

# A Discriminative 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 methods was recently proposed as a speech classifier with modest results. Fashioned along a “best representation” approach, the absence of valuable interclass information in the speech model, however, impairs the ability of the classifier to distinguish between phonetically alike classes. This letter proposes an improved classifier that implements interclass information. Specifically, a measure of the discriminative quality of individual class elements is defined and determined for all class elements. The discrimination measures thus obtained are subsequently applied in the classification procedure. Simulation results of the proposed signal subspace classifier in an isolated digit speech recognition problem reveal an improved performance over its predecessor.

**Index Terms**—Class discrimination information, classification, speech modeling, speech recognition, subspace methods.

## I. INTRODUCTION

THE underlying assumption in signal subspace speech modeling lies in the premise that speech signals are strictly nonstationary and consist of short-term dynamics that are slowly changing [1, Chaps. 2–3], [2, Chap. 4]. In fact, popular speech processing techniques like the short-time Fourier transform (STFT), linear prediction coding (LPC) [3]–[5], and cepstral methods [3], [6], [7] are developed on the basis that the spectral content is slowly changing across the entire speech signal. These methods generally function to specify the relevant acoustic events in the speech signal in terms of a compact and efficient set of speech parameters.

A signal subspace speech model is a characterization of the speech signal in terms of its subspace information. The signal subspace approach to speech processing was originally applied for speech enhancement techniques (such as [8]–[10]), and it has only been used recently as a speech classifier [11]. This letter aims to improve on the latter. The signal subspace classifier in [11] was generally designed according to a “best representation” approach, i.e., finding the best fit from within the set of class prototypes. However, owing largely to the absence of discriminative elements in the speech model, it was inefficient when separating between phonetically alike classes. In such cases, it is often necessary to revise the original (best representation) modeling strategy to include these measures (compare [12] and [13]). Thus, recognizing the insufficiency of the earlier classifier, the proposed classifier compensates in the same manner. The performance of the proposed signal subspace classifier is subsequently tested against its predecessor in an isolated digit speech recognition problem, and results appear promising.

Manuscript received July 3, 2006; revised July 12, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Maja Bystrom.

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.2006.882091

The rest of this letter is structured as follows: A brief review on signal subspace modeling and classification is covered in Section II. Section III develops the key ideas behind the proposed signal subspace classifier. In particular, a classification scheme that enables class discrimination information is introduced. Simulation results and discussions are recorded in Section IV, and Section V provides the major conclusions.

## 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 a Toeplitz matrix of the form [8], [10]

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

where  $n > m$  and the matrix dimension constrained by  $K = n + m - 1$ . The measurement matrix is typically rank deficient [8], that is, the actual signal content lies in a *signal subspace*  $\mathcal{S}$  of a lower rank, i.e.,  $\mathcal{S} \subset \mathbb{R}^p, p < m$ .

The thin SVD (or economy-size SVD) of the measurement matrix  $X$  is defined as [14, 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

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

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

A least-squares estimate of  $X$  by a matrix of a lower rank  $p < m$  is

$$X_p = \sum_{k=1}^p \mathbf{u}_k \sigma_k \mathbf{v}_k^T \quad (2)$$

with the estimation error  $\|X - X_p\|_2 = \sigma_{p+1}$  [14, p. 72]. In practice, however, there are no definite ways to estimate  $p$  to effect (2), as a clearly defined spectral gap is often absent in the singular spectra due to the stochastic nature of speech signals. It is nevertheless reasonable to assume that the general quality of speech is tied to the formants, represented in the singular spectrum by the dominant singular values [10]. For the rest of this

letter, these assumptions are tacitly assumed. The rank-revealing property of the singular spectrum will be especially useful later when distinguishing between the various subspaces of  $X$ .

### B. Subspace Decomposition and Selection

A running rectangular window is used to acquire the analysis frames of the speech signal. The window has length  $K$  and advances every  $K_1$  samples. The  $t$ th frame would therefore consist of the samples  $K_1(t-1)+1, K_1(t-1)+2, \dots, K_1(t-1)+K$ . 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  $x$ . The samples contained in each frame are then organized into the Toeplitz structure resembling (1). Following this procedure, we obtain the set of measurement matrices  $\{X(t) : t \in \mathbb{N}_T\} = \mathcal{T}(\mathbf{x})$  of the signal  $\mathbf{x}$ . Here,  $\mathcal{T}$  denotes the frame operator, and  $\mathbb{N}_i$  is the subset of natural numbers  $\{1, 2, \dots, i\}$ .

The general notion of a signal subspace model lies in the assumption that the signal subspace is slowly changing and that it is possible to specify the signal subspace as a composite of subspace trajectories. In [11], a *subspace trajectory*  $\psi(t)$  is defined as a vector-valued function of the right singular vectors of a particular signal subspace. Two right singular vectors  $\mathbf{v}_t$  and  $\mathbf{v}_{t+1}$  of successive frames belong to the same trajectory if  $|\cos^{-1}(\mathbf{v}_t^T \mathbf{v}_{t+1})| \leq \theta_{\text{th}}$ , where  $\theta_{\text{th}} < \pi/4$  is the *transition bound*. Collectively, the *family of subspace trajectories*  $\Psi$  characterizes the entire signal subspace.

*Algorithm 1:* Subspace trajectory decomposition algorithm.

- 1) Given signal  $\mathbf{x}$ , construct the set of measurement matrices  $\{X(t) : t \in \mathbb{N}_T\} = \mathcal{T}(\mathbf{x})$ .
- 2)
  - a) Obtain the set of singular values  $\{\sigma_k : k \in \mathbb{N}_m\}$  and the set of right singular vectors  $\{\mathbf{v}_k : k \in \mathbb{N}_m\}$  from the SVD of  $X(1)$ .
  - b) For  $k = 1$  to  $m$ , set  $\psi_k(1) \leftarrow \sigma_k \mathbf{v}_k$ .
  - c) Set  $\mathcal{I} \leftarrow \mathbb{N}_m, \mathcal{I}' \leftarrow \emptyset$ , and  $M \leftarrow m$ .
- 3) For  $t = 2$  to  $T$ , do the following.
  - a) Obtain the set of singular values  $\{\sigma_k : k \in \mathbb{N}_m\}$  and the set of right singular vectors  $\{\mathbf{v}_k : k \in \mathbb{N}_m\}$  from the SVD of  $X(t)$ .
  - b) For  $k = 1$  to  $m$ , do the following.
    - i) For  $\forall i \in \mathcal{I}$ , determine  $\theta_i \leftarrow |\cos^{-1}(\mathbf{v}_k^T \psi_i(t-1))|$ .
    - ii) Find  $\theta_0 \leftarrow \min_{i \in \mathcal{I}} \{\theta_i\}$  and  $i_0 \leftarrow \arg \min_{i \in \mathcal{I}} \{\theta_i\}$ .
    - iii) If  $\theta_0 \leq \theta_{\text{th}}$ , then set  $\psi_{i_0}(t) \leftarrow \sigma_k \mathbf{v}_k$  and  $\mathcal{I}' \leftarrow \mathcal{I}' \cup i_0$ . Otherwise, set  $\psi_{M+1}(t) \leftarrow \sigma_k \mathbf{v}_k, \mathcal{I}' \leftarrow \mathcal{I}' \cup (M+1)$  and  $M \leftarrow M+1$ .
  - c) Update  $\mathcal{I} \leftarrow \mathcal{I}'$  and  $\mathcal{I}' \leftarrow \emptyset$ .
- 4) Construct the family of subspace trajectories  $\Psi \leftarrow \{\psi_j : j \in \mathbb{N}_M\}$ .

As a consequence to the rank degeneracy of the measurement matrices, the family of subspace trajectories  $\Psi$  obtained in Algorithm 1 contains also trajectories generated in the *noise subspace*  $\mathcal{N} = \mathcal{S}^\perp$ . It is vital, therefore, to adopt a selection strategy to pick a *minimal set of subspace trajectories*  $\check{\Psi} \subset \Psi$  that discards all but the signal related trajectories of  $\Psi$ . For the same purpose, an *energy ratio*  $E_{\text{th}}$  is predetermined.

*Algorithm 2:* Subspace trajectory selection algorithm.

- 1) Given the family of subspace trajectories,  $\Psi = \{\psi_j : j \in \mathbb{N}_M\}$ .
- 2) For  $j = 1$  to  $M$ , compute the energy map  $E_j \leftarrow \sum_{t=1}^T \|\psi_j(t)\|^2$ .
- 3) Pick the minimal set of indexes  $\mathcal{J} \subset \mathbb{N}_M$  such that  $\sum_{k \in \mathcal{J}} E_k > E_{\text{th}} \sum_{l \in \mathbb{N}_M} E_l$ .
- 4) Construct the minimal set of subspace trajectories  $\check{\Psi} \leftarrow \{\psi_j : j \in \mathcal{J}\}$ .

### C. Subspace Classification

Let  $\mathcal{C} = \{1, 2, \dots, C\}$  denote the set of the  $C$  known classes. For every class  $c \in \mathcal{C}$ , let the class prototype  $\check{\Psi}_c = \{\psi_{c_j} : j \in \mathcal{J}_c\}$  denote the minimal set of subspace trajectories obtained by Algorithm 2. Given an unknown signal  $\mathbf{x}_0$ , we desire to classify it into one of the classes in  $\mathcal{C}$ .

In [11], the function

$$d(\psi_{c_j}, X) \triangleq \frac{\|\psi_{c_j}\| - \|X\psi_{c_j}\|}{\|\psi_{c_j}\|} \quad (3)$$

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

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

and  $w_{c_j}(t) = \|\psi_{c_j}(t)\|$ . Following that, a minimum-distance rule is applied to pick the class label of the prototype most similar to  $\mathbf{x}_0$ .

*Algorithm 3:* Subspace classification algorithm.

- 1) Given unknown signal  $\mathbf{x}_0$  and prototypes  $\{\check{\Psi}_c : c \in \mathcal{C}\}$ .
- 2) For  $c = 1$  to  $C$ , set  $v_c \leftarrow \mathcal{A}_w(\check{\Psi}_c, \mathbf{x}_0)$  according to (4).
- 3) Classify  $\mathbf{x}_0$  according to the minimum-distance rule, i.e.,  $\arg \min_{c \in \mathcal{C}} \{v_c\}$ .

## III. DISCRIMINATIVE CLASSIFICATION STRATEGY

The choice of the trajectory norms as the weighting coefficients, in a sense, steers the classifier into favouring trajectories with higher signal content, i.e., such trajectories becoming the primary candidates in the classification task. The drawback of this approach, however, is that if two distinct classes have matching trajectories that are spectrally dominant, it then becomes difficult to separate the two classes using Algorithm 3. A more effective classification strategy, in such a case, would be to substitute the trajectory norms with other qualities that are more discriminative in nature.

### A. Class Discrimination Measure

Let the discriminative potential of a trajectory  $\psi_{c_j} \in \check{\Psi}_c$  of class  $c$  with respect to the trajectories of another class  $\bar{c}$  that is contained in the prototype element  $\check{\Psi}_{\bar{c}} = \{\psi_{\bar{c}k} : k \in \mathcal{J}_{\bar{c}}\}$  be defined as

$$d_1(\psi_{c_j}, \check{\Psi}_{\bar{c}}) \triangleq \left| \|\psi_{c_j}\| - \sqrt{\sum_{k \in \mathcal{J}_{\bar{c}}} \|\psi_{\bar{c}k}\|^2 \cos^2 \beta_{c_j \bar{c}k}} \right| \quad (5)$$

where

$$\cos \beta_{c_j \bar{c} k} = \frac{\psi_{c_j}^T \psi_{\bar{c} k}}{\|\psi_{c_j}\| \|\psi_{\bar{c} k}\|}. \quad (6)$$

The *class discrimination measure*  $d_1$  of (5) resembles the dissimilarity measure  $d$  of (3). Lemma 1 will show that the two measures are equivalent if  $\mathbf{x}_0$  generates  $\Psi_{\bar{c}}$  by Algorithm 1. It can further be shown that, if  $\bar{c} = c$ , then  $d_1(\psi_{c_j}, \Psi_{\bar{c}}) = d_1(\psi_{c_j}, \Psi_c) = 0$ .

*Lemma 1:* Consider a signal  $\mathbf{x}_{\bar{c}}$  belonging to class  $\bar{c}$ . Let  $\mathbf{x}_{\bar{c}}$  generate  $\Psi_{\bar{c}}$  by Algorithm 1 and suppose that  $X \in \mathcal{T}(\mathbf{x}_{\bar{c}})$  is one of the measurement matrices of  $\mathbf{x}_{\bar{c}}$ . Then, for any trajectory  $\psi_{c_j} \in \Psi_c$  of class  $c$ , we have

$$d(\psi_{c_j}, X) = d_1(\psi_{c_j}, \Psi_{\bar{c}}).$$

*Proof:* Let the SVD of the measurement matrix be  $X = \sum_{k=1}^m \mathbf{u}_k \sigma_k \mathbf{v}_k^T$ . We have [11]

$$\|X \psi_{c_j}\| / \|\psi_{c_j}\| = \sqrt{\sum_{k=1}^m \sigma_k^2 \cos^2 \phi_{c_j k}} \quad (7)$$

where  $\cos \phi_{c_j k} = \mathbf{v}_k^T \psi_{c_j} / \|\psi_{c_j}\|$ . It suffices to show that the sums in (5) and (7) are similar. As  $\mathbf{x}_{\bar{c}}$  generates  $\Psi_{\bar{c}}$  by Algorithm 1, then, for all  $k \in \mathbb{N}_m$ , there must exist  $\psi_{\bar{c} k} = \sigma_k \mathbf{v}_k$ , and subsequently

$$\sum_{k=1}^m \sigma_k^2 \cos^2 \phi_{c_j k} = \sum_{k=1}^m \|\psi_{\bar{c} k}\|^2 \cos^2 \beta_{c_j \bar{c} k}$$

with  $\cos \beta_{c_j \bar{c} k}$  defined in (6). Furthermore, Algorithm 1 will show that there are precisely  $m$  active trajectories at any instance. Thus, the sum in (5) can be written as

$$\sum_{k \in \mathcal{J}_{\bar{c}}} \|\psi_{\bar{c} k}\|^2 \cos^2 \beta_{c_j \bar{c} k} = \sum_{k=1}^m \|\psi_{\bar{c} k}\|^2 \cos^2 \beta_{c_j \bar{c} k}$$

thereby completing the proof.  $\blacksquare$

For every class  $c \in \mathcal{C}$ , the measures in (5) are computed for all trajectories of the class with respect to all other classes  $\bar{c} \in \mathcal{C} - c$  according to Algorithm 4. In particular, all trajectories are apportioned a measure that equals its discriminative potential. Between two distinct classes  $c$  and  $\bar{c}$ , a trajectory of  $c$  is said to be highly discriminative if it is largely dissimilar with all other trajectories of  $\bar{c}$ . The set of measures obtained using Algorithm 4, i.e.,  $\{\Delta_{c\bar{c}} : c \in \mathcal{C}, \bar{c} \in \mathcal{C} - c\}$ , is deemed the *class discrimination information*.

*Algorithm 4:* Constructing the set of class discrimination measures.

- 1) Given the prototypes  $\check{\Psi}_c = \{\psi_{c_j} : j \in \mathcal{J}_c\}$  and  $\check{\Psi}_{\bar{c}}$  of two distinct classes  $c$  and  $\bar{c}$ , respectively.
- 2) For  $\forall j \in \mathcal{J}_c$ , set  $\delta_{c_j \bar{c}} \leftarrow d_1(\psi_{c_j}, \check{\Psi}_{\bar{c}})$  according to (5).
- 3) Construct the set of class discrimination measures  $\Delta_{c\bar{c}} \leftarrow \{\delta_{c_j \bar{c}} : j \in \mathcal{J}_c\}$ .

## B. Discriminative Subspace Classification

To accommodate the class discrimination information in the classification procedure, the weighted average of (4) is amended as follows, i.e.,

$$\mathcal{A}_{\delta}(\check{\Psi}_c, \mathbf{x}_0) = \frac{\sum_{j \in \mathcal{J}_c} \sum_{t=1}^T \delta_{c_j}(t) d(\psi_{c_j}(t), X(t))}{\sum_{j \in \mathcal{J}_c} \sum_{t=1}^T \delta_{c_j}(t)} \quad (8)$$

where

$$\delta_{c_j}(t) \triangleq \frac{1}{C-1} \sum_{\bar{c} \in \mathcal{C}-c} \delta_{c_j \bar{c}}(t).$$

Essentially, the weighting coefficients of (8) are the means of the class discrimination measures obtained by Algorithm 4. A signal subspace classifier that implements the class discrimination information is given in Algorithm 5.

*Algorithm 5:* Discriminative subspace classification algorithm.

- 1) Given unknown signal  $\mathbf{x}_0$ , prototypes  $\{\check{\Psi}_c : c \in \mathcal{C}\}$  and the class discrimination information  $\{\Delta_{c\bar{c}} : c \in \mathcal{C}, \bar{c} \in \mathcal{C} - c\}$ .
- 2) For  $c = 1$  to  $C$ , set  $v_c \leftarrow \mathcal{A}_{\delta}(\check{\Psi}_c, \mathbf{x}_0)$  according to (8).
- 3) Classify  $\mathbf{x}_0$  according to the minimum-distance rule, i.e.,  $\arg \min_{c \in \mathcal{C}} \{v_c\}$ .

It is interesting to note that, in the event of the null discrimination, then  $d_1(\psi_{c_j}, \emptyset) = \|\psi_{c_j}\|$  and  $\delta_{c_j} = \|\psi_{c_j}\| = w_{c_j}$ , or equivalently, the weighted average of (8) reduces to (4).

## IV. RESULTS AND DISCUSSIONS

This section examines the performance of the proposed signal subspace classifier as a speech recognizer in two test cases.

### A. Isolated Digit Speech Recognition

Speech recordings, at a sampling frequency of 10 kHz, are collected from a pool of 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. Next, a recording of each digit is selected randomly to build the set of class prototypes (Algorithms 1 and 2) and the class discrimination information (Algorithm 4), while the other nine recordings are used as the testing data. The signal subspace classifier obtained is then evaluated on the testing data with the average recognition rate (of the three speakers) as the yardstick for performance.

In the simulation, we have used  $K = 160$  and  $K_1 = 40$  and chosen  $(m = 20, \theta_{\text{th}} = 25^\circ, E_{\text{th}} = 0.9)$  and  $(m = 28, \theta_{\text{th}} = 35^\circ, E_{\text{th}} = 0.925)$  as the parameter sets of two signal subspace classifiers, dubbed DSub1 and DSub2, respectively. For comparison purposes, the same parameter sets are also used to construct two subspace classifiers, dubbed Sub1 and Sub2, respectively, which implement Algorithm 3 (see [11]). We have also 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.

TABLE I  
RECOGNITION RATE, IN PERCENTAGE (%), OF THE VARIOUS  
SPEECH RECOGNIZERS

| Classifier | Recognition rate (%) |
|------------|----------------------|
| DSub1      | 86.9                 |
| DSub2      | 92.3                 |
| Sub1       | 85.2                 |
| Sub2       | 89.6                 |
| LPCC1      | 89.2                 |
| LPCC2      | 92.9                 |
| MFCC       | 85.2                 |

Two variants of the LPCC, each with a different distortion measure, are tested: LPCC1 for the Euclidean distance and LPCC2 for the cepstral projection measure [15]. For the LPCC and MFCC recognizers, a Hamming window ( $K = 240, K_1 = 80$ ) is applied to the data, and 12 cepstral coefficients, filtered with  $w_{\text{lift}}(k) = 1 + 6 \sin(\pi k/12)$ , are retained as the cepstral vector [1, Chap. 4]. The main results of the simulation are displayed in Table I.

It is apparent, from the results in Table I, that the proposed signal subspace classifiers (DSub1 and DSub2) are superior in performance to their predecessors (Sub1 and Sub2). The improvement is especially evident in DSub2 (over Sub2) as compared to DSub1 (over Sub1). It has been shown in [11] that Sub2 generates more trajectories than Sub1, i.e., 47 against 28.8 on the average. Thus, more reliable interclass information is available to Algorithm 4 for constructing DSub2 and consequently the marked improvement.

### B. White Noise Robustness

The second part of our simulation investigates the white noise robustness of the proposed classifier. The signal-to-noise ratio (SNR) of the original recorded speech data is approximately 24 dB. Artificial stationary white noise is introduced into the testing data, and the classifier's performance at various levels of SNR is recorded [15]. As in [11], the proposed signal subspace classifier displays remarkable robustness to additive white noise (see Fig. 1). This is almost in complete contrast to the LPCC and MFCC recognizers, which fare poorly in noisy environments. The white noise robustness of subspace classifiers owes much to the noise filtering quality inherent in subspace-based methods.

### V. CONCLUSION

This letter examines a signal subspace speech classifier that improves on [11]. The "best representation" approach taken in [11] is restrictive in the sense that it made no reference to valuable interclass information. Due to that, it is prohibitive when distinguishing between phonetically alike classes. Addressing these issues, this letter defines a set of class discrimination measures, deemed the class discrimination information, to characterize discriminative qualities of individual class elements that

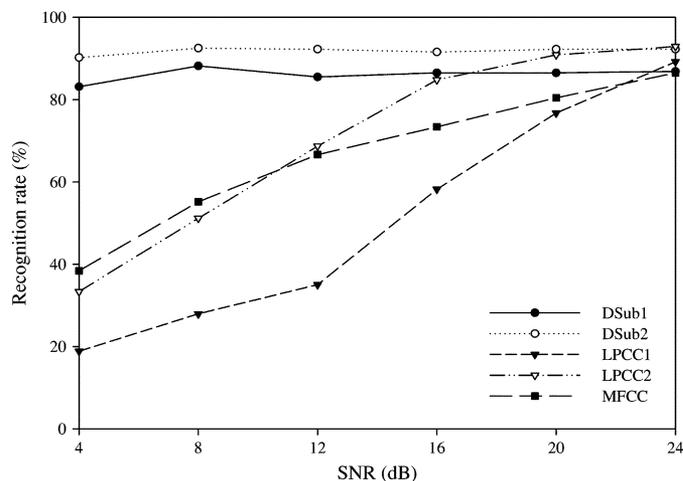


Fig. 1. Performance of various speech recognizers as SNR varies.

are later applied in the classification procedure. Results from simulation affirm the theoretical analysis.

### REFERENCES

- [1] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [2] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [3] J. W. Picone, "Signal modeling techniques in speech recognition," *Proc. IEEE*, vol. 81, no. 9, pp. 1215–1247, Sep. 1993.
- [4] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. PROC-63, no. 4, pp. 561–580, Apr. 1975.
- [5] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-23, no. 1, pp. 67–72, Feb. 1975.
- [6] A. V. Oppenheim and R. W. Schafer, "Homomorphic analysis of speech," *IEEE Trans. Audio Electroacoust.*, vol. AE-16, no. 2, pp. 221–226, Jun. 1968.
- [7] Y. Tokhura, "A weighted cepstral distance measure for speech recognition," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 35, no. 10, pp. 1414–1422, Oct. 1987.
- [8] 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.
- [9] 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.
- [10] P. S. K. Hansen, "Signal subspace methods for speech enhancement," Ph.D. dissertation, Tech. Univ. Denmark, Lyngby, 1997.
- [11] 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.
- [12] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 713–718, Mar. 1992.
- [13] N. Saito and R. R. Coifman, "Local discriminant bases," in *Proc. SPIE Wavelet Applications Signal Image Processing II*, 1994, vol. 2303, pp. 2–14.
- [14] G. H. Golub and C. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: The John Hopkins Univ. Press, 1989.
- [15] 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. 37, no. 11, pp. 1659–1671, Nov. 1989.