Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

# Local reconstruction method and voice system

R. Radha[*], S. Sivananthan

*Department of Mathematics, Indian Institute of Technology Madras, Chennai 600036, India*

ARTICLE INFO

*Article history:*
Received 10 February 2008
Accepted 13 January 2009

*Keywords:*
Non uniform sampling
Probability distribution
Speech recognizer
Wavelet

ABSTRACT

It is shown that a local reconstruction method from a nonuniform sampled data along with discrete wavelet transform and a simple statistical method is applicable in a voice system.

## 1. Introduction and background

Speech recognizers are used in some graphics work stations as input devices to accept voice commands. The voice system input can be used to initiate graphics operations or to enter data (see [1]). These systems operate by matching an input against a predefined dictionary of words and phrases.

A dictionary is set up for a particular operator by having the operator speak the command words to be used into the system. Each word is spoken several times, and the system analyses the word and establishes a frequency pattern for that word in the dictionary along with the corresponding function to be performed. Later, when a voice command is given, the system searches the dictionary for a frequency-pattern match.

**Definition 1.1.** Let $\psi \in L^2(\mathbb{R})$, $\psi_{j,k}(x) = 2^{-\frac{j}{2}}\psi(2^{-j}x - k),\ j, k \in \mathbb{Z}$. If $\{\psi_{jk} : j, k \in \mathbb{Z}\}$ forms an orthonormal basis for $L^2(\mathbb{R})$, then $\psi$ is called a wavelet and $\{\psi_{jk} : j, k \in \mathbb{Z}\}$ a wavelet basis in $L^2(\mathbb{R})$. Using this wavelet basis, any $f \in L^2(\mathbb{R})$ can be written as

$$f = \sum_{j,k} \langle f, \psi_{jk}\rangle \psi_{jk}. \tag{1.1}$$

The sequence $\{\langle f, \psi_{jk}\rangle : j, k \in \mathbb{Z}\}$ is called discrete wavelet transform (DWT) of $f$. The inversion formula of discrete wavelet transform (IDWT) is given by the Eq. (1.1).

**Definition 1.2.** A orthogonal multi resolution analysis (MRA) is a sequence of closed subspaces $V_j,\ j \in \mathbb{Z}$, in $L^2(\mathbb{R})$ such that

1. $V_j \subset V_{j-1}$ for all $j \in \mathbb{Z}$
2. $\overline{\bigcup V_j} = L^2(\mathbb{R})$ and $\bigcap V_j = \{0\}$
3. $f \in V_j \Leftrightarrow f(2\cdot) \in V_{j-1}$
4. $f \in V_j \Leftrightarrow f(\cdot - k) \in V_j$ for all $k \in \mathbb{Z}$.
5. There exists a scaling function $\varphi \in V_0$ such that $\{\varphi(\cdot - k) : k \in \mathbb{Z}\}$ forms an orthogonal basis for $V_0$.

[*] Corresponding author.
*E-mail addresses:* radharam@iitm.ac.in (R. Radha), ssiva_math@yahoo.co.in (S. Sivananthan).

0898-1221/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2009.01.036

Then $\varphi$ satisfies

$$\varphi(x) = \sum_k c_k \varphi(2x-k).$$

This is called scaling identity.

**Theorem 1.3.** *Let $\{V_n, \varphi\}$ denote a multiresoltion analysis. Then there exists a wavelet $\psi$ defined by*

$$\psi(x) = \sqrt{2} \sum_{k=-\infty}^{\infty} \beta_k\, \varphi(2x-k), \quad \text{where } \beta_k = (-1)^k \bar{\alpha}_{1-k}. \tag{1.2}$$

*where $\alpha_k$ satisfies $\varphi(x) = \sqrt{2} \sum_{k=-\infty}^{\infty} \alpha_k\, \varphi(2x-k)$.*

We refer to Daubechies [2] for a detailed study of wavelets.

Assume that $\psi$ comes from an orthogonal MRA. Then one can implement the DWT repeatedly as follows: If $c_j(k) = \langle f, \varphi_{jk}\rangle$, $d_j(k) = \langle f, \psi_{jk}\rangle$ denote the coefficients associated with the scaling function and the wavelet at the $j$th resolution ($j$th level), then the corresponding coefficients at $j+1$th resolution ($j+1$th level) can be obtained using the following formulae:

$$c_{j+1}(k) = \sum_{l\in\mathbb{Z}} h(l-2k) c_j(l) \tag{1.3}$$

$$d_{j+1}(k) = \sum_{l\in\mathbb{Z}} (-1)^k \bar{h}(2k-l-1) c_j(l) \tag{1.4}$$

where $h(k)$ satisfies the following: $\varphi(\frac{x}{2}) = 2^{\frac{1}{2}} \sum_k h(k)\varphi(x-k)$.

The coefficients $c_j(k)$ at resolution $j$ (level $j$) can be obtained from the coefficients $c_{j+1}$ and $d_{j+1}$ at a coarser resolution (level $j+1$) by the reconstruction algorithm

$$c_j(k) = \sum_{l\in\mathbb{Z}} h(2l-k) c_{j+1}(l) + \sum_{l\in\mathbb{Z}} (-1)^k \bar{h}(k-2l-1) d_{j+1}(l) \tag{1.5}$$

which is essentially the computational algorithm for inverting the DWT associated with MRAs. We refer to [3] for further details.

**Theorem 1.4** (*Central Limit Theorem*). *If $\overline{X}$ is the mean of a sample of size $n$ taken from a population having the mean $\mu$ and the finite variance $\sigma^2$, then $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is a random variable whose distribution function approaches that of the standard normal distribution as $n \to \infty$.*

**Remark 1.5.** 1. In practice, the normal distribution provides an excellent approximation to the sampling distribution of the mean $\overline{X}$ for $n$ as small as 25 or 30, with hardly any restrictions on the shape of the population.

2. The sample mean $\overline{X}$ and population mean $\mu$ differ from each other as follows. We can assert with probability $1-\alpha$ that the inequality $-z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}$ will be satisfied for the large sample with size $n$ when $n \ge 30$. In other words $|\overline{X}-\mu| \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = E$ with probability $1-\alpha$. The most practical values for $1-\alpha$ are 0.95 and 0.99. Graphically $\frac{\alpha}{2}$ denotes the area under the normal curve to the right of $z_{\alpha/2}$.

We refer to [4] for background in statistics.

The local reconstruction from a finite number of nonuniform samples is one of the most desirable properties for many applications in signal processing. However, the local reconstruction problem has not been investigated except for the spline space and shift invariant space with compactly supported generator ([5,6]). The natural question is to obtain a local reconstruction method for functions belonging to shift invariant spaces with other generators. Recently in [7], the authors discuss a local reconstruction method for functions belonging to a shift invariant space with the generator having polynomial decay in time domain and moderate decay in frequency domain such that its Fourier transform is non vanishing on a unit interval. In fact, the following results are proved in [7].

Let $E_l(\mathbb{R})$ $(l \ge 2)$ denote the class of all complex valued continuous functions $\varphi$ defined on $\mathbb{R}$ satisfying the following conditions

(i) $\varphi(x) = o\left(\frac{1}{x^l}\right)$ and $\exists\, \alpha > 1$ such that $\hat{\varphi}(x) = \mathcal{O}\left(\frac{1}{x^\alpha}\right)$, where $\hat{\varphi}$ denotes the Fourier transform of $\varphi$.

(ii) $\hat{\varphi}(x) \neq 0$ for all $x \in [b, b+1]$, for some $b \in \mathbb{R}$.

Then one has the following results.

**Theorem 1.6.** *Let $\varphi \in E_{2l}(\mathbb{R})$ and $f \in V(\varphi)$. Let $[a,b]$ be an interval in $\mathbb{R}$. Given $\epsilon > 0$, there exists a positive integer $M$ such that*

$$\left| f(x) - \sum_{k\in(a-M,b+M)\cap\mathbb{Z}} c_k\varphi(x-k) \right| < \frac{\epsilon}{M^l} \quad \forall\, x \in [a,b].$$

*In other words, $f$ restricted to an interval $[a,b]$ can be approximately determined by a finite number of coefficients $c_k$ locally.*

**Theorem 1.7.** *Fix $l \geq 2$. Let $\varphi \in E_{2l}(\mathbb{R})$, $f \in V(\varphi)$, $[a,b]$ an interval in $\mathbb{R}$ and $\epsilon > 0$. Let $M$ be a positive integer satisfying Theorem 1.6. Let $X$ denote a sample set $\{x_j\}$, $x_j \in [a,b]$ such that $2M+b-a-1 \leq \#X \leq M^{2l}$, where $\#X$ denotes the number of elements in $X$. Define $U_{jk} = \varphi(x_j - k)$, $1 \leq j \leq \#X$, $k \in [a-M+1, B+M-1] \cap \mathbb{Z}$. If $U$ possesses full column rank then $\exists\, g_r \in V(\varphi)$ (the reconstructed function for $f|_{[a,b]}$ from the nonuniform sample $X$) such that*

$$\|f|_X - g_r|_X\|_2 < \epsilon(1 + \|U\|\|(U^TU)^{-1}U^T\|) + \mathcal{O}(\epsilon^2).$$

In this paper, we show that the above local reconstruction method along with a simple statistical method can be applied to find a pattern match in voice system. Since in practice voice signals occur with noise, we use discrete wavelet transform for the purpose of denoising. We also provide an algorithm and illustrations using Matlab. This method is advantageous compared to the usage of statistical method alone as it reduces storage space a lot. In fact, for representing the signals shown in Fig. 3 one originally uses around 10,000 sample data. On the other hand, we require only 250 samples.

## 2. The main result

Let us assume that the predefined dictionary consists of, say, $n$ signal patterns, where in each signal pattern corresponds to a particular word. In order to store a frequency pattern for a particular word, an operator has to speak several times. At the same time we cannot expect the operator to speak the same word in exactly the same fashion. Therefore we have to collect at least 30 different signal patterns (30 samples) for a particular word and then choose an *ideal* frequency pattern for each word. In this way, we save $n$ words.

We assume that these frequency patterns are members of $V(\varphi)$. The original frequency pattern of an ideal word is obtained from $m$ different signal patterns of the word. More precisely, each function value of the ideal word is taken as the sample mean arising from these samples. Notice that these sample means are not true means. But we know that the sample mean differs from the population mean at the most $E$ mentioned in Remark 1.5.

Now the problem is stated as follows: Input a signal and test whether this signal finds a match with a signal in the database which consists of $n$ words. We assume that all these signals are in $V(\varphi)$. Each word (a signal $h$) need not be stored in the database as such. We can save only some function values of $h$ which will help us to reconstruct $h$. These function values need not be taken uniformly. A nonuniform sample of $h$ alone needs to be saved in the database so that $h$ can be reconstructed from these values with '$\epsilon$' error (which is a priori fixed based on the application). In other words we do the following. Discretize $h$ (a voice signal). i.e., Find a large number of $u_j$ so that $h$ can be reconstructed more exactly from these $u_j$. These values $h(u_j)$, $j = 1, 2, \ldots, N'$ generate a sample drawn from a normal population when $N' \geq 30$. However in the practical situation $N'$ will be very large. But we need not take all these $u_j$'s to obtain a sample. We choose the nonuniform sample $x_j$, $j = 1, 2, \ldots, N_1$, from the sample $u_j, j = 1, 2, \ldots, N'$, so that $h$ can be reconstructed from these $h(x_j)$ using Theorem 1.7. Let $\zeta_1$ denote the sample $\zeta_1 : y_j = h(x_j), j = 1, 2, \ldots, N_1$. Thus we save only $\zeta_1$ in the database instead of $h$. Let $g$ denote an input signal. Similarly find $t_j$'s using which $g$ can be reconstructed from $g(t_j)$. Let $\zeta_2$ (the sample of $g$) be written as $\zeta_2 : z_j = g(t_j), j = 1, 2, \ldots, N_2$. i.e., we input $\zeta_2$ only. Now, in order to say that $h$ and $g$ are matched, we need to say that $\zeta_1, \zeta_2$ are drawn from the populations $P_1, P_2$ having the same mean $\mu_1 = \mu_2$ and variance $\sigma_1^2 = \sigma_2^2$. Towards this end, we adopt the following tests in statistics: Let $\zeta_1 : y_1, y_2, \ldots, y_{N_1}$ and $\zeta_2 : z_1, z_2, \ldots, z_{N_2}$ denote two samples.

**Test for equality of mean**

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$
Compute

$$z = \frac{\bar{y} - \bar{z}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{2.1}$$

where $\bar{y} = \frac{y_1+y_2+\cdots+y_{N_1}}{N_1}$, $\bar{z} = \frac{z_1+z_2+\cdots+z_{N_2}}{N_2}$ and $s_1^2, s_2^2$ denote the sample variances. If the calculated value of $z$ lies between $-z_{\frac{\alpha}{2}} \leq z \leq z_{\frac{\alpha}{2}}$, then we conclude that with $(1-\alpha)100\%$ confidence that $\zeta_1, \zeta_2$ are drawn from the populations having the same mean.

**Table 1**
Comparison with signal 4(a).

| Signal in database | $z$ | $F$ |
|---|---|---|
| (a) | −1.7027 | 2.0866 |
| (b) | −2.3274 | 1.2433 |
| (c) | −2.1959 | 1.0195 |
| (d) | 3.2885e−004 | 1.0003 |

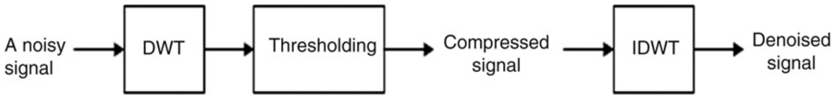
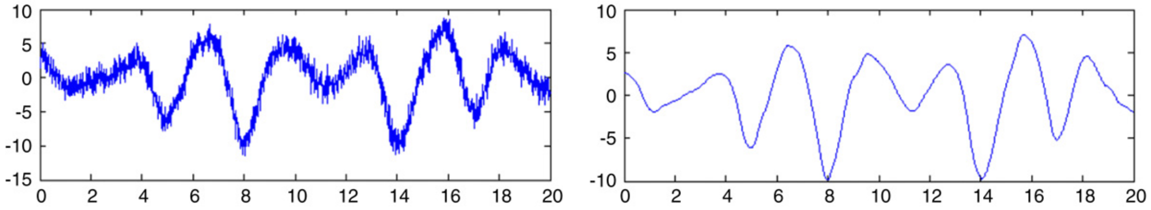


**Fig. 1.**



(a) Noisy signal.

(b) Denoised signal using Db5 wavelet with 16 levels.

**Fig. 2.**

**Test for equality of variance**
$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 \neq \sigma_2^2$
Compute

$$F = \frac{s_{\max}^2}{s_{\min}^2} \tag{2.2}$$

where $s_{\max} = \max\{s_1, s_2\}$, $s_{\min} = \min\{s_1, s_2\}$ where $s_1^2, s_2^2$ denote the sample variances.

If the calculated value of $F$ is less than $f_{\frac{\alpha}{2},N-1,N-1}$ then we conclude with $(1-\alpha)100\%$ confidence that $\zeta_1, \zeta_2$ are drawn from the populations having the same variance.

We can notice an advantage here. Originally one has to work with all $u_j$'s but the procedure is carried out using $x_j$'s only. This is possible because the reconstruction is unique. Thus **storage space** for signals in the database as well as for input signal is **reduced**.

Based on these tests if $g$ is matched with a function $h$ in the database, $h$ is reconstructed and shown.

We also adopt the following strategy. Since the frequency patterns are obtained on different domains, for the sake of convenience in order to find matching, we do the following: First, shift and scale the frequency patterns from their domains to the interval $[0, 1]$ and then store uniformly all of them on an interval $[a, b]$. In the same way, the input signal will also be represented on the same interval $[a, b]$.

We provide the algorithm in the next section. As mentioned earlier, in practice there may be an occurrence of noise along with a voice signal. In such cases, we use denoising as pre-processing before inputting the signal or storing it in the database.

Given a signal, we apply discrete wavelet transform using (1.3) and (1.4). The resulting signal has two parts consisting of low frequency components and high frequency components. In general, it is assumed that the appearance of noise is exhibited in high frequency components only and hence we apply thresholding to the high frequency components to get the compressed signal. Then we apply the inverse discrete wavelet transform using (1.5). The resulting signal may contain less amount of noise compared to the original signal. Thus in order to obtain a denoised signal it may be necessary to apply DWT Eqs. (1.3) and (1.4) repeatedly (by varying $j$) before applying thresholding. This method is illustrated in Fig. 1.

## 3. Algorithm and illustrations

**Step-0.** To reconstruct $g_r$ from the nonuniform samples of $f$ using Theorems 1.6 and 1.7.

(1) Let $y_i = f(x_i),\ i = 1, 2, \ldots, N$. Let $y = (y_1\, y_2 \ldots y_N)^T$, where $T$ denotes the transpose and $N = \#X$.
(2) Define the matrix $U$ as $U_{jk} = \varphi(x_j - k)$ for $1 \leq j \leq N$, $k \in I$. Write $y = Uc$. Here $I = [a - M + 1, b + M - 1] \cap \mathbb{Z}$ and $c = (c_k)_{k \in I}$.

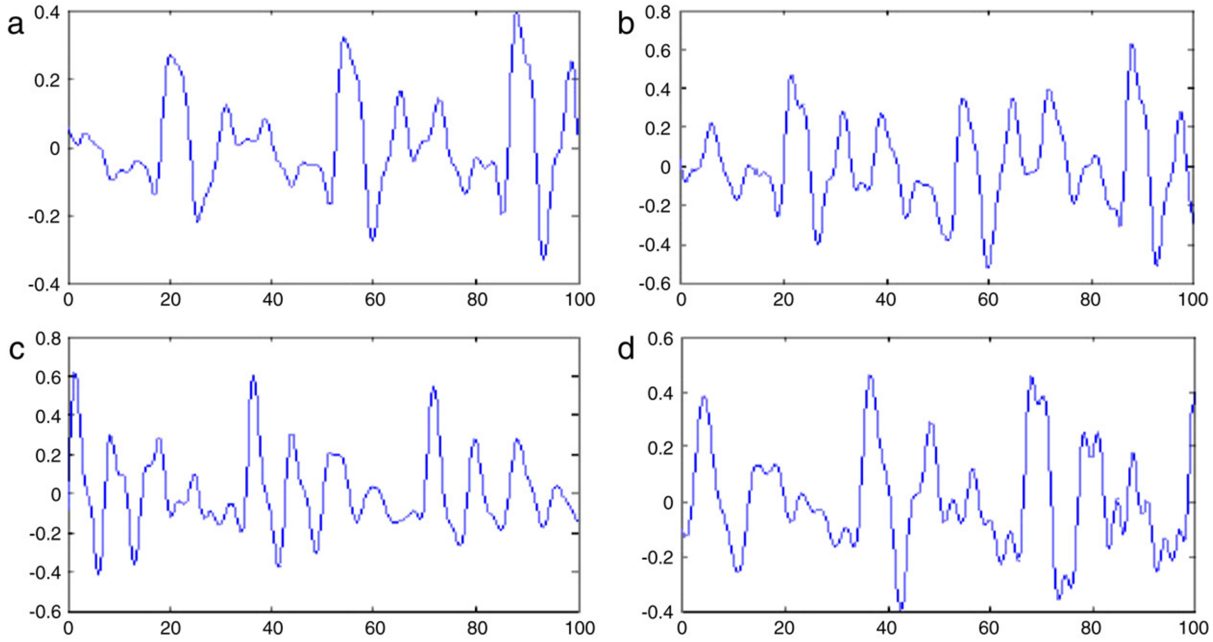
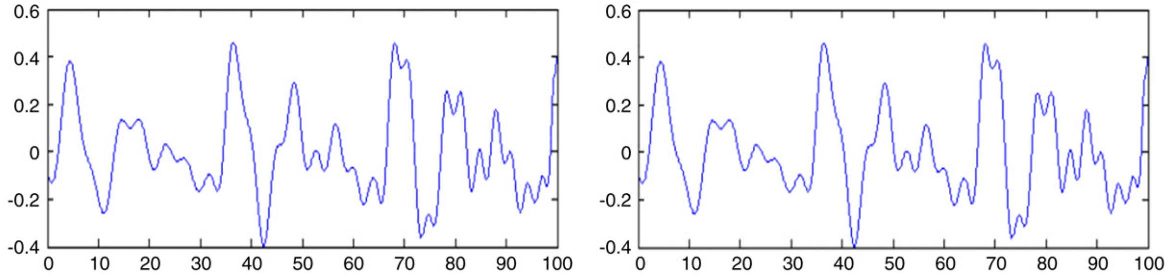


**Fig. 3.** Signals in database.



(a) Input signal.

(b) Reconstructed signal.

**Fig. 4.**

(3) Compute $\tilde{c}$ as $\tilde{c} = (U^T U)^{-1} U^T y$ and calculate

$$g_r(x) = \sum_{k \in I} \tilde{c}_k \, \varphi(x - k)$$

**Step-1.** To select an "ideal" word.

(i) Various frequency patterns denoting same word are input. Here $n(\geq 30)$ such frequency patterns $h_1, h_2, \ldots, h_n$ need to be collected.

(ii) Applying scaling and translation wherever necessary and collect $h_1, h_2, \ldots, h_n$ on a particular time domain $[a, b]$.

(iii) Let $h = \frac{h_1 + h_2 + \cdots + h_n}{n}$. Then consider the nonuniform sample $h(x_j)$, where the sample set $X = \{x_1, x_2, \ldots, x_{N_1}\}$ satisfies the required conditions as in Theorem 1.7, from which $h$ can be reconstructed with the error $\epsilon$.

(iv) Now store the nonuniform sample $h(x_j)$ with the sample points $x_j$ in the database. This represents our ideal word.

**Step-2.** Apply if necessary scaling and translation and store all *ideal* frequency patterns (denoting different words in a predefined dictionary) on a fixed time domain $[a, b]$. Collect $N$(say 10) such signals in a data base.

**Step-3.** Input a new signal $g$ which is to be matched. First shift it to the domain $[a, b]$, after applying scaling if necessary.

**Step-4.** Now choose the sample set $Y = \{y_1, y_2, \ldots, y_{N_2}\}$ such that $g$ can be reconstructed from the samples $\{g(y_j) : y_j \in Y\}$ with the error $\epsilon$.

**Step-5.** Calculate $z$ and $F$ (see Eqs. (2.1) and (2.2)) for the input $\{g(y_j) : y_j \in Y\}$ and for each ideal word in the database $\{h(x_j) : x_j \in X\}$. Apply the test for equality of mean and variance. Based on these tests, if for some ideal word $h$ in the database and input signal $g$ we can conclude that $h$ and $g$ are drawn from the population with the same mean and variance,

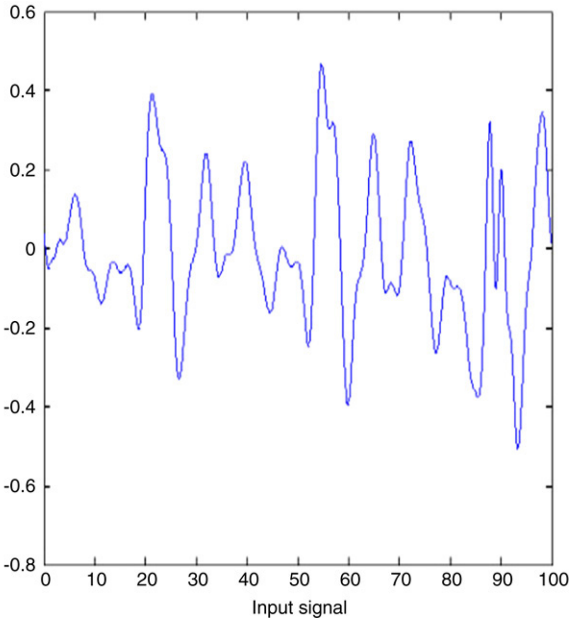


**Fig. 5.**

**Table 2**
Comparison with signal 5.

| Signal in database | $z$ | $F$ |
|---|---:|---:|
| (a) | 0.6538 | 2.2131 |
| (b) | −0.3977 | 1.1722 |
| (c) | −0.1652 | 1.0404 |
| (d) | 2.0083 | 1.0609 |

we decide that $h$ and $g$ are matched and we display the original signal. Otherwise (if none of $h$ matches with $g$) we print the message "Not matched with any of the signals in the database".

A signal in the presence of noise is shown in Fig. 2(a). The denoised signal is shown in Fig. 2(b), using discrete wavelet transform with Db5 wavelet.

The Fig. 3(a), (b), (c) and (d) show the original (voice) signals, which are stored in the database as nonuniform samples (as discussed in algorithm) with approximately 250 sample values (originally there were 10,000 sample values to describe each of these functions). The input signal is shown in Fig. 4(a). This is being compared with each signal in the database. Out of these signals one signal is matched with the input signal. The matched signal which is reconstructed from the database is shown in Fig. 4(b). Similarly Fig. 5 is an input signal which does not find a match in the data base. The calculated values of $z$ and $F$ are tabulated in Tables 1 and 2 respectively. Here we take $1-\alpha = 0.95$ for the testing the equality of means and $\alpha = 0.01$ for testing the equality of variance.

## Acknowledgment

We thank National Board for Higher Mathematics, Department of Atomic Energy for the project fund with the project No: 48/4/2004/R&D-II/2119.

## References

[1] D. Hearn, M. Pauline Baker, Computer Graphics, Prentice-Hall of India, 1997.
[2] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, SIAM, Philadelphia, PA, 1992.
[3] A. Aldroubi, The wavelet transform: a surfing guide, in: A. Aldroubi, M. Unser (Eds.), Wavelets in medicine and biology, CRC, Boca Raton, FL, 1996, pp. 3–36.
[4] I. Miller, J.E. Freund, Probability and Statistics for Engineers, Prentice-Hall of India, 1981.
[5] K. Gröchenig, H. Schwab, Fast local reconstruction methods for nonuniform sampling in shift-invariant spaces, SIAM J. Matrix Anal. Appl. 24 (4) (2003) 899–913.
[6] Q. Sun, Local reconstruction for sampling in shift-invariant spaces, preprint, 2007.
[7] R. Radha, S. Sivananthan, Local reconstruction of a function from a non-uniform sampled data, Appl. Numer. Math. 59 (2) (2009) 393–403.