BINAURAL DEREVERBERATION BASED ON A DUAL-CHANNEL WIENER FILTER WITH OPTIMIZED NOISE FIELD COHERENCE

Marco Jeub and Peter Vary

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
{jeub,vary}@ind.rwth-aachen.de

ABSTRACT

In this paper a novel speech enhancement algorithm for binaural dereverberation is proposed. It is based on a multichannel Wiener filter approach, which is optimized for the application to digital hearing aids and binaural telephony headsets. This is mainly done by two different modifications. First, an optimized model for the binaural coherence which takes the shadowing effects of the head into account. Second, a binaural input-output structure which does not affect the most important binaural cues, i.e., interaural time difference (ITD) and interaural level difference (ILD), and hence, keeps the localization ability. Evaluations with measured binaural room impulse responses (BRIR) show that this approach is capable of reducing reverberation especially in highly reverberant environments.

Index Terms— Binaural, dereverberation, binaural cue preservation, coherence, speech enhancement

1. INTRODUCTION

The effects of room reverberation on the human perception and on speech recognition systems have been studied for many years. It is well-known that the speech quality and intelligibility is degraded by room reverberation. The adverse effects are noticeable especially in the context of digital hearing aids and binaural telephony headsets. Several contributions have been made in the past to reduce the effects in reverberant rooms and hence to increase the intelligibility.

An overview of existing methods as well as a low-delay single-channel dereverberation concept can be found in [1] and the references therein. Most of these techniques are suitable for bilateral processing only, which means that each side of the device is performing monaural enhancement without taking spatial information into account. Here, we focus on binaural algorithms, since a full-data link between both sides of the headset and in future between both sides of the hearing aid can be assumed.

This paper proposes a new postfilter concept which is based on the well-known multichannel Wiener filter approach. We employ an improved model for describing the noise field coherence which occur in reverberant rooms, taking the shadowing effect of the head into account (binaural coherence). Additionally, we use an algorithm structure which does not affect the binaural cues interaural time difference (ITD) and interaural level difference (ILD). By doing so, the effects of reverberation can be reduced without affecting the sound localization ability. The proposed technique has a low computational complexity and can be easily applied to real-time applications.

2. BINAURAL COHERENCE MODEL

The effects of head shadowing on the input signals to both ears will be described in terms of the spectral coherence. The coherence between the two signals $x_{l|r}(k)$ is defined as

$$\Gamma_{x_l x_r}(\Omega) = \frac{\Phi_{x_l x_r}(e^{j\Omega})}{\sqrt{\Phi_{x_r x_r}(e^{j\Omega}) \cdot \Phi_{x_l x_l}(e^{j\Omega})}}, \tag{1}$$

where $\Phi_{x_l x_l}(e^{j\Omega})$ and $\Phi_{x_r x_r}(e^{j\Omega})$ represent the auto-spectral densities (APSD) of $x_l(k)$ and $x_r(k)$ respectively. The cross-power spectral density (CPSD) between $x_l(k)$ and $x_r(k)$ is denoted by $\Phi_{x_l x_r}(e^{j\Omega})$. The coherence between two microphones of an ideal spherically isotropic (diffuse) noise field can be expressed as [2]

$$\Gamma^{(\text{diff})}_{x_l x_r}(f) = \text{sinc}\left(\frac{2\pi f d_{mic}}{c}\right), \tag{2}$$

with distance d_{mic} between two omnidirectional microphones with a line-of-sight and frequency f. The noise field in a reverberant room can be approximated by a diffuse noise field, cf., [2, 3].

It is well-known that the coherence between two microphones changes (compared to Eq. 2) when an object is in the line-of-sight. This has been shown theoretically and in experiments on measured data with a dummy head in a crowded cafeteria in [4]. Investigations in reverberant rooms have recently been published in [3].

This subsection describes an improved coherence model for a diffuse noise field compared to Eq. 2 which takes the shadowing effect of the head into account. In order to describe the complex geometry of the human head, a simplified configuration with two circular plates (S_l and S_r) as depicted in Fig. 1 will be assumed in the following [4]. The corresponding coherence of the 3D sound field can now be calculated by integration over all azimuth and elevation angles (θ, φ) according to

$$\Gamma^{(\text{head})}_{x_l x_r}(f) = \frac{\left| \int\limits_{\varphi=0}^{2\pi} \int\limits_{\theta=0}^{\pi} H_l(f,\theta,\varphi) H_r^*(f,\theta,\varphi) \sin\theta d\theta d\varphi \right|}{\sqrt{\int\limits_0^{2\pi}\int\limits_0^{\pi} |H_l(f,\theta,\varphi)|^2 \sin\theta d\theta d\varphi \int\limits_0^{2\pi}\int\limits_0^{\pi} |H_r(f,\theta,\varphi)|^2 \sin\theta d\theta d\varphi}}. \tag{3}$$

Here, $\{\cdot\}^*$ denotes the complex conjugate and H_l and H_r represent the transfer functions between a punctual sound source at the position $\underline{r}_q$ and the two microphones M_l and M_r. It is also assumed that the distance of the sound source is large compared to the microphone distance (far-field assumption).

This work was supported by Infineon Technologies, Sophia-Antipolis, France.

978-1-4244-4296-6/10/$25.00 ©2010 IEEE

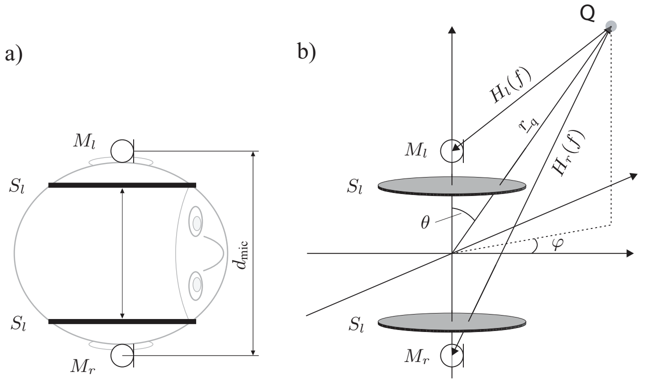

Fig. 1. Simplified geometrical model of the human head: a) head with the two plates, b) geometrical model.

A solution of this equation invokes the use of the Helmholtz-Kirchhoff Integral theorem and is described in greater detail in [4]. Since this requires the calculation of the integrals over every angle θ and φ, a simple curve-fitting is proposed as an efficient alternative. Based on the sum of Gaussians, an approximation of the noise field coherence can be expressed by

$$\hat{\Gamma}^{(\mathrm{head})}_{x_l x_r}(f) = \sum_{p=1}^{P} a_p \cdot \exp\left(-\left(\frac{f-b_p}{c_p}\right)^2\right), \tag{4}$$

with constants a_p, b_p, c_p and the model order P. Since a natural ear spacing of $d_{mic} = 0.15 - 0.17\,\mathrm{m}$ is assumed, this coherence function needs to be evaluated only once. The coefficients for $d_{mic} = 0.17\,\mathrm{m}$ and a mixture of $P = 3$ Gaussians are calculated using the MATLAB Curve Fitting Toolbox and given in Table 1. The root mean squared error between the solution of Eq. 3 and the approximation in Eq. 4 is RMSE $= 2.4 \cdot 10^{-3}$ in the frequency range $1 - 48000\,\mathrm{Hz}$. Figure 2 shows the corresponding curves for two microphones at a distance of $d_{mic} = 0.17\,\mathrm{m}$. The functions are plotted as the squared magnitudes of the coherence function $\Gamma^2_{x_l x_r}(\Omega)$ bounded above 0.99. The theoretical curves represent the ideal diffuse sound field without head (Eq. 2) and the sound field with head shadowing (Eq. 4). The measured curves have been obtained by a set of measured binaural room impulse responses of a lecture room, with and without a dummy head. It could be assumed that the influence of the head can be modeled by scaling d_{mic} of the ideal diffuse coherence in Eq. 2. However, it turned out in several experiments that this does not lead to a sufficient solution compared to the model of Eq. 4.

p	a_p	b_p	c_p
1	1	18.97	291.1
2	$14.5 \cdot 10^{-3}$	875.2	105.7
3	$2.38 \cdot 10^{-3}$	1371	151.5

Table 1. Coefficients of the binaural coherence model for $d_{mic} = 0.17\,\mathrm{m}$ using a non-linear least-squares fitting.

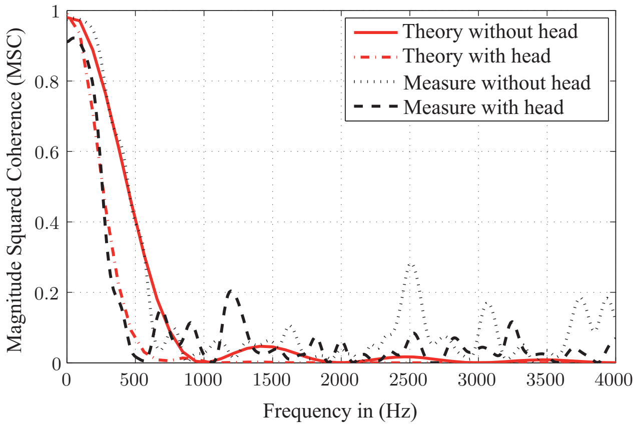

Fig. 2. Magnitude squared coherence of ideal diffuse sound field and shadowing influence. Plotted are the theoretical curves and results from measurements in a reverberant environment ($\mathrm{T}_{60} = 0.72\,\mathrm{s}$).

3. BINAURAL CUE PRESERVING DEREVERBERATION

The considered speech dereverberation algorithm is realized by short-term spectral weighting using the weighted overlap-add method. For the transformation into the frequency domain, the disturbed input signals $x_{l|r}(k)$ are first segmented into overlapping frames of length L. After windowing (e.g., applying a Hann window), these frames are transformed via Fast Fourier Transform (FFT) of length M. At discrete frequency bins μ, the distorted signals for right and left channel are $X_{l|r}(\lambda, \mu)$. Late reverberation is assumed as uncorrelated noise. The enhanced spectra $\hat{S}_{l|r}(\lambda, \mu)$ can be obtained by multiplying the coefficients $X_{l|r}(\lambda, \mu)$ with weighting gains $G(\lambda, \mu)$

$$\hat{S}_l(\lambda, \mu) = X_l(\lambda, \mu) \cdot G(\lambda, \mu), \tag{5a}$$

$$\hat{S}_r(\lambda, \mu) = X_r(\lambda, \mu) \cdot G(\lambda, \mu). \tag{5b}$$

The enhanced time domain signals $\hat{s}_{l|r}(k)$ are obtained by using the Inverse Fast Fourier Transform (IFFT) and overlap-add. Applying different weighting gains to each channel can cause unwanted modifications in the spatial impression. The main idea here is to apply the same weighting gains to both channels which ensures no modification of the binaural ILD cue, cf., [5]. The ITD is also not affected since the algorithm keeps the phase of the input signals. The calculation of the weighting gains $G(\lambda, \mu)$ has to be performed on the time-aligned input signals. By applying the gains to the non time-aligned spectra, we ensure that the algorithm works effectively for different azimuth angles as shown later. Since the maximum time difference between the right and left channel is limited by the head geometry (c.f., [6]), the maximum ITD range of ±750 µs is small compared to a typical frame length of $10 - 30\,\mathrm{ms}$. The time-alignment is performed by means of the Generalized Cross-Correlation with Phase Transform (GCC-PHAT), c.f., [7].

3.1. Dereverberation based on Noise Field Coherence

A common framework for multichannel speech enhancement is based on the optimal Minimum Mean Square Error (MMSE) criterion. It turns out that the optimal weighting vector is given by the multichannel Wiener solution. Furthermore, it can be shown that the resulting filter can be decomposed into a Minimum Variance

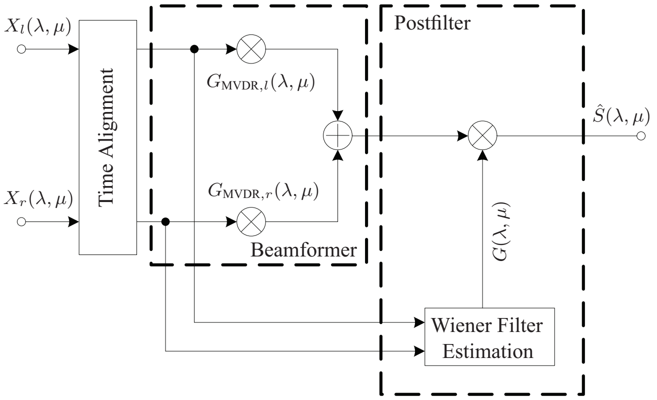

Fig. 3. Frequency domain representation of the conventional beamformer and Wiener postfilter with dual-channel input and single-channel output.

Distortionless Response (MVDR) beamformer and a single-channel Wiener filter operating on the beamformer output [8]. The resulting gains are calculated by

$$G_{\text{opt}}(\lambda,\mu) = \underbrace{\frac{\Phi_{ss}(\lambda,\mu)}{\Phi_{ss}(\lambda,\mu)+\Phi_{nn}(\lambda,\mu)}}_{\textbf{Wiener postfilter}} \cdot \underbrace{\frac{\underline{\Phi}_{nn}^{-1}(\lambda,\mu)\,\mathbf{d}}{\mathbf{d}^H\,\underline{\Phi}_{nn}^{-1}(\lambda,\mu)\,\mathbf{d}}}_{\textbf{MVDR beamformer}}, \qquad (6)$$

where $\Phi_{ss}(\lambda,\mu)$ denotes the APSD of the original (undisturbed) signal, $\Phi_{nn}(\lambda,\mu)$ the APSD of the additive noise component and $\underline{\Phi}_{nn}(\lambda,\mu)$ the noise APSD matrix. The propagation vector $\mathbf{d}$ describes the acoustic path from the desired source to the microphones.

The corresponding block diagram for the dual-channel case is depicted in Fig. 3. The time-aligned input spectra are multiplied with the MVDR gains $\mathbf{G}_{\text{MVDR}}(\lambda,\mu)$ and added afterwards. As mentioned above, the Wiener filter is estimated on the beamformer input signals and applied to the single-channel output of the beamformer. In the following we focus on the estimation of the Wiener postfilter gains $G(\lambda,\mu)$.

In order to estimate the optimal postfilter coefficients, several approaches have been presented in the past. They all have in common that the estimation procedure is optimized for a specific noise field. A well-known technique by Zelinski assumes a perfectly incoherent noise field and hence, uncorrelated noise at different sensors [9]. Since this assumption does not hold in real noise fields, a further approach was presented by McCowan in [10] who proposed to use the coherence model of a spherically isotropic (diffuse) noise field. Since the head-shadowing has a severe impact on the coherence between the microphone signals, we propose to use the coherence model for a binaural spherically isotropic noise field in the following.

The calculation of the weighting gains $G(\lambda,\mu)$ comprises an estimation of the APSDs $\Phi_{x_lx_l}(\lambda,\mu)$, $\Phi_{x_rx_r}(\lambda,\mu)$ and CPSD $\Phi_{x_lx_r}(\lambda,\mu)$ of the two time-aligned input channels. This is performed by means of an recursive periodogram approach according to

$$\begin{aligned}\hat{\Phi}_{x_lx_l|x_rx_r}(\lambda,\mu) &= \alpha\,\hat{\Phi}_{x_lx_l|x_rx_r}(\lambda-1,\mu)\\ &\quad+\,(1-\alpha)\,|X_{l|r}(\lambda,\mu)|^2,\end{aligned} \qquad (7)$$

$$\begin{aligned}\hat{\Phi}_{x_lx_r}(\lambda,\mu) &= \alpha\,\hat{\Phi}_{x_lx_r}(\lambda-1,\mu)\\ &\quad+\,(1-\alpha)\,X_l(\lambda,\mu)\cdot X_r^*(\lambda,\mu),\end{aligned} \qquad (8)$$

with smoothing factor $0\le\alpha\le 1$ and periodograms $|X_{l|r}(\lambda,\mu)|^2$. Afterwards, an estimate of the original (undistorted) signal APSD is calculated by [10]

$$\hat{\Phi}_{ss}(\lambda,\mu) = \frac{\text{Re}\{\hat{\Phi}_{x_lx_r}(\lambda,\mu)\} - \frac{1}{2}\text{Re}\{\Gamma_{x_lx_r}(\Omega)\}\Big(\hat{\Phi}_{x_lx_l}(\lambda,\mu)+\hat{\Phi}_{x_rx_r}(\lambda,\mu)\Big)}{1-\text{Re}\{\Gamma_{x_lx_r}(\Omega)\}}. \qquad (9)$$

The function $\text{Re}\{\cdot\}$ returns the real part of its argument. Since the estimate of the signal APSD may not be negative or singular, a maximum threshold Γ_{max} for the coherence function has to be applied to ensure that $1-\text{Re}\{\Gamma_{x_lx_r}(\Omega)\}>0$ holds for the denominator.

Taking into account the improved binaural coherence model of Eq. 4 and the estimate of Eq. 9, the resulting spectral weights of the Wiener postfilter can now be calculated by

$$G(\lambda,\mu) = \frac{\hat{\Phi}_{ss}(\lambda,\mu)}{\frac{1}{2}\cdot\Big(\hat{\Phi}_{x_lx_l}(\lambda,\mu)+\hat{\Phi}_{x_rx_r}(\lambda,\mu)\Big)}. \qquad (10)$$

The spectral weights are further confined by a lower threshold G_{min} in order to control the trade-off between the amount of dereverberation and musical tones. For the experiments, we use the weighting gains of the dual-channel algorithm by Allen et al. [11] as a reference. This algorithm is related to the aforementioned method since it uses directly the estimated coherence. The corresponding gains are calculated by

$$G_{\text{allen}}(\lambda,\mu) = \frac{|\hat{\Phi}_{x_lx_r}(\lambda,\mu)|}{\sqrt{\hat{\Phi}_{x_lx_l}(\lambda,\mu)\cdot\hat{\Phi}_{x_rx_r}(\lambda,\mu)}} \qquad (11)$$

and applied to each channel according to the proposed binaural dereverberation concept.

4. EXPERIMENTS AND RESULTS

The experiments have been performed with measured binaural room impulse responses taken from the Aachen Impulse Response (AIR) database [3]. All selected BRIRs are measured with a dummy head in different acoustical environments at a microphone distance $d_{mic} = 0.17\,\text{m}$. The BRIR of an office and lecture room have been convolved with utterances from the NTT database. Reverberation times RT_{60}, loudspeaker-microphone distances d_{LM} and azimuth angle θ between head and loudspeaker are as follows

- Office room: $\text{RT}_{60} = 0.37\,\text{s}$, $d_{\text{LM}} = 1\,\text{m}$, $\theta = 90°$ (frontal),
- Lecture room: $\text{RT}_{60} = 0.72\,\text{s}$, $d_{\text{LM}} = 5.5\,\text{m}$, $\theta = 90°$.

Three different algorithms are compared. The postfilter of Sec. 3.1 assuming the ideal diffuse noise field of Eq. 2 is named *Diffuse*. The same postfilter taking the new coherence model of Eq. 4 into account is termed as *Proposed*. Additionally, the binaural version of the Allen algorithm with the gain function given by Eq. 11 is used (*Allen*). The experiments are performed by means of a non-intrusive measurement based on the Speech to Reverberation Modulation energy Ratio (SRMR) [12]. It is calculated by means of a gammatone filterbank analysis of temporal envelopes of the speech signal. In order to rate the amount of speech distortion, we use the Bark Spectral Distortion (BSD) measure. Further simulation parameters are listed in Table 2. For each channel, the measurements are calculated separately and averaged. The ΔSRMR gives the enhancement compared to the reverberant speech, averaged over all dereverberated files as listed in Table 3a).

Parameter	Value
Sampling frequency	$f_s = 16\,\text{kHz}$
Smoothing factor	$\alpha = 0.8$
Frame length, FFT length	$L = 256$, $M = 256$
Frame overlap	50 % overlap (Hann window)
Coherence threshold	$\Gamma_{max} = 0.9$
Gain factor threshold	$G_{min} = 0.2$

Table 2. Main simulation parameters.

	ΔSRMR	
a)	Office room	Lecture room
Postfilter (Allen, [11])	+1.02	+1.52
Postfilter (Diffuse, [10])	+1.07	+1.81
Postfilter (Proposed)	+1.77	+2.87

	BSD	
b)	Office room	Lecture room
Reverberant speech	0.17	0.88
Postfilter (Allen, [11])	0.12	0.43
Postfilter (Diffuse, [10])	0.11	0.37
Postfilter (Proposed)	0.13	0.39

Table 3. Evaluation results for different postfilter techniques.

The results in terms of speech distortion can be found in Table 3b). It can be seen that in terms of the SRMR measure, the proposed postfilter shows the highest amount of dereverberation. The more reverberant the room, the more enhancement can be obtained. The BSD improvement is approximately the same for the ideal diffuse and the proposed coherence model. In all experiments, the Allen algorithm shows the lowest amount of enhancement.

In a further experiment, we show the need for the time-alignment block. Figure 4 depicts the dereverberation performance in terms of the SRMR in dependency of the azimuth angle. The corresponding BRIRs have been measured in a stairway hall. It can be seen, that without the time-alignment, a sufficient enhancement can only be obtained in the range $0^\circ \leq \theta \leq 30^\circ$. The time-alignment ensures a similar dereverberation performance over the entire azimuth range.

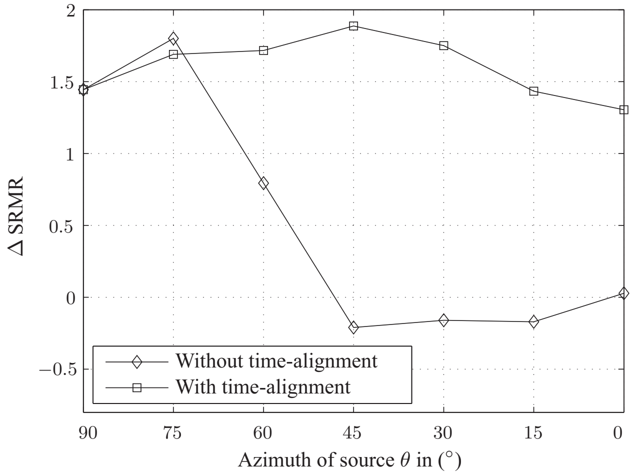

Fig. 4. Influence of time-alignment on dereverberation performance.

5. CONCLUSIONS

This paper proposes a novel speech enhancement technique for binaural dereverberation. An improved coherence model taking the shadowing effects of the head into account is used for a binaural input binaural output algorithm. Additionally, the algorithm ensures a dereverberation performance independent of the azimuth angle of the speech source and preserves the binaural cues. Experiments with measured binaural room impulse responses have shown that this algorithm is capable of reducing significantly the effects of reverberation especially in highly reverberant rooms. The algorithm has a low computational complexity and can further be combined with the methods discussed in [1]. A further enhancement, especially in rooms with moderate reverberation, can be obtained by means of an adaptive coherence model based on a measure of the “diffusiveness”.

6. REFERENCES

[1] H.W. Löllmann and P. Vary, “Low delay noise reduction and dereverberation for hearing aids,” *EURASIP Journal on Applied Signal Processing*, vol. 1, 2009.

[2] H. Kuttruff, *Room Acoustics*, Taylor & Francis, London, 2000.

[3] M. Jeub, M. Schäfer, and P. Vary, “A binaural room impulse response database for the evaluation of dereverberation algorithms,” in *Proc. Int. Conference on Digital Signal Processing (DSP)*, Santorini, Greece, 2009.

[4] M. Dörbecker, *Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen*, Ph.D. thesis, RWTH Aachen University, Aachen, Germany, 1998.

[5] T. Lotter and P. Vary, “Dual-channel speech enhancement by superdirective beamforming,” *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–14, 2006.

[6] J. Blauert, *Spatial Hearing - The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, USA, rev. edition, 1996.

[7] N. Madhu and R. Martin, “Acoustic source localization with microphone arrays,” in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds. Wiley&Sons, Chichester, 2008.

[8] U. Simmer and J. Bitzer, “Post-filtering techniques,” in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. Springer, Berlin, 2001.

[9] R. Zelinski, “A microphone array with adaptive post-filtering for noise reduction in reverberant rooms,” in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New York, USA, 1988, pp. 2578–2581 vol.5.

[10] I.A. McCowan and H. Bourlard, “Microphone array post-filter based on noise field coherence,” *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 709–716, 2003.

[11] J.B. Allen, D.A. Berkley, and J. Blauert, “Multimicrophone signal-processing technique to remove room reverberation from speech signals,” *J. Acoust. Soc. Am.*, vol. 62, no. 4, pp. 912–915, 1977.

[12] T.H. Falk and W.-Y. Chan, “A non-intrusive quality measure of dereverberated speech,” in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, 2008.