APROPOSED APPROACH TO PROTECTION OF SYNTHETIC SPEECH DATA TRANSMITTED OVER NOISY CHANNELS

Rateb, M.* Abulshohoud, W.** Elmoghazy, A.*
* M.T.C. ** Armament Authority, Cairo, Egypt.

RESUME

Le côdage de parole base sur le principe de "côdage lineaire predective" (CLP), permit la transmission simitanné de plasieur cannaux.
Les errours du canal, quant'il s'trouent dans une partie particulaire du Signal Code en CLP, sont un source tres efficase de la distortion.
On expose ici une methode de la correction des erreours pour le signal code en CLP. La methode prend en considerationl' effet pondre des errours sur la qualite du signal, et il, est nomée "Supervision multiple de codage". La methode proposée est realisée par une structure de codage en reseaux.Elle permet la protection contre les erreurs en permettant, au meme temps, un rytme de transmission relativement élevé.

SUMMARY

LPC speech signals can achieve greater number of multiplexed channels than PCM and DM at nearly the same speech quality when TDM is considered. Channel errors falling in specific segments of the transmitted signal yielding to harmful distortion made LPC less attractive for practical communication systems application.
This paper introduces a proposed weightened FEC technique in which the LPC parameter data are protected against channel errors according to the effect of those errors on the received speech signal. This approach referred to as "Multiple-Code Supervision" yielded to a "Lattice Coding Structure"which provides the required protection at relatively high transmission rate.

1- INTRODUCTION

The quality of speech when using waveform coders significantly deteriorates if bit rate falls below 16 kb/s. On the other hand vocoders achieve maximum quality at a bit rate of approximately 4.8 kb/s and a negligible quality improvement is attained by further bit-rate increase. Reduction of bit rate to the order of 2.4 kb/s and even smaller will cause a nonremarkable degradation of speech quality. The basic concept behind the linear prediction is that speech could be modelled as the outcome of a linear time-varying system excited by voiced or unvoiced speech. The voiced speech is in the quasi-periodic pulses while the unvoiced is in the noise-like form. The system could be uniquely identified as an all-pole linear system. The formulation of linear prediction analysis yield to nearly equivalent methods like covariance, autocorrelation, lattice, inverse filter, maximum likelihood, spectral estimation,...,etc. Attention will be devoted to the autocorrelation method which is the most suitable in case of short-time averages processing.
According to the simplified model in Fig.1, the speech samples s(n) are related to the excitation u(n) by the difference equation:

$$s(n)=\sum_{k=1}^{p} a_k\ s(n-k)+Gu(n) \qquad (1)$$

with p poles, u(n) is the proper input excitation, a_k is the set of predictor coefficients and G is the filter gain.

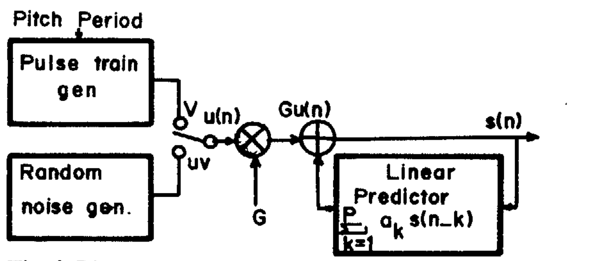

Fig:1 Block diagram For Speech Production.

On applying the autocorrelation the waveform segment under consideration is windowed (usually by Hamming window) such that:

$$\grave{s}_n=\begin{cases} s_n w_n & o\leq n\leq N-1 \\ o & \text{otherwise}\end{cases} \qquad (2)$$

The autocorrelation function R(j) is defined by

$$R(j)=\sum_{n=o}^{N-i-1} \grave{s}_n \grave{s}_{n+j} \qquad (3)$$

Then the LPC coefficients are achieved from the following normal equations:

$$\sum_{k=1}^{p} a_k R(j-k)= -R(j) \qquad ,\ p\geq j\geq 1 \qquad (4)$$

The solution of normal equations (4) could be achieved either by director iterative techniques. The Durbin's method is preferable since it is twice times faster as Levinson's and requires only 2p storage locations and p^2 operations. The Durbin's recursive procedure takes place as follows:

$$E_o= R(o) \qquad (5\text{-a})$$

$$k_j= -\ \left[R(j)+\sum_{i=1}^{j-1} a_i^{(j-1)} R(j-i)\right] /A_{j-1} \qquad (5\text{-b})$$

$$a_j^{(j)}= k_j \qquad (5\text{-c})$$

$$a_i^{(j)}= a_i^{(j-1)}+ k_j\ a_{j-i}^{(j-1)}\ ,\ j-1\geq i\geq 1 \qquad (5\text{-d})$$

$$E_j= (1-k_j^2)\ E_{j-1} \qquad (5\text{-e})$$

Equations (5a-5e) are recursively solved for j=1,2, ...,p. The final solution is:

$$a_i= a_i^{(p)} \qquad ,\ p\geq i\geq 1 \qquad (6)$$

There exist many possible sets of parameters that can uniquely characterize the all-pole filter H(z) and its inverse A(z) where:

$$H(z)= G/\left[1+\sum_{k=1}^{p} a_k\ z^{-k}\right] \text{ and } A(z)= G/H(z) \qquad (7)$$

Those sets are the poles of H(z) or zeros of A(z), impulse response of either H(z) or A(z), autocorrelation coefficients, spectral coefficients of A(z) and reflection coefficients k_j, $p\geq j\geq 1$. The last ones (thereflection or PARCOR coefficients) are the most

suitable for quantization since they secure the simultaneous filter stability upon quantization and the natural ordering of those parameters.
For filter stability it must be fulfilled:

$$|k_j| < 1, \quad p \geqslant j \geqslant 1 \qquad (8)$$

which is obviously satisfied by (5-e)
Pitch detection techniques utilizes the time-domain, frequency-domain or time and frequency-domain properties of the speech signals. Various quantization approaches for quantizing the LPC reflection coefficients exist. The most attractive one is the log area ratio (LAR) quantization method. The LAR coding scheme transforms k_j into a new coefficient g_j given by:

$$g_j = \text{Log}\left[(1+k_j)/(1-k_j)\right], \quad p \geqslant j \geqslant 1 \qquad (9)$$

Consequently, the coding scheme uses efficiently a relatively small number of bits for each coefficient. However from point of view of hardware realization the piecewise linear quantization is much simpler.

2- BASIC CONCEPTS

2-1 Considerations Affecting the Selection of Frame Structure and LPC Bit Rate. It has been found that the pitch and gain parameters should be geometrically interpolated [2] . This geometrical interpolation takes place linearly on a Log scale.
The PARCOR coefficients are bounded ($|k_j| < 1$) and they could be polated directly while the filter stability is maintained. Similarly the resulting Log area ratio (LAR) coefficients can be interpolated yielding to a stable system if the original reflection coefficients preserve this stability. Due to the fact that the PARCOR coefficients have highly skewed distribution, LAR coefficients given by (9) are the most suitable for transmission and 5-6 bits per log area ratio are sufficient to achieve the same quality synthetic speech as obtained from uncoded parameters. The choice of the number of LAR coefficients is tightly connected with the variation of prediction error with the predictor order "p" specially for voiced speech. For the autocorrelation method it was found by [3] that the normalized mean-squared error "V_n" remains at a value 0.1 for $p \geqslant 7$ for short windows (N=60) which is our typical case.

$$\text{where} \quad V_n = \sum_{m=0}^{N+p-1} e_n^2(m) \Big/ \sum_{m=0}^{N-1} s_n^2(m) \qquad (10)$$

$e_n(m)$ is the output of the prediction error filter corresponding to the speech segment $s_n(m)$. Thus, selecting p=10 will hand over a satisfactory result. On the basis of all the previously-mentioned considerations the frame structure will consist of 7 LAR coefficients each of 5 bits and the last three sach of 4 bits, the pitch period is represented by 6 bits, the gain by 5 bits and v/uv by one bit. In addition a permanent synch. bit ateach frame begining yielding to a total sum of 60 bits per frame. When transmitting LPC speech data over multiplexed PCM systems we have two alternatives. The first is to decrease the LPC vocoder actual bit rate to 2 KHz then rise it to 4KHz to allow the addition of equal number of redundant bits. In the second alternative the actual bit rate is doubled to achieve better speech quality on the expense of decreasing the number of multiplexed channels by one half. Consequently, 512 channels for the first and 256 channels (including signalling and synchronization) for the second version could be multiplexed instead of 32 channels for the European 2.048 MHz PCM systems. For the Bell T-1 carrier 1.544 Mb/s system, 384 channel for the first and 192 channel for the second version instead of 24 channels in case of PCM.

2-2 Basic Idea Of Proposed Coding Approach.
In most cases the transmitted data inside a message are of different levels of imprtance. Channel errors falling in specific critical segments of the message may lead to the loss of the whole message. This is a typical case when LPC data transmission is considered. Channel errors falling into the first two PARCOR coefficients in the LAR form or in the pitch period will yield to the complete distortion of the data contained in the whole frame [4] .
The newly proposed weightened FEC based on "Muliple Code Supervision" of the critical data segments, hands over an effective solution to this problem on the expense of the least possible redundancy. In this technique the critical segment is sharing more than one error correcting and detecting code encoding other data of lower levels of importance. The number of supervising codes is determined according to the immunity required and to the available redundancy. According to Fig.2.a, the pitch period (PP) with an added parity is encoded with the 8th (g_8),9th(g_9)and 10th(g_{10}) LAR coefficients with 5 parity checks for each to establish the codes (16,11) with mimimum distance d=4 c_1,c_2, and c_3 respectively. The (PP) segment with added parity is supervised by 3 codes.

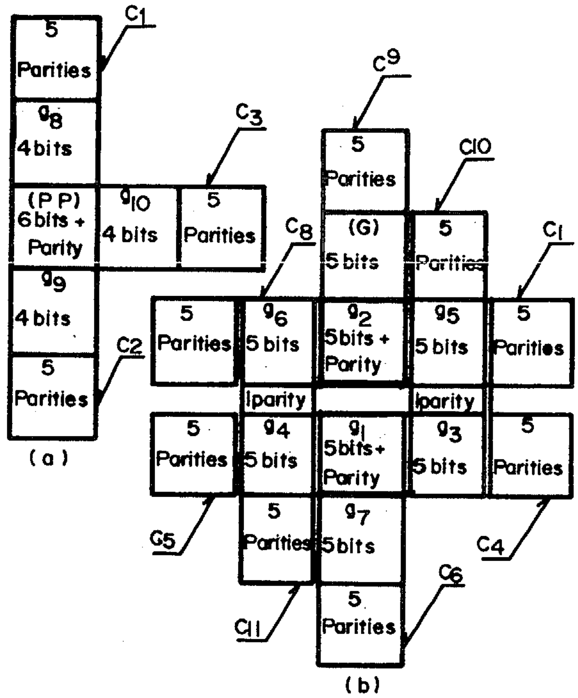

Fig:2 Proposed coding lattice For LPC

In Fig 2-b the first LAR coefficient (g_1) is supervised by the codes c_4,c_5 and c_6 while (g_2) is supervised by c_7,c_8 and c_9. Each of them has an added parity and all codes are (16,11) with d=4, (g_3) with added parity is supervised by c_4 and c_{10} while (g_4) with added parity is supervised by c_5 and c_{11}. (g_5) is supervised by c_7 and c_{10} while (g_6)is supervised by c_8 and c_{11} without added parity. The "Lattice Coding Structure" introduced secures a coding strength proportional to the importance of those parameters.

3- DECODING CRITERIA AND PROBABILITIES

3-1 Decoding Criteria

The selected block code with d=4 is a single error-correcting code with an additional overall parity. This parity enables the detection of all even error patterns except those yielding to a permissible code word.
For an arbitrary code segment having an additional parity supervised by R code blocks the following criteria are valid:

a- Basic criteria for correct decoding

1- No error is detected in all supervising code blocks and local parity detects no error.
2- All code blocks detect even errors while the local parity detects no error.
3- When at least two code blocks locate the same error inside the segment and local parity detects an error, while the rest of code blocks locate different errors.

b- Criteria for error detection

1- Local parity detects an error while supervising codes point to different error locations inside or outside the segment.
2- All supervising codes detect even errors and local parity detects an error.

3-2 Associated Probabilities

Assume a BSC having channel probability of error P. Let the length of the supervised code segment by n_s and the length of the supervising code be n_j, $j=1,2,..,R$. On the basis of the upper mentioned critéria the probability of correct decoding will be:

$$P_c(n_s)=(1-P)^{n_s}+n_sP(1-P)^{n_s-1}\Bigg[1-\Bigg\{\prod_{j=1}^{R}\Big\{\sum_{i_j=1}^{n_j-n_s}\binom{n_j-n_s}{i_j}P^{i_j}(1-P)^{n_j-n_s-i_j}\Big\}\Bigg\}-\prod_{j=1}^{R}(n_s/n_j)\Bigg[\sum_{k=1}^{R}(1-P)^{n_k-n_s}\prod_{\substack{j=1\\ j\neq k}}^{R}\Big\{\sum_{i_j=1}^{n_j-n_s}\binom{n_j-n_s}{i_j}P^{i_j}(1-P)^{n_j-n_s-i_j}\Big\}^{j-1}\prod_{\substack{j=1\\ j\neq k}}^{R}\Big\{\sum_{i_j=1}^{n_o}\binom{n_j-n_s}{2i_j-1}P^{2i_j-1}(1-P)^{n_j-n_s-2i_j+1}\Big\}\Bigg]\Bigg] \quad (11)$$

The penalty payed for the application of those decoding criteria will yield to erroneous decision with a very small probability given by:

$$P_n(n_s)=\sum_{r=1}^{n_e}\binom{n_s}{2r+1}P^{2r+1}(1-P)^{n_s-2r-1}\Bigg[\prod_{j=1}^{R}(n_s/n_j)\Bigg[\sum_{k=1}^{R}\sum_{i_k=o}^{n_e}\binom{n_k-n_s}{2i_k}P^{2i_k}(1-P)^{n_k-n_s-2i_k}\prod_{\substack{j=1\\ j\neq k}}^{R}\Big\{\sum_{i_j=1}^{n_o}\binom{n_j-n_s}{2i_j-1}P^{2i_j-1}(1-P)^{n_j-n_s-2i_j+1}\Big\}\Bigg]\Bigg]+\sum_{r=1}^{n_o}\binom{n_s}{2r}P^{2r}(1-P)^{n_s-2r}\Bigg[\prod_{j=1}^{R}\Big\{\sum_{i_j=o}^{n_e}\binom{n_j-n_s}{2i_j}P^{2i_j}(1-P)^{n_j-n_s-2i_j}\Big\}\Bigg] \quad (12)$$

and the probability of error detection is simply given by:

$$P_d(n_s)=1-\left[P_c(n_s)+P_d(n_s)\right] \quad (13)$$

where

$$n_o=\begin{cases}(n_j-n_s+1)/2 & \text{if } (n_j-n_s) \text{ is odd}\\ (n_j-n_s)/2 & \text{if } (n_j-n_s) \text{ is even}\end{cases} \quad (14)$$

and

$$n_e=\begin{cases}(n_j-n_s-1)/2 & \text{if } (n_j-n_s) \text{ is odd}\\ (n_j-n_s)/2 & \text{if } (n_j-n_s) \text{ is even}\end{cases} \quad (15)$$

If all supervising codes have equal length, ie, $n_1=n_2=n_j=n$ the above equations will reduce to:

$$P_c(n_s)=(1-P)^{n_s}+n_sP(1-P)^{n_s-1}\Bigg[1-\Big\{\sum_{i=1}^{n-n_s}\binom{n-n_s}{i}P^i(1-P)^{n-n_s-i}\Big\}^R-(n_s/n)^R\Bigg[R(1-P)^{n-n_s}\Big\{\binom{n-n_s}{i}P^i(1-P)^{n-n_s-i}\Big\}^{R-1}-\Big\{\sum_{j=1}^{n_o}\binom{n-n_s}{2j-1}P^{2j-1}(1-P)^{n-n_s-2j+1}\Big\}^{R-1}\Bigg]\Bigg] \quad (16)$$

and

$$P_n(n_s)=\sum_{r=1}^{n_e}\binom{n_s}{2r+1}P^{2r+1}(1-P)^{n-n_s-2r-1}\Bigg[(n_s/n)^R R\cdot\Big[\Big\{\sum_{i=o}^{n_e}\binom{n-n_s}{i}P^{2i}(1-P)^{n-n_s-2i}\Big\}\Big\{\sum_{j=1}^{n_o}\binom{n-n_s}{2j-1}P^{2j-1}(1-P)^{n-n_s-2j+1}\Big\}\Big]\Bigg]+\sum_{r=1}^{n_o}\binom{n_s}{2r}P^{2r}(1-P)^{n_s-2r}\cdot\left[\sum_{i=o}^{n_e}\binom{n-n_s}{2i}P^{2i}(1-P)^{n-n_s-2i}\right]^r \quad (17)$$

As the number of supervising code blocks increases expressions (11) and (16) will reduce to:

$$P_o(n_s)=(1-P)^{n_s}+n_sP(1-P)^{n_s-1}=\sum_{i=o}^{1}\binom{n_s}{i}P^i(1-P)^{n_s-i} \quad (18)$$

This means that the supervised short data segment will have the capability of an independent error correcting code of the segment length without additional redundant bits. This will yield to a substential high drop in the probability of error.

3-3 The Complete Decoding Algorithm

The complete decoding algorithm is interpreted by the simple flowchart Fig.3.

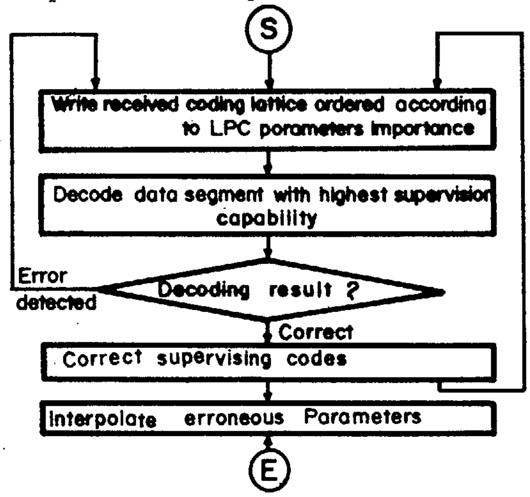

Fig.3 Simplified Flowchart of the decoding algorithm.

The received lattice coding structure is stored in a buffer and the decoding begins by the segment having the highest supervision capabilities. The decoding is accompliched on the basis of logic introduced in 3-1. If the decoding result will yield to a correct decision the supervising codes are afterwards corrected, otherwise the following data segment is processed. After all segments are being decoded, the erroneous parameters which have error beyond the correction capabilities are interpolated.

4- RESULTS ANALYSIS AND CONCLUSIONS

Fig.4 illustrates the probabilities of corret decision "P_c" and of error "P_n" for a code segment (6,5) supervised by three codes (16,11) which is in direct correspondance with the proposed coding structure. The results computed on the basis of (16) and (17) show a dramatic increase in P_c and decrease in P_n when compared with the capabilities handed over by the supervising code. This deviation increases P is increased. However there is an implicit improvement in capabilities of the supervising code handed over by the proposed coding lattice structure. This additional capabilities occure when one supervising code has two channel errors one falling inside and the other outside the supervised segment. Both errors in such cases are correctable. The overall probability of correct decoding for this supervising code will be:

$$P_c(n)=\sum_{i=o}^{1}\binom{n}{i}P^i(1-P)^{n-i}+n_s(n-n_s)P_c(n_s)\cdot P^2(1-P)^{n-2} \quad (19)$$

The interpolation of parameters having detected errors adds an implicit improvement which can not be analytically evaluated.

The efficiency of this approach increases as the supervising code length is increased with respect to the supervised data segment and as their number increases. This new approach of multiple code supervision yielding to the lattice coding structure is not only devoted to synthetic speech protection but also it is applicable to many types of messages met in practical communication. However protected synthetic or synthesized speech is an urgent future need since reliable speech communication is achived at a capacity many times more than that achieved now by PCM or DM.

5- REFERENCES

1- L.R. Rabiner and R.W. Schafer, "Digital Processing of Speech Signals", Prentice Hall, NewJersy, 1978.

2- B.S Atal and S.L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", Journal Acoust. Soc.Am., Vol.50, pp. 637-655, 1971.

3- S. Candra and W.C. Lin, "Experimental Comparison Between Stationary and Non-stationary Formulations of Linear Prediction Applied to Speech", IEEE Trans. Acousties, Speech and Signal Proc., Vol. ASSP-22, pp.403-415, 1974.

4- C.K. Un and S.J.Lee, "Effect of Channel Errors on the Performance of LPC Vocoder", IEEE Trans. Acoustics, Speech and Signal Proc., Vol. ASSP-31, No.1, pp.234-236, Feb.1983.

5- W.W. Peterson and E.J. Weldon, Jr., "Error-Correcting Codes (2nd Ed.)", MIT Press, Cambridge, Mass., 1972.

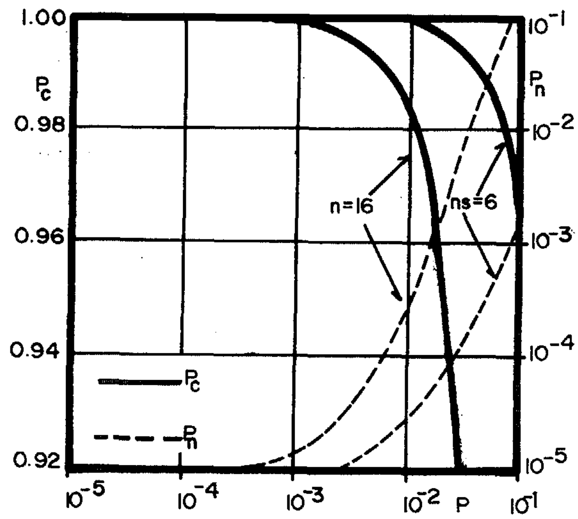

Fig:4 P_c , P_n =F(P) For n_s=6, n=16 and R=3