Robust Signal Subspace Speech Classifier

Alan W. C. Tan, *Student Member, IEEE*, M. V. C. Rao, and B. S. Daya Sagar, *Senior Member, IEEE*

***Abstract*—A speech model inspired by the signal subspace approach was recently proposed as a speech classifier with modest results. The method entails, in general, the assemblage of a set of subspace trajectories that consist of the right singular vectors of measurement matrices of the signal under consideration. Given an unknown signal, a simple distortion measure then applies in the classification procedure to pick the best matched class prototype. This letter examines the issue of robustness in the subspace classification scheme. Borrowing an important result on noisy measurement matrices, this letter formally establishes the notion of robustness in subspace classification and proceeds to propose a class of robust distortion measures for signal subspace models. Simulation results of subspace classifiers implementing the new distortion measures in an isolated digit speech recognition problem reveal no degradation in recognition accuracy, even under low SNR conditions.**

***Index Terms*—Robust distortion measures, speech modeling, speech recognition, subspace methods.**

I. Introduction

The signal subspace approach to speech processing has traditionally been confined to speech enhancement problems; some examples include [1] and [2]. Recently, the authors have applied the same method in speech modeling and classification problems with modest results (e.g., [3] and [4]). From a series of eigendecomposition or SVD on the measurement matrices, the row space (or column space, if one so chooses) of the speech signal is broken down into signal subspace trajectories. Subspace information contained in these trajectories has been shown to be reliable in characterizing the slowly changing speech signal, and as a feature extraction preprocessing step, this method is comparable to the more popular cepstral-derived techniques like the LPCC and MFCC (e.g., [5]–[7] and [8]). Simulation results also reveal signal subspace-based classifiers to be fairly robust at moderate levels of signal-to-noise ratio (SNR) [3].

This letter extends the work in [3]. In particular, we will formally introduce the notion of robustness in the context of signal subspace classification. We will show, specifically, that the distortion measure used in [3] is not robust in the sense that an artifact term is always present in the function due to noise. To overcome the problem, we will formulate a novel class of robust distortion measures that is specially designed to be immune to additive stationary white noise. Results from computer simulation on an isolated digit speech recognition problem affirm the theoretical analysis.

The rest of this letter is structured as follows: A brief review on signal subspace modeling and classification is covered in Section II. The section reproduces an important result on noisy measurement matrices that appeared in [9] and that will serve as a basis for developing robust distortion measures in the next section. Section III, which contains the main ideas of this letter, also examines an experimental analysis on the effect of additive white noise in signal subspace modeling and then proceeds to propose a class of robust distortion measures. Simulation results and discussions are recorded in Section IV, and Section V provides the major conclusions.

Manuscript received November 15, 2006; revised April 10, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Aleksandra Mojsilovic.

The authors are with the Faculty of Engineering and Technology, Multimedia University, Melaka 75450, Malaysia (e-mail: wctan@mmu.edu.my).

Digital Object Identifier 10.1109/LSP.2007.900036

II. Subspace Modeling

A. Signal Model and the SVD

A measurement matrix $X \in \mathbb{R}^{n\times m}$ is constructed by organizing samples of the measurement data $x_1, x_2, \dots, x_K$ into the Toeplitz matrix of the form [9], [2]

$$X = \begin{bmatrix} x_m & x_{m-1} & \cdots & x_1 \\ x_{m+1} & x_m & \cdots & x_2 \\ \vdots & \vdots & & \vdots \\ x_K & x_{K-1} & \cdots & x_{K-m+1} \end{bmatrix} \quad (1)$$

where $n > m$ and the matrix dimension constrained by $K = n + m - 1$. The measurement matrix is typically rank deficient [9], that is, the actual signal content lies in a *signal subspace* $\mathcal{S}$ of a lower dimension, i.e., $\dim(\mathcal{S}) = p < m$.

The thin SVD (or economy-size SVD) of the measurement matrix X is defined as [10, p. 72]

$$X = U\Sigma V^T = \sum_{k=1}^{m} \mathbf{u}_k \sigma_k \mathbf{v}_k^T$$

where the columns of $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m] \in \mathbb{R}^{n\times m}$ are mutually orthonormal, $V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m] \in \mathbb{R}^{m\times m}$ is a unitary matrix, and $\Sigma \in \mathbb{R}^{m\times m}$ has the form

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_m).$$

The diagonal elements of Σ are the *singular values* of X and are ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m$. The columns of U and V are, respectively, called the *left* and *right singular vectors*.

Let us assume that X is made up of a rank p signal matrix $S \in \mathbb{R}^{n\times m}$ and a matrix $N \in \mathbb{R}^{n\times m}$ of the noise process, i.e.,

$$X = S + N.$$

Let us further suppose that the SVD of the rank p signal matrix be given as

$$S = \begin{bmatrix} U_{s1} & U_{s2} \end{bmatrix} \begin{bmatrix} \Sigma_{s1} & \\ & 0 \end{bmatrix} \begin{bmatrix} V_{s1}^T \\ V_{s2}^T \end{bmatrix}$$

where $U_{s1} \in \mathbb{R}^{n\times p}$, $U_{s2} \in \mathbb{R}^{n\times (m-p)}$, $V_{s1} \in \mathbb{R}^{m\times p}$, $V_{s2} \in \mathbb{R}^{m\times (m-p)}$, and $\Sigma_{s1} \in \mathbb{R}^{p\times p}$. Under the assumption that the noise is white and uncorrelated with the signal of interest, i.e.,

$N^TN = \nu^2 I_m$ and $S^TN = U_{s1}^T N = 0$, an explicit expression of X in terms of the SVD of S can be formulated as [9]

$$\begin{aligned} X &= U_{s1}\Sigma_{s1}V_{s1}^T + N\left(V_{s1}V_{s1}^T + V_{s2}V_{s2}^T\right) \\ &= \left[\,(U_{s1}\Sigma_{s1} + NV_{s1})\left(\Sigma_{s1}^2 + \nu^2 I_p\right)^{-1/2} \quad \nu^{-1}NV_{s2}\,\right] \\ &\quad \cdot \begin{bmatrix} \left(\Sigma_{s1}^2 + \nu^2 I_p\right)^{1/2} & \\ & \nu I_{m-p} \end{bmatrix} \begin{bmatrix} V_{s1}^T \\ V_{s2}^T \end{bmatrix}. \end{aligned} \tag{2}$$

It is not difficult to show that (2) is an SVD of X of the form

$$X = [\,U_{x1} \quad U_{x2}\,] \begin{bmatrix} \Sigma_{x1} & \\ & \Sigma_{x2} \end{bmatrix} \begin{bmatrix} V_{x1}^T \\ V_{x2}^T \end{bmatrix}.$$

One result that follows immediately from (2) is that the row space (or equivalently, the right singular vectors) of X are precisely those of S since $V_{x1} = V_{s1}$ and $V_{x2} = V_{s2}$. In the colored noise case, a prewhitening matrix $R_n^{-1/2} \in \mathbb{R}^{m\times m}$, where $R_n = N^TN$ is the noise covariance matrix, can always be applied to the measurement matrices, and one can still recover V_{s1} and V_{s2} from $XR_n^{-1/2}$ after a bit of work [9] (see also [1] and [11]). Thus, we will assume, without the loss of generality, that the noise is white, and the aforementioned assumptions will be tacitly implied for the rest of this letter.

B. Subspace Decomposition and Selection

A running rectangular window is used to acquire the analysis frames of the speech signal $\mathbf{x}$. The window is K length and advances every K_1 samples. The tth frame, therefore, consists of the samples $K_1(t-1)+1, K_1(t-1)+2, \ldots, K_1(t-1)+K$, and there are, in total, $T = \lceil (L-K)/K_1 + 1\rceil$ frames, where L is the number of samples in the speech signal and the operator $\lceil x \rceil$ returns the smallest integer greater than or equal to x. The K samples contained in each frame, say, $x_1, x_2, \ldots, x_K$, are then organized into a measurement matrix of the form as in (1). Following this procedure, we obtain a set of measurement matrices $\{X(t) : t \in \mathbb{N}_T\} = \mathcal{T}(\mathbf{x})$ from the signal $\mathbf{x}$. Here, $\mathcal{T}$ denotes the frame operator and $\mathbb{N}_i$ is the subset of natural numbers $\{1, 2, \ldots, i\}$.

The general notion of signal subspace modeling lies in the assumption that the signal subspace is slowly changing and that it is possible to specify it as a composite of subspace trajectories. In the case here, the *subspace trajectory* $\psi(t)$ is a vector-valued function of the right singular vectors of successive measurement matrices. It is nonzero (or *active*) in some frame interval $t_1 \le t \le t_2$, and two right singular vectors $\mathbf{v}_t$ and $\mathbf{v}_{t+1}$ of successive frames belong to the same trajectory if $|\cos^{-1}(\mathbf{v}_t^T\mathbf{v}_{t+1})| \le \theta_{\mathrm{th}}$, where $\theta_{\mathrm{th}} < \cos^{-1}(m^{-1/2})$[1] is the *transition bound*. The bound θ_{th} ensures that no more than one vector $\mathbf{v}_{t+1}$ (from the set of orthonormal right singular vectors) lies within θ_{th} of vector $\mathbf{v}_t$. Collectively, the *family of subspace trajectories* $\Psi = \{\psi_j(t) : j \in \mathcal{J}, t \in \mathbb{N}_T\}$, where $\mathcal{J}$ represents the set of trajectory indices, characterizes the entire signal subspace.

Due to the rank degeneracy of the measurement matrices, it becomes necessary to separate the signal-related trajectories from the ones in the noise subspace $\mathcal{N} = \mathcal{S}^\perp$. For this purpose, a simple selection algorithm, which attaches to each trajectory an energy measure and thereafter picks a *minimal set of subspace trajectories* $\breve{\Psi} \subset \Psi$ such that the accumulated energies in $\breve{\Psi}$ exceed a preset threshold E_{th}, is applied. Implementation details of the subspace trajectory decomposition and selection algorithms can be found in [3].

[1]In the $m = 2$ case, and supposing $\mathbf{v}_1$ and $\mathbf{v}_2$ are the two orthonormal vectors spanning the Euclidean plane, then the upper bound on θ_{th}—let us call that θ_{ub}—is the angle $\cos^{-1}(\mathbf{u}^T\mathbf{v}_1)$ [or equivalently $\cos^{-1}(\mathbf{u}^T\mathbf{v}_2)$], where $\mathbf{u} = 2^{-1/2}(\mathbf{v}_1 + \mathbf{v}_2)$. This angle is readily seen as $\theta_{\mathrm{ub}} = \cos^{-1}(2^{-1/2})$. In the Euclidean m-space, it follows that $\theta_{\mathrm{ub}} = \cos^{-1}(m^{-1/2})$.

C. Subspace Classification

For every class $c \in \mathcal{C}$, where $\mathcal{C}$ denotes the set of all known classes, let $\breve{\Psi}_c = \{\psi_{cj}(t) : j \in \mathcal{J}_c, t \in \mathbb{N}_T\}$ be the class prototype obtained through the subspace decomposition and selection algorithms. Given an unknown signal $\mathbf{x}_0$, we desire to classify it into one of the classes in $\mathcal{C}$. In [3], the function

$$d(\psi_{cj}, X) = \big|\|\psi_{cj}\| - \|X\psi_{cj}\|/\|\psi_{cj}\|\big| \tag{3}$$

where $X \in \mathcal{T}(\mathbf{x}_0)$ is used to define the measure of distortion. In particular, (3) calculates the deviation of the content of X in ψ_{cj} from the actual value. The weighted average of (3), with the trajectory norms as the weighting coefficients, is thereafter evaluated across all trajectories in $\breve{\Psi}_c$, i.e.,

$$\mathcal{A}_w(\breve{\Psi}_c, \mathbf{x}_0) = \frac{\sum_{j\in\mathcal{J}_c}\sum_{t=1}^{T} w_{cj}(t)\, d(\psi_{cj}(t), X(t))}{\sum_{j\in\mathcal{J}_c}\sum_{t=1}^{T} w_{cj}(t)} \tag{4}$$

where $w_{cj}(t) = \|\psi_{cj}(t)\|$, and afterward, the class label of the prototype that minimizes (4) is picked as the class label of $\mathbf{x}_0$.

III. Robust Subspace Classifiers

To examine the notion of robustness in signal subspace modeling, let us consider for the moment the distortion measure of (3) in the following special case.

Lemma 1: Let $\mathbf{x}_c$ denote the *clean* signal of class c with which the subspace prototype $\breve{\Psi}_c = \{\psi_{cj}(t) : j \in \mathcal{J}_c, t \in \mathbb{N}_T\}$ is associated and suppose $\mathbf{x}_c^m = \mathbf{x}_c + \mathbf{n}$, where $\mathbf{n}$ is a stationary white noise process with variance $\nu^2 > 0$, is the noisy measurement. Then, the distortion measure in (3) is nonzero, i.e., $d(\psi_{cj}, X) > 0$, for any active $\psi_{cj} \in \breve{\Psi}_c$ and $X \in \mathcal{T}(\mathbf{x}_c^m)$ of the same frame instant.

Proof: Let the SVD of the measurement matrix be given as $X = U\Sigma V^T = \sum_{k=1}^{m} \mathbf{u}_k\sigma_k\mathbf{v}_k^T$ and suppose $\psi_{cj} = \sigma_{cj}\mathbf{v}_{cj}$. We may then write

$$\|X\psi_{cj}\|/\|\psi_{cj}\| = \|U\Sigma V^T\mathbf{v}_{cj}\| = \|\Sigma V^T\mathbf{v}_{cj}\|.$$

With this, (3) becomes

$$d(\psi_{cj}, X) = |\sigma_{cj} - \|\Sigma V^T\mathbf{v}_{cj}\||.$$

By (2), it turns out that $\mathbf{v}_{cj}$ must be exactly one of the right singular vectors of X, i.e., there exists an l such that $\mathbf{v}_l = \mathbf{v}_{cj}$ and $\sigma_l = (\sigma_{cj}^2 + \nu^2)^{1/2}$. Henceforth, we have

$$d(\psi_{cj}, X) = |\sigma_{cj} - \sigma_l| > 0.$$

■

Lemma 1 clearly demonstrates the insufficiency of (3) as a robust distortion measure. An undesirable artifact, whose magnitude increases with the noise power, is always present in (3), and as a result, recognition accuracy will tend to deteriorate as SNR drops. The robust classifier should, despite the inclusion of the noise terms, perform comparably to the noise-free case, i.e., the distortion measure is zero under the conditions described in Lemma 1. This section explores the issue in greater detail.

A. Qualitative Analysis

Given $\mathbf{x}_c$ and $\mathbf{x}_c^m$ as the clean and noisy signals defined in Lemma 1, we will assess empirically the departure of the right singular vectors of the measurement matrices $X \in \mathcal{T}(\mathbf{x}_c^m)$ from that of the subspace trajectories $\psi_{cj} \in \breve{\Psi}_c$ of $\mathbf{x}_c$. Recall that each

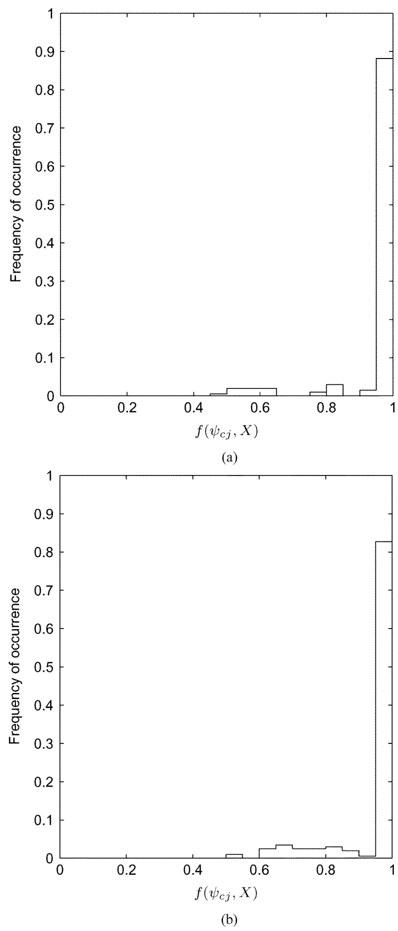

Fig. 1. Histograms of $f(\psi_{cj}, X)$ of the utterance "Eight" at (a) SNR $= 16$ dB, and (b) SNR $= 4$ dB. $[K = 160, K_1 = 40, m = 20, \theta_{\text{th}} = 25°, E_{\text{th}} = 0.9]$.

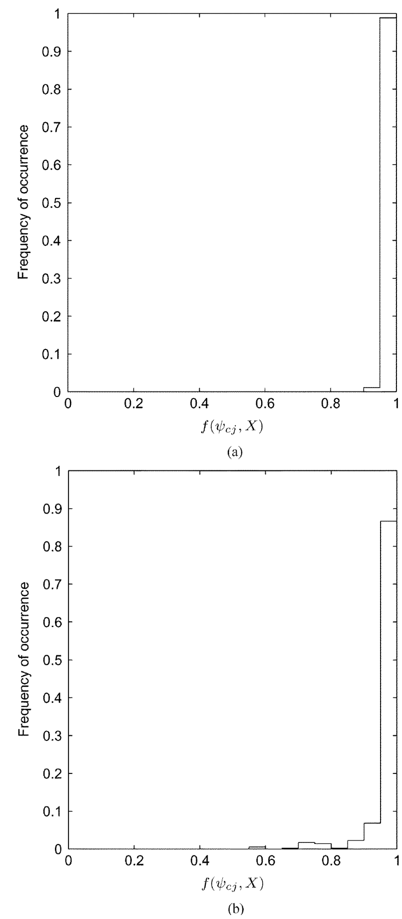

Fig. 2. Histograms of $f(\psi_{cj}, X)$ of the utterance "Five" at (a) SNR $= 16$ dB and (b) SNR $= 4$ dB. $[K = 160, K_1 = 40, m = 20, \theta_{\text{th}} = 25°, E_{\text{th}} = 0.9]$.

trajectory ψ_{cj} is a function of the right singular vectors of the measurement matrices of $\mathbf{x}_c$. Thus, our analysis will, in effect, compare the right singular vectors of the measurement matrices of the two signals.

To that end, we define, for every active trajectory $\psi_{cj} \in \check{\Psi}_c$, the measure

$$f(\psi_{cj}, X) = \max_i |\mathbf{v}_i^T \psi_{cj}| / \|\psi_{cj}\|$$

where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are the right singular vectors of X of the same frame instant. The function $f(\psi_{cj}, X)$, in general, calculates the cosine of the angle between ψ_{cj} with the best matched right singular vector of X. Due to (2), we expect this value to be one, but simulation results will show otherwise. As shown in Figs. 1 and 2, there are some instances at which $f(\psi_{cj}, X) < 1$ and the frequency of such occurrences generally increases as SNR drops. Nevertheless, it is fair to conclude from the results that the right singular vectors (or row space) of the two sets of measurement matrices are almost identical, even if they are not exactly alike.

B. Robust Distortion Measures

The above findings lay out some fundamental considerations in the attempt to define robust distortion measures for signal subspace classifiers. Suppose $\check{\Psi}_c = \{\psi_{cj}(t) : j \in \mathcal{J}_c, t \in \mathbb{N}_T\}$ is the class $c \in \mathcal{C}$ subspace prototype obtained through the usual subspace decomposition and selection algorithms. We desire to classify the unknown signal $\mathbf{x}_0$ into one of the classes in $\mathcal{C}$.

Let $X \in \mathcal{T}(\mathbf{x}_0)$ be one of the measurement matrices of $\mathbf{x}_0$ and suppose there exists the active trajectory $\psi_{cj} = \sigma_{cj}\mathbf{v}_{cj}$ in the same frame instant. We will define the term

$$l = \arg\max_i \left|\mathbf{v}_i^T \psi_{cj}\right| / \|\psi_{cj}\| = \arg\max_i \left|\mathbf{v}_i^T \mathbf{v}_{cj}\right|$$

such that $\mathbf{v}_l$ and σ_l, which are obtained from the SVD of X, are the best matched right singular vector of ψ_{cj} and the corresponding singular value, respectively. It is readily seen that the following *robust subspace difference measure*

$$d_Q(\psi_{cj}, X) = 1 - \left|\mathbf{v}_l^T \mathbf{v}_{cj}\right|^Q \tag{5}$$

where $Q > 0$ satisfies the condition for robustness expressed previously. A global dissimilarity measure mirroring (4), i.e.,

$$\mathcal{A}_w(\check{\Psi}_c, \mathbf{x}_0) = \frac{\sum_{j \in \mathcal{J}_c} \sum_{t=1}^{T} w_{cj}(t)\, d_Q(\psi_{cj}(t), X(t))}{\sum_{j \in \mathcal{J}_c} \sum_{t=1}^{T} w_{cj}(t)}$$

with weighting coefficients chosen as $w_{cj} = \sigma_l \sigma_{cj}$, henceforth applies to pick the class label of the unknown signal.

Lemma 2: For signals $\mathbf{x}_c$ and $\mathbf{x}_c^m$ defined in Lemma 1, the distortion measure in (5) is always zero, i.e., $d_Q(\psi_{cj}, X) = 0$, for any active $\psi_{cj} \in \check{\Psi}_c$ and $X \in \mathcal{T}(\mathbf{x}_c^m)$ of the same frame instant.

One possible extension of (5) is obtained by writing

$$\check{d}_Q(\psi_{cj}, X) = -\sigma_l \sigma_{cj} \left|\mathbf{v}_l^T \mathbf{v}_{cj}\right|^Q . \tag{6}$$

If $Q = 1$, the function in (6) has the form of an inner product, i.e., $\check{d}_1(\psi_{cj}, X) = -|\langle \sigma_l \mathbf{v}_l, \psi_{cj} \rangle|$, or equivalently, the projection of $\sigma_l \mathbf{v}_l$ onto ψ_{cj}. For this reason, we shall refer to (6) as the *robust subspace projection measure*. In the case of the projection measure, the unknown signal $\mathbf{x}_0$ is assigned the class label of the prototype that minimizes the simple sum

$$\mathcal{A}(\breve{\Psi}_c, \mathbf{x}_0) = \sum_{j \in \mathcal{J}_c} \sum_{t=1}^{T} \check{d}_Q(\psi_{cj}(t), X(t)).$$

IV. Results and Discussions

In this section, we investigate the white noise robustness of the proposed classifier in an isolated digit speech recognition problem. Speech recordings, at a sampling frequency of 10 kHz and SNR of approximately 24 dB, are collected from three male speakers. For every digit between 1–9 and two different utterances of the digit 0, i.e., "Zero" and "Oh," ten recordings are obtained, thereby yielding a total of $11 \times 10 = 110$ recordings per speaker. A recording of each digit is then randomly selected to build the set of class prototypes, while the other nine recordings are used as the testing data. To gauge the white noise robustness of the proposed classifier, artificial stationary white noise is introduced into the testing data at different levels of SNR [8]. Then, incorporating the proposed robust distortion measures (particularly the functions d_1, d_2, $\check{d}_1$ and $\check{d}_2$) into the classification scheme, the average recognition rate (of the three speakers) applies as the yardstick for performance. In the simulation, we have chosen $K = 160$ and $K_1 = 40$ following [2] and [3] and used $m = 20$, $\theta_{\rm th} = 25^\circ$ and $E_{\rm th} = 0.9$ as the parameters in the subspace decomposition and selection algorithms.

For comparison purposes, we also present simulation results on the original signal subspace classifier (see [3]) and included in the simulation two widely used speech recognizers, i.e., the LP-derived cepstral coefficients (LPCC) recognizer and the Mel-frequency-derived cepstral coefficients (MFCC) recognizer, both implementing dynamic time warping. For the LPCC, the cepstral projection measure, which was originally designed to combat additive white noise [8], applies as the distortion measure, whereas in the MFCC, the Euclidean distance is used. In both the recognizers, a Hamming window $(K = 240, K_1 = 80)$ is applied to the data and 12 cepstral coefficients, lifted with $w_{\rm lift}(k) = 1 + 6\sin(\pi k/12)$, are retained as the cepstral vector [12, Ch. 4]. The main results of the simulation are presented in Fig. 3.

It is clear from the figure that the four subspace classifiers implementing the proposed distortion measures recorded relatively consistent performance over the range of SNR tested. At SNR of 0 dB, we note that the recognition accuracy of these classifiers is only slightly lower (by about 3%) from when the SNR is 24 dB. The original signal subspace classifier, while fairly robust for SNR above 8 dB due to the noise-filtering quality inherent in the subspace approach, declines rapidly in performance as SNR falls below this level. Of the four new subspace classifiers evaluated, the second-order robust projection measure, or $\check{d}_2$, achieves the highest recognition accuracy, and this result, in a way, parallels that of [8], where it has been shown that projection measures are often more reliable than difference (or distance) measures in adverse ambient conditions. It is important to note, however, that unlike the distortion measure in [3], the pair of robust distortion measures proposed here requires the computationally expensive SVD. Thus, in resource-constrained applications, it may be desirable to approximate the SVD with alternative decomposition techniques that are cheaper computation-wise (e.g., [11]).

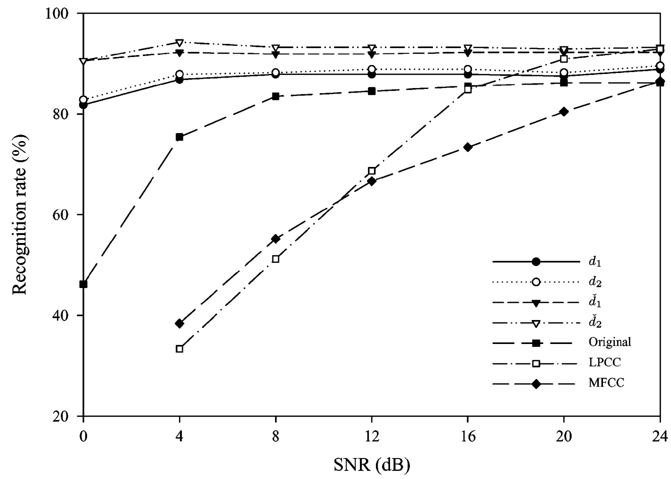

Fig. 3. Performance of various speech recognizers as SNR varies.

V. Conclusion

In this letter, we established the notion of robustness in signal subspace classifiers. We have experimented with adding stationary white noise to the measurement data and obtained results that corroborated with earlier theoretical findings in [9]. Based on these results, we have proposed a class of robust distortion measures for signal subspace classifiers, and preliminary experimentation on an isolated digit speech recognition problem reveals promising results.

References

[1] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[2] P. S. K. Hansen, "Signal subspace methods for speech enhancement," Ph.D. dissertation, Technical Univ. Denmark, Lyngby, Denmark, 1997.

[3] A. W. C. Tan, M. V. C. Rao, and B. S. D. Sagar, "A signal subspace approach for speech modelling and classification," *Signal Process.*, vol. 87, no. 3, pp. 500–508, Mar. 2007.

[4] A. W. C. Tan, M. V. C. Rao, and B. S. D. Sagar, "A discriminative signal subspace speech classifier," *IEEE Signal Process. Lett.*, vol. 14, no. 2, pp. 133–136, Feb. 2007.

[5] J. W. Picone, "Signal modeling techniques in speech recognition," *Proc. IEEE*, vol. 81, no. 9, pp. 1215–1247, Sep. 1993.

[6] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. PROC-63, no. 4, pp. 561–580, Apr. 1975.

[7] Y. Tokhura, "A weighted cepstral distance measure for speech recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1414–1422, Oct. 1987.

[8] D. Mansour and B.-H. Juang, "A family of distortion measures based upon projection operation for robust speech recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, no. 11, pp. 1659–1671, Nov. 1989.

[9] B. De Moor, "The singular value decomposition and long and short spaces of noisy matrices," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2826–2838, Sep. 1993.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The John Hopkins Univ. Press, 1996.

[11] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 6, pp. 439–448, Nov. 1995.

[12] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.