DYNAMIC PROGRAMMING BASED OPTIMUM NON-UNIFORM SAMPLES FOR SPEECH RECONSTRUCTION AND CODING

Prasanta Kumar Ghosh ⋆ *and* *T.V. Sreenivas* †

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore-560 012, INDIA
E-mail: ⋆ prasanta@ece.iisc.ernet.in † tvsree@ece.iisc.ernet.in

ABSTRACT

Non-uniform sampling of a signal is formulated as an optimization problem which minimizes the reconstruction signal error. Dynamic programming (DP) has been used to solve this problem efficiently for a finite duration signal. Further, the optimum samples are quantized to realize a speech coder. The quantizer and the DP based optimum search for non-uniform samples (DP-NUS) can be combined in a closed-loop manner, which provides distinct advantage over the open-loop formulation. The DP-NUS formulation provides a useful control over the trade-off between bitrate and performance (reconstruction error). It is shown that 5-10 dB SNR improvement is possible using DP-NUS compared to extrema sampling approach. In addition, the close-loop DP-NUS gives a 4-5 dB improvement in reconstruction error.

1. INTRODUCTION

Signal reconstruction from nonuniform samples (NUS) is a widely studied problem. There have been many approaches to signal reconstruction from non-uniform samples, namely, from zero-crossings [1], from level crossings [2], signal reconstruction from periodically non-uniform samples [3], or through iterative methods [4]. However, very few attempts have been made [5] to analyze the quantization properties of NUS and use them for signal compression. [6, 7] put some light on the aspect on nonuniform sampling for coding of speech waveform. In [8] we have presented quantization properties of extrema samples (ES) for speech signal reconstruction and speech coder based on this is proposed. However, the resulting coder is a variable rate coder since the number of ES (being nonuniform in nature) varies with time. Also, the bitrate of the ES based coder is not easily scalable in the sense that ES have been chosen as a fixed NUS and no measure is incorporated to increase or decrease the number of NUS.

The aim of this paper is to find optimum NUS to reconstruct a finite duration signal which minimizes certain cost function. The optimum NUS are obtained efficiently by dynamic programming (DP). It is found that for a particular choice of interpolating function, the extrema are not necessarily optimum NUS to minimize the reconstruction signal error. The dynamic programming provides the advantage of choosing the number of NUS to achieve a specific performance. We also incorporate the quantization of NUS location and amplitude while minimizing the cost function using DP approach and we find that it improves reconstructed signal quality over that obtained using DP and quantization separately.

2. ANALYSIS BY SYNTHESIS APPROACH FOR OPTIMUM NONUNIFORM SAMPLES

Let $x[n]$, $0 \leq n \leq N-1$ be the signal segment for resampling and reconstruction. Let $\{\eta_i\}_{i=1}^{M}$ be the NUS locations of $x[n]$. The reconstructed signal $\hat{x}[n]$ is in general a function of $\{\eta_i\}_{i=1}^{M}$ and $\{x[\eta_i]\}_{i=1}^{M}$, i.e.,

$$\hat{x}[n] = \mathcal{F}(\{\eta_k\}, \{x[\eta_k]\}), k = i, i+1, ..., i+\Delta$$

i.e., a subset of $(\Delta+1)$ NUS are used to reconstruct $x[n]$. Specifically, for Δ=1, we can write:

$$\hat{x}[n] = x[\eta_i] + (x[\eta_{i+1}] - x[\eta_i]) F_j\left(\frac{n-\eta_i}{\eta_{i+1}-\eta_i}\right), \quad \eta_i \leq n < \eta_{i+1} \qquad (1)$$

where $F_j(X)$ is the local interpolation function. We consider three local interpolation functions as in [8]:

$$\begin{aligned} \text{Linear interpolation} &: F_1(X) = X. \\ \text{Polynomial interpolation} &: F_2(X) = X^2(3-2X). \\ \text{Sinusoidal interpolation} &: F_3(X) = \left(\sin\frac{\pi X}{2}\right)^2. \end{aligned}$$

The error in reconstruction is denoted by:

$$e[n] = x[n] - \hat{x}[n]. \qquad (2)$$

We seek to minimize the energy of $e[n]$ to obtain optimum NUS. Thus we can pose the optimization problem to determine the optimum NUS as follows:

$$\left\{\eta_i^{opt}\right\} = \arg\min_{\{\eta_i\}} \left(\frac{1}{N}\sum_{n=0}^{N-1}\{x[n]-\hat{x}[n]\}^2\right) \; 1 \leq i \leq M \qquad (3)$$

The above cost function to be minimized is in general not a quadratic function of $\{\eta_i\}$, rather it is dependent on F_j. A close form solution of (3) is not possible and hence, we resort to the analysis by synthesis (AbS) approach, which is efficiently solved using dynamic programming. Fig. 1 shows a block diagram of finding the optimum NUS using this approach. To check the performance of DP in selection of NUS,

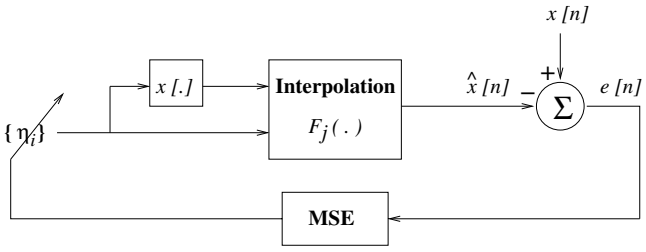

Fig. 1. *AbS approach to optimum NUS selection.*

we consider two synthetic signals - triangular and sinusoidal shown in Fig. 2 and obtain optimum DP-NUS for different choices of interpolation functions. It is clear that when the local signal property matches the interpolation function, the extrema are exactly identified and when there is no match, there is deviation in NUS from extrema.

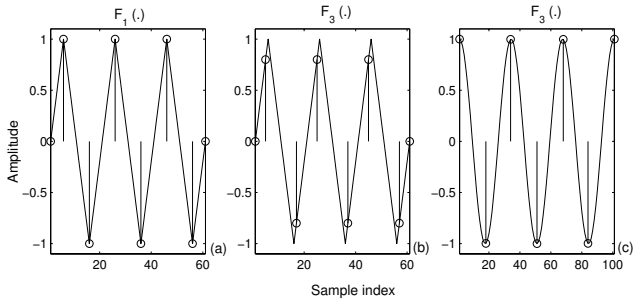

Fig. 2. *Optimum NUS for synthetic signals (a)-(b) optimum NUS for triangular wave using F_1 and F_3 (c) optimum NUS for sinusoidal signal using F_3.*

Though in [8] extrema have been used for good quality reconstruction of speech signals, we would like to explore other NUS for coding. The new NUS is based on signal reconstruction criterion, rather than the original signal property. This criterion can be controlled resulting in a coder of either fixed bitrate or a particular performance. The solution of (3) using DP is possible because of the local nature of the interpolation functions. This permits to expand (3) as a trellis of successive cost with increasing number of NUS. We can stop the trellis recursive search with a pre-fixed number of NUS and back-track to find the optimum NUS positions.

Fig. 3 illustrates the difference between optimum DP-NUS and extrema of a signal. The signal considered in Fig. 3 has 29 extrema, which are shown in Fig. 3 (d); Fig. 3 (c) shows optimum DP-NUS obtained using interpolation function F_2 where M=30. It can be observed that some NUS

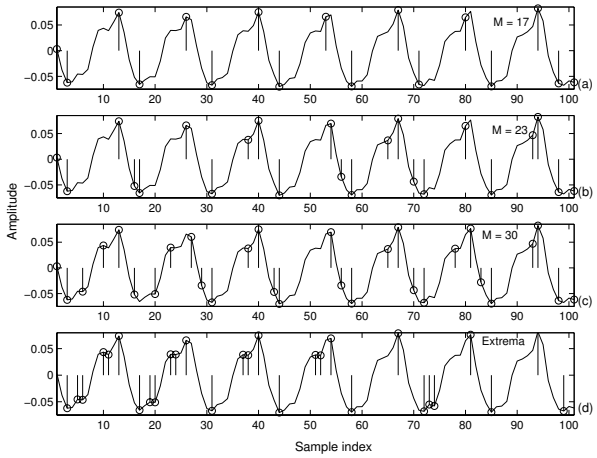

Fig. 3. *(a)-(c) Optimum DP-NUS of a signal segment using F_2 as interpolation function for M=17, 23 and 30 respectively, (d) 29 extrema of the same signal.*

are placed where there is no extrema in the signal. However, a closer look reveals that they are placed where the extrema of the first derivative of the signal (i.e. zeros of the second derivative) occur. Thus the turning points are selected by DP to minimize the cost function. Fig. 3 (a) and (b) plot the optimum DP-NUS for M=17 and 23. It can be seen that the optimum NUS for these cases are placed near the major turning points of the signal so that it preserves the basic structure of the waveform.

Next, we consider a speech segment and choose the frame length as 20 msec i.e. N=160. We obtain $M(m)$ optimum NUS for each frame where $M(m)$ is set to the number of extrema in m^{th} frame. Fig. 4 (a) shows the segmental SNR ($Seg\_SNR[m] = \frac{\sum_{n=0}^{N-1} x^2[n+mN]}{\sum_{n=0}^{N-1}\{x[n+mN]-\hat{x}[n+mN]\}^2}$) for both the extrema based reconstruction and the reconstruction from DP based NUS. Although the number of NUS is the same in both cases, it clear that DP based NUS is always better and provides an improvement of about 5-10 dB. Fig. 4 (b) and (c) show the average segmental SNR for various choice of M for male and female speakers respectively for different interpolation functions. The average segmental SNR shown in the figure is computed by averaging over utterances of five different speakers taken from TIMIT database. It is seen that the interpolation functions F_2 and F_3 show almost similar performance and both of them show a consistently higher SNR over that of F_1 for both the male and female speakers particularly at higher M. Thus we choose F_2 for all subsequent experiments of DP-NUS. Fig. 4 (d) compares the average segmental SNR for male and female speaker for various choices of M using F_2 as interpolation function. It is seen that for any choice of M the SNR of speech signal of a male speaker is

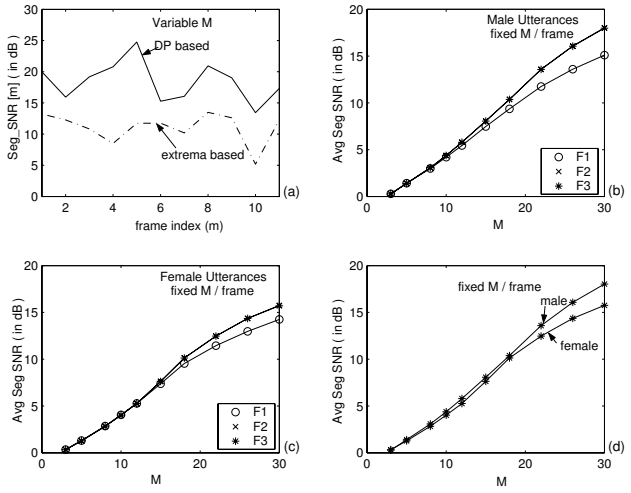

Fig. 4. *(a) $Seg\_SNR[m]$ for signal reconstruction using DP based NUS and extrema based signal reconstruction, (b)-(c) Average segmental SNR of male and female utterances for different interpolation functions, (d) for same number of NUS, speech of male speaker gives a better reconstruction than that of female speaker.*

higher than that of a female speaker. This can be attributed to the higher pitch frequency of the female speaker resulting in more number of periods in the signal frame and hence it requires more NUS to meet the same performance as that of the male utterances.

3. DP BASED ABS APPROACH USING QUANTIZED NUS

We are interested in quantizing the NUS location and amplitude for the purpose of coding. Therefore, we need to quantize both NUS location $\{\eta_i\}$ and NUS amplitude (NUSA) $\{x[\eta_i]\}$. Optimum quantizers Q_2 and Q_1 are designed for the NUS location interval (NLI) ($\delta\eta_i \stackrel{\triangle}{=} \eta_i - \eta_{i-1}$) and NUSA based on their probability density functions (pdf). The pdfs are generated from the NUS obtained from the 'open loop' formulation of section 2. These quantizers are incorporated in the AbS loop as shown in Fig. 5. The reconstructed signal amplitude $\hat{x}[n]$ is computed using $\{\eta_i^q\}$ and $\{x^q[\eta_i]\}$. We can view Fig. 5 as a 'closed-loop' formulation similar to DPCM wherein the quantization error is included into the NUS optimization which will lead to better performance in terms of overall reconstruction error.

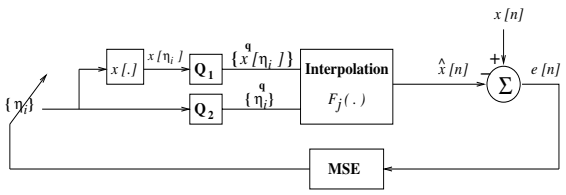

Fig. 5. *Block diagram of the 'closed-loop' formulation.*

Since the pdf of NLI and NUSA differ as M is changed, we consider three different quantizers for different values of M=8, 22, 30. The number of levels for each quantizer is decided based on the trade-off between the performance and the bitrate. This approach is similar to the design of optimum quantizer for the ES in [8]; 6 bits (i.e. 2^6=64 levels) are used to code NUSA while 5 bits (i.e. 32 levels) are used to code NLI information. To investigate the performance of closed-loop DP+quantizer, we compare the average segmental Signal to Quantization Noise Ratio (SQNR) (see Table 1) of the reconstructed signal, in both the open-loop and closed-loop schemes of Fig. 5 and Fig. 1 respectively, while using the same quantizers for both the schemes. For these experiments, we fix F_2 as the local interpolation function and average segmental SNR is computed by averaging over speech utterances of five different speakers (male and female speakers separately) taken from TIMIT database. From Table 1 we

Table 1. *Comparison of the performance of joint DP+quantization based NUS (closed-loop) and open-loop DP followed by quantization.*

Nonuniform Samples per 20 msec (M)	Average Seg SQNR (in dB)			
	Male Speaker		Female Speaker	
	Open loop	Closed loop	Open loop	Closed loop
8	0.83	3.05	0.79	2.81
22	8.15	13.05	7.28	12.22
30	14.10	17.96	12.16	15.34

see that for all three choices of M (bitrate increases with M), average segmental SQNR of the closed-loop configuration is significantly better than the open-loop configuration.

It may be noted that for a fixed choice of M (fixed bitrate) we achieve a certain performance in terms of average segmental SNR. However, it is possible to obtain identical SNR performance in each frame (i.e. constant $Seg\_SQNR[m]$) by varying M in each frame, to meet the required SQNR; DP based NUS formulation permits this, unlike many other compression schemes. In this case the number of NUS in each frame varies resulting in a variable bitrate coder.

From Fig. 6 we find that the optimum NUS for closed-loop scheme are different from that of open-loop and also the error in reconstructed signal is almost uncorrelated to the original signal.

4. NUS BASED SPEECH CODER

To implement a speech coder based on NUS we note that for frames belonging to unvoiced and silence regions in the speech signal, reconstruction using NUS is unnecessary and expensive in bitrate. Also, due to perceptual properties of speech signal, such frames need not be reconstructed exactly. Hence, we go for voiced-unvoiced-silence classification of

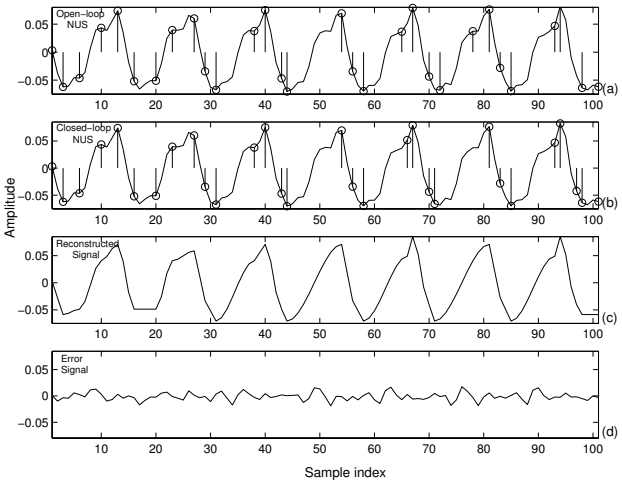

Fig. 6. *Performance of closed-loop DP based optimum NUS with quantization for the same example in Fig. 3 (a)-(b) optimum NUS for open-loop and close-loop scheme (c) reconstructed signal in closed-loop scheme (d) error signal.*

the signal using zero-crossing and short-time energy of the speech signal and use the NUS model only for voiced and transition segments. Using half-frame voicing decision we categorize each frame into one of the following: silence, unvoiced (UU), voiced (VV) and mixture of voiced and unvoiced (UV/VU); two bits are used to code the category information for each frame. When a frame is UU, the zero-crossing rate (ZCR) in that frame is computed; based on the observation that the fricatives with decreasing zero-crossing rate are /sh/, /s/, /f/, /ch/, /z/, the fricative information and its envelope is coded and such frames are synthesized based on the fricative-ID and the envelope. For a frame detected as silence, no parameter is coded. Frames belonging to VV and (UV/VU) categories are coded using nonuniform samples obtained from the solution of the AbS scheme of Fig. 5. This coding scheme results in a variable bitrate speech coder.

To make a comparison with the ES based speech coder of [8] we set $M(m)$ for the DP search as the number of extrema in each frame of the signal. We also use the quantizers of extrema location interval (ELI) and extrema amplitude (EA) designed in [8], for solving the optimum NUS. We choose only VV and (UV/VU) frames to compute the average segmental SQNR for the reconstructed signal using either coding schemes. Table 2 compares the two coder performances in terms of average segmental SQNR and also indicates the respective bitrate for two male and two female sentences taken from TIMIT database.

From Table 2 it is evident that the closed-loop DP-NUS speech coder shows a significant improvement over ES based coder, for utterances of both male and female speakers. Informal listening test of the generated speech samples also supports this performances.

Table 2. *Comparison of the performance of ES based coder and closed-loop DP-NUS based coder*

Sentences	Average Seg SQNR (in dB)		Bitrate
	ES based	DP based	(in kbps)
Sen 1 (male)	8.71	13.51	14.3
Sen 2 (male)	9.06	13.73	15.1
Sen 3 (female)	8.93	12.97	16.8
Sen 4 (female)	8.45	13.21	16.2

5. CONCLUSION

Our goal of finding optimum NUS for a given a speech signal, is not perfect reconstruction; rather, the emphasis is to obtain optimum samples to achieve best performance in quantization and coding of the signal. Accordingly, we have formulated DP based AbS approach in which the energy of the reconstruction error signal is chosen as the cost function to be minimized. We also show that a closed-loop DP which includes quantization into the optimization, results in significant performance advantage. One major advantage of using DP based AbS approach is that we can easily control the trade-off between bitrate and performance of the NUS based speech coder.

6. REFERENCES

[1] F. A. Marvasti, "A unified approach to zero-crossings and non-uniform sampling", *Nonuniform Press*, 1987.

[2] A. Zakhor, A. Oppenheim , "Sampling schemes for reconstruction of multidimensional signals from multiple level threshold crossing", *Int. Conf. Acoust., Speech, and Signal Proc.*, vol 2, pp 721-724, April 1988.

[3] P.L. Butzer, G. Hiusen, "Reconstruction of bounded signals from pseudo-periodic, irregularly spaced samples", *Signal Processing*, vol. 17, pp 1-17, 1989.

[4] F. Marvasti, M. Analoui, M. Gamshadzahi, "Recovery of signals from nonuniform samples using iterative methods", *IEEE Trans. Signal Proc.*, Vol. 39, issue 4, pp 872-878, Apr 1991.

[5] J.W. Mark, T.D. Todd, "A nonuniform sampling approach to data compression", IEEE Trans Comm., vol. COM-29, No. 1, pp 24-32, 1981.

[6] M. Bae, W. Lee and D. Kim, "On a new vocoder technique by the nonuniform sampling", Proc. Military Comm. Conf. MILCOM, pp. 649-652, Oct. 1996.

[7] B. Honary, W. He and M. Darnell, "Adaptive-rate sampling applied to the efficient encoding of speech waveforms", Second IEE National conf. on Telecomm., pp. 352-357, 2-5 Apr, 1989.

[8] Prasanta Kumar Ghosh and T.V. Sreenivas, "Waveform Reconstruction from Nonuniform Samples with Application to Speech Coding", IEEE-EURASIP workshop on Nonlinear Signal and Image Processing (NSIP), May 2005.