

Study of a Non-Linear DPCM Voice Compression Algorithm Optimised Using Neural Network Concepts

Radu Arsinte, Attila Ferencz *, Costin Miron ***

*Software ITC Cluj-Napoca-Str.Gh.Bilascu 109-3400 Cluj-Napoca-ROMANIA
Phone:+40-64-197681 Fax:+40-64-196787
E-mail:sitc@bavaria.utcluj.ro

**Technical University of Cluj-Napoca-Faculty of Electronics and
Telecommunications-Str.Gh.Barituu 28,3400 Cluj-Napoca-ROMANIA

Abstract

The paper presents the fundamentals, experimental study and the implementation of a voice compression algorithm, using DPCM and a non-linear predictor. The unusual aspect of the algorithm is the fact that the optimisation of coefficients is done using a training technique similar to those used in neural networks. Experiments revealed a better performance of the optimised predictor compared with both the linear predictor or fixed coefficients non-linear predictor.

1. Introduction

DPCM ([1][2]) and his variants is a general method of compression useful in various fields of the signal processing technology. Even today, when other tools offer a larger compression ratio, simplicity of the DPCM is a major advantage that results in a cheap implementation and in a relative high compression speed. First, let's examine the possible choices in implementing a PCM voice compression scheme. Waveform coders attempt to reproduce the input signal's waveform. They are generally designed to be signal independent so they can be used to code a wide variety of signals. They also exhibit a slight degradation in the presence of noise and transmission errors. However, to be effective they can only be used for medium bit rates. Waveform coding can be carried out in either the time or the frequency domains. We are focusing this short analysis on time

domain coding.

Pulse Code Modulation (PCM)

Pulse code modulation is the simplest type of waveform coding. It is essentially just a quantisation process. As each sample enters the coder it is quantized to one of a finite set of reconstruction levels, each level is assigned a unique sequence of binary digits and it is this sequence that is transmitted to the receiver. Any form of scalar quantisation can be used with this scheme, but the most common form of quantisation used is logarithmic quantisation. In fact the International Telegraph and Telephone Consultative Committee's (CCITT) G.711 standard defines 8 bit A-law and -law PCM as the standard method of coding telephone speech. This scheme is very popular because of its ability to endure several encoding and decoding stages, its low complexity and its low encoding delay. The low delay allows some analogue stages to be used in the network.

Differential Pulse Code Modulation (DPCM)

PCM makes no assumptions about the nature of the waveform to be coded, hence it works very well for non-speech signals. However, when coding speech there is a very high correlation between adjacent samples. This correlation could be used to reduce the resulting bit rate. One simple method of doing this is to transmit only the differences between each sample. This difference signal will have a much lower dynamic range than the original speech, so it can be effectively quantized using a quantizer with fewer reconstruction levels. In the above

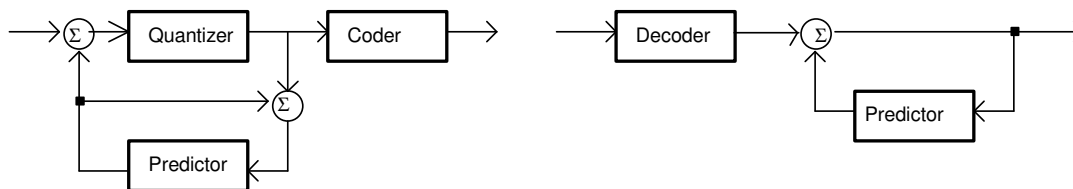


Fig.1. Block diagram of DPCM compression

method the previous sample is being used to predict the value of the present sample. Obviously the prediction would be improved if a much larger block of the speech is used to make the prediction.

Normally the predicted value $x'(n)$, is a linear combination of a finite number of past speech samples, $x(n)$ and

$$\epsilon(n) = x(n) - x'(n) \quad (1)$$

The difference signal, $\epsilon(n)$, is known as the residual and it is this residual that is quantized and transmitted to the receiver. The predictor's coefficients are chosen to minimise the mean squared prediction error.

Adaptive Differential Pulse Code Modulation (ADPCM)

With DPCM both the predictor and the quantizer remain fixed in time. Greater efficiency could be achieved if the quantizer adapted to the changing statistics of the prediction residual. Further gains could be made if the predictor, itself, could adapt to the speech signal. This would ensure that the mean squared prediction error was being continually minimised independently of the speaker and the speech signal. There are two methods for adapting quantizers and predictors, namely feedforward and feedback adaptation. With feedforward adaptation the

reconstruction levels and the prediction coefficients are calculated at the transmitter, using a block of speech. They are then, themselves, quantized and transmitted to the receiver as side information. Both the transmitter and the receiver use these quantized values to make the predictions and quantize the residual. For feedback adaptation the reconstruction levels and predictor coefficients are calculated using the coded signal. Since this signal is known to both the transmitter and the receiver there is no need to transmit any side information, so the predictor and quantizer can be updated for every sample.

Feedback adaptation can produce lower bit rates but it is more sensitive to transmission errors than feedforward adaptation techniques.

Adaptive differential pulse code modulation (ADPCM) is very useful for coding speech in medium bit rates. The CCITT have formalised a standard for coding telephone speech in 32 kb/s([1],[2]). This is the G.721 standard. It uses a feedback adaptation scheme for both the quantizer and the predictor. The predictor has two poles and six zeros so it will produce reasonable a quality output for a non-speech inputs.

Non-linear predictors are not often explored in technical literature and few practical applications exists. Neural network concepts, in principal the possibility of coefficients optimisation and error minimisation by 'training' the entire scheme, opens new windows for another approach, and consequently, with other results in

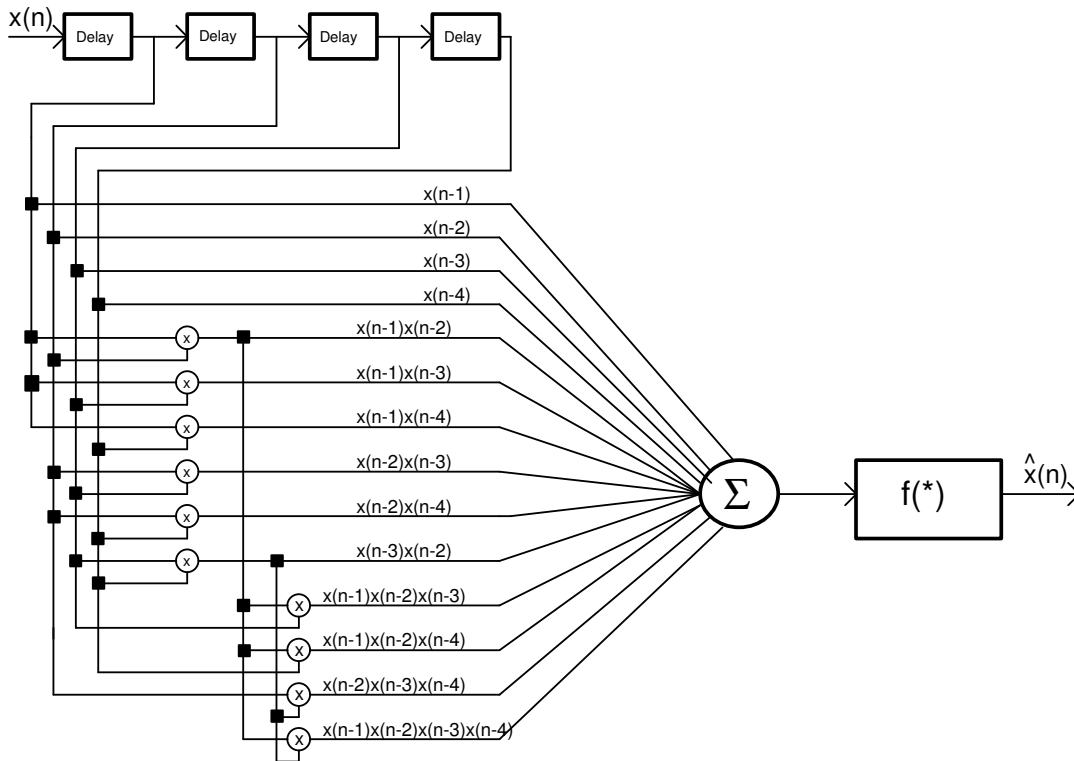


Fig.2. Non-linear DPCM predictor 4-th order

the family of DPCM algorithms coding

2. Description of algorithm

The basis of DPCM compression is , as discussed, quantisation and coding of the difference between the predicted sample and the real sample .The predicted sample is often a linear combination of the preceding samples. One approach is to use a statistical model of the data to derive a function which relates the value of the neighboring samples to that of the current one in an optimal manner. An autoregressive model (AR) is a model successfully applied. In most applications the prediction is limited at the few closest preceding samples of the predicted value. Shortly, the linear variant of DPCM algorithm can be described in the following formula:

$$x(n) = \sum_i w_i x(n-i) + \varepsilon_n \quad (2)$$

where {w_j} is the set of AR coefficients, and {ε} is a set of zero-mean independent and identically distributed (iid) random variables. In this case the predicted value is a linear sum of the neighboring samples as in the relationship:

$$\hat{x}(n) = \sum_i w_i x(n-i) \quad (3)$$

For the non-linear predictor the general formula is as following:

$$x(n) = \sum_i w_i x(n-i) + \sum_i \sum_j w_{ij} x(n-i)x(n-j) + \sum_i \sum_j \sum_k w_{ijk} x(n-i)x(n-j)x(n-k) + \dots + \varepsilon_n \quad (4)$$

In the experiments related to the compression scheme proposed ,the non-linear high order terms (3,4,..) are omitted:

$$x(n) = \sum_i w_i x(n-i) + \sum_i \sum_j w_{ij} x(n-i)x(n-j) + \varepsilon_n \quad (5)$$

where: w_i-coefficients of the predictor
x(n)-the n-th sample value
ε-residual error

We have experimented the 4-stage non-linear predictor, similar to the predictor presented in [3] and [5] (in the last reference he was used for image compression), and this predictor is presented in the block scheme from fig.2. Choosing the coefficients is often the main problem of the algorithm, and can be done for error minimisation and/or to obtain a higher compression ratio. Coefficients were optimised by "training" using real voice data. The "training" is done on relatively large blocks (1Ksamples and more).

An other problem in the algorithm evaluation is to establish measures of fidelity.

The most common indicator is the Signal to Noise Ratio (SNR) defined as:

$$SNR = 20 \lg \frac{E_{max}}{\sigma} \quad (6)$$

where: E_{max}- maximum value of the signal
σ – average error

In the experiments we also use this parameter.

An other form to compute the SNR is the following formula:

$$SNR = 10 \lg \frac{\sigma_s^2}{\sigma_n^2} \quad (7)$$

with σ_s,σ_n,-effective value of the signal and noise, respectively

The average (square) error can be determined with the following formula:

$$\sigma^2 \cong \frac{\sum_{i=0}^{N-1} (x(i) - x'(i))^2}{N} \quad (8)$$

with:

N-dimension of the reference block

x(i)-original sample

x'(i)-reconstructed sample

3. Experimental system and results

The algorithm described was experimented using the MATLAB environment and then implemented in a DSP based system ,similar to the system described in [4],with some improvements. The central element of DSP system is the accelerator board (coprocessor board) called DSPxx25, around him being developed several acquisition modules specialised for various domains: audio frequency, high frequency and video. The DSPxx25 board can be equipped with one of the following fixed point DSP: TMS320C25,TMS320C50, TMS320C2xx. Performances of the DSP board used are the following:

-DSP type	TMS320C25,TMS320C50 or TMS320C2xx
-max.clock frequency	40MHz
-Wait states	maximum 2
-memory	64KWords program/data
-extensions	local DSP bus DSP serial
-Acquisition speed	25KSamples/sec
-Resolution	12 bit

We have used two waveforms for the experiments (fig.3).First waveform is the vowel A , the second a random noise ("shuffle").

Compared with the linear predictor , for a compression factor of 4 , the gain in SNR for the non-linear predictor is 2,3 dB in case of first waveform and 3,1dB for the second waveform.

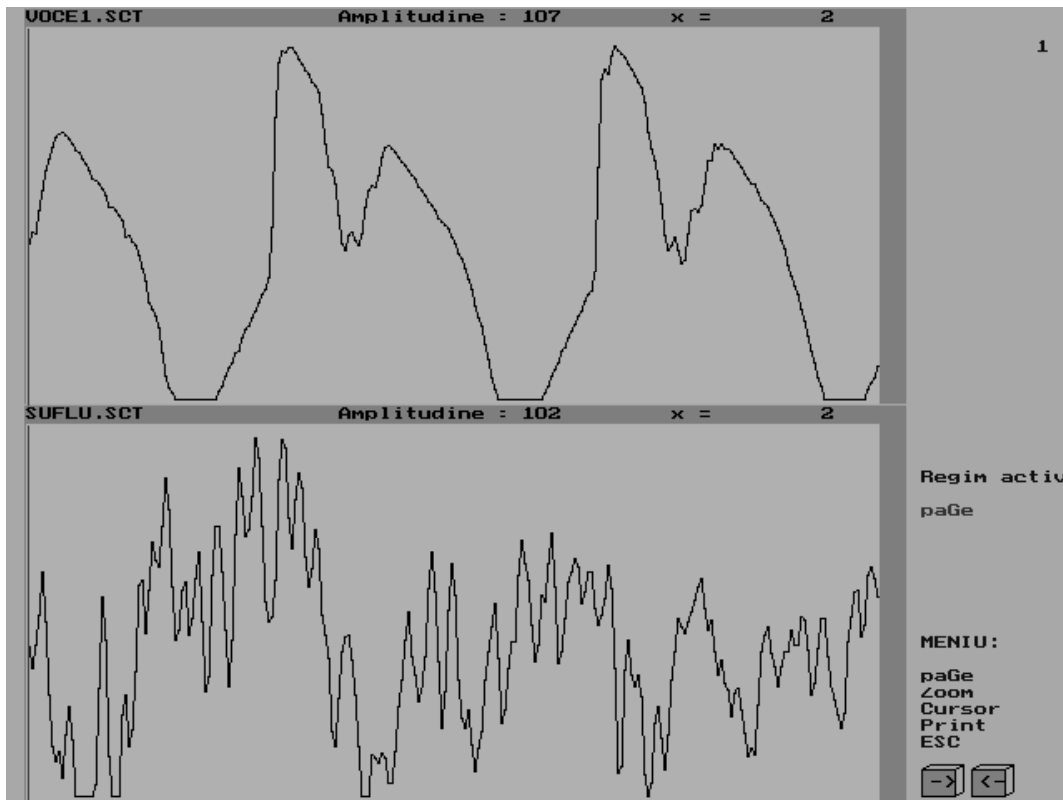


Fig.3.Waveforms used for experiments

The compression factor can be further increased (to 6-8) ,still preserving a good quality of reproduced sound.

4.Conclusion

The use of non-linear predictors (with 3-4 stages),as the experiments revealed, open the possibility to achieve compression rates of at least 6-8. The use of non-linear predictors of higher order is less efficient ,real time implementation and "training" being difficult. The main advantage is the simplicity and easiness in implementation.

In fact the system presented is a variant of an adaptive system with the optimisation improved using the new studies and concepts on neural networks.

Of course, we didn't made all the studies possible on this type of predictor, especially the influence of different training techniques on the efficiency of coding (compression) and improvement in SNR. This is a large field of experiments in the future.

5.References

[1] J.Raymer,M.MacMahan,M.Arjmand-32-kbit/s
ADPCM with the TMS32010 -In Digital Signal

Processing Applications with the TMS320 Family -
Texas Instruments-Edition 1994

- [2] M.Randolph-*The Design of an Adaptive Predictive Coder Using a Single-Chip Digital Signal Processor* -In *Digital Signal Processing Applications with the TMS320 Family*-Texas Instruments-Edition 1994
- [3] Robert Dony,Simon Haykin-*Neural Network Approaches to Image Compression* -Proceedings of the IEEE Vol.83. No.2.Feb.1995
- [4] Radu Arsinte,Attila Ferencz,Costin Miron-*DSP Based System for Real-Time Voice Synthesis Applications Development*-Proceedings of SPECOM'96-St.Petersburg-Russia -Nov.1996
- [5] Radu Arsinte-*Studies and Experiments on Data Acquisition with Applications in Visual Intelligent Sensors Implementation* -PhD Thesis-Technical University of Cluj-Napoca-April 1997