

Multiband Subspace Tracking Speech Enhancement for In-Car Human Computer Speech Interaction

CHUNG-HSIEN YANG, JIA-CHING WANG, JHING-FA WANG,
HSIAO-PING LEE, CHUNG-HSIEN WU* AND KAI-HSING CHANG

Department of Electrical Engineering

**Department of Computer Science and Information Engineering*

National Cheng Kung University

Tainan, 701 Taiwan

In this paper, a new subspace-based speech enhancement algorithm for in-car human computer speech interaction is presented. We first incorporate a perceptual filterbank which is derived from psycho-acoustic model with signal subspace approach to effectively suppress in-car noises of engine. Second, for real-time applications, a new subspace tracking algorithm is derived by modifying PASTd algorithm to solve the data dependent hazard of tracking algorithm. Six different types of in-car noises in TAICAR database are used in our evaluation. The experimental results demonstrate that our approach is superior to conventional subspace and spectral subtraction methods.

Keywords: human computer interaction, in-car noise, speech enhancement, PASTd algorithm, perceptual filterbank, subspace tracking

1. INTRODUCTION

With the increased use of Global Positioning System (GPS), mobile phones, wireless internet and electronic controls inside cars, there is a greater need of hands-free human computer interaction via speech. In a car environment, the speech signal is corrupted by the ambient noise coming from the car. The distance between a hands-free car microphone and the speaker makes this problem more serious. It is therefore essential to enhance noisy speech for robust human computer speech interaction in car environments.

Traditionally, the signal subspace approach has been successfully applied to frequency estimation, direction of arrival estimation, and system identification. It is only recently that it was applied to speech enhancement. The idea behind it is to project the noisy signal onto two subspaces: the signal-plus-noise subspace, or simply signal subspace (since the signal dominates this subspace), and the noise subspace. The noise subspace contains signals from the noise process only; hence an estimate of the clean signal can be made by removing the components of the signal in the noise subspace and retaining only the components of the signal in the signal subspace. The decomposition of the space into two subspaces can be done using either the singular value decomposition [1, 2] or the eigenvalue decomposition [3, 4].

Ephraim and Van Trees [3] proposed a subspace-based speech enhancement which it seeks for an optimal estimator that would minimize the speech distortion subject to the constraint that the residual noise fell below a preset threshold. Using the eigenvalue de-

Received August 16, 2005; accepted January 17, 2006.

Communicated by Jhing-Fa Wang, Pau-Choo Chung and Mark Billinghurst.

composition of the covariance matrix, it shows that the decomposition of the vector space of the noisy speech into a signal and noise subspace can be obtained by applying the Karhunen-Loeve transform (KLT) to the noisy speech. The KLT components representing the signal subspace were modified by a gain function determined by the estimator, while the remaining KLT components representing the noise subspace were nulled. The enhanced speech was obtained from the inverse KLT of the modified components.

In this paper, a new speech enhancement technique based on perceptual filterbank and subspace-based method is proposed. The perceptual filterbank is obtained by adjusting the decomposition tree structure of the conventional wavelet packet transform in order to approximate the critical bands of the psycho-acoustic model as close as possible. The reason is that the signal subspace methods do not operate in the frequency domain where the available hearing models are developed. Combining multiband processing with subspace decomposition, our algorithm is more capable to suppress color engine noise than conventional subspace approaches in car noisy environments.

Moreover, the subspace decomposition is achieved by using a subspace tracking algorithm to reduce the computational load. The tracking method is a normalized least-mean-square (NLMS) adaptive filter. It has computational complexity of linear order and is suitable for real time applications. Implementations of these techniques, however, have been based on batch eigenvalue decomposition of the sample correlation matrix or on singular value decomposition of data matrix. This approach is unsuitable for adaptive processing because the task is very time consuming. In order to overcome this difficulty, a number of algorithms have been proposed for tracking the dominant subspace [5-8]. Among the most robust and most efficient methods are the projection approximation subspace tracking (PAST) algorithm and its variant, the projection approximation subspace tracking deflation (PASTd) algorithm [7]. However, both algorithms are not suitable for fast pipeline computation because of their data dependant hazard. This motivates us to design an improved PASTd algorithm without data dependant hazard.

This paper is organized as follows. Section 2 gives an overview of the conventional subspace-based speech enhancement algorithm. In section 3, the proposed speech enhancement system for in-car noisy environments is introduced. It contains a perceptual filterbank which is used for subband processing and the new subspace tracking algorithm without data dependant hazard. The experimental results of speech enhancement are presented in section 4 and finally conclusion remarks are given in section 5.

2. SUBSPACE-BASED SPEECH ENHANCEMENT

The speech enhancement problem will be described as a clean speech signal \bar{x} being transmitted through a distortionless channel that is corrupted by additive noise \bar{n} . The resulting noisy speech signal \bar{y} can be expressed as

$$\bar{y} = \bar{x} + \bar{n}, \quad (1)$$

where $\bar{x} = [x_1, x_2, \dots, x_M]^H$, $\bar{n} = [n_1, n_2, \dots, n_M]^H$, and $\bar{y} = [y_1, y_2, \dots, y_M]^H$. The observation period has been denoted as M . Henceforth, the vectors \bar{x} , \bar{n} , \bar{y} will be considered as part of complex space C^M .

If it is assumed that clean speech is confined to a subspace of dimensionality K , where $K < M$, then C^M can be decomposed into two subspaces: a signal subspace and a noise subspace. Ephraim and Van Trees [3] realized this partitioning by postulating a linear model for the speech frame under analysis. The range and the null space were characterized as the signal and noise subspaces, respectively. The linear model for the clean speech assumes that every M -sample frame can be represented using the model:

$$\bar{x} = V\bar{s} = \sum_{i=1}^K s_i v_i, \quad K \leq M, \tag{2}$$

where $\bar{s} = [s_1, s_2, \dots, s_K]^H$ is a sequence of zero mean complex random variables. $V \in R^{M \times K}$ is known as the model matrix. Assuming that the columns of V are linearly independent, and then the rank of V is K . The range of V defines the signal subspace. The noise subspace is the null space of the model matrix. This subspace has rank $M - K$ and only contains vectors resulting from the noise process.

The subspace decomposition can be achieved using KLT, i.e. eigenvector matrix. Let R_x and R_y denote the covariance matrix of the \bar{x} and \bar{y} , respectively. The eigen-decomposition is performed on the covariance matrix R_x and the following form is obtained

$$R_x = [U_1 U_2] \begin{bmatrix} A_{x1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix}, \tag{3}$$

where A_{x1} is a $K \times K$ diagonal matrix with eigenvalues $\lambda_x(1), \lambda_x(2), \dots, \lambda_x(K)$ as diagonal elements. The eigenvector matrix U has been partitioned into two sub-matrices, U_1 and U_2 . The matrix U_1 contains eigenvectors corresponding to non-zero eigenvalues. These eigenvectors form a basis for the signal subspace. Meanwhile, U_2 contains the eigenvectors which span the noise subspace.

Let I_1 and I_2 represent the identity matrices $I_{K \times K}$ and $I_{(M-K) \times (M-K)}$, respectively. Similar to Eq. (3), the eigen-decomposition of R_y is given by

$$R_y = [U_1 U_2] \begin{bmatrix} A_{y1} & 0 \\ 0 & \sigma_n^2 I_2 \end{bmatrix} \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix} = [U_1 U_2] \begin{bmatrix} A_{x1} + \sigma_n^2 I_1 & 0 \\ 0 & \sigma_n^2 I_2 \end{bmatrix} \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix}, \tag{4}$$

where A_{y1} is a $K \times K$ diagonal matrix with eigenvalues $\lambda_y(1), \lambda_y(2), \dots, \lambda_y(K)$ as diagonal elements.

As indicated by Eq. (4), the clean speech lies only within the signal subspace while the noise spans the entire space. Therefore, only the contents of the signal subspace are used to estimate the clean speech signal.

The clean speech can be estimated using a linear estimator

$$\hat{x} = H\bar{y}, \tag{5}$$

which H is a $K \times K$ matrix. The residual signal, \bar{e} , can then be represented as

$$\bar{e} = \hat{x} - \bar{x} = H\bar{y} - \bar{x} = H(\bar{x} + \bar{n}) - \bar{x} = (H - I)\bar{x} + H\bar{n} = \bar{e}_x + \bar{e}_n \tag{6}$$

where \bar{e}_x refers to the signal distortion while \bar{e}_n denotes the residual noise. The energy of the signal distortion can be calculated from Eq. (6)

$$\varepsilon_x^2 = \text{tr}E\{\bar{e}_x\bar{e}_x^H\} = \text{tr}\{(H-I)R_x(H-I)^H\}. \quad (7)$$

Similarly, the energy of the residual noise can be derived from Eq. (7)

$$\varepsilon_n^2 = \text{tr}E\{\bar{e}_n\bar{e}_n^H\} = \sigma_n^2 \text{tr}\{HH^H\}. \quad (8)$$

The energy of the total error, ε thus can be calculated as

$$\varepsilon^2 = \varepsilon_x^2 + \varepsilon_n^2. \quad (9)$$

The time domain constrained estimator minimizes signal distortion while constraining the average residual noise power to be less than $\alpha\sigma_n^2$. Thus

$$H_{opt} = \arg \min_H \varepsilon_x^2$$

$$\text{subject to: } \frac{1}{M} \varepsilon_n^2 \leq \alpha\sigma_n^2 \quad (10)$$

where $0 \leq \alpha \leq 1$. The resulting filter from the TDC estimation has the form

$$H_{opt} = R_x(R_x + \gamma\sigma_n^2 I)^{-1}, \quad (11)$$

where γ is the Lagrange multiplier. Applying the eigen-decomposition Eq. (3) of R_x to Eq. (11), we can rewrite the optimal linear estimator as

$$H_{opt} = U \begin{bmatrix} G_\gamma & 0 \\ 0 & 0 \end{bmatrix} U^H \quad (12)$$

where

$$G_\gamma = \Lambda_{x1}(\Lambda_{x1} + \gamma\sigma_n^2 I_1)^{-1}. \quad (13)$$

Hence, the signal estimate $\hat{x} = H_{opt}\bar{y}$ is obtained by applying the KLT to the noisy signal, appropriately modifying the components of the KLT $U^H\bar{y}$ by a gain function, and by inverse KLT of the modified components.

3. PROPOSED SPEECH ENHANCEMENT SYSTEM

The proposed speech enhancement system based on subspace tracking is shown in Fig. 1. The input signal is divided into subband time series by the analysis filterbank. Following subband analysis, the vector of subband signal is presented to the subspace tracking algorithm to extract eigenvectors. Then, the gain adaptation is performed to estimate the clean speech. To reconstruct the enhanced full-band speech, the subband

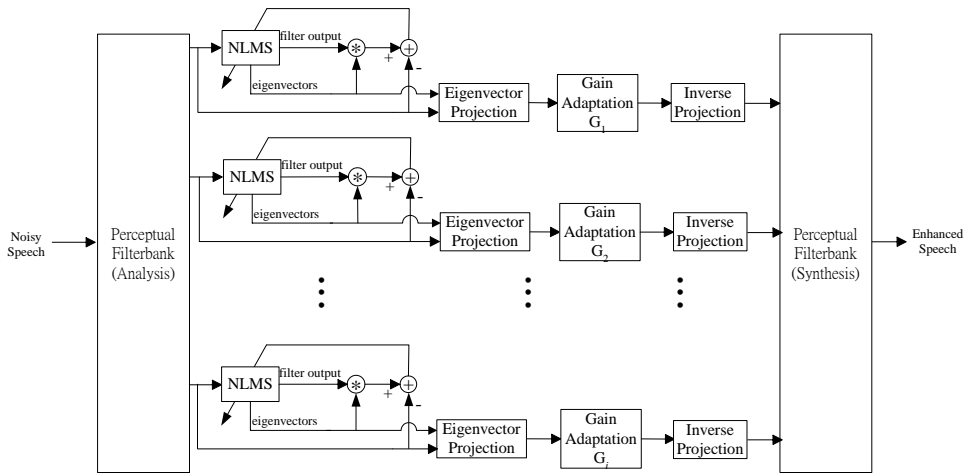


Fig. 1. Block diagram of proposed speech enhancement system.

synthesizer is applied to the gain-modified vector of subband signal. The following sections will describe more details of the proposed method.

3.1 Subspace-Based Speech Enhancement Using Perceptual Filterbank

The perceptual filterbank is obtained by adjusting the decomposition tree structure of the conventional wavelet packet transform in order to approximate the critical bands of the psycho-acoustic model as close as possible. The primary reason for embedding the psycho-acoustic model in the filterbank is that humans are capable of detecting the desired speech in a noisy environment without prior knowledge of the noise. One class of critical band scales is called Bark scale. The Bark scale z can be approximately expressed in terms of the linear frequency by

$$Z(f) = 13\arctan(7.6 \times 10^{-4}f) + 3.5\arctan(1.33 \times 10^{-4}f)^2, \tag{14}$$

where f is the linear frequency in Hertz. The corresponding critical bandwidth (CBW) of the center frequencies can be expressed by

$$CBW(f_c) = 25 + 75(1 + 1.4 \times 10^{-6}f_c^2)^{0.69}, \tag{15}$$

where f_c is the center frequency (unit: Hertz). Theoretically, the range of human auditory frequency spreads from 20 to 20000 Hz and covers approximately 25 Barks. In this paper, the underlying sampling rate was chosen to be 8 kHz, yielding a bandwidth of 4 kHz. Within this bandwidth, there are approximately 17 critical bands as listed in Table 1.

According to the specifications of center frequencies, CBW, lower and upper cutoff frequencies given in Table 1, the tree structure of the perceptual wavelet packet transform can be constructed as shown in Fig. 2 (a). The corresponding frequency bandwidth of the wavelet packet tree is shown in Fig. 2 (b). It contains 16 decomposition cells with 5 decomposition stages to approximate these 17 critical bands which are corresponding

Table 1. Characteristics of critical bands under 4 kHz.

Critical Band Number	Center Frequency (Hz)	CBW	Lower Cutoff Frequency (Hz)	Upper Cutoff Frequency (Hz)
1	50	-	-	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1000	160	920	1080
10	1170	190	1080	1270
11	1370	210	1270	1480
12	1600	240	1480	1720
13	1850	280	1720	2000
14	2150	320	2000	2320
15	2500	380	2320	2700
16	2900	450	2700	3150
17	3400	550	3150	3700

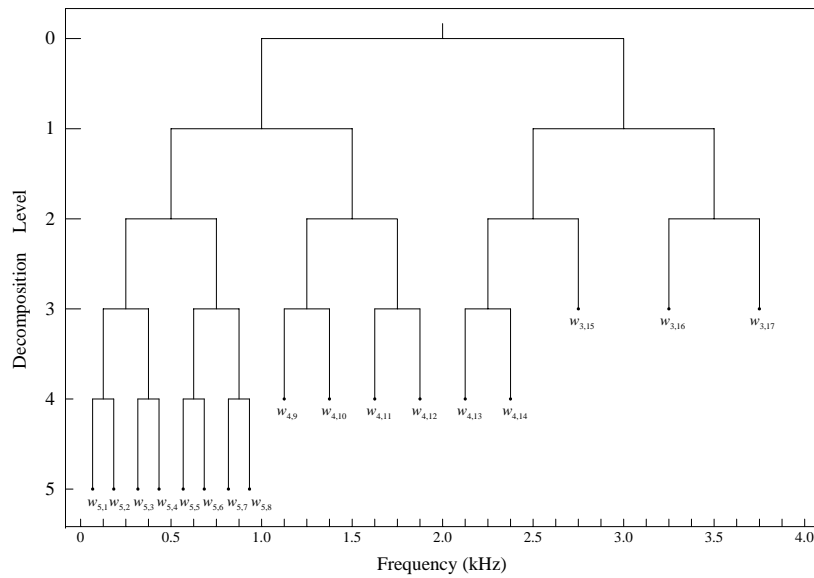


Fig. 2. (a) Tree structure of the perceptual filterbank.

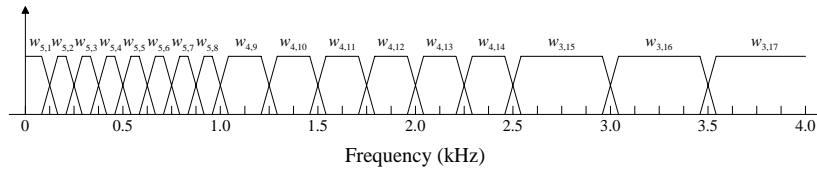


Fig. 2. (b) The frequency bandwidths for the perceptual filterbank.

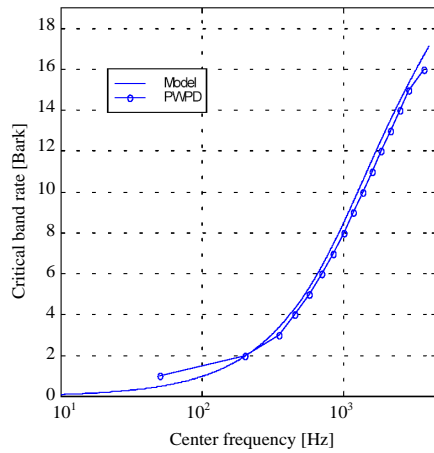


Fig. 3. Bark scale as a function of center frequency.

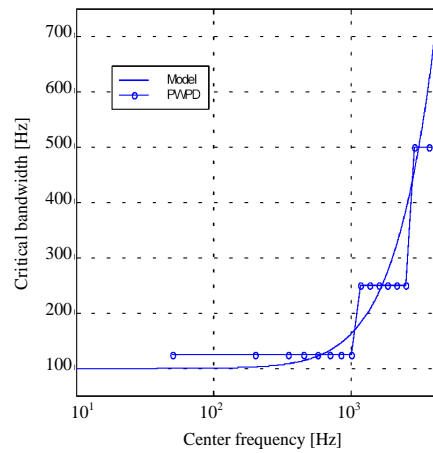


Fig. 4. Critical bandwidth as a function of center frequency.

to wavelet packet coefficient sets $w_{j,m}$, where $j = 3, 4, 5$, $m = 1, \dots, 17$. The resulting 17-band perceptual wavelet packet transform of the Bark scale and the CBW are also plotted in Figs. 3 and 4, respectively.

To suppress the car noise of each band, decomposing noisy signal into critical band signals will be able to choose different gain values for each subband to minimize speech distortions.

3.2 Subspace Tracking Using Improved PASTd Algorithm

Projection approximation subspace tracking (PAST) algorithm is an adaptive method that tracks eigenvectors of covariance matrix using a recursive least square (RLS) type algorithm. The performance of this method is extensively analyzed in [9, 10]. The PAST algorithm requires $3nr + O(r^2)$ operations every updating. A variant PAST algorithm named projection approximation subspace tracking deflation (PASTd) is designed based on the deflation technique. The basic idea of the deflation technique is the sequential estimation of the eigencomponents. Although the PASTd algorithm is normalization free and able to update eigenvalues and eigenvectors every $4nr + O(r)$ operations, its algorithm flow has data dependant hazard occurred at (a.1) and (a.2). While updating x_{i+1} , its previous content is still required in next loop. This hazard makes the PASTd algorithm not suitable for pipeline computation.

Algorithm The PASTd Algorithm

Choose $d_i(0)$ and $\bar{u}_i(0)$ suitably

For $n = 1, 2, \dots$, Do

$$\bar{x}_1(n) = \bar{x}(n)$$

For $i = 1, 2, \dots, r$ Do

$$v_i(n) = \bar{u}_i^H(n-1)\bar{x}_i(n)$$

$$d_i(n) = \beta d_i(n-1) + |v_i(n)|^2 \quad (\text{a.1})$$

$\begin{aligned}\bar{e}_i(n) &= \bar{x}_i(n) - \bar{u}_i(n-1)v_i(n) \\ \bar{u}_i(n) &= \bar{u}_i(n-1) + \bar{e}_i(n)[v_i(n)/d_i(n)] \\ \bar{x}_{i+1}(n) &= \bar{x}_i(n) - \bar{u}_i(n)v_i(n)\end{aligned}$ <p style="text-align: right;">(a.2)</p> <p>end</p> <p>end</p>
--

In this paper, an improved PASTd algorithm without data dependant hazard is proposed. For the derivation of the improved PASTd algorithm, (a.1) is rewritten as

$$\begin{aligned}v_i(n) &= \bar{u}_i^H(n-1)\bar{x}_i(n) \\ &= \bar{u}_i^H(n-1)(\bar{x}_{i-1}(n) - \bar{u}_{i-1}(n)v_{i-1}(n)) \\ &= \bar{u}_i^H(n-1)\bar{x}_{i-1}(n) - \bar{u}_i^H(n-1)\bar{u}_{i-1}(n)v_{i-1}(n), \quad \text{for } i \geq 2 \\ &\because U(n) \text{ converges to orthogonal matrix} \\ &\therefore \bar{u}_i^H(n-1)\bar{u}_{i-1}(n) \approx 0.\end{aligned}\tag{16}$$

Since $U(n)$ converges to an orthogonal matrix, the $\bar{u}_i^H(n-1)\bar{u}_{i-1}(n)$ will approach to zero in steady state and the equation is rewritten as follows

$$\begin{aligned}v_i(n) &= \bar{u}_i^H(n-1)\bar{x}_{i-1}(n) \\ &= \bar{u}_i^H(n-1)\bar{x}_{i-2}(n) \\ &= \bar{u}_i^H(n-1)\bar{x}_{i-3}(n) = \dots \\ &= \bar{u}_i^H(n-1)\bar{x}_1(n), \quad \text{for } i \geq 2.\end{aligned}\tag{17}$$

Hence, $\bar{u}_i^H(n-1)\bar{x}(n)$ depends only on the input data at time n .

Now, it implies the relation between $\bar{e}_i(n)$ and $\bar{x}_{i+1}(n)$,

$$\bar{x}_{i+1}(n) = \bar{x}_i(n) - \bar{u}_i(n)v_i(n)\tag{18}$$

and

$$\bar{u}_i(n) = \bar{u}_i(n-1) + \bar{e}_i(n)[v_i^*(n)/d_i(n)],\tag{19}$$

then $\bar{x}_{i+1}(n)$ is obtained as

$$\bar{x}_{i+1}(n) = \bar{e}_i(n) \left(1 - \frac{v_i^2(n)}{d_i(n)} \right),\tag{20}$$

where

$$1 - \frac{v_i^2(n)}{d_i(n)} = \frac{d_i(n) - v_i^2(n)}{d_i(n)} = \frac{\beta d_i(n-1)}{\beta d_i(n-1) + v_i^2(n)}.\tag{21}$$

Let $\mu_i(n-1)$ denote as

$$\mu_i(n-1) = \frac{1}{d_i(n-1)}, \quad (22)$$

$$\frac{\beta}{\beta + \mu_i(n-1)v_i^2(n)} = \beta\alpha_i(n), \quad (23)$$

$$\bar{x}_{i+1}(n) = \beta\alpha_i(n)\bar{e}_i(n), \quad (24)$$

where

$$\alpha_i(n) = \frac{1}{\beta + \mu_i(n-1)v_i^2(n)}. \quad (25)$$

For $i \geq 2$,

$$\begin{aligned} \bar{e}_{i+1}(n) &= \bar{x}_{i+1}(n) - \bar{u}_{i+1}(n-1)v_{i+1}(n) \\ &= \beta\alpha_i(n)\bar{e}_i(n) - \bar{u}_{i+1}(n-1)v_{i+1}(n), \quad \text{substitute } i \text{ to } i-1. \end{aligned} \quad (26)$$

$$\bar{e}_i(n) = \beta\alpha_{i-1}(n)\bar{e}_{i-1}(n) - \bar{u}_i(n-1)v_i(n), \quad \text{for } i \geq 2. \quad (27)$$

The principal term $d_i(n)$ is

$$\mu_i(n) = \frac{1}{d_i(n)} = \frac{1}{\beta d_i(n-1) + v_i^2(n)} = \frac{\mu_i(n-1)}{\beta + \mu_i(n-1)v_i^2(n)}. \quad (28)$$

From the above derivations, the inner loop of improved PASTd algorithm is summarized as follows. The data dependant hazard is solved, i.e. $\bar{e}_i(n)$ can be computed recursively without waiting $\bar{u}_i(n)$.

Algorithm The Improved PASTd Algorithm

For $i = 1, \dots, r$

$$v_i(n) = \bar{u}_i^H(n-1)\bar{x}(n) \quad (b.1)$$

$$z_i(n) = \mu_i(n-1)|v_i(n)|^2 \quad (b.2)$$

$$\alpha_i(n) = \frac{1}{\beta + z_i(n)} \quad (b.3)$$

$$\bar{e}_i(n) = \begin{cases} \bar{x}_i(n) - \bar{u}_i(n-1)v_i(n) & i = 1 \\ \beta\alpha_{i-1}(n)\bar{e}_{i-1}(n) - \bar{u}_i(n-1)v_i(n) & i \geq 2 \end{cases} \quad (b.4)$$

$$\mu_i(n) = \alpha_i(n)\mu_i(n-1) \quad (b.6)$$

$$\bar{u}_i(n) = \bar{u}_i(n-1) + \mu_i(n)\bar{e}_i(n)v_i(n) \quad (b.7)$$

end

The data flow graphs of PASTd and improved PASTd algorithms are described in Figs. 5 and 6, respectively. It can be seen that each iteration of the improved PASTd algorithm can be scheduled based on pipeline computation. Let $x(n)$ be a noisy speech with 8-bit samples and 8 kHz sampling rate. The sample correlation matrix R_x is updated via

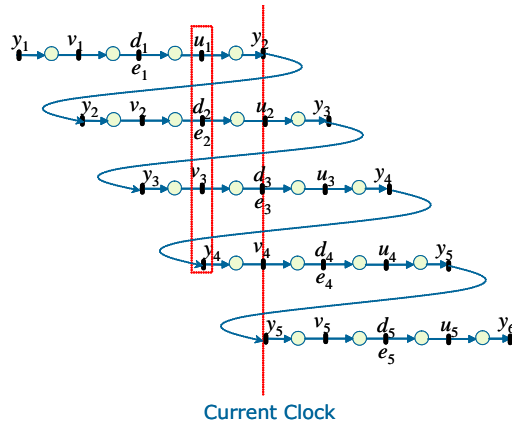


Fig. 5. Data dependant hazard of PASTd algorithm.

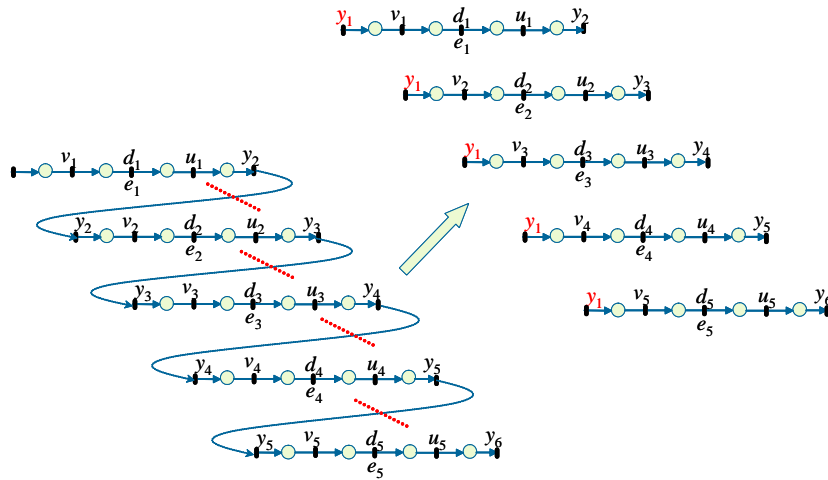


Fig. 6. Pipeline computation of proposed PASTd algorithm.

$$\hat{R}_x(n) = \sum_{i=1}^n \beta^{n-i} \bar{x}(i) \bar{x}^H(i). \tag{29}$$

Its complete eigen-decomposition is computed by a standard batch method for obtaining the signal subspace. The mean-square error (MSE) between the estimated and the actual largest eigenvector was used as the comparison criterion. Let β be equal to 0.98 and U is an identity matrix. The learning curves of original and improved PASTd algorithms are shown in Figs. 7 and 8, respectively. The curves were obtained by averaging 100 runs. It can be seen from the simulation that proposed PASTd algorithm can reach the steady-state MSE same as PASTd algorithm. With the same steady-state MSE as PASTd algorithm, the proposed PASTd algorithm is more attracted for real-time hardware implementation because its data flow can be realized in fast pipeline computation.

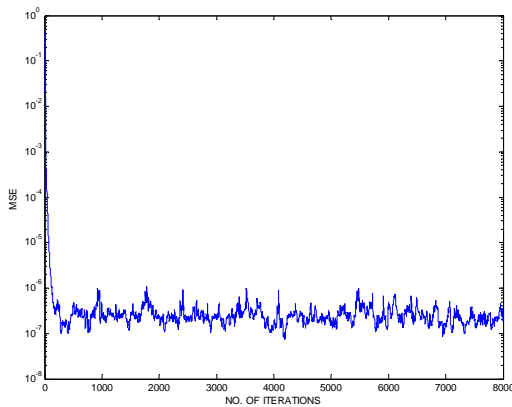


Fig. 7. Learning curve of PASTd algorithm.

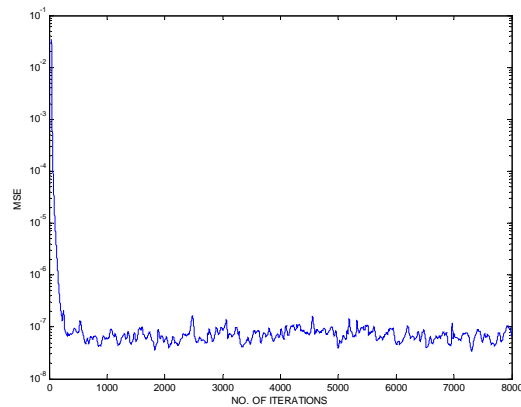


Fig. 8. Learning curve of proposed PASTd algorithm.

4. THE EXPERIMENTAL RESULTS

In this section, an analysis of the performance of proposed method is presented. For evaluation purposes, we used three sentences sampled at 8 kHz. For car interior noise, six noises measured from different cars in TAICAR database [11] were adopted. The experiment was performed using natural speeches corrupted by additive in-car noises. For comparative purposes, we also implemented and evaluated the spectral subtraction method of Berouti *et al.* [12] and the conventional subspace method in [3]. Fig. 9 shows the waveforms and spectrograms of degraded speech and enhanced speech processed by three algorithms: (1) spectral subtraction; (2) conventional subspace method; (3) the proposed approach.

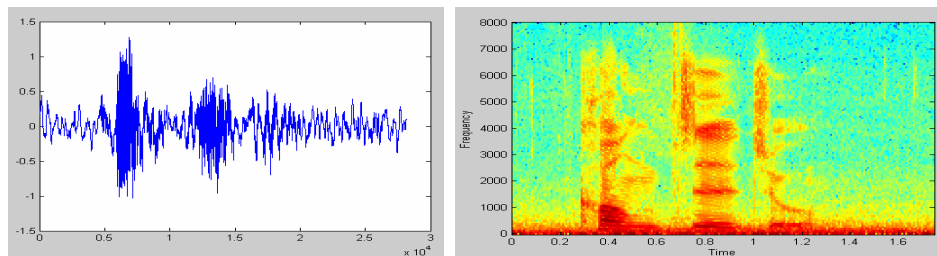


Fig. 9. (a) The waveform and spectrogram of noisy speech signal.

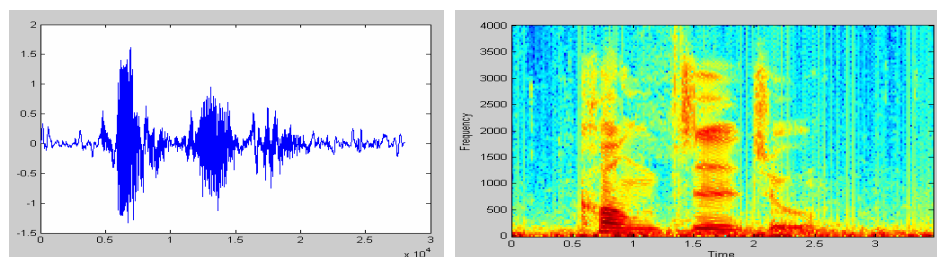


Fig. 9. (b) The waveform and spectrogram of enhanced speech using spectral subtraction method.

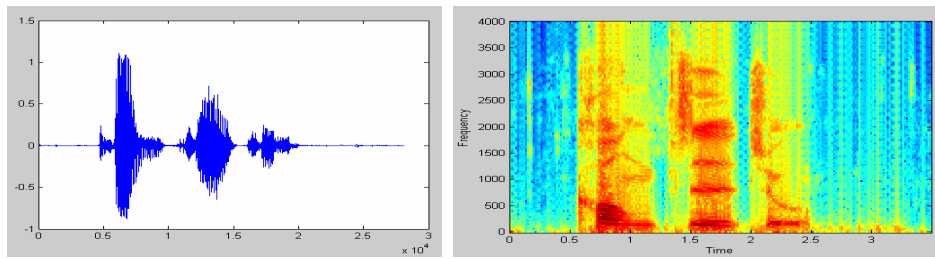


Fig. 9. (c) The waveform and spectrogram of enhanced speech using conventional subspace method.

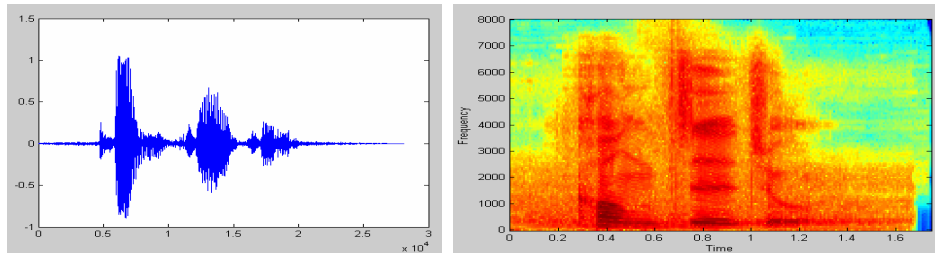


Fig. 9. (d) The waveform and spectrogram of enhanced speech using the proposed method.

Table 2. Performance comparison in SNR (dB) for sentences corrupted by six different in-car noises.

Noise Source	Noisy Speech	Spectral Subtraction	Conventional Subspace	Proposed
Honda	0.3261	4.0408	9.0472	13.7227
Toyota	-0.5933	3.9228	5.8111	7.8719
Opel	-0.2285	1.1338	8.3881	9.6808
Volvo	-0.8155	-1.2760	7.3737	12.5982
Nissan	-0.0630	2.9627	9.7108	11.5947
Ford	0.1348	1.0998	8.6697	10.3389
Average	-0.2066	1.9807	8.1668	10.9679

For objective evaluation, the overall (global) SNR measure was used to evaluate these speech enhancement algorithms. The calculation of SNR is given by

$$SNR = 10 \log_{10} \left(\frac{\sum_{n=1}^N |x(n)|^2}{\sum_{n=1}^N |x(n) - \hat{x}(n)|^2} \right), \quad (29)$$

where $x(n)$ and $\hat{x}(n)$ denote the clean and enhanced speech signals, respectively.

The performance comparison using SNR is given in Table 2. The SNR of speech signals corrupted by additive in-car noise is ranged from -1 to 1 dB and averagely -0.2066 dB. The SNR of enhanced speech using the proposed approach is 10.9679 dB in average. In other words, our method has SNR improvement averagely about 11 dB from

noisy speech. Compared with the spectral subtraction and the conventional subspace methods, the proposed approach has significant improvements about 8.9 dB and 2.8 dB, respectively. These experimental results demonstrate the superiority of the proposed speech enhancement algorithm.

5. CONCLUSIONS

For human computer speech interaction in car noisy environment, a new subspace-based speech enhancement algorithm is presented. The proposed method incorporates psycho-acoustic model (perceptual filterbank) by adjusting the decomposition tree structure of the conventional wavelet packet transform. Applying the perceptual filterbank, we are able to suppress varied types of in-car noises. Moreover, this paper uses subspace tracking strategy to reduce the computational load. An improved PASTd algorithm without data dependant hazard is also presented so that the subspace tracking procedure can be performed in pipeline fashion. Future research on real-time hardware implementation for the proposed system is also underway.

REFERENCES

1. M. Dendrinis, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise: a regenerative approach," *Speech Communication*, Vol. 10, 1991, pp. 45-57.
2. S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Transactions on Speech and Audio Processing*, Vol. 3, 1995, pp. 439-448.
3. Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, Vol. 3, 1995, pp. 251-266.
4. A. Rezaeey and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, Vol. 9, 2001, pp. 87-95.
5. I. Karasalo, "Estimating the covariance matrix by signal subspace averaging," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 34, 1986, pp. 8-12.
6. R. DeGroat, "Noniterative subspace tracking," *IEEE Transactions on Signal Processing*, Vol. 40, 1992, pp. 571-577.
7. B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, Vol. 43, 1995, pp. 95-107.
8. P. Strobach, "Bi-iteration SVD subspace tracking algorithms," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 45, 1997, pp. 1222-1240.
9. V. Solo and X. Kong, "Performance analysis of adaptive eigenanalysis algorithms," *IEEE Transactions on Signal Processing*, Vol. 46, 1995, pp. 636-646.
10. T. Gustafson, "Instrumental variable subspace tracking using projection approximation," *IEEE Transactions on Signal Processing*, Vol. 46, 1998, pp. 669-681.
11. H. C. Wang, C. H. Yang, J. F. Wang, C. H. Wu, and J. T. Chien "TAICAR-the collection and annotation of an in-car speech database created in Taiwan," *International Journal of Computational Linguistics and Chinese Language Processing*, Vol. 10, 2005, pp. 237-250.

12. M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1979, pp. 208-211.



Chung-Hsien Yang (楊宗憲) received the B.S. degree in Computer Science and Information Engineering from Tunghai University in 1997 and the M.S. degree in Computer Science and Information Engineering from National Cheng Kung University, Tainan, Taiwan, in 1999. He is a Ph.D. candidate in the Department of Electrical Engineering at National Cheng Kung University. His research areas include speech recognition and speech enhancement.



Jia-Ching Wang (王家慶) received the M.S. and Ph.D. degrees in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan, in 1997 and 2002, respectively. His research interests include signal processing and VLSI architecture design. Dr. Wang is a member of Phi Tau Phi. He is also a member of IEEE, ACM, and International Speech Communication Association (ISCA).



Jhing-Fa Wang (王駿發) is now a Chair Professor in National Cheng Kung University, Tainan, Taiwan. He received his Master and Bachelor degrees in the Department of Electrical Engineering from National Cheng Kung University, Taiwan in 1979 and 1973, respectively and Ph.D. degree in the Department of Computer Science and Electrical Engineering from Stevens Institute of Technology, U.S.A. in 1983. He was elected as an IEEE Fellow in 1999 and now the Chairman of IEEE Tainan Section. He got outstanding awards from Institute of Information Industry in 1991 and National Science Council of Taiwan in 1990, 1995, and 1997, respectively. He has been invited to give keynote speech in PACLIC 12 (Pacific Asia Conference on Language, Information and Computation), Singapore and served as the general chairman of International Symposium on Communication (ISCOM 2001), Taiwan. He has developed a Mandarin speech recognition system called Venus-Dictate known as a pioneering system in Taiwan. He was an associate editor for IEEE Transaction on Neural Networks and VLSI System. He is currently leading a research group of different disciplines for the development of "Advanced Ubiquitous Media for

Created Cyberspace.” He has published about 91 journal papers and 217 conference papers and obtained 5 patents since 1983. His research areas include wireless content-based media processing, speech recognition and natural language understanding.



Hsiao-Ping Lee (李曉屏) received the B.S. and M.S. degrees in Electrical Engineering from National Central University, Chungli, Taiwan, in 1999 and 2001, respectively. She is currently working toward the Ph.D. degree in Electrical Engineering at National Cheng Kung University. Her current research interests include speech enhancement and independent component analysis applied to speech processing.



Chung-Hsien Wu (吳宗憲) received the B.S. degree in Electronics Engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1981, and the M.S. and Ph.D. degrees in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan, R.O.C., in 1987 and 1991, respectively. Since August 1991, he has been with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan. He became a professor in August 1997. From 1999 to 2002, he served as the Chairman of the Department. He also worked at Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, in summer 2003 as a visiting scientist. His research interests include speech recognition, text-to-speech, multimedia information retrieval, spoken language processing and sign language processing for hearing-impaired. Dr. Wu is a senior member of IEEE and a member of International speech communication association (ISCA) and ROCLING.



Kai-Hsing Chang (張凱行) received the B.S. degree in Computer and Communication Engineering from National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan, in 2002, and the M.S. degree in Computer Science and Information Engineering from National Cheng Kung University, Tainan, Taiwan, in 2004. His research areas include adaptive filter theory and VLSI design.