

# A DCT-BASED FAST ENHANCEMENT TECHNIQUE FOR ROBUST SPEECH RECOGNITION IN AUTOMOBILE USAGE

Jun Huang<sup>1</sup>

Yunxin Zhao<sup>2</sup>

Stephen Levinson<sup>3</sup>

<sup>1,3</sup>Beckman Institute and Dept. of ECE, University of Illinois, Urbana, IL 61801, USA

<sup>2</sup>Department of CECS, University of Missouri, Columbia, MO 65211, USA

<sup>1</sup>jhuang@ifp.uiuc.edu    <sup>2</sup>zhao@cecs.missouri.edu    <sup>3</sup>sel@ifp.uiuc.edu

## ABSTRACT

In this paper, a fast computational method is proposed to approximate the Karhunen-Loève transform (KLT) in signal-subspace-based speech enhancement algorithm. The discrete cosine transform (DCT) is shown to be a good approximation of KLT for the covariance matrix of the autoregressive process of order  $p$  (AR( $p$ )). A fast algorithm which reduces the computation of eigenvalues of an  $N \times N$  symmetric Toeplitz matrix from  $O(N^3)$  in KLT to  $O(N^2)$  is developed. Experiment results demonstrate that the performance of the fast algorithm is very close to that of the KLT-based method in robust speech recognition in car environment while significantly reduces the computation time. An acoustic normalization scheme is also found to be useful to compensate the mismatch between the training and test conditions and thus further improves the recognition performance.

Keywords: Karhunen-Loève transform, discrete cosine transform, autoregressive process, signal subspace, cepstral mean normalization

## 1. INTRODUCTION

Recently, Karhunen-Loève transform (KLT) has attracted attention for its applications in speech enhancement and robust automatic speech recognition. A signal subspace method was proposed by Ephraim et al for speech enhancement under additive white noise condition [1]. Huang and Zhao further improved the signal subspace method by introducing an energy constraint to better enhance the low-energy, noise-like segments in speech signal, and by introducing a prewhitening procedure to deal with colored noise corruption [2]. In both methods, Karhunen-Loève transform is needed for the eigen decomposition of the short-time covariance matrices of the speech signal. For an analysis frame length of  $N$ , the cost of computing the eigenvalues of the signal correlation matrix is  $O(N^3)$ , which is too high for real-time implementation of the signal subspace based methods. In this paper, we introduce a fast algorithm to approximate the Karhunen-Loève transform of an AR( $p$ ) process, and show its applicability in speech enhancement. We prove that discrete cosine transform (DCT) is a good approximation to KLT for the covariance matrix of an AR( $p$ ) process. We derive a fast algorithm of computational cost  $O(N^2)$  for computing eigenvalues based on DCT. Furthermore, an acoustic normalization method is used to compensate the mismatched training and test conditions [3]. Experiment results demonstrate that this fast algorithm achieved results very close to that of KLT for noisy speech recognition in car environment. The acoustic normalization method can further improve the recognition performance under various driving conditions. This paper is organized

as follows. In Section 2, the noncausal minimum variance representation (MVR) of an AR( $p$ ) process is derived and a fast algorithm to approximate KLT is obtained. In Section 3, the system diagram of the fast DCT-based energy-constrained signal-subspace (ECSS) speech enhancement algorithm is described. In Section 4, experiment results are provided for connected-digit recognition in real car environment. A conclusions is given in Section 5.

## 2. A FAST ALGORITHM TO APPROXIMATE KL TRANSFORM

In this section, we first prove that DCT is a good approximation to KLT for an AR( $p$ ) process. The basic idea is to use the noncausal minimum variance representation (MVR) [4] to represent the AR( $p$ ) process. The inverse of the covariance matrix of the AR( $p$ ) process given the boundary values is proven to be a banded symmetric Toeplitz matrix whose first  $p + 1$  entries in the first column are determined by the AR coefficients and the rest entries in the first column are all zeros. It can be shown that DCT is a good approximation to KLT for this type of matrix. Furthermore, a fast algorithm which requires only  $N^2$  multiplications is derived to compute the eigenvalues of the symmetric Toeplitz covariance matrix using the DCT approximation.

Consider an AR( $p$ ) process given by:

$$u(n) = \sum_{k=1}^p a(k)u(n-k) + \epsilon(n) \quad (1)$$

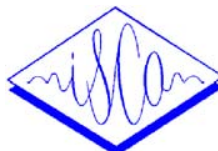
$$E\{\epsilon(n)\} = 0; E\{\epsilon^2(n)\} = \beta^2; E\{\epsilon(n)u(m)\} = 0, m < n \quad (2)$$

The PSD of the AR( $p$ ) process is given by:

$$S(e^{j\omega}) = \frac{\beta^2}{(1 - \sum_{k=1}^p a(k)e^{-jk\omega})(1 - \sum_{k=1}^p a(k)e^{jk\omega})} \quad (3)$$

so that:

$$\begin{aligned} \frac{1}{S(e^{j\omega})} &= \frac{1}{\beta^2} \left[ \left(1 - \sum_{k=1}^p a(k)e^{-jk\omega}\right) \left(1 - \sum_{k=1}^p a(k)e^{jk\omega}\right) \right] \\ &= \frac{1}{\beta^2} \left[ \sum_{k=0}^p |a(k)|^2 + 2 \sum_{k=1}^p \sum_{n=0}^{p-k} a(n)a(n+k) \cos(k\omega) \right] \\ &= \sum_{k=-\infty}^{+\infty} s(k)e^{-jk\omega} \end{aligned} \quad (4)$$



From Eq. (4), we get:

$$s(k) = \begin{cases} \frac{1}{\beta^2} \sum_{n=0}^{p-|k|} a(n)a(n+|k|) & 1 \leq |k| \leq p \\ \frac{\sum_{n=0}^p |a(n)|^2}{\beta^2} = \frac{C^2}{\beta^2} & k = 0 \\ 0 & |k| > p \end{cases} \quad (5)$$

Applying the noncausal MVR theorem in [4], we can get the noncausal MVR of the AR( $p$ ) process as following:

$$u(n) = \sum_{k=1}^p \alpha(k)[u(n-k) + u(n+k)] + v(n) \quad (6)$$

$$r_v(n) = -\sigma_v^2 \alpha(n), \quad n = 0, 1, 2, \dots \quad (7)$$

$$\alpha(k) = -\sum_{n=0}^{p-k} \frac{a(n)a(n+k)}{C^2}, \quad k = 1, 2, \dots, p \quad (8)$$

where  $C^2 = \sum_{n=0}^p |a(n)|^2$ ,  $a(0) = -1$ ,  $\sigma_v^2 = \frac{\beta^2}{C^2}$ .

Now consider an  $N \times 1$  vector  $\vec{u}$  consisting of samples of an AR( $p$ ) process  $\{u(n), 1 \leq n \leq N\}$  and an  $N \times 1$  vector  $\vec{v}$  consisting of samples of the noncausal prediction error  $\{v(n), 1 \leq n \leq N\}$ . Here we assume  $N > 2p$ , where  $p$  is the order of the AR process. We can write the noncausal MVR in Eq. (6) into the matrix form:

$$Q\vec{u} = \vec{v} + \vec{b} \quad (9)$$

where:

$$Q = \begin{bmatrix} 1 & -\alpha(1) & \dots & -\alpha(p) & 0 & \dots & 0 \\ -\alpha(1) & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ -\alpha(p) & \ddots & \ddots & \ddots & \ddots & \ddots & -\alpha(p) \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & -\alpha(1) \\ 0 & \dots & 0 & -\alpha(p) & \dots & -\alpha(1) & 1 \end{bmatrix} \quad (10)$$

$$\vec{b} = \begin{bmatrix} \sum_{k=1}^p \alpha(k)u(1-k) \\ \dots \\ \alpha(p)u(0) \\ 0 \\ \dots \\ 0 \\ \alpha(p)u(N+1) \\ \dots \\ \sum_{k=1}^p \alpha(k)u(N+k) \end{bmatrix} \quad (11)$$

where  $\alpha(k), 1 \leq k \leq p$  are given in Eq. (8).

It can be seen from Eq. (10) that  $Q$  is a banded symmetric Toeplitz matrix; its entries are all zero except within the band  $|i-j| \leq p$ . We will prove in the following that DCT is a good approximation to KLT for this type of matrix and thus is a good approximation to KLT for the covariance matrix of an AR( $p$ ) process given the boundary values.

#### Lemma 1

DCT is a good approximation to KLT for the symmetric Toeplitz matrix  $Q$  given in Eq. (10).

Proof:

Define  $Q_k$  to be a symmetric Toeplitz matrix with its only nonzero elements located on the upper and lower  $k^{\text{th}}$  subdiagonals,

We can decompose the matrix  $Q$  into a sum of  $p+1$  matrices:

$$Q = I - \sum_{k=1}^p \alpha(k)Q_k \quad (12)$$

Now consider matrix  $Q_1$ . Define vectors  $\vec{h}_j, 1 \leq j \leq N$  as following:

$$\vec{h}_j = \begin{bmatrix} \cos(\frac{(j-1)\pi}{2N}) \\ \cos(\frac{(j-1)3\pi}{2N}) \\ \vdots \\ \cos(\frac{(j-1)(2N-1)\pi}{2N}) \end{bmatrix} \quad (13)$$

So that:

$$Q_1 \cdot \vec{h}_j = \begin{bmatrix} \cos(\frac{(j-1)3\pi}{2N}) \\ 2\cos(\frac{(j-1)\pi}{N})\cos(\frac{(j-1)3\pi}{2N}) \\ \vdots \\ 2\cos(\frac{(j-1)\pi}{N})\cos(\frac{(j-1)(2N-3)\pi}{2N}) \\ \cos(\frac{(j-1)(2N-3)\pi}{2N}) \end{bmatrix} \\ \approx 2\cos(\frac{(j-1)\pi}{N}) \cdot \vec{h}_j = \lambda_j \vec{h}_j \quad (14)$$

The equality of the Eq. (14) is hold except for the first and last elements. This means that  $\vec{h}_j, 1 \leq j \leq N$  are approximate eigenvectors of  $Q_1$ , with the approximate eigenvalues of  $2\cos(\frac{(j-1)\pi}{N})$ . In the same way, we can prove that  $\vec{h}_j, 1 \leq j \leq N$  are approximate eigenvectors of  $Q_2, \dots, Q_p$ .

Define the matrix

$$C = \frac{1}{\sqrt{N}} [\vec{h}_1 \quad \sqrt{2}\vec{h}_2 \quad \dots \quad \sqrt{2}\vec{h}_N]^T \quad (15)$$

Then  $C^T$  is approximately the eigenvector matrix of  $Q_1, Q_2, \dots, Q_p$ . Define  $Q_j C^T \approx C^T \Lambda_j, 1 \leq j \leq p$ , where  $\Lambda_j$  denotes the eigenvalue matrix of  $Q_j$ . As such,

$$QC^T = C^T [I - \sum_{j=1}^p \alpha(j)Q_j] \approx C^T [I - \sum_{j=1}^p \alpha(j)\Lambda_j] = C^T \Lambda \quad (16)$$

where  $\Lambda = I - \sum_{j=1}^p \alpha(j)\Lambda_j$  is the eigenvalue matrix of  $Q$ . Therefore  $C^T$  is also approximately the eigenvector matrix of  $Q$ . Furthermore, from the definition of DCT [5], we can see that  $C$  is the DCT matrix. This proves that DCT is approximately the KLT for the matrix  $Q$  defined in Eq. (10).

Consider the noncausal MVR of the AR( $p$ ) process in Eq. (9), the covariance matrix of the noncausal prediction error vector  $\vec{v} = [v_1 \ v_2 \ \dots \ v_N]^T$  is obtained from Eq. (7) as  $R_v = E[\vec{v}\vec{v}^T] = \sigma_v^2 Q$ . Multiplying both sides of the Eq. (9) by  $Q^{-1}$ , we get:

$$\vec{u} = \vec{u}_0 + \vec{u}_b \quad (17)$$

where  $\vec{u}_0 = Q^{-1}\vec{v}$ ,  $\vec{u}_b = Q^{-1}\vec{b}$ .

Note that  $\vec{u}_b$  is orthogonal to  $\vec{u}_0$  due to the orthogonality condition between  $\vec{b}$  and  $\vec{v}$ , and  $\vec{u}_b$  is therefore completely determined by the boundary variables  $u(1-p), u(2-p), \dots, u(0), u(N+1), \dots, u(N+p)$ , which are assumed known. This means that Eq. (17) is an orthogonal decomposition of  $\vec{u}$ . The covariance matrix of  $\vec{u}_0$  is given by:

$$R_0 = E[\vec{u}_0 \vec{u}_0^T] = Q^{-1} \sigma_v^2 Q Q^{-1} = \sigma_v^2 Q^{-1} \quad (18)$$

Since  $\vec{u}_0$  and  $\vec{u}_b$  are orthogonal and  $E[\vec{u}_0] = 0$ , the conditional mean of  $\vec{u}$  given the boundary variables  $u(1-p), u(2-p), \dots, u(0), u(N+1), \dots, u(N+p)$  is simply  $\vec{u}_b$ . Therefore,

$$\begin{aligned} R &= \text{cov}[\vec{u}|u(1-p), \dots, u(0), u(N+1), \dots, u(N+p)] \\ &= E[(\vec{u} - \vec{u}_b)(\vec{u} - \vec{u}_b)^T] = E[\vec{u}_0 \vec{u}_0^T] = \sigma_v^2 Q^{-1} \end{aligned} \quad (19)$$

This means that the covariance matrix of  $\vec{u}$  is the same as that of  $\vec{u}_0$  given the boundary values. Because  $\sigma_v^2$  is a scalar, the eigenvectors of  $R_0$  are the same as those of  $Q^{-1}$ , and hence the same as the eigenvectors of  $Q$ . According to Lemma 1, the eigenvectors of the banded symmetric Toeplitz matrix  $Q$  can be approximated by the DCT matrix, which means that the KLT of the covariance matrix of the AR( $p$ ) process can be approximated by DCT. As such, the eigenvalue matrix can be approximately computed by:

$$\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_N) = CRC^T \quad (20)$$

where  $\text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_N)$  denotes a diagonal matrix whose diagonal elements are  $\hat{\lambda}_1, \dots, \hat{\lambda}_N$ , respectively.

In practice, the covariance matrix  $R$  is estimated from data samples and is constructed as a symmetric Toeplitz matrix:

$$R = \begin{bmatrix} r(0) & r(1) & \dots & r(N-1) \\ r(1) & r(0) & \dots & r(N-2) \\ \dots & \dots & \dots & \dots \\ r(N-1) & r(N-2) & \dots & r(0) \end{bmatrix} \quad (21)$$

In calculating the approximate eigenvalue matrix  $\hat{\Lambda}$  using DCT, the computation cost to perform DCT for a column of  $R$  is  $O(N \times \log_2(N))$  [4], and thus the total computation cost is  $O(N^2 \times \log_2(N))$ . In the next theorem, we derive a faster algorithm to compute the approximate eigenvalues of the symmetric Toeplitz matrix  $R$  in Eq. (21) using  $O(N^2)$  computations.

### Theorem 1

Consider the eigenvalue matrix  $\hat{\Lambda}$  in Eq. (20) and the covariance matrix  $R$  in Eq. (21). Define:

$$\vec{\lambda} = [\hat{\lambda}_1 \quad \hat{\lambda}_2 \quad \dots \quad \hat{\lambda}_N]^T \quad (22)$$

$$\vec{r} = [r(0) \quad r(1) \quad \dots \quad r(N-1)]^T \quad (23)$$

Then:

$$\vec{\lambda} = B\vec{r} \quad (24)$$

where  $B = [b_{ij}]_{N \times N}$  is a matrix defined by:

$$b_{ij} = \begin{cases} \sum_{k=1}^N c_{i,k}^2, & j=1 \\ 2 \sum_{k=1}^{N-j} c_{i,k} c_{i,k+j}, & 2 \leq j \leq N \end{cases} \quad (25)$$

and  $c_{i,j}$  are the elements of the DCT matrix defined in Eq. (15).

Proof: From Eq. (20), we can get:

$$\begin{aligned} \hat{\lambda}_i &= [c_{i,1} \quad c_{i,2} \quad \dots \quad c_{i,N}] \\ &\begin{bmatrix} r(0) & r(1) & \dots & r(N-1) \\ r(1) & r(0) & \dots & r(N-2) \\ \dots & \dots & \dots & \dots \\ r(N-1) & r(N-2) & \dots & r(0) \end{bmatrix} \begin{bmatrix} c_{i,1} \\ c_{i,2} \\ \dots \\ c_{i,N} \end{bmatrix} \\ &= \sum_{k=1}^N \sum_{l=1}^N c_{i,k} c_{i,l} r(|l-k|) \end{aligned} \quad (26)$$

So that:

$$\begin{aligned} \hat{\lambda}_i &= r(0) \sum_{k=1}^N c_{i,k}^2 + 2r(1) \sum_{k=1}^{N-1} c_{i,k} c_{i,k+1} + \dots \\ &+ 2r(N-2) \sum_{k=1}^2 c_{i,k} c_{i,k+N-2} + 2r(N-1) c_{i,1} c_{i,N} \\ &= [b_{i,1} \quad b_{i,2} \quad \dots \quad b_{i,N}] \begin{bmatrix} r(0) \\ r(1) \\ \dots \\ r(N-1) \end{bmatrix} \end{aligned} \quad (27)$$

where  $b_{i,j}$ ,  $1 \leq i, j \leq N$  is given by Eq. (25).

Writing out Eq. (27) for  $1 \leq i \leq N$ , we get  $\vec{\lambda} = B\vec{r}$ . This proves Theorem 1.

Since the DCT matrix  $C$  is fixed for a given analysis length  $N$ , the matrix  $B$  in Eq. (25) is also fixed. The  $B$  matrix can be precomputed and stored, and therefore the cost for computing the  $N$  eigenvalues of matrix  $R$  is only  $O(N^2)$  by using Eq. (24). This fast computation method is referred to as FDCT, which stands for fast eigenvalue computation based on DCT. The computation cost of FDCT method compares favorably with those of the direct DCT method ( $O(N^2 \log_2(N))$ ) and the KLT method ( $O(N^3)$ ).

### 3. FAST DCT-BASED ECSS SPEECH ENHANCEMENT SYSTEM

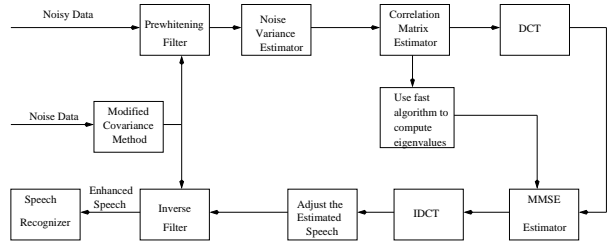


Figure 1. System diagram

Fig. 1 shows the system diagram of the fast DCT-based energy-constrained signal subspace (ECSS) speech enhancement system for speech recognition in car. The car noise was extracted during the silence period of the speech and was modeled by an AR process. A modified covariance method was used to estimate the AR parameters of the car noise and a prewhitening filter was constructed based on the estimated AR parameters. The vector space of the noisy speech signal was decomposed into a signal-plus-noise subspace and a noise subspace. The clean speech signal was estimated using a minimum mean square error (MMSE) estimator from the signal-plus-noise subspace. In order to better estimate the noise-like speech segments, an energy constraint was introduced to match the short-time energy of the enhanced speech signal to the unbiased estimate of the short-time energy of the clean speech. Details of the KLT-based ECSS algorithm was described in [2].

### 4. EXPERIMENT RESULTS

In this experiment, the effectiveness of the fast DCT algorithm for robust speech recognition in real car environment is investigated. The training data consisted of 4235 digit strings spoken by 55 speakers from the TIDIGITS database. The test data set was the noisy speech collected in car, including 4935 digit strings spoken by 25 speakers

under three driving conditions (idle, 30 miles/hour and 55 miles/hour). The speech recognizer was based on hidden Markov models (HMM) of digit units. Each HMM was modeled by eight states and each state was modeled by a Gaussian mixture density of size nine. The analysis feature vector included 12 MFCCs, 1 normalized log energy, 12 delta MFCCs and 1 delta log energy (total dimension was 26).

In this experiment, both the original KLT-based ECSS algorithm (ECSS-KLT) and the fast DCT-based ECSS algorithm (ECSS-FDCT) were investigated for enhancement of car noisy speech. The recognition performance on test data was evaluated under both cases of known string-length (KL) and unknown string-length (UL). Table 1 shows the baseline word recognition accuracy (WRA). The WRAs of the digit strings enhanced by the ECSS-KLT and ECSS-FDCT algorithms are listed in Table 2. It can be seen from these tables that the ECSS-KLT and ECSS-FDCT methods can significantly improve the WRA under various driving conditions and the recognition results using ECSS-FDCT method are very close to those using ECSS-KLT method. The WRA under the driving conditions of idle and 30 miles/hour using fast DCT are even slightly higher, though not significant, than the WRA using KLT.

Table 1. WRA of the baseline

Driving Condition	Idle	30 mph	55 mph
Unknown Length	67.81%	43.98%	35.32%
Known Length	74.28%	49.91%	41.46%

Table 2. WRA of ECSS-KLT and ECSS-FDCT methods

Driving Condition	Idle	30 mph	55 mph
ECSS-KLT method			
Unknown Length	89.60%	73.39%	65.96%
Known Length	92.63%	78.02%	70.54%
ECSS-FDCT method			
Unknown Length	89.89%	75.05%	65.11%
Known Length	92.78%	79.86%	70.47%

In our experiment, an acoustic normalization method is also used to compensate the speakers' difference and the mismatch between the training and the test data [3]. The linear distortion part of the mismatch is modeled by a linear transformation in the linear spectral domain which result in a bias term in the logarithmic spectral domain. The acoustic normalization is given by:

$$\tilde{x}_{i,t}^{(q)} = x_{i,t}^{(q)} - h^{(q)} \quad (28)$$

where  $x_{i,t}^{(q)}$  and  $\tilde{x}_{i,t}^{(q)}$  denote the cepstra of the phone unit  $i$ , at time  $t$ , from the speaker  $q$  before and after the normalization, respectively.  $h^{(q)}$  is called the cepstral bias and is estimated by:

$$h^{(q)} = \frac{1}{T^{(q)}} \sum_{t=1}^{T^{(q)}} x_t^{(q)} - \sum_{i=1}^K \frac{N_i}{N} \mu_i = \bar{x}^{(q)} - \bar{\mu} \quad (29)$$

where  $T^{(q)}$  denotes the number of the frames of the utterance by speaker  $q$ ,  $N_i$  is the sample size of the  $i^{th}$  phone unit,  $N = \sum_{i=1}^K N_i$ ,  $\bar{x}^{(q)}$  denotes the mean cepstra of the utterance by speaker  $q$  and  $\bar{\mu}$  denotes the mean cepstra of the entire training set.

Table 3. WRA of ECSS-KLT and ECSS-FDCT methods combined with cepstral mean normalization

Driving Condition	Idle	30 mph	55 mph
ECSS-KLT method			
Unknown Length	90.92%	75.01%	70.35%
Known Length	93.58%	79.09%	74.58%
ECSS-FDCT method			
Unknown Length	90.17%	74.98%	68.59%
Known Length	93.19%	79.95%	73.15%

Table 3 lists the WRAs of the ECSS-KLT and ECSS-FDCT methods combined with the acoustic normalization. We can see from this table that the cepstral mean normalization consistently improved the recognition performance under various driving conditions. The WRA improvements in the ECSS-KLT case is slightly better than the ECSS-FDCT case.

## 5. CONCLUSION

In this study, a fast algorithm is proposed to approximate KLT which entails heavy computations in signal-subspace based speech enhancement algorithms. It is proven that the DCT is a good approximation to KLT for the covariance matrix of an AR( $p$ ) process. For computing the approximate eigenvalues of a symmetric Toeplitz matrix, a fast algorithm is further derived to reduce the computation cost from  $O(N^3)$  in KLT to  $O(N^2)$ . This fast eigen-decomposition algorithm is incorporated into the ECSS speech enhancement algorithm [2] for robust speech recognition in car. Experiment results show that the speech recognition performance of the fast DCT-based ECSS algorithm was very close to that of the KLT-based ECSS algorithm while the fast DCT-based ECSS algorithm significantly reduced processing time.

## ACKNOWLEDGMENT

This work is supported in part by the National Science Foundation under the grants NSF-IRI-9502074 and a grant from the Whitaker Foundation. Jun Huang is the recipient of a Motorola Partnerships in Research grant. The authors would like to thank Dr. Mark Randolph for providing Motorola's car noisy speech data.

## REFERENCES

- [1] Y. Ephraim and H.L. Van Trees, "A Signal Subspace Approach for Speech Enhancement," *IEEE Trans. on Speech and Audio Processing*, Vol. 3, No. 4, pp. 251-266, Jul., 1995.
- [2] J. Huang and Y. Zhao, "An Energy-Constrained Signal Subspace Method for Speech Enhancement and Recognition in White and Colored Noises," *Speech Communication*, Vol. 26, pp. 165-181, 1998.
- [3] Y. Zhao, "Self-learning Speaker and Channel Adaptation Based on Spectral Variation Source Decomposition," *Speech Communication*, Vol. 18, pp. 65-77, 1996.
- [4] A. K. Jain, *Fundamentals of Digital Image Processing* Prentice-Hall, 1989.
- [5] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE Trans. on Communication*, pp. 90-93, Jan. 1974.