Reconstruction of Speech Signals with Lost Samples

Farokh A. Marvasti, *Senior Member, IEEE,* Peter M. Clarkson, *Senior Member, IEEE,* Miroslav V. Dokic, Ut Goenchanart, and Chuande Liu

***Abstract*—In various applications of speech transmission and processing, there is a possibility of lost samples. In this paper, we examine a simple but effective technique to recover a speech signal from the remaining samples. The algorithm is implemented in real time using a low-cost DSP processor, and subjective tests verify the theoretical and objective simulation results.**

I. Introduction

In digital recording and playback of voice signals, and in the digital transmission of speech, there is always a possibility of loss of samples. The missing samples have to be interpolated somehow if the quality of the speech signal is not to be sacrificed. Various techniques have been employed for this purpose; they range from simple low-pass filtering to more sophisticated estimation of the missing values using previous and future samples [1]–[4]. Another common approach is to add redundancy bits for error and erasure correction; this method has the advantage of detecting and correcting errors in samples that would otherwise be assumed to be correct. However, this has the disadvantage of requiring additional bandwidth.

In this paper we propose a different approach consisting of the application of a simple nonlinear system which can be implemented in real-time on a low-cost microprocessor system. We shall show that we can obtain significant improvement relative to simple interpolation schemes. By contrast, more sophisticated techniques are unsuitable for real-time implementation due to their computational complexity.

II. Method and Analysis

The proposed method is shown in Fig. 1, and is a development of an approach outlined in [5], [6] and discussed in [7]. The degraded signal (nonuniform samples of speech) is represented by $x_s(t)$. This is subsequently low-pass filtered giving $x_{s_{lp}}(t)$. In parallel, $x_s(t)$ is hard limited, rectified, and low-pass filtered to yield $x_{p_{lp}}(t)$, which is a representation of the positions of the missing samples. Division of $x_{s_{lp}}(t)$ by $x_{p_{lp}}(t)$ yields $\hat{x}(t)$, an estimate of the original speech signal.

Manuscript received May 10, 1990; revised September 26, 1991.

F. A. Marvasti is with the Communications Research Group, Department of Electrical Engineering, King's College, University of London, Strand, London WC 2R 2LS, U.K.

P. M. Clarkson, M. V. Dokic, and U. Goenchanart are with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616.

C. Liu is with U.S. Robotics, Skokie, Chicago, IL.

IEEE Log Number 9203237.

A set of uniform samples with some missing is a special case of a set of nonuniform samples. For a review of nonuniform sampling theory, see [7]. According to nonuniform sampling theory, if the average rate of sampling is greater than the Nyquist rate, then we can uniquely recover the original band-limited signal. This theorem implies that, in general, if the sampling rate is $a\%$ higher than the Nyquist rate, we can afford to lose up to $(a/100)/(1 + a/100)$ of the samples and still satisfy the uniqueness theorem. For exact recovery, iterative methods can be used [8]–[10]. The method shown in Fig. 1 is an approximation which works quite well despite its simplicity. The analysis is similar to the general model for the nonuniform sampling case [5]. For our special case, from Fig. 1, we have

$$x_s(t) = x(t)x_p(t) \tag{1}$$

where

$$x_p(t) = \sum_k \delta(t - t_k); \qquad t_k = kT_1, \qquad k \neq \{i\} \tag{2}$$

where we have denoted the nonuniform positions t_k as uniform positions $t_k = kT_1$ except when the samples are lost; we denote the positions of the lost samples as $\{i\}$. The sampling rate $(1/T_1)$ is assumed to be $a\%$ higher than the Nyquist rate $(1/T)$, so that

$$\frac{1}{T_1} = \frac{(1 + a/100)}{T}.$$

We assume that the percentage of lost samples is $b < (a/(1 + a/100))$ % giving an average sampling rate of

$$\frac{1}{T_2} = \frac{1}{T_1} - \frac{b/100}{T_1} = \frac{(1 - b/100)(1 + a/100)}{T} > \frac{1}{T}. \tag{3}$$

From the theory of generalized functions [11], we can write

$$\delta(t - t_k) = |\dot{g}(t)|\delta[g(t)] \tag{4}$$

provided that $g(t_k) = 0$ and $\dot{g}(t_k) \neq 0$, and that $g(t)$ has no zeros other than $\{t_k\}$. One possible $g(t)$ can be written as

$$g(t) = t - kT_2 - \theta(t) \tag{5}$$

where $\theta(t)$ is any function such that $g(t_k) = t_k - kT_2 - \theta(t_k) = 0$, that is, $\theta(t_k) = t_k - kT_2$ is the deviation of the nonuniform samples from the uniform position of an av-

1053-587X/93$3.00

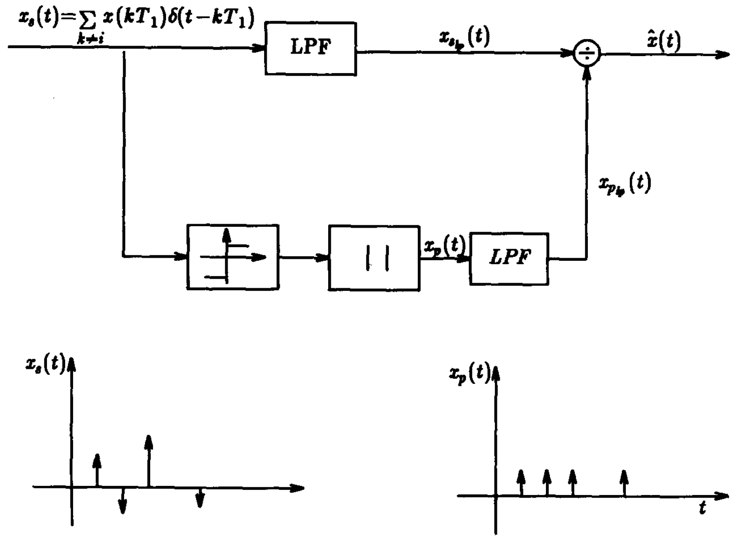


Fig. 1. Block diagram of the reconstruction scheme.

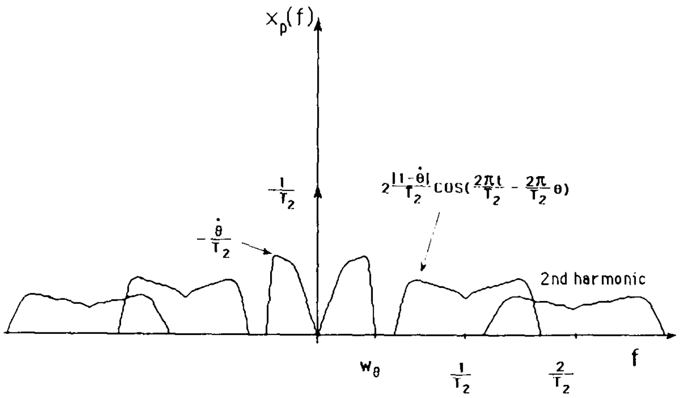


Fig. 2. Spectrum of $x_s(t)$ (see (8)).

erage sampling period[1] of

$$x_p(t) = [1 - \dot\theta(t)] \sum_{k=-\infty}^{\infty} \delta(\Phi(t) - kT_2) \tag{6}$$

where $\Phi(t) = t - \theta(t)$, and where we assume that $1 > \dot\theta(t)$. The Fourier expansion of (6) is[2]

$$\begin{aligned} x_p(t) &= [1 - \dot\theta(t)] \frac{1}{T_2} \sum_{k=-\infty}^{\infty} e^{jk2\pi/T_2\Phi(t)} \\ &= \frac{[1 - \dot\theta(t)]}{T_2}\left[1 + 2\sum_{k=1}^{\infty} \cos\left(\frac{2\pi kt}{T_2} - \frac{2\pi k\theta(t)}{T_2}\right)\right]. \end{aligned} \tag{7}$$

[1]Since the average sampling rate of the nonuniform samples of $\theta(t_k)$ is $(1/T_2)$, we can uniquely determine a band-limited function $\theta(t)$ with a bandwidth of $(1/2T_2)$ or less.

[2]Note that a similar analysis has been conducted by Soumekh [12] for the problem of nonuniform samples of a two-dimensional signal.

Equation (7) reveals that $x_p(t)$ has a DC component $[1 - \dot\theta(t)]/T_2$ plus harmonics that resemble phase modulated (PM) signals. The index of modulation is $(2\pi k)/T_2$; the bandwidth is proportional to the index of modulation, the bandwidth of $\theta(t)$ and the maximum amplitude of $\theta(t)$, which in this case is related to $\theta(t_k) = t_k - kT_2$. The spectrum of $x_p(t)$ is sketched in Fig. 2. The nonuniform samples can be expanded using (1) and (7), that is,

$$x_s(t) = x(t)x_p(t) = x(t)\frac{[1 - \dot\theta(t)]}{T_2} \cdot \left[1 + 2\sum_{k=1}^{\infty} \cos\left(\frac{2\pi kt}{T_2} - \frac{2\pi k\theta(t)}{T_2}\right)\right]. \tag{8}$$

The spectrum of (8) is obtained by convolving the spectrum of $x(t)$ with the spectrum shown in Fig. 2. Now if we assume $(1/T_2) >> (1/T)$, and given that the bandwidth of $\theta(t)$ is less than $(1/2T_2)$, the PM signal at the carrier frequency $(1/T_2)$ is a narrow-band PM and has a bandwidth of approximately twice the bandwidth of $\theta(t)$, that is less than $(1/T_2)$. Therefore, low-pass filtering $x_p(t)$ (see Fig. 2) yields

$$x_{p_{lp}}(t) = \frac{[1 - \dot{\theta}(t)]}{T_2}. \quad (9)$$

If the bandwidth of $\theta(t)$ is taken to be W_θ, then as long as $(1/T_2) - W_\theta - W > W + W_\theta$, there is no overlap between the narrow band PM signal and $X(f) * \mathfrak{F}\{(-\dot{\theta}(t)/T_2)\}$ where $\mathfrak{F}$ is the Fourier transform operator. Thus, low-pass filtering $x_s(t)$ (with a bandwidth of $W + W_\theta$), we get

$$x_{s_{lp}}(t) = x(t) \frac{[1 - \dot{\theta}(t)]}{T_2}. \quad (10)$$

Comparing (9) to (10), we can recover $x(t)$ by dividing $x_{s_{lp}}(t)$ by $x_{p_{lp}}(t)$, that is,

$$x(t) = \frac{x_{s_{lp}}(t)}{x_{p_{lp}}(t)}. \quad (11)$$

(The division is possible if $\dot{\theta}(t) \neq 1$ and hence the denominator of (11) is nonzero. It can be shown [13] that a sufficient condition for this to be true is that $|\theta(t)| < T_2/\pi$ for all values of t.) In practice, we normally choose the bandwidth of the low-pass filter for $x_{s_{lp}}(t)$ to be equal to $(1/2T)$. This assumption creates some minor distortion.

III. Results

A number of off-line simulations of the enhancement system have been conducted. These simulations employed two phonetically balanced sentences: "The Navy attacked the big task force" and "See the cat glaring at the scared mouse." The speech data was recorded under low noise, anechoic conditions. Two speakers were used, one male and one female. The speech was subsequently antialias filtered at 3.5 kHz and digitized using a 12-bit A/D at a sampling rate of 8 kHz. Sample rate increases were achieved by simple interpolation. Degradation of this signal was synthetically simulated by setting samples to zero at random intervals. Objective evaluation was achieved using an overall (average) signal-to-noise ratio (S/N) measure:

$$S/N = 10 \log_{10} \left(\frac{\sum_k x^2(t_k)}{\sum_k (x(t_k) - \hat{x}(t_k))^2} \right) \quad (12)$$

where $x(t_k)$ and $\hat{x}(t_k)$ are the samples of the original undegraded speech and of the estimate recovered from the degraded speech, respectively. In each case the summation is taken over the entire data length. In all cases the measure was computed for the degraded (missing samples) speech, for the results of a simple low-pass filtering operation, and for the proposed method. All low-pass filtering operations were implemented using 31 point linear phase FIR filters designed using a Remez exchange algorithm, and a corresponding 15 point delay was introduced into the signal $x(t)$ prior to S/N computation to ensure alignment.

The off-line results obtained by applying the proposed method using the speech data sampled at 16 kHz are displayed in Fig. 3. The figure shows the S/N results for this case plotted against the percentage of missing samples. It is clear from the plot that the proposed method produces very significant improvements (in excess of 5 dB). A typical result with 15% of samples removed is illustrated in Fig. 4 which shows, in order, the original speech (Fig. 4(a)), the degraded signal (Fig. 4(b)), the signal recovered by simple low-pass filtering (Fig. 4(c)), and the recovery using the proposed method (Fig. 4(d)). Each plot corresponds to a time-frequency decomposition of a short segment (0.16 s) of the utterance. The improvement produced by the system is clearly a function of the percentage of samples missing. A further series of trials was conducted using data sampled at 8 kHz. The results are shown in Fig. 5. It is apparent that for all corruption levels tested the proposed method produced some improvements over the degraded signal. Furthermore, the results were also improved relative to simple low-pass filtering; though in no case did the improvement exceed 2 to 3 dB.

A final series of off-line trials involved selective application of the proposed method (using the division only at identified zeros) and comparison with selective use of linear and cubic spline interpolation. The results of the experiments are shown in Fig. 6. We see that the proposed method outperforms selective linear interpolation for loss rates of up to 20% at this 16 kHz sample rate. At higher loss rates linear interpolation becomes superior. This is because at high rates the probability of loss of two or more consecutive samples becomes significant. This is detrimental to a technique employing division because the likelihood of division by small values is increased. The division method produces inferior results to cubic spline. However, as we indicated in the introduction, we have sought a method which is suitable for real-time implementation with a simple processor. While linear interpolation and the proposed method satisfy this constraint, the cubic spline does not. For the cubic spline method, all the data samples must be available before any interpolation takes place. Additionally, for the 40 000 samples employed, and using a 10% loss rate, the cubic spline requires almost four times the number of multiplications and additions compared to the proposed method. (Other methods such as Lagrange interpolation are even worse. For a similar example, Lagrange interpolation requires about 8000 times the computation of the division method.)

To facilitate testing in a realistic environment and to allow subjective evaluation, the system represented in Fig. 1 was implemented on a small-scale DSP microprocessor. The implementation was achieved using the Texas Instruments TMS320-C25 system which is a fixed-point 16-b processor which has 10 million operations/s. Trials were conducted using sample rates of 8 and 16 kHz. The removal of randomly chosen samples was achieved using a second input source derived from a Gaussian noise gen-

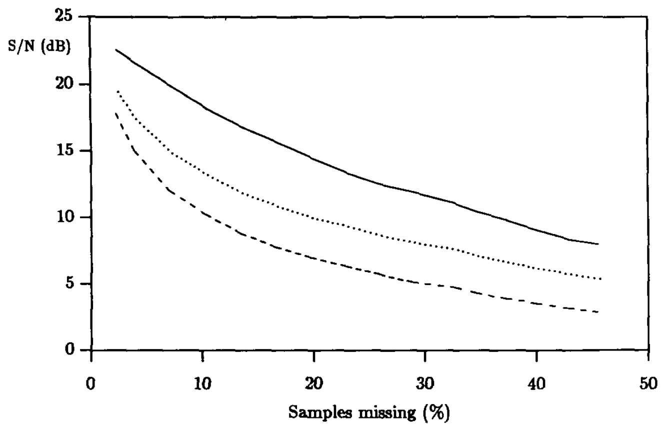
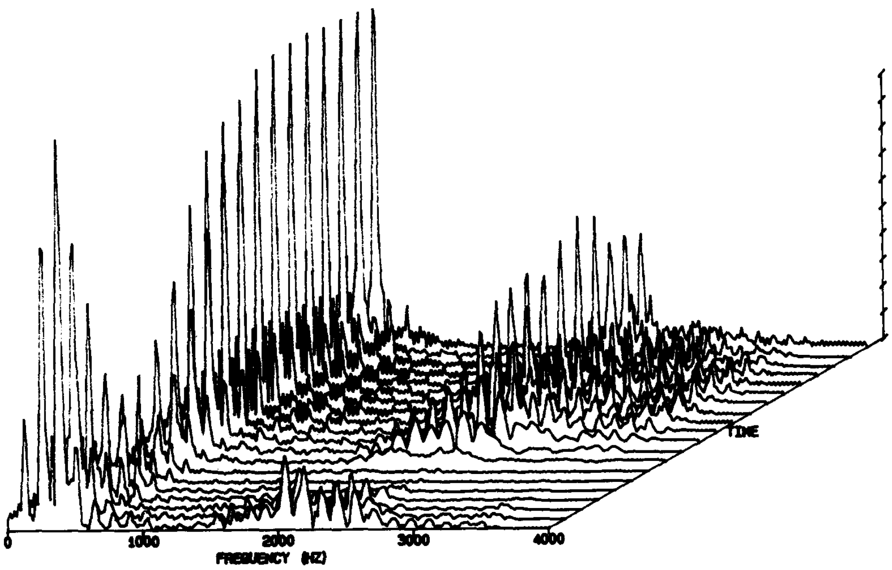


Fig. 3. S/N results versus percentage loss rate (f_s = 16 kHz). The dashed line represents raw (degraded) speech; the dotted line is the result for low-pass filtering, and the solid line shows the result obtained by processing using the proposed system.



(a)

Fig. 4. Time-frequency decomposition of a segment of 1280 samples from the utterance "The navy attacked the big task force." (a) Original speech. (*Continued on next page.*)

erator. After antialias filtering and A/D conversion, the samples of this noise were compared to a preselected threshold. Samples below this threshold were set to zero along with the corresponding samples in the speech. This secondary input also provided the source for the required impulse train. This was obtained in a straightforward manner by applying a hardlimiter and taking the absolute value. The low-pass filtering was achieved using a 64 point FIR filter, generated off-line using a Remez exchange algorithm. The stopband attenuation was approximately 60 dB. The cutoff was nominally 4 kHz. The final output was then obtained by division of the two results. This division poses difficulties due to the fixed-point processing involved. These effects can be controlled, however, by the expedient of adding a small positive constant to the denominator and by careful scaling before and after division.

Informal listening by the authors and colleagues at the Illinois Institute of Technology produced conclusions consistent with the objective evaluations discussed above: The system produced definite improvements in signal quality at all levels and all the sampling rates tested. The improvements were greatest at higher sample rates with near perfect recovery for 16 kHz sample rate and 10% loss. Although definitive statements about intelligibility changes require systematic listening trials, there is no question in the authors' minds that these quality gains were matched by significant improvements in intelligibil-

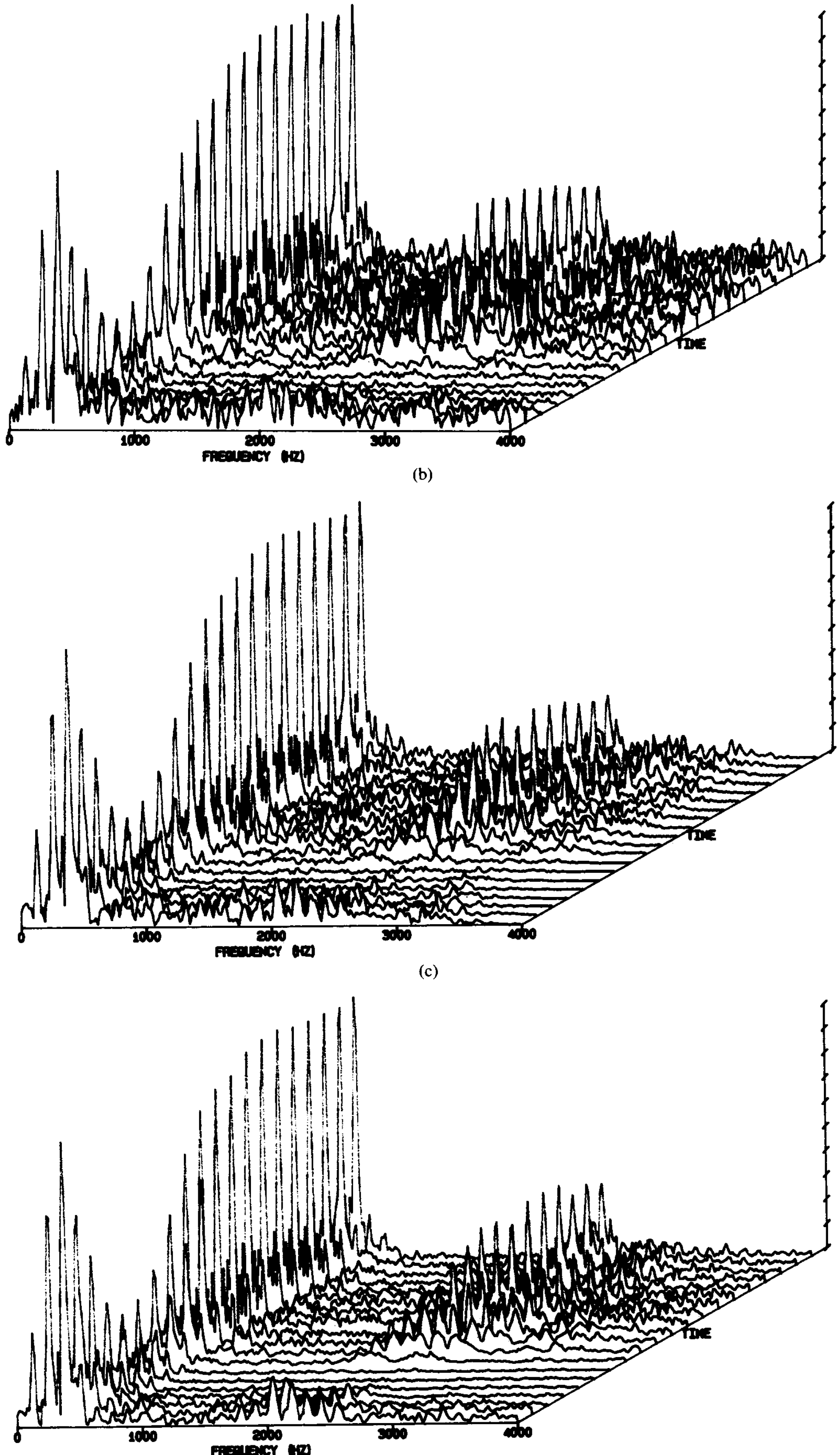


Fig. 4. (*Continued.*) (b) Degraded speech (approximately 15% of samples removed), (c) speech reconstruction using low-pass filtering, (d) speech reconstruction using the proposed method.

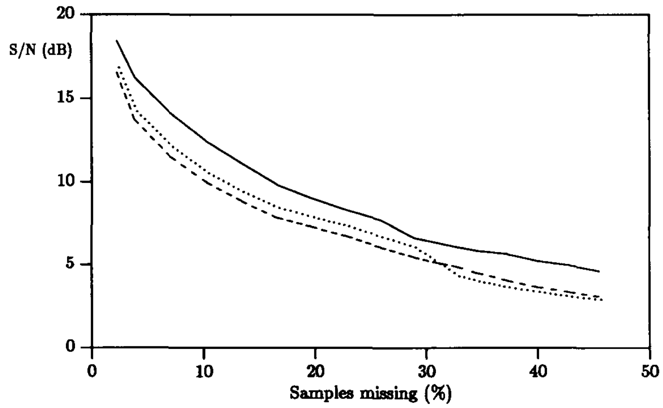


Fig. 5. S/N results versus percentage loss rate (f_s = 8 kHz). (For comparison see Fig. 3.)

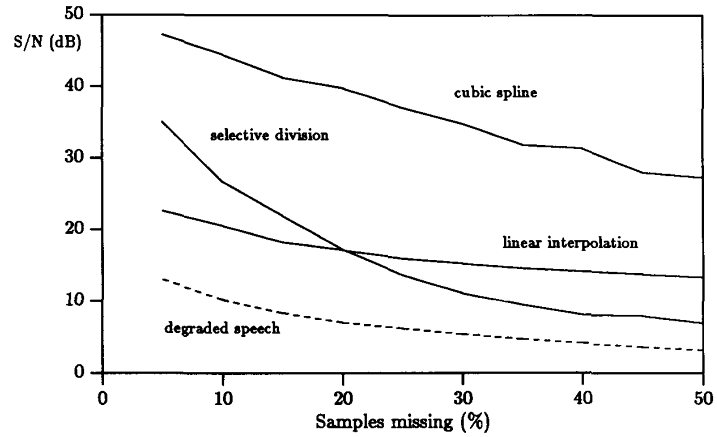


Fig. 6. S/N results for several interpolation methods (f_s = 16 kHz).

ity. Subjective assessments were also made when the speech was corrupted by random noise. It was desired to ascertain whether in this noisy condition, the nonlinearity of the processing system might produce any unexpected, subjectively disturbing result. No such result was observed; the system degraded gracefully and remained superior to simple low-pass filtering at all corruption levels tested.

References

[1] M. Yuito and N. Matsuo, "A new sample-interpolation method for recovering missing speech samples in packet voice communications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1989, pp. 381-384.

[2] R. J. Marks, II, "Restoring lost samples from an oversampled band-limited signal," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 752-755, 1983.

[3] R. J. Marks, II, and D. Radbel, "Error in linear estimation of lost samples in an oversampled band-limited signal," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 648-654, 1984.

[4] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete-time signals that can be modelled as autoregressive process," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 2, pp. 317-330, 1986.

[5] F. A. Marvasti, "Spectrum of nonuniform samples," *Electron. Lett.*, vol. 20, pp. 896-897, 1984.

[6] E. Plotkin, L. Rogtman, and M. Swamy, "Reconstruction of nonuniformly sampled band-limited signals and jitter error reduction," *Signal Processing*, vol. 7, pp. 151-160, 1984.

[7] F. A. Marvasti, *A Unified Approach to Zero-Crossings and Nonuniform Sampling of Single and Multidimensional Signals and Systems.* Oak Park, IL: Nonuniform, 1987.

[8] R. G. Wiley, H. Schwarzlander, and D. D. Weiner, "Demodulation procedure for very wide-band FM," *IEEE Trans. Commun.*, vol. COM-25, pp. 318-327, 1977.

[9] S. Yeh and H. Stark, "Iterative and one-step reconstruction from nonuniform samples by convex projections," *J. Opt. Soc. Amer. A*, vol. 7, pp. 491-499, 1990.

[10] F. A. Marvasti *et al.*, "Recovery of signals from nonuniform samples using iterative method," *IEEE Trans. Signal Processing*, vol. 39, pp. 872-877, 1991.

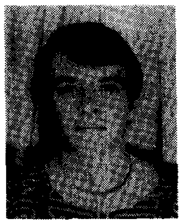
[11] H. E. Rowe, *Signals and Noises in Communication Systems.* Princeton, NJ: Van Nostrand, 1965.

[12] M. Soumekh, "Band-limited interpolation from unevenly spaced sampled data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 110-122, 1988.

[13] F. A. Marvasti and T. J. Lee, "Analysis and recovery of sample-and-hold and linearly interpolated signals with irregular samples," *IEEE Trans. Signal Processing*, vol. 40, pp. 1884-1891, 1992.

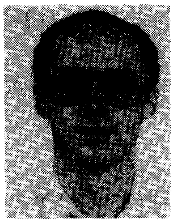
Farokh A. Marvasti (S'72–M'74–SM'83) received the B.S., M.S., and Ph.D. degrees, all in electrical engineering, from Rensselaer Polytechnic Institute in 1973.

Since 1972 he has worked for Graphic Sciences, Singer-Kearfott, and AT&T Bell Laboratories. He was a Professor at Sharif University of Technology, Tehran, Iran, from 1976 to 1985 where he did extensive consulting to telecommunication and power companies. He was a visiting Professor at the University of California, Davis, from 1984 to 1985. He was an Associate Professor at the Illinois Institute of Technology from 1987 to 1991. From 1991 to 1992, he was a consultant on video compression for satellite applications. He is now the head of the Communications Signal Processing Laboratory of the Communications Research Group, Department of Electrical Engineering, King's College, University of London. He is the author of 60 journal and conference papers. He is also the author of the monograph *A Uniform Approach to Zero-Crossings and Nonuniform Sampling of Single and Multidimensional Signals and Systems* (Nonuniform Publications, Oak Park, IL). He has also contributed two chapters for the book, *Advanced Topics in Shannon Sampling and Interpolations Theory*, edited by R. Marks, II (Springer-Verlag). He has also contributed a chapter for the *Telecommunication Encyclopedia*, edited by F. Froelich (Marcel Dekker). He is also the Editor on Data Communications for the IEEE TRANSACTIONS ON COMMUNICATIONS.



Peter M. Clarkson (M'86–SM'91) was born in Altrincham, England, in 1958. He obtained the B.Sc., with first class honors, from Portsmouth Polytechnic, England, in 1980, and the Ph.D. degree from the University of Southampton, England, in 1984.

From 1980 to 1987 he was associated with the Institute of Sound and Vibration Research at the University of Southampton; from 1980 to 1984 as a Research Assistant, from 1984 to 1985 as a postdoctoral fellow of the U.K. Science and Engineering Research Council, and from 1985 to 1987 as a Research Lecturer funded by the U.K. Admiralty Research Establishment. Since 1987 he has been with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, where he currently is an Associate Professor. His research interests are in digital signal processing, particularly adaptive signal processing and its application to speech and other acoustic signals.



Miroslav V. Dokic was born in Bosanski Samac, Bosnia, on July 1, 1962. He received the B.S. (with highest honors) and M.S. degrees in electrical engineering from the University of Sarajevo, Bosnia, in 1986 and 1989, respectively. He is currently working towards the Ph.D. degree at the Illinois Institute of Technology (IIT), Chicago.

From 1986 to 1988 he worked as a Research Assistant at Unis Institute, Sarajevo. During 1988–1989 he was an IREX Research Fellow in the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago. Since 1989 he has been working both as a Teaching and Research Assistant at IIT. His current research areas include nonlinear adaptive filters and real-time speech processing.

Mr. Dokic is also a 1991–1992 IEEE Comsoc scholarship recipient.

Ut Goenchanart was born in Bangkok, Thailand, in 1955. He received the B.S. degree in food science and technology from Chiangmai University, Thailand, in 1980, the B.S. degree in electrical engineering technology from Devry Institute of Technology, Chicago, in 1986, and the M.S. degree in electrical engineering from Illinois Institute of Technology in 1989. He is now working towards the Ph.D. degree at Illinois Institute of Technology. His areas of interest include digital signal processing, communication, and computer systems.

Chuande Liu was born in China in 1962. He received the B.S. degree in petroleum engineering from East China Petroleum Institute, Shandong, China, in 1983 and the M.S. degree in electrical engineering from Illinois Institute of Technology in 1988. He is currently working towards the Ph.D. degree.

He is now a Software Engineer with U.S. Robotics, Skokie, Chicago, IL. His interests include signal processing, coding, and computer software design.