seems to perform better in the Lecture Hall. The greater improvement in terms of BSD has been obtained by the FK method, as seen in Fig. 4 (b).

In order to evaluate the subjective performance of the presented algorithms, a modified version of the ITU P.835 test has been conducted [21]. The subjects were asked to rate in a 1-5 scale (i) the speech signal naturalness (Sp. Nat), (ii) the reverberation intrusiveness (Rev. Intr.) and (iii) the overall signal quality (Ov. Qual.) [22, 23]. For the subjective test, four phrases from two male and two female speakers along with three BRIRs measured in a Stairway Hall ($RT_{60} = 0.69$ sec) at a source-receiver distance of $3m$

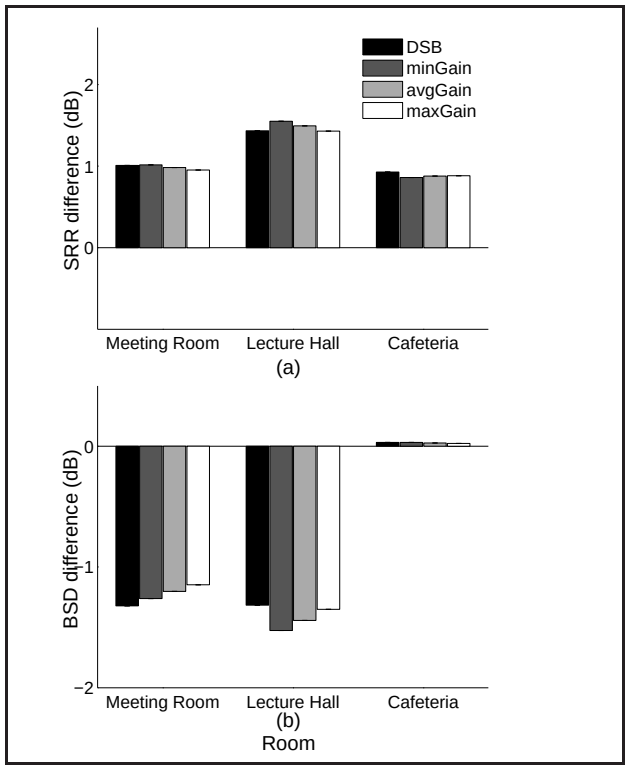


Figure 2. Objective results for the LB method: (a) SRR difference and (b) BSD difference

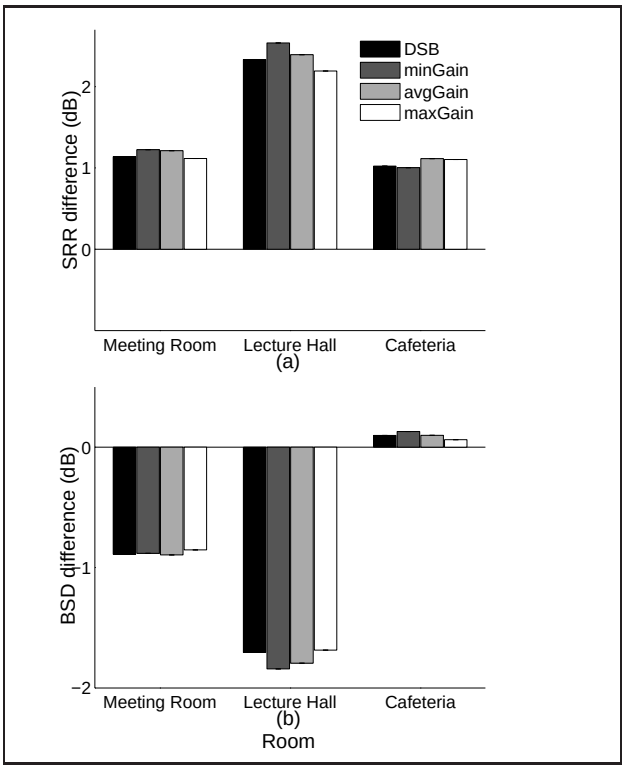


Figure 4. Objective results for the FK method: (a) SRR difference and (b) BSD difference

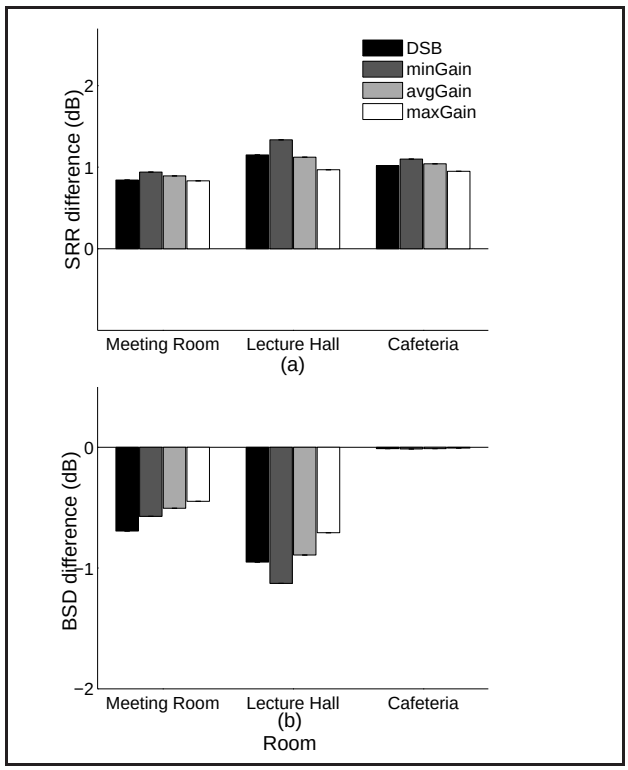


Figure 3. Objective results for the WW method: (a) SRR difference and (b) BSD difference

and azimuth angles of $0, 45$ and $90^\circ$ were used [15]. In order to reduce the experimental conditions the authors conducted unofficial listening tests to choose the optimum gain adaptation scheme for each dereverberation method. Hence, the avgGain adaptation has been chosen for the LB and WW methods while the maxGain has been used for the FK method. Twenty self-reported normal hearing subjects participated in the tests and a training session preceded the formal experiment.

Fig. 5 presents the subjective scores in terms of speech naturalness, reverberation intrusiveness and overall signal quality for the proposed binaural dereverberation techniques. The results were subjected to an analysis of variance (ANOVA) and a highly significant effect for the tested method was revealed for the speech naturalness ($F(3,228)=112.7$, $p<0.001$), for the reverberation intrusiveness ($F(3,228)=62.1$, $p<0.001$) and for the overall quality ($F(3,228)=38.8$, $p<0.001$). No significant effect was found for the tested azimuth angles. Following the ANOVA multiple Tukey' s, HSD tests were made to reveal significant differences between algorithms. In all cases, listeners rated that the unprocessed reverberant signals were significantly more natural than the dereverberated signals ($p<0.001$). This was due to the artifacts introduced from the dereverberation processing. On the other hand the FK method performed significantly worse than the other two methods in terms of speech naturalness. No significant difference was noticed between the LB and WW methods ($p>0.05$).

Furthermore, the three dereverberation methods have significantly reduced the reverberation intrusive-

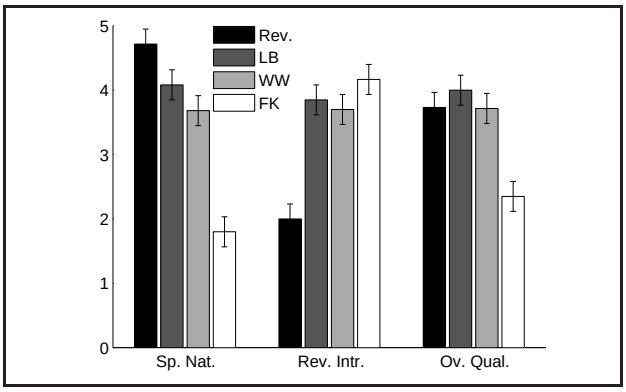


Figure 5. Subjective results for the tested methods evaluating the speech naturalness, the reverberation intrusiveness and the overall signal quality

ness ($p<0.001$). The FK method method performed significantly better than the WW method ($p<0.05$) in terms of subjective reverberation suppression. However, no significant difference between the FK and the LB or the LB and WW methods was found ($p>0.05$). Finally, the LB, the WW methods and the reverberant signals were rated significantly better in terms of overall quality than the FK method ($p<0.001$), but no significant quality difference was found between the LB method, the WW method and the reverberant signals ($p>0.05$).

From the objective and subjective results it appears that all methods significantly suppress reverberation, but the introduced processing artifacts reduce the naturalness of the speech signals. The FK method achieves greater reverberation suppression than the LB and WW methods; however, it appears that the produced signals are more degraded. On the other hand, the LB and WW dereverberation methods moderately reduce the reverberation but they preserve the subjective signal quality.

## 4. Conclusions

A generalized framework for binaural spectral subtraction dereverberation has been presented. The presented framework adapts single-channel dereverberation methods in the binaural scenario. The proposed technique preserves the binaural cues through bilateral gain adaptation and three alternative gain adaptation schemes were investigated. The objective and subjective results reveal that the LB and WW methods utilizing both the avgGain adaptation scheme achieve significant reverberation suppression without compromising the signal's quality. Nevertheless, the processed signals are somewhat less natural than the unprocessed reverberant speech. Hence, it appears that there is a practical limit in the potential performance of such an approach and further improvements can be achieved by developing novel signal processing algorithms that take into account many aspects of the binaural hearing mechanism.

**Acknowledgement**

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. MICRO2-38/E-II-A.

## Appendix

For the estimation of the late reverberation short time spectrum, Lebart et al. [10] proposed a method (LB) based on the RIR modeling. The short time spectral magnitude of the reverberation is estimated as:

$$|R(\omega,j)| = \frac{1}{\sqrt{|SNR_{pri}(\omega,j)|+1}}|Y(\omega,j)|$$

where $|SNR_{pri}(\omega,j)|$ is the a priori Signal to Noise Ratio that can be approximated by a moving average of the a posteriori Signal to Noise Ratio $|SNR_{post}(\omega,j)|$ in each frame:

$$|SNR_{pri}(\omega,j)| = \beta|SNR_{pri}(\omega,j-1)|+ (1-\beta)max(0,|SNR_{post}(\omega,j)|)$$

where $\beta$ is a constant taking values close to 1.

The method proposed by Wu and Wang [11] (WW) is motivated by the observation that the smearing effect of late reflections produces a smoothing of the signal spectrum in the time domain. Hence, the late reverberation power spectrum is considered a smoothed and shifted version of the power spectrum of the reverberant speech:

$$|R(\omega,j)|^2 = \gamma w(j-\rho) * |Y(\omega,j)|^2$$

where $\rho$ is a frame delay, $\gamma$ a scaling factor. The term $w(j)$ represents an assymetrical smoothing function given by the Rayleigh distribution:

$$w(j) = \begin{cases} \frac{j+\alpha}{\alpha^2}\exp\left(\frac{-(j+\alpha)^2}{2\alpha^2}\right) & \text{if } j < -\alpha \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha$ represents a constant number of frames.

Alternatively, Furuya and Kataoka [12] proposed a method (FK) where the short time power spectrum of late reverberation in each frame can be estimated as the sum of filtered versions of the previous frames of the reverberant signal's short time power spectrum:

$$|R(\omega,j)|^2 = \sum_{l=1}^{K}|a_{late}(\omega,j)|^2|Y(\omega,j-l)|^2$$

where $K$ is the number of frames that corresponds to an estimation of the $RT_{60}$ and $a_{late}(\omega, j)$ are the coefficients of late reverberation. The coefficients of late reverberation are derived from:

$$a_{late}(\omega, j) = E\left\{ \frac{Y(\omega, j)Y^*(\omega, j-l)}{|Y(\omega, j-l)|^2} \right\}$$

## References

[1] J L Flanagan and Lummis R.C. Signal processing to reduce multipath distortion in small rooms. *Journal of the Acoustical Society of America*, 47:1475–1481, 1970.

[2] O M M Mitchell and D A Berkley. Reduction of long time reverberation by a center clipping process. *Journal of the Acoustical Society of America*, 47:84, 1970.

[3] R P Lippmann. Speech recognition by machines and humans. *Speech Communication*, 22(1):1–15, July 1997.

[4] T Wittkop and V Hohmann. Strategy-selective noise reduction for binaural digital hearing aids. *Speech Communication*, 39:111–138, 2003.

[5] H W Löllmann and P Vary. Low delay noise reduction and dereverberation for hearing aids. *EURASIP Journal on Advances in Signal Processing*, 2009:1–9, 2009.

[6] V Hamacher, J Chalupper, J Eggers, E Fischer, U Kornagel, H Puder, and U Rass. Signal Processing in High-End Hearing Aids: State of the Art, Challenges, and Future Trends. *EURASIP Journal on Applied Signal Processing*, pages 2915–2929, 2005.

[7] Yiteng Huang, Jingdong Chen, and J. Benesty. Immersive audio schemes. *IEEE Signal Processing Magazine*, 28(1):20 –32, 2011.

[8] J.-H. Lee, S.-H. Oh, and Lee S.-Y. Binaural semi-blind dereverberation of noisy convoluted speech signals. *Neurocomputing*, 72:636–642, 2008.

[9] M Jeub and P Vary. Binaural dereverberation based on a dual-channel Wiener filter with optimized noise field coherence. In *Proc. of the IEEE ICASSP*, pages 4710–4713, 2010.

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

[11] 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.

[12] K Furuya and A Kataoka. Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction. *IEEE Transactions on Audio, Speech and Language Processing*, 15:1571–1579, 2007.

[13] 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(3):1439–1451, 2011.

[14] 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.

[15] M Jeub, M Schafer, T Esch, and P Vary. Model-Based Dereverberation Preserving Binaural Cues. *IEEE Transactions on Audio, Speech, and Language Processing*, 18:1732–1745, 2010.

[16] A Tsilfidis, K E Kokkinis, and J Mourjopoulos. Suppression of late reverberation at multiple speaker positions utilizing a single impulse response measurement. In *Forum Acusticum*, Aalborg, Denmark, 2011.

[17] E Kokkinis, A Tsilfidis, E Georganti, and J. Mourjopoulos. Joint noise and reverberation suppression for speech applications. In *Proc. of the 130th Convention of the Audio Engineering Society*, 2011.

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

[19] H Kayser, S D Ewert, J Anemuller, T Rohdenburg, V Hohhmann, and B Kollmeier. Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses. *EURASIP Journal on Applied Signal Processing*, 2009:1–10, 2009.

[20] E Georganti, A Tsilfidis, and J Mourjopoulos. Statistical Analysis of Binaural Room Impulse Responses. In *Proc. of the 130th Convention of the Audio Engineering Society*, May 2011.

[21] International Telecommunications Union (ITU-T, P.835), Geneva, Switzerland. *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*, 2003.

[22] Yi Hu and P C Loizou. Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, 49(7):588–601, 2007.

[23] T.H. Falk, Chenxi Zheng, and Wai-Yip Chan. A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(7):1766 –1774, 2010.