Study of the frequency-domain multichannel noise reduction problem with the householder transformation

Conference Paper · March 2017

DOI: 10.1109/ICASSP.2017.7952203

3 authors:

Gongping Huang
Technion - Israel Institute of Technology. הטכניון

Jacob Benesty
Institut National de la Recherche Scientifique

Jingdong Chen
Institute of Electrical and Electronics Engineers

STUDY OF THE FREQUENCY-DOMAIN MULTICHANNEL NOISE REDUCTION PROBLEM WITH THE HOUSEHOLDER TRANSFORMATION

Gongping Huang[1], *Jacob Benesty*[2], *and Jingdong Chen*[1]

[1]CIAIC
Northwestern Polytechnical University
Xi'an, Shaanxi 710072, China

[2]INRS-EMT, University of Quebec
800 de la Gauchetiere Ouest, Suite 6900
Montreal, QC H5A 1K6, Canada

ABSTRACT

This paper presents an approach to the multichannel noise reduction problem. It first transforms the multichannel noisy speech signals into the frequency domain. A Householder transformation is then constructed, which converts the multichannel coefficients in each frequency bin into two components: one dominated by speech and the other dominated by noise. A Wiener filter is subsequently formed to achieve an estimate of the noise in the speech dominated component from the noise dominated component. The enhanced speech is then obtained by subtracting the noise estimate from the speech dominated component. This approach consists of two critical steps: construction of the Householder transformation and formation of the noise reduction Wiener filter. If the source incidence angle is known a priori, the Householder transformation can be directly constructed using the steering vector and the optimal estimate of the signal of interest can then be obtained by applying the Wiener filter. If the source incidence angle is not known a priori, the Householder transformation can be constructed from a hypothesized incidence angle. Then, the optimal signal estimate is obtained by searching the maximum of the variance of the enhanced signal with the Wiener filter in the interested range of the incidence angle.

Index Terms— Noise reduction, speech enhancement, microphone arrays, Householder transformation, Wiener filter.

1. INTRODUCTION

Noise reduction, also called speech enhancement, refers to the process of recovering a speech signal of interest from noisy observations. It has a wide range of applications in voice communications and human-to-machine interfaces [1–4]. Most early efforts focused on the single-channel case primarily because communication devices at that time were equipped with only a single microphone [5, 6]. However, although they are able to improve the signal-to-noise ratio (SNR) and speech quality, all the single-channel techniques achieve noise reduction at the cost of adding speech distortion [7, 8]. Generally, more noise reduction comes with more speech distortion [6]. Recently, multichannel noise reduction with an array of microphones has attracted much research and engineering attention [9–14]. In comparison with the single-channel noise reduction techniques, the multichannel ones have the potential to achieve better performance, with less speech distortion or/and more noise reduction [8, 15]. Moreover, they can also better deal with nonstationary noise, which is a big issue with the single-channel methods. Because of these great potentialities, multichannel methods have now been more and more used in real-world applications and it is common today to see smart terminals (e.g., smartphones and tablets) that contain at least two microphones.

This paper investigates the problem of multichannel noise reduction. We first apply the Householder transformation [16–18] to the multichannel noisy signals. This transformation naturally separates the multichannel signals into two components: one dominated by speech and the other dominated by noise (or even consisting of noise-only if there is no reverberation). Noise reduction is then achieved by applying an optimal filter that predicts the noise in the speech dominated component from the noise dominated one. The major contributions of this paper are as follows. 1) We apply the Householder transformation to the multichannel noise reduction problem in the frequency domain. This transformation separates the noisy signals into two components with one either consisting of noise-only or dominated by noise. 2) We develop a multichannel Wiener filter that can achieve noise reduction based on the output of the Householder transformation and the knowledge of the source incidence angle. 3) We also derive a Wiener filter that can work when the source incidence angle is not known *a priori*. 4) The performance of the new approach is validated in both the presence and absence of reverberation.

2. SIGNAL MODEL AND PROBLEM FORMULATION

We consider a uniform linear microphone array consisting of M omnidirectional microphones where the distance between two successive sensors is equal to δ. The received (microphones') signals, at the time index t, are written as

$$\begin{aligned} y_m(t) &= g_m(t) * s(t) + v_m(t) \\ &= x_m(t) + v_m(t),\ m = 1, 2, \ldots, M, \end{aligned} \tag{1}$$

where $*$ stands for linear convolution, $g_m(t)$ is the acoustic impulse response from the position of the unknown speech source $s(t)$ to the mth microphone, $x_m(t) = g_m(t) * s(t)$ is the convolved speech signal at the mth sensor, and $v_m(t)$ is the additive noise at the mth sensor. The source signal, $s(t)$, is assumed to be uncorrelated with the noise terms, $v_m(t),\ m = 1, 2, \ldots, M$, and all signals are considered to be real, zero mean, and broadband.

In the frequency domain, at the frequency index f, (1) can be expressed as [3, 19]

$$\begin{aligned} Y_m(f) &= G_m(f)S(f) + V_m(f) \\ &= X_m(f) + V_m(f),\ m = 1, 2, \ldots, M, \end{aligned} \tag{2}$$

where $Y_m(f)$, $S(f)$, $G_m(f)$, $X_m(f)$, and $V_m(f)$ are the frequency-domain representations of $y_m(t)$, $s(t)$, $g_m(t)$, $x_m(t)$, and $v_m(f)$, respectively. Let us choose microphone 1 as the reference. The objective of noise reduction in the frequency domain with multiple mi-

This work was supported in part by the NSFC "Distinguished Young Scientists Fund" under grant No. 61425005.

crophones is then to estimate the desired signal, $X_1(f)$, from the M observations $Y_m(f),\ m = 1, 2, \ldots, M$, the best way we can.

It is more convenient to write the M frequency-domain microphone signals in a vector form:

$$\begin{aligned} \mathbf{y}(f) &= \mathbf{g}(f)S(f) + \mathbf{v}(f) \\ &= \mathbf{x}(f) + \mathbf{v}(f), \end{aligned} \tag{3}$$

where

$$\begin{aligned} \mathbf{y}(f) &= \begin{bmatrix} Y_1(f) & Y_2(f) & \cdots & Y_M(f) \end{bmatrix}^T, \\ \mathbf{x}(f) &= \begin{bmatrix} X_1(f) & X_2(f) & \cdots & X_M(f) \end{bmatrix}^T \\ &= S(f)\mathbf{g}(f), \\ \mathbf{g}(f) &= \begin{bmatrix} G_1(f) & G_2(f) & \cdots & G_M(f) \end{bmatrix}^T, \\ \mathbf{v}(f) &= \begin{bmatrix} V_1(f) & V_2(f) & \cdots & V_M(f) \end{bmatrix}^T, \end{aligned}$$

and the superscript ^{T} denotes the transpose operator. Expression (3) can also be expressed as

$$\mathbf{y}(f) = X_1(f)\mathbf{d}(f) + \mathbf{v}(f), \tag{4}$$

where $\mathbf{d}(f) = \mathbf{g}(f)/G_1(f)$ is the frequency-domain signal propagation vector.

In the particular case of an anechoic environment, each sensor output can be modeled as a delayed copy of the source signal contaminated by some additive noise. In this case, the signal propagation vector is [14]

$$\mathbf{d}(f, \theta) = \begin{bmatrix} 1 & e^{-\jmath 2\pi f \tau_0 \cos\theta} & \cdots & e^{-\jmath 2(M-1)\pi f \tau_0 \cos\theta} \end{bmatrix}^T, \tag{5}$$

where $\jmath$ is the imaginary unit with $\jmath^2 = -1$, $\tau_0 = \delta/c$ is the delay between two successive sensors at the angle 0°, with $c = 340$ m/s being the speed of sound in air, and θ is the source incidence angle. Note that (5) becomes the steering vector if one replace the signal incidence angle θ with the steering angle.

3. NOISE REDUCTION WITH THE HOUSEHOLDER TRANSFORMATION

We first derive and study the optimal filter in the ideal situation where there is no reverberation. Let us assume that the source incidence angle θ is known and is equal to 0°. In this case, we have $\mathbf{d}(f) = \mathbf{d}(f, 0)$. We define the Householder transformation [16] associated with the steering vector, $\mathbf{d}(f, 0)$, as

$$\mathbf{T}(f) = \mathbf{I}_M - \frac{2}{\mathbf{b}^H(f)\mathbf{b}(f)}\mathbf{b}(f)\mathbf{b}^H(f), \tag{6}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix, and

$$\mathbf{b}(f) = \mathbf{d}(f, 0) + \sqrt{M}\, \mathbf{i}_1, \tag{7}$$

with $\mathbf{i}_1$ being the first column of $\mathbf{I}_M$. It can be checked that $\mathbf{T}(f)$ is Hermitian and unitary. It is also easy to verify that $\mathbf{T}(f)\mathbf{d}(f) = -\sqrt{M}\, \mathbf{i}_1$. Now, by left-multiplying both sides of (3) by $-\mathbf{T}(f)/\sqrt{M}$, we get

$$\mathbf{y}'(f) = -\frac{1}{\sqrt{M}}\mathbf{T}(f)\mathbf{y}(f) = X_1(f)\mathbf{i}_1 + \mathbf{v}'(f), \tag{8}$$

or, equivalently,

$$\begin{bmatrix} Y_1'(f) \\ \mathbf{y}_2'(f) \end{bmatrix} = \begin{bmatrix} X_1(f) \\ \mathbf{0}_{(M-1)\times 1} \end{bmatrix} + \begin{bmatrix} V_1'(f) \\ \mathbf{v}_2'(f) \end{bmatrix}. \tag{9}$$

One can see how the Householder transformation gives a clear noise reference signal. Indeed, $Y_1'(f) = X_1(f) + V_1'(f)$ is the sum of the desired signal and noise, while the $(M-1)$-dimensional vector $\mathbf{y}_2'(f) = \mathbf{v}_2'(f)$ contains noise only. Furthermore, $V_1'(f)$ and $\mathbf{v}_2'(f)$ are (partially) coherent. Based on this transformation, we can deduce an estimate of $X_1(f)$ as follows:

$$\begin{aligned} Z(f) &= Y_1'(f) - \mathbf{h}'^H(f)\mathbf{y}_2'(f) \\ &= X_1(f) + V_1'(f) - \mathbf{h}'^H(f)\mathbf{v}_2'(f), \end{aligned} \tag{10}$$

where $\mathbf{h}'(f)$ is a complex-valued filter of length $M-1$. Since $X_1(f)$ in $Z(f)$ is not affected by the filter $\mathbf{h}'(f)$, this approach does not add any distortion to the desired speech signal in this ideal situation where there is no reverberation and $\mathbf{d}(f, 0)$ is known precisely; otherwise, the filter $\mathbf{h}'(f)$ may introduce some distortion.

The variance of $Z(f)$ is

$$\begin{aligned} \phi_Z(f) &= \phi_{Y_1'}(f) - \mathbf{\Phi}^H_{\mathbf{y}_2' Y_1'}(f)\mathbf{h}'(f) \\ &\quad - \mathbf{h}'^H(f)\mathbf{\Phi}_{\mathbf{y}_2' Y_1'}(f) + \mathbf{h}'^H(f)\mathbf{\Phi}_{\mathbf{y}_2'}(f)\mathbf{h}'(f), \end{aligned} \tag{11}$$

where $\phi_{Y_1'}(f)$ is the variance of $Y_1'(f)$, $\mathbf{\Phi}_{\mathbf{y}_2' Y_1'}(f) = E\left[\mathbf{y}_2'(f)Y_1'^*(f)\right]$, the superscript * denotes the complex-conjugate operator, and $\mathbf{\Phi}_{\mathbf{y}_2'}(f) = E\left[\mathbf{y}_2'(f)\mathbf{y}_2'^H(f)\right]$ is the correlation matrix of $\mathbf{y}_2'(f)$.

From the minimizing of $\phi_Z(f)$ with respect to $\mathbf{h}'^H(f)$, we easily find the optimal Wiener filter:

$$\mathbf{h}'_{\mathrm{W}}(f) = \mathbf{\Phi}^{-1}_{\mathbf{y}_2'}(f)\mathbf{\Phi}_{\mathbf{y}_2' Y_1'}(f). \tag{12}$$

The optimal estimate of $X_1(f)$ in the Wiener sense is then

$$Z_{\mathrm{W}}(f) = Y_1'(f) - \mathbf{h}'^H_{\mathrm{W}}(f)\mathbf{y}_2'(f). \tag{13}$$

The estimate from this Wiener filter based on the Householder transformation is similar to that of the well-known generalized sidelobe canceler (GSC) [20]. It can also be shown that this Wiener filter is equivalent to the minimum variance distortionless response (MVDR) filter given in [19]. Although equivalence exists between the two filters, the Wiener filter based on the Householder transformation is preferable to the MVDR one in practice as $\mathbf{\Phi}_{\mathbf{y}_2'}(f)$ is much better conditioned than $\mathbf{\Phi}_{\mathbf{y}}(f) = E\left[\mathbf{y}(f)\mathbf{y}^H(f)\right]$ (see Section 5).

In practice, reverberation is inevitable. In the presence of reverberation, the signal propagation vector can be modeled as $\mathbf{d}(f) = \mathbf{d}(f, 0) + \mathbf{e}(f)$, where $\mathbf{e}(f)$ is an error vector. In this case, we do not have a clear separation between the speech-plus-noise and noise components as in the ideal case. Nevertheless, if no *a priori* information on $\mathbf{e}(f)$ is available, we can still estimate $X_1(f)$ by following the same principles. Now, the desired signal is distorted and the degree of distortion depends on the level of reverberation.

4. NOISE REDUCTION WITHOUT THE A PRIORI KNOWLEDGE OF THE DIRECTION OF THE DESIRED SIGNAL

In this section, we study the more general case where the desired source signal propagates from the direction θ ($0^\circ \le \theta < 180^\circ$); but the angle θ is not known. Let us consider the ideal scenario where the signal propagation vector is given in (5) and the signal model becomes

$$\mathbf{y}(f) = \mathbf{x}(f) + \mathbf{v}(f) = \mathbf{d}\left(f, \theta\right)X_1(f) + \mathbf{v}(f). \tag{14}$$

Now, let us take any angle θ_1 ($0^\circ \leq \theta_1 < 180^\circ$). The Householder transformation associated with the steering vector, $\mathbf{d}(f, \theta_1)$, is defined as

$$\mathbf{T}(f, \theta_1) = \mathbf{I}_M - \frac{2}{\mathbf{b}^H(f, \theta_1)\,\mathbf{b}(f, \theta_1)} \mathbf{b}(f, \theta_1)\,\mathbf{b}^H(f, \theta_1), \tag{15}$$

where $\mathbf{b}(f, \theta_1) = \mathbf{d}(f, \theta_1) + \sqrt{M}\ \mathbf{i}_1$. Similarly, left-multiplying both sides of (14) by $-\mathbf{T}(f, \theta_1)/\sqrt{M}$, we get

$$\begin{bmatrix} Y_1'(f, \theta_1) \\ \mathbf{y}_2'(f, \theta_1) \end{bmatrix} = \begin{bmatrix} X_1'(f, \theta_1) \\ \mathbf{x}_2'(f, \theta_1) \end{bmatrix} + \begin{bmatrix} V_1'(f, \theta_1) \\ \mathbf{v}_2'(f, \theta_1) \end{bmatrix}. \tag{16}$$

As before, the desired signal, $X_1(f)$, is estimated by

$$Z(f, \theta_1) = Y_1'(f, \theta_1) - \mathbf{h}'^H(f, \theta_1)\,\mathbf{y}_2'(f, \theta_1), \tag{17}$$

where $\mathbf{h}'(f, \theta_1)$ is a complex-valued filter of length $M - 1$. The minimization of the variance of $Z(f, \theta_1)$ leads to the Wiener filter:

$$\mathbf{h}'_{\mathrm{W}}(f, \theta_1) = \mathbf{\Phi}^{-1}_{\mathbf{y}'_2}(f, \theta_1)\,\mathbf{\Phi}_{\mathbf{y}'_2 Y'_1}(f, \theta_1). \tag{18}$$

As a result, the optimal estimate of $X_1(f)$ in the Wiener sense is

$$Z_{\mathrm{W}}(f, \theta_1) = Y_1'(f, \theta_1) - \mathbf{h}'^H_{\mathrm{W}}(f, \theta_1)\,\mathbf{y}_2'(f, \theta_1). \tag{19}$$

Let $\phi_{Z_{\mathrm{W}}}(f, \theta_1)$ be the variance of $Z_{\mathrm{W}}(f, \theta_1)$. It is clear that

$$\theta = \arg\max_{\theta_1} \phi_{Z_{\mathrm{W}}}(f, \theta_1). \tag{20}$$

From (20), we can get the direction of the source by scanning the space from 0° to 180° and compute each time the corresponding Wiener filter; then, the optimal filter for noise reduction corresponds to the angle that maximizes $\phi_{Z_{\mathrm{W}}}(f, \theta_1)$.

The angle θ can also be estimated according to [21] as

$$\theta = \arg\min_{\theta_1} \phi_{Y'_m}(f, \theta_1),\ m = 2, 3, \ldots, M \tag{21}$$

or

$$\theta = \arg\max_{\theta_1} \phi_{Y'_1}(f, \theta_1), \tag{22}$$

where $\phi_{Y'_m}(f, \theta_1)$ is the variance of $Y'_m(f, \theta_1)$. Therefore, in practice, one can obtain the estimate of the source incidence angle with the estimators from (20) to (22). It is worth mentioning that the estimators from (20) to (22) are defined on a narrowband basis. But they can be extended to process broadband speech signals by combining the narrowband estimators from different frequency bands as shown in [21].

5. SIMULATIONS

In this section, we assess the performance of the developed Wiener filters through simulations. We consider a room of size $4\ \mathrm{m} \times 4\ \mathrm{m} \times 3\ \mathrm{m}$. A loudspeaker is placed at (2.5, 2.0, 2.0), which plays back a speech signal prerecorded from a female speaker in a quiet office room, and eight omnidirectional microphones are located, respectively, at (x, 2.0, 1.6), where $x = 1.30 : 0.04 : 1.58$. The acoustic channel impulse responses from the source to the microphones are generated with the image model method [22]. Then, the microphone signals are generated by convolving the source signal with the corresponding simulated impulse responses and adding some noise. We consider two types of noise: white and diffuse [23]. All the signals are 30 seconds long and the sampling frequency is 8 kHz. The

Fig. 1. $X_1'(f)$ and $X_m'(f)$ after the Householder transformation in an anechoic environment: (a) amplitude of $X_1'(f)$ and $X_m'(f)$ and (b) zoomed plot of the amplitude of $X_m'(f)$, $m = 2, 3, \ldots, M$. $M = 8$, $f = 1000$ Hz, and the source incidence angle is 0°.

Fig. 2. Comparison of the condition numbers of $\mathbf{\Phi}_{\mathbf{y}'_2}(f)$ and $\mathbf{\Phi}_{\mathbf{y}}(f)$: (a) in white noise and (b) in diffuse noise. $f = 1000$ Hz, iSNR = 10 dB and $M = 8$.

Wiener filters are implemented in the STFT domain as follows. The microphone array signals are partitioned into overlapping frames with a frame size of $K = 128$ and an overlapping factor of 75%. A Kaiser window is then applied to each frame and the windowed frame signal is subsequently transformed into the STFT domain using a 128-point FFT. The Householder transformation is then constructed and applied to the noisy STFT coefficients and the Wiener filter is subsequently used to reduce the noise. Finally, the inverse FFT (with overlap add technique) is used to obtain the time-domain clean speech estimate. In the implementation, all the correlation ma-

Fig. 3. SNR Performance of the Wiener filter as a function of M: (a) in diffuse noise and (b) in white noise. iSNR = 0 dB.

trices are computed using a recursive method [3].

We first show the behavior of the signal vector after the Householder transformation in an anechoic environment. We use $\mathbf{x}_2'(f)$ to denote the signal leakage after the Householder transformation in practice (theoretically, $X_1'(f) = X_1(f)$ and $\mathbf{x}_2'(f) = \mathbf{0}$ in anechoic environments with known direction-of-arrival (DOA) information). We set $M = 8$ and the desired source propagates from the angle 0°. Figure 1 plots the amplitudes of $X_1'(f)$ and $X_m'(f)$, $m = 2, 3, \ldots, M$, as a function of the frame index at one subband ($f = 1000$ Hz) after the Householder transformation. It is clearly seen that the amplitudes of $X_m'(f)$ are much smaller than that of $X_1'(f)$ and they can be considered as zero in comparison with $X_1'(f)$. This is consistent with the discussion in Section 3, where we have shown that the Householder transformation projects $\mathbf{x}(f)$ into a vector that has zeros in all positions but one. Notice that $X_m'(f)$, $m = 2, 3, \ldots, M$, are not completely zero. This is mainly due to the windowing effect in the overlap-add process.

While theoretical analysis show that the Wiener filter based on the Householder transformation is equivalent to the traditional MVDR filter in nonreverberant environments, the former is preferable to use in practice, particularly in fixed-point implementations. The underlying reason can be explained as follows. The implementation of the Wiener filter involves the computation of the inverse of $\mathbf{\Phi}_{\mathbf{y}_2'}(f)$, which is of size $(M-1) \times (M-1)$. Since the vector $\mathbf{y}_2'(f)$ consists of either noise-only or dominated by noise, $\mathbf{\Phi}_{\mathbf{y}_2'}(f)$ is generally well conditioned, so its inverse can be computed reliably. In comparison, the vector $\mathbf{y}(f)$ consists of both the speech of interest and noise, so the eigenvalue spread of $\mathbf{\Phi}_{\mathbf{y}}(f)$ is much larger than that of $\mathbf{\Phi}_{\mathbf{y}_2'}(f)$, indicating that $\mathbf{\Phi}_{\mathbf{y}}(f)$ is ill conditioned and, therefore, its inverse is numerically less reliable to compute. To illustrate this, we show the condition numbers of $\mathbf{\Phi}_{\mathbf{y}_2'}(f)$ and $\mathbf{\Phi}_{\mathbf{y}}(f)$, where the l_2-norm condition number of a matrix is defined as

$$\chi(\mathbf{A}) = \|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2. \tag{23}$$

Figure 2 plots the results with $f = 1000$ Hz as a function of the frame index in both white and diffuse noises (note that similar results were observed across the interested frequency range for both types of noise). It is clearly seen that the condition number of $\mathbf{\Phi}_{\mathbf{y}}(f)$ is much larger than that of $\mathbf{\Phi}_{\mathbf{y}_2'}(f)$ in both noise conditions while the difference is more significant in diffuse noise. This result certainly shows that the Wiener filter is numerically more reliable to implement than the MVDR filter based on the use of the $\mathbf{\Phi}_{\mathbf{y}}(f)$ matrix.

The performance of the Wiener filter as a function of the number of microphones in this moderate reverberant environment (the reverberation time, T_{60}, is approximately 200 ms) is plotted in Fig. 3, where we showed the results in two types of noise: diffuse and white. For the purpose of comparison, results obtained in the same noisy but nonreverberant environments are also plotted in Fig. 3. It is seen from Fig. 3 that the output SNR increases with the number of microphones. So, the more the number of microphones, the better the noise reduction performance. Comparing the two noise conditions, we see that the SNR improvement in diffuse noise is slightly higher than that in white noise. In white noise, the SNR improvement (the difference between the input and output SNRs in decibel) is approximately 3 dB if two microphones are used and this improvement increases if more microphones are used. Theoretically, it can be checked that the SNR improvement in white noise is approximately $10 \log_{10} M$ dB, which is corroborated by simulations in Fig. 3(b).

6. CONCLUSIONS

This paper presented a two-step approach to the problem of multichannel noise reduction in the frequency domain. In the first step, a Householder transformation is constructed. In the second step, a Wiener filter is formed by using the noise-only or noise dominated component to estimate the noise in the speech-plus-noise component, thereby achieving noise reduction. The construction of the Householder transformation in each frequency bin requires the knowledge of the DOA information. If this information is not known *a priori*, which is often the case in practice, we derived a method that combines the noise reduction and DOA estimation together into one process. Simulations were conducted to validate the performance of the developed approach in both non-reverberant and reverberant environments. The results showed that it can improve the SNR significantly and this improvement increases with the number of microphones regardless whether there is reverberation or not.

7. RELATION TO PRIOR WORK

Noise reduction has long been a major problem in signal processing for voice communications and human-machine interfaces [1–10]. A significant number of efforts have been devoted to this problem in the literature [11–13, 15]. Most early efforts focused on the single-channel case primarily as most communication devices at that time were equipped with only one microphone [5, 6]. Recently, multichannel noise reduction with an array of microphones have been intensively investigated [8–10, 24], which has been found to be more flexible in dealing with noise, i.e., more noise reduction with more microphones [6, 8].

The Householder transformation has been widely used in many applications such as in numerical linear algebra and in adaptive signal processing [17, 25, 26]. Recently, we investigated how to use this transformation to deal with the problem of direction-of-arrival (DOA) estimation of acoustic sources [21]. This paper presents our continued effort in the study of the Householder transformation and we showed how this transformation can be used to cope with the problem of multichannel noise reduction. We applied the Householder transformation to the multichannel noise reduction problem in the frequency domain, and developed a multichannel Wiener filter that can achieve noise reduction with or without the knowledge of the source incidence angle.

8. REFERENCES

[1] M. R. Schroeder. US Patent No. 3,180,936, filed Dec. 1, 1960, issued Apr. 27, 1965.

[2] M. R. Schroeder. US Patent No. 3,1403,224, filed May. 28, 1965, issued Sept. 24, 1968.

[3] J. Benesty, J. Chen, and E. A. Habets, *Speech Enhancement in the STFT Domain*. Berlin, Germany: Springer-Verlag, 2011.

[4] M. Souden, S. Araki, K. Kinoshita, T. Nakatani, and H. Sawada, "A multichannel MMSE-based framework for joint blind source separation and noise reduction," in *Proc. IEEE ICASSP*, 2012, pp. 109–112.

[5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, 1985.

[6] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, pp. 1218–1234, Jul. 2006.

[7] J. Yu, J. Benesty, G. Huang, and J. Chen, "Examples of optimal noise reduction filters derived from the squared Pearson correlation coefficient," in *Proc. IEEE ICASSP*, 2014, pp. 1552–1556.

[8] G. Huang, J. Benesty, T. Long, and J. Chen, "A family of maximum SNR filters for noise reduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 12, pp. 2034–2047, 2014.

[9] S. Gannot and I. Cohen, "Speech enhancement based on the general transfer function GSC and postfiltering," *IEEE Trans. Speech, Audio Process.*, vol. 12, no. 6, pp. 561–571, 2004.

[10] N. Ito, H. Shimizu, N. Ono, and S. Sagayama, "Diffuse noise suppression using crystal-shaped microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2101–2110, 2011.

[11] R. Talmon, D. Kushnir, R. R. Coifman, I. Cohen, and S. Gannot, "Parametrization of linear systems using diffusion kernels," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1159–1173, 2012.

[12] S. E. Nordholm, I. Claesson, and N. Grbic, "Performance limits in subband beamforming," *IEEE Trans. Speech, Audio Process.*, vol. 11, no. 3, pp. 193–203, 2003.

[13] A. Pezeshki, B. D. Van Veen, L. L. Scharf, H. Cox, and M. L. Nordenvaad, "Eigenvalue beamforming using a multirank MVDR beamformer and subspace selection," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1954–1967, 2008.

[14] G. Huang, J. Benesty, and J. Chen, "Superdirective beamforming based on the Krylov matrix," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 12, pp. 2531–2543, 2016.

[15] Y. Peled and B. Rafaely, "Objective performance analysis of spherical microphone arrays for speech enhancement in rooms," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1473–1481, 2012.

[16] G. H. Golub and C. F. V. Loan, *Matrix Computations. Third Edition*. Baltimore, Maryland: The Johns Hopkins University Press, 1996.

[17] R. McIlhenny and M. D. Ercegovac, "On the design of an on-line complex Householder transform," in *Proc. 40th Asilomar conference on signals, systems and computers*, pp. 318–322, 2006.

[18] A. O. Steinhardt, "Householder transforms in signal processing," *IEEE ASSP Magazine*, vol. 5, no. 3, pp. 4–12, 1988.

[19] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, vol. 1. Springer Science & Business Media, 2008.

[20] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.

[21] G. Huang, J. Chen, and J. Benesty, "Direction-of-arrival estimation of passive acoustic sources in reverberant environments based on the Householder transformation," *J. Acoust. Soc. Am.*, vol. 138, no. 5, pp. 3053–3060, 2015.

[22] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[23] F. Jacobsen, *The Diffuse Sound Field: Statistical Considerations Concerning the Reverberant Field in the Steady State*. Acoustics Laboratory, Technical University of Denmark, 1979.

[24] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multi-microphone noise reduction preserving binaural cues," in *Proc. IEEE ICASSP*, 2016, pp. 460–464.

[25] M. L. De Campos, S. Werner, and J. A. Apolinário Jr, "Householder-transform constrained LMS algorithms with reduced-rank updating," in *Proc. IEEE ICASSP*, 1999, vol. 4, pp. 1857–1860.

[26] M. L. De Campos, S. Werner, and J. A. Apolinário Jr, "Constrained adaptation algorithms employing Householder transformation," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2187–2195, 2002.