COHERENT CHANNEL BASED SUBBAND MULTICHANNEL DEREVERBERATION

JeeSok Lee, Sejin Oh, Hong-Goo Kang*

Yonsei University,
Department of Electrical and Electronics Engineering,
50 Yonsei-Road, Seodaemungu, Seoul, South Korea

ABSTRACT

This paper presents a multichannel dereverberation algorithm that only uses coherent acoustic channels. In the framework of multi-input/output inverse theorem (MINT), the equalization performance varies depending on the length of the input acoustic channels. However, only the portion of observed channel that resemble the true acoustic channel contributes to performance enhancement when measurement error is accounted. Hence, the proposed algorithm derives the frequency dependent viable channel length (VCL) from the coherence analysis of Monte Carlo observations of a single acoustic channel. The VCL of the room impulse response (RIR) is determined by the portion where the stochastic characteristic of multiple observations is highly coherent. Experiments are conducted to compare the equalization performance of the subband MINT algorithm depending on the length of the input RIR. The equalization performance using frequency dependent VCL is as good as the one obtained using the maximum length of the measured channel, while its complexity is significantly reduced.

Index Terms— Dereverberation, multichannel equalization, subband MINT, coherent acoustic channel, viable channel length

1. INTRODUCTION

Reverberation is known to provide spaciousness and intimacy to the sound produced by distorting the original signal [1, 2]. Although it can be used to provide immersive user experience [3,4], it is also known to degrade the intelligibility of speech signals [5]. Hence, a removal of the effect caused by the pre-existing reverberation needs to be considered in scenarios where such effect cannot be disregarded.

Various dereverberation algorithms have been developed to equalize or alleviate the effect of reverberation, and of those algorithms, multi-input/output inverse theorem (MINT) is best-known for its simplicity [6]. MINT is a multichannel time-domain dereverberation algorithm that equalizes the acoustic channels, described by room impulse response (RIR) filters. In ideal condition, MINT produces a set of exact equalization filters to a predefined set of input RIRs. Hence it is extremely sensitive to the fluctuation of measured RIRs and its computational complexity restricts increasing the length of input RIRs. Such drawbacks limit the usage of MINT algorithm in real environments.

Variations of MINT algorithms have been developed to overcome such limitations. Modification of cost function, various equalized responses, singular value decomposition, and weighting functions are used to stabilize the equalization filter and to improve the performance under the influence of measurement error [7-10]. To reduce its computational complexity, subband based approaches have been implemented [11, 12]. Using subband approach, the length of input RIRs can be increased by ten-folds compared to that of the fullband approach while it only requires small amount of increase in complexity [12].

Although in-depth analysis have been conducted on how to enhance the equalization performance given the measured RIRs, not much discussion is made on how to correctly determine the length of the input RIRs. It is general belief that the better equalization performance can be obtained as longer RIRs are used. Such statement is true in ideal environments, but not in real environments where measurement error cannot be avoided. The equalization performance saturates when the measurement error becomes dominant in the observed RIR. Hence, by inverting only the portion where the true RIR is dominant, the complexity of MINT can be significantly reduced while achieving near optimal equalization performance. However, the length of the viable acoustic channel, which is the portion of measured RIR that resemble the true RIR, should be determined by various environmental factors such as reverberation time, noise level, and so on.

In this paper, we propose a novel algorithm to estimate the viable channel length (VCL) of the measured RIRs in each frequency band based on Monte Carlo measurements. By measuring the coherence of multiple observations of the same RIR, the RIR is segmented into two parts; one where true RIR is dominant, and the other where the measurement error is dominant. Experimental results verify that the equalization performance of MINT algorithm varies depending on the length of the input RIRs used. The results also show that the equalization performance obtained by using the proposed frequency varying VCLs is as good as the one obtained using the maximum possible channel length.

2. BACKGROUND THEORY

For sound transmission in enclosed environment, multiple paths exist from a source to a receiver. The sum of the paths, RIR, can be described by a finite impulse response (FIR) filter $h(n)$. For M separate channels, M distinct sets of RIRs exist, and for a given source signal $s(n)$, the observed signal $x_m(n)$ at m^{th} channel can be described using convolution operation,

$$x_m(n) = h_m(n) * s(n), \tag{1}$$

where $m = 1, 2, \cdots, M$ and $h_m(n)$ is assumed to have a length of L_h.

*This project is supported by Intelloid under project funded by Small and Medium Business Administration (SMBA) of South Korea.

2.1. Multi-input/output Inverse Theorem (MINT)

The room effect is equalized using the inverse filter $g_m(n)$, which has a length of L_g. The objective function for equalization is

$$\hat{s}(n) = g_m(n) * x_m(n) = (h_m(n) * g_m(n)) * s(n) = s(n - n_d), \quad (2)$$

where $\hat{s}(n)$ is the equalized signal. Disregarding the source signal, the objective function can be simplified as $d(n) = h_m(n) * g_m(n)$, where $d(n)$ is the equalized response of length $L_h + L_g - 1$. An n_d-tap delayed Dirac's delta function is often replaced by $d(n)$, as done in equation (2).

In vector notations, the simplified objective function for m^{th} channel is

$$\mathbf{d} = \mathbf{H}_m \mathbf{g}_m, \quad (3)$$

where the vectors $\mathbf{g}_m = [g_m(0)\, g_m(1) \cdots g_m(L_g - 1)]$ and $\mathbf{d} = [d(0)\, d(1) \cdots d(L_h + L_g - 2)]$ denote equalization and desired response filters. Matrix $\mathbf{H}_m$ is the Sylvester matrix generated using $h_m(n)$, of size $(L_h + L_g - 1) \times (L_g)$. It is well known that a unique solution to equation (3) does not exist because the system matrix $\mathbf{H}_m$ is an overdetermined matrix and thus brings an ill-posed problem [6]. However, a full-rank matrix can be obtained by utilizing auxiliary channels, where the objective equation now becomes

$$\mathbf{d} = \sum_{m=1}^{M} \mathbf{H}_m \mathbf{g}_m = \mathbf{H}\mathbf{g}, \quad (4)$$

where $\mathbf{H} = [\mathbf{H}_1\, \mathbf{H}_2 \cdots \mathbf{H}_M]$ and $\mathbf{g} = \left[\mathbf{g}_1^H\, \mathbf{g}_2^H \cdots \mathbf{g}_M^H\right]^H$, with $\mathbf{d}$ being equivalent to that in equation (3) and $\bullet^H$ denoting Hermitian transpose operator. Equation (4) is known to have a unique solution when $L_g = \lfloor (L_h - 1)/M \rfloor$, where $\lfloor \cdot \rfloor$ denotes flooring operation [13]. The equalization filter can be obtained using the least-square method as follows

$$\mathbf{g} = \left(\mathbf{H}^H\mathbf{H} + \delta\mathbf{I}\right)^{-1} \mathbf{H}^H\mathbf{D}, \quad (5)$$

where the regularization term δ is used to stabilize the equalization filter by suppressing its norm [7].

2.2. Subband based MINT

Adopting MINT to critically sampled subbands, where the decimation factor N is equivalent to the number of subbands K, brings aliasing effect [11]. Due to this effect, each subband cannot be processed separately, and neighboring subbands have to be taken into consideration into the inversion process [14]. In oversampled subbands, where $N < K$, however, it becomes possible to process each subband independently since the guard band relieves the impact caused by adjacent-channel interference [15, 16].

Generalized discrete Fourier transform (GDFT) filter banks are employed to generate analysis and synthesis filter banks in oversampled subband processing [15]. For a given prototype filter bank $p(n)$ having a low-pass characteristic, and of which length is L_p, RIRs are decomposed into K subbands using the following analysis filter banks

$$u_k(n) = p(n)\, e^{j2\pi(n+n_0)(k+k_0)/K}, \quad (6)$$

where $k = 0, 1, \cdots, K/2 - 1$ is the subband index, and $n_0 = 0$ and $k_0 = 0.5$ are usually used. RIRs for each channel are then decomposed into subband filters $c_{k,m}(n) = h_m(n) * u_k(n)$, each with length $L_c = L_h + L_p - 1$. The downsampled subband filters $\mathbf{h}_{k,m} = \left[h_{k,m}(0)\, h_{k,m}(1) \cdots h_{k,m}(0)\right]$ are obtained by

$$\mathbf{h}_{k,m} = \mathbf{U}_{k,N}^{+} \mathbf{c}_{k,m,N}, \quad (7)$$

where $\mathbf{U}_{k,N}^{+}$ is the pseudo-inverse of the Sylvester matrix of the decimated analysis filter $\mathbf{u}_{k,N} = \left[u_k(0)\, u_k(N) \cdots u_k(NL_{p,N})\right]$ and $\mathbf{c}_{k,m,N} = \left[c_{k,m}(0)\, c_{k,m}(N) \cdots c_{k,m}(NL_{c,N})\right]$. $L_{p,N} = \lfloor (L_p - 1)/N \rfloor$ and $L_{c,N} = \lfloor (L_c - 1)/N \rfloor$ denote the length of the decimated filters. The inverse filters $g_{m,k}(n)$ are then obtained using eq. (5) for each subband.

The full-band equalization filter $g_m(n)$ for each channel is synthesized using its subband components as

$$g_m(n) = \sum_{k=0}^{K-1} g_{k,m,N}(n) * v_k(n), \quad (8)$$

where $v_k(n) = u_k^*(L_p - n - 1)$, with $\cdot^*$ denoting complex conjugate. Here, $g_{k,m,N}(n)$ is obtained by upsampling $g_{k,m,u}(n)$ as

$$g_{k,m,N}(n) = \begin{cases} g_{k,m,u}\,(n/N) & n = lN,\ l \in \mathbb{Z} \\ 0 & otherwise, \end{cases}$$

where $g_{k,m,u}(n) = g_{k,m}(n) * u_k(Nn)$.

3. PROPOSED CHANNEL TRUNCATION METHOD

3.1. RIR as deterministic and stochastic components

Measured RIRs, unless measurement conditions are ideal, cannot be regarded as true RIRs, and are generally assumed as

$$\hat{h}_m(n) = h_m(n) + \epsilon_m(n), \quad (9)$$

where $h_m(n)$, $\hat{h}_m(n)$, and $\epsilon_m(n)$ denote true RIR, measured RIR, and measurement noise at the m^{th} channel respectively. We assume that the location of the source and the receiver remains the same and the characteristics of the measurement error does not change throughout P observations. Given such assumptions, it can be said that there is no fluctuation in $h_m(n)$ and that $\epsilon_m(n)$ can be modelled as an ergodic stationary random process. Hence, we may assume that the measured RIR is composed of deterministic and stochastic components, where $h_m(n)$ falls into the former and $\epsilon_m(n)$ falls into the latter component. We also assume that $\epsilon_m(n)$s from each measurement process are uncorrelated to $h_m(n)$ and to each other.

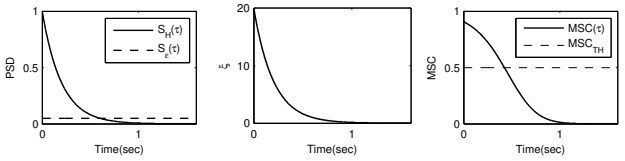
Since $\epsilon_m(n)$ is assumed to be a stationary random process, we define its variance as a constant value which is same across all channels as

$$\sigma_{\epsilon_m}^2(n) = E\left[|\epsilon_m(n)|^2\right] = \sigma_\epsilon^2, \quad (10)$$

where $E[\cdot]$ denotes the expectation operator. Since it is a stationary process, we also assume that its power spectral density (PSD) is stationary and constant, which depends only on the frequency f as

$$S_{\epsilon_m}(\tau, f) = E\left[|\mathbb{F}\{\epsilon_m(\tau)\}|^2\right] = S_\epsilon(f), \quad (11)$$

for some time instance τ, where $\mathbb{F}$ denotes the Fourier transform. On the other hand, RIR is a non-stationary process, where its energy decays as time evolves. The energy decaying rate is highly



(a) (b) (c)

Fig. 1: Estimated evolution of (a) PSD, (b) SNR, and (c) MSC

dependent on the characteristics of the room. Polack models the RIR as

$$h_{Polack}(n) = \begin{cases} b(n)\, e^{-\Delta n} & n \geq 0 \\ 0 & otherwise, \end{cases} \tag{12}$$

where $b(n)$ is a zero-mean white Gaussian noise and $\Delta = 3 \ln 10 / T_r$ indicates how fast the energy of RIR decays based on the reverberation time T_r [5]. Many other literatures also describe that the energy envelope of RIR decays exponentially as time evolves [1, 3]. Hence the exponentially decaying model

$$S_{H_m}(\tau, f) = e^{-\Delta(f)\tau} \tag{13}$$

is employed for time-frequency representation of energy envelope of $h_m(n)$. We assume that the room modeling parameter $\Delta(f)$ is solely dependent on the frequency f and is uniform across all channels.

Regarding deterministic and stochastic components as desired and noise signals respectively, we can presume the instantaneous signal-to-noise ratio (SNR), $\xi_m(\tau, f)$, to be

$$\xi_m(\tau, f) = \frac{S_{H_m}(\tau, f)}{S_{\epsilon_m}(\tau, f)} = \frac{e^{-\Delta(f)\tau}}{S_\epsilon(f)} = e^{-\Delta(f)\tau - \alpha(f)}, \tag{14}$$

where $\alpha(f) = \log S_\epsilon(f)$ is a coefficient which describes the recording environment characteristics.

3.2. VCL estimation based on coherence of RIR

The similarity between two signals Y and Z is measured by the magnitude squared coherence [17, 18], which is defined as

$$MSC_{YZ}(\tau, f) = \frac{|S_{YZ}(\tau, f)|^2}{S_{YY}(\tau, f) S_{ZZ}(\tau, f)}. \tag{15}$$

$S_{YZ}(\tau, f)$ is the cross-PSD (CPSD) function, which is defined as

$$S_{YZ}(\tau, f) = E\left[Y(\tau, f) Z^*(\tau, f)\right], \tag{16}$$

where $Y(\tau, f)$ and $Z(\tau, f)$ are STFT of $y(n)$ and $z(n)$ [19, 20].

Let $H_m(\tau, f)$ and $\hat{H}_{m,i}(\tau, f)$ denote the STFT of $h_m(n)$ and $\hat{h}_{m,i}(n)$ respectively, where subscript i denotes observation index of the measurement process. For each measured RIRs, the dominant portion of direct path and early reflections are considered as the deterministic component, whereas that of late reverberation is considered as the stochastic component. Moreover, using the assumptions given in equations (9), (11), and (13), spectral density estimations in (16) is assumed to be

$$S_{\hat{H}_{m,i}\hat{H}_{m,j}}(\tau, f) = \begin{cases} S_{H_m}(\tau, f) & i \neq j \\ S_{H_m}(\tau, f) + S_{\epsilon_m}(\tau, f) & i = j, \end{cases} \tag{17}$$

where i and j denote observation index. Then the MSC in eq. (15) becomes

$$MSC_{\hat{H}_m} = \left(\frac{S_{H_m}}{S_{H_m} + S_{\epsilon_m}}\right)^2 = \left(\frac{1}{1 + \xi_m^{-1}}\right)^2, \tag{18}$$

where time-frequency indices (τ, f) has been ignored for simplicity. Note that MSC in eq. (15) is a squared logistic function. Hence the coherence of the measured RIR for a single channel decreases as time evolves, where the rate of decrease and the initial coherence are determined by room parameters $\Delta(f)$ and $\alpha(f)$ respectively.

Ideally, $S_H(\tau, f)$ and $S_\epsilon(\tau, f)$ evolves as depicted in Fig. 1 (a), where the corresponding SNR and MSC obtained are shown in Fig. 1 (b) and 1 (c), respectively, for a single frequency. Since eq. (18) and Fig. 1 (c) implies that $\hat{h}_m \approx h_m(n)$ only up to a certain time, the length of the viable channel, $\tau_{f,m}$, for each channel m and frequency f, is defined as

$$\tau_{f,m} = \underset{\tau}{\operatorname{argmax}} \left\{ MSC_{\hat{H}_m}(\tau, f) \geq MSC_{TH} \right\}, \tag{19}$$

for some predefined threshold value MSC_{TH}. For all $\tau_{f,m}$s obtained, the true VCL for each subband is chosen as

$$\tau_k = \max_{f \in f_k} \left(\max_m \tau_{f,m} \right), \tag{20}$$

where f_k is the frequency region of the k^{th} subband.

4. SIMULATIONS AND RESULTS

In this section, experiments are performed to verify the existence of VCL. Equalization performance of subband MINT algorithm, for a set of given RIRs of various lengths, is used for verification. Results show that near optimal performance can be achieved even when the length of the input RIRs are much shorter than the maximum length.

4.1. Simulation settings and performance evaluation criteria

A single measurement is generated by adding the true RIR $h(n)$ and the measurement noise $\epsilon(n)$. The RIRs from MARDY database are regarded as true RIRs [21]. RIRs used in this experiments are measured using two sources, eight microphones, in three environments. The environments are classified into the cases where its reverberation time is short, medium, and long. Each RIR has 65,536 taps and is recorded at 48 kHz. One measurement process is generated from each RIR, where $P = 20$ observations are made for each process. For each observation, $\epsilon(n)$ is generated using a zero-mean white Gaussian noise with variance σ. The value of the σ is set by controlling the channel mismatch

$$\eta = 10 \log_{10} \left(\sum_n |\epsilon(n)|^2 \Big/ \sum_n |h(n)|^2 \right), \tag{21}$$

and it does not change within a measurement. -5, -10, -15, and -20 decibels (dB) are used for setting the value of η.

VCLs for each subband is obtained using $MSC_{TH} = 0.5$ in (19). Averaged observation from each measurement process, $\tilde{h}_m(n) = E[\hat{h}_m(n)]$, is used as input RIRs, and only the location of the source is different in the two input RIRs. For subband decomposition, $k_0 = 0.5$, $n_0 = 0$, $K = 32$, $N = 24$, and $L_p = 512$ are used to obtain $u_k(n)$ in eq. (6). In this experiment, the maximum value of VCL is set to 30,000 taps.

To verify the effect of VCL, equalization using the regularized MINT (RMINT) given in eq. (5) with various lengths of input RIRs is employed. The length of the input RIRs, τ_{in},

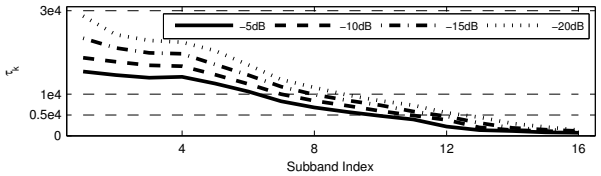


Fig. 2: Average VCL for each η vs. fixed channel length

τ_{in}	5,000 taps	10,000 taps	30,000 taps
RMINT	2.53	12.27	323.02

(a)

η	-5 dB	-10dB	-15dB	-20dB
RMINT	14.74	23.92	40.00	64.28
MSC+RMINT	61.70	70.89	86.97	111.25

(b)

Table 1: Averaged *cputime* comparison for (a) conventional and (b) proposed algorithm

are defined by VCL in the proposed algorithm, and are set to [5,000 10,000 30,000] taps for the conventional algorithm. To quantitatively measure the equalization performance, clarity and definition indices, defined as follows

$$C_{50} = \frac{\sum_{n=0}^{N_{50}} \left|\hat{d}(n)\right|^2}{\sum_{n=N_{50}}^{\infty} \left|\hat{d}(n)\right|^2},\ D_{50} = \frac{\sum_{n=0}^{N_{50}} \left|\hat{d}(n)\right|^2}{\sum_{n=0}^{\infty} \left|\hat{d}(n)\right|^2}, \tag{22}$$

are employed [22]. In (22), N_{50} denotes sample index that corresponds to 50 milliseconds and $\hat{d}(n)$ denotes the equalized response. $\delta = 10^{-\eta/10}$ is used for the regularization parameter given in eq. (5). Also, *cputime* function in MATLAB is used to compare the complexity of the proposed and conventional algorithms.

4.2. Simulation results

Results in Fig. 2 depict VCL obtained for each subband when different ηs are used. The thin dashed line in the figure denotes τ_{in} for the conventional algorithm. It is clear that the VCL decreases when η increases or frequency band increases. The energy of the true RIR tends to decrease in higher frequency bands, whereas that of the noise is uniform. Consequently, η increases and VCL becomes shorter in higher frequency bands.

Fig. 3 show the average C_{50} and D_{50} of the equalized responses obtained from RMINT. The results are segmented into three groups along the x-axis depending on the reverberation time. From the figures, it can be noted that τ_{in} = 30,000 taps yields the maximum equalization performance. The equalization performance degrades as η decreases when 5,000 taps of input RIRs are used whereas it is enhanced when 30,000 taps are used. C_{50} drops 5 dB when τ_{in} is 5,000 taps whereas it increases 5 dB when τ_{in} is 30,000 as η decreases from -5 to -20 dB. On the other hand, D_{50} plummets as η decreases when T_r is long and τ_{in} is short. Also, the figures show that the performance of the proposed algorithm is as good as the highest performance can be obtained in the conventional method.

By combining the results depicted in Fig. 2 and Fig. 3, it can be concluded that using only RIRs with high coherence in the measurement process provides near-optimal performance. Even

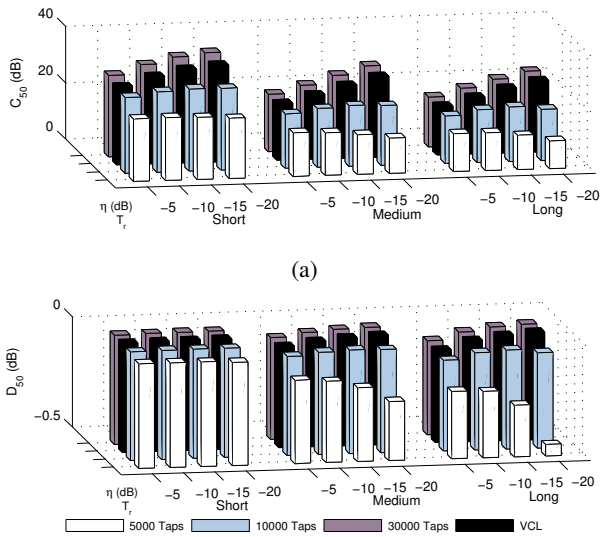


Fig. 3: Average (a) C_{50}, and (b) D_{50} for channel length and environment variation

though most of the RIRs are truncated in higher frequency subbands, the performance degradation is insignificant compared to the performance of the best case scenario. Also, when the measured RIRs are more reliable, longer RIRs are required to achieve the near-optimal performance.

Finally, results in Table 1 show that the time elapsed in the proposed algorithm is remarkably shorter than that elapsed in conventional algorithm. The time elapsed to compute RMINT algorithm ranges from 15 to 64 seconds using VCL, whereas it takes over 300 seconds when τ_{in} is 30,000 taps. Altogether, the complexity of RMINT can be reduced significantly with only small loss in equalization performance if VCL is employed.

5. DISCUSSION AND CONCLUSION

Two conclusions can be drawn from the simulations conducted regarding the effect of temporal truncation of channel. Firstly, the length of the viable channel should be considered separately for each subband. Reverberation time varies for each frequency due to the frequency dependent characteristics of the reflection coefficients of the walls. Secondly, the length of viable channel depends on the recording environment. As reverberation becomes longer and the environment becomes quieter, the length of viable channel becomes longer. Hence, the optimal method would be to estimate the length of the viable channel upon the measurement of RIRs.

Relation to prior work — Many literatures describe on improving the performance of MINT algorithm for given RIRs, or on improving RIR measurement techniques. However, not many literatures have focused on the effect that the channel length has on the equalization performance of MINT algorithm and how they should be defined. The work presented in this paper provides insights to how and why the viable channel length should be defined. Since the VCL can be estimated upon the measurement of the RIR, it is applicable to scenarios of any environments regardless of their reverberation time.

6. REFERENCES

[1] Heinrich Kuttruff, *Room Acoustics*, Spon Press, 4th edition, 2000.

[2] J.R. Hyde, "Acoustical intimacy in concert halls: Does visual input affect the aural experience?," *Proceedings of the Institute of Acoustics*, vol. 24, 2002.

[3] Jean-Marc Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *IEEE Intl. Conference on Acoust., Speech, and Signal Process. (ICASSP) 1992*, Mar 1992, vol. 2, pp. 221–224.

[4] Enzo De Sena, Huseyin Hacihabiboglu, and Zoran Cvetkovic, "Scattering delay network: An interactive reverberator for computer games," in *Audio Engineering Society Conference: 41st Intl. Conference: Audio for Games*, Feb 2011.

[5] K Lebart and JM Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acoustica*, vol. 87, no. 3, pp. 359–366, MAY 2001.

[6] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. on Acoust., Speech and Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb 1988.

[7] Takafumi Hikichi, Marc Delcroix, and Masato Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 62–62, Jan. 2007.

[8] I Kodrasi and S. Doclo, "Robust partial multichannel equalization techniques for speech dereverberation," in *IEEE Intl. Conference on Acoust., Speech and Signal Process. (ICASSP) 2012*, Kyoto, Japan, March 2012, pp. 537–540.

[9] Wancheng Zhang, Emanuel A. P Habets, and Patrick A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, 2010.

[10] I Kodrasi, S. Goetze, and S. Doclo, "A perceptually constrained channel shortening technique for speech dereverberation," in *IEEE Intl. Conference on Acoust., Speech and Signal Process. (ICASSP) 2013*, Vancouver, Canada, May 2013, pp. 151–155.

[11] H. Yamada, H. Wang, and F. Itakura, "Recovering of broadband reverberant speech signal by sub-band mint method," in *IEEE Intl. Conference on Acoust., Speech and Signal Process. (ICASSP) 1991*, Apr 1991, vol. 2, pp. 969–972.

[12] N.D. Gaubitch and P.A Naylor, "Equalization of multichannel acoustic systems in oversampled subbands," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 17, no. 6, pp. 1061–1070, Aug 2009.

[13] G. Harikumar and Y. Bresler, "Fir perfect signal reconstruction from multiple convolutions: minimum deconvolver orders," *IEEE Trans. on Signal Process.*, vol. 46, no. 1, pp. 215–218, Jan 1998.

[14] Andre Gilloire and Martin Vetterli, "Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. on Signal Process.*, vol. 40, no. 8, pp. 1862–1875, Aug 1992.

[15] Stephen Weiss, *On Adaptive Filtering in Oversampled Subbands*, Ph.D. thesis, University of Strathclyde, 1998.

[16] J.P. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, "The complex subband decomposition and its application to the decimation of large adaptive filtering problems," *IEEE Trans. on Signal Process.*, vol. 50, no. 11, pp. 2730–2743, Nov 2002.

[17] I Santamaria and J. Via, "Estimation of the magnitude squared coherence spectrum based on reduced-rank canonical coordinates," in *IEEE Intl. Conference on Acoust., Speech and Signal Process. (ICASSP) 2007*, Hawaii, USA, April 2007, vol. 3, pp. III–985–III–988.

[18] K. Shi and Xiaoli Ma, "A frequency domain step-size control method for lms algorithms," *IEEE Signal Process. Letters*, vol. 17, no. 2, pp. 125–128, Feb 2010.

[19] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, Jul 1989.

[20] L. Stankovic, S. Stankovic, and M. Dakovic, "From the stft to the wigner distribution [lecture notes]," *Signal Processing Magazine, IEEE*, vol. 31, no. 3, pp. 163–174, May 2014.

[21] Emanuel A.P. Habets Tony Myatt Jimi Y.C. Wen, Nikolay D. Gaubitch and Patrick A. Naylor, "Evaluation of speech dereverberation algorithms using the mardy database," in *Proc. Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, 2006.

[22] ISO 3382-1:2009, *Acoustics - Measurement of room acoustic parameters -I: Performance spaces*, International Standards Organization, 2009.