

Theory and Applications of Constrained Linear Predictive (LP) models

Igor N. Presnjakov, Leonid I. Nefedov, Stanislaw A. Krivenko, and Alexander P. Stativka

Abstract — The present paper relates generally to speech encoding and decoding in voice communication systems; and, more particularly, it relates to various techniques used with code-excited linear prediction coding to obtain high quality speech reproduction through a limited bit rate communication channel.

Index Terms — AMR, Speech coding, CELP

I. INTRODUCTION

SIGNAL modeling and parameter estimation play significant roles in communicating voice information with limited bandwidth constraints. To model basic speech sounds, speech signals are sampled as a discrete waveform to be digitally processed. In one type of signal coding technique called LPC (linear predictive coding), the signal value at any particular time index is modeled as a linear function of previous values. A subsequent signal is thus linearly predictable according to an earlier value. As a result, efficient signal representations can be determined by estimating and applying certain prediction parameters to represent the signal.

Applying LPC techniques, a conventional source encoder operates on speech signals to extract modelling and parameter information for communication to a conventional source decoder via a communication channel. Once received, the decoder attempts to reconstruct a counterpart signal for playback that sounds to a human ear like the original speech.

A certain amount of communication channel bandwidth is required to communicate the modelling and parameter information to the decoder. In embodiments, for example

Manuscript received February 29, 2008.

I. N. Presnjakov is with the Department of Communications, Kharkiv National University of Radio Electronics, Kharkiv, 61166 Ukraine (phone: 057-702-1429; fax: 057-702-1429; e-mail: tkvt@kture.kharkov.ua).

L. I. Nefedov is with Department of Automation, Kharkiv National Auto Road University, Kharkiv, and 61022 Ukraine (e-mail: nefedov@mail.ru).

S. A. Krivenko is with the Department of Communications, Kharkiv National University of Radio Electronics, Kharkiv, 61166 Ukraine (phone: 057-702-1429; cell phone: 067-723-7551; e-mail: stanislas@ukr.net).

A. P. Stativka was with the Department of Communications, Kharkiv National University of Radio Electronics, Kharkiv, 61166 Ukraine. He is now with the Department of system performer, Telesystems of Ukraine, Kyiv, 04080 (e-mail: stativka@people.net.ua).

where the channel bandwidth is shared and real-time reconstruction is necessary, a reduction in the required bandwidth proves beneficial. However, using conventional modeling techniques, the quality requirements in the reproduced speech limit the reduction of such bandwidth below certain levels.

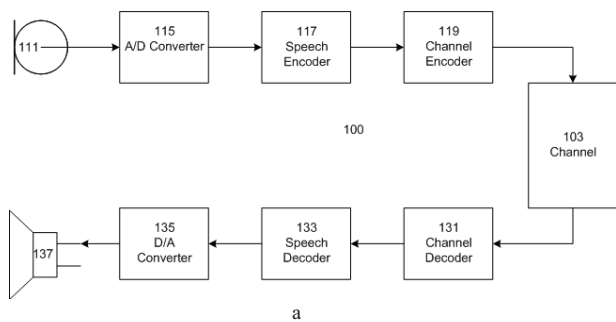
Speech encoding becomes increasingly difficult as transmission bit rates decrease. Particularly for noise encoding, perceptual quality diminishes significantly at lower bit rates. Straightforward code-excited linear prediction (CELP) is used in many speech codec, and it can be very effective method of encoding speech at relatively high transmission rates. However, even this method may fail to provide perceptually accurate signal reproduction at lower bit rates. One such reason is that the pulse like excitation for noise signals becomes more sparse at these lower bit rates as less bits are available for coding and transmission, thereby resulting in annoying distortion of the noise signal upon reproduction.

Many communication systems operate at bit rates that vary with any number of factors including total traffic on the communication system. For such variable rate communication systems, the inability to detect low bit rates and to handle the coding of noise at those lower bit rates in an effective manner often can result in perceptually inaccurate reproduction of the speech signal. This inaccurate reproduction could be avoided if a more effective method for encoding noise at those low bit rates were identified.

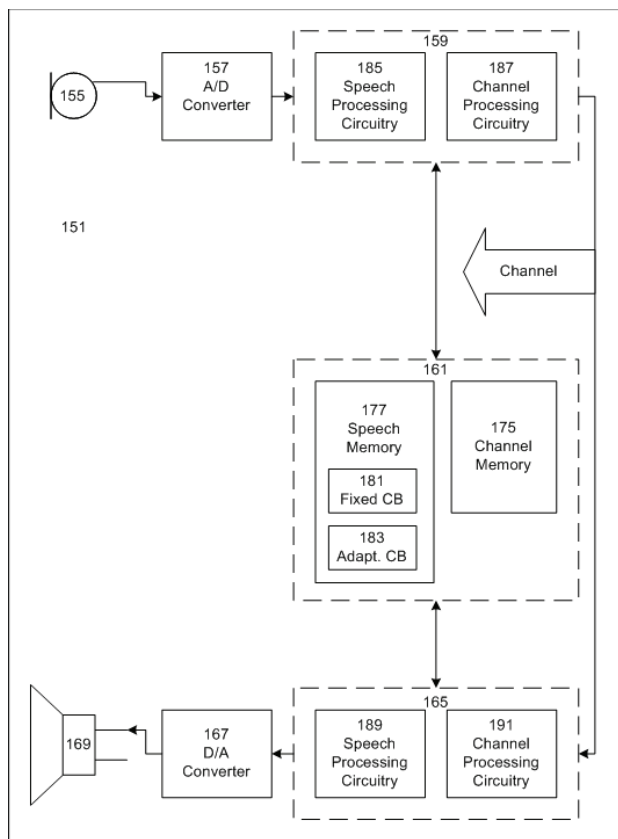
Additionally, the inability to determine the optimal encoding mode for a given noise signal at a given bit rate also results in an inefficient use of encoding resources. For a given speech signal having a particular noise component, the ability to selectively apply an optimal coding scheme at a given bit rate would provide more efficient use of an encoder processing circuit. Moreover, the ability to select the optimal encoding mode for type of noise signal would further maximize the available encoding resources while providing a more perceptually accurate reproduction of the noise signal.

II. DETAILED DESCRIPTION MODELS

Fig. 1a is a schematic block diagram of a speech communication system illustrating the use of source encoding and decoding in accordance with the present model.



a



b

Fig. 1: a - There is a speech communication system;
b - A schematic block diagram illustrates several variations of an exemplary communication device employing the functionality of Fig. 1a.

Therein, a speech communication system 100 supports communication and reproduction of speech across a communication channel 103. Although it may comprise for example a wire, fiber or optical link, the communication channel 103 typically comprises, at least in part, a radio frequency link that often must support multiple; simultaneous speech exchanges requiring shared bandwidth resources such as may be found with cellular telephony embodiments.

Although not shown, a storage device may be coupled to the communication channel 103 to temporarily store speech information for delayed reproduction or playback, e.g., to

perform answering machine functionality, voiced email, etc. Likewise, the communication channel 103 might be replaced by such a storage device in a single device embodiment of the communication system 100 that, for example, merely records and stores speech for subsequent playback.

In particular, a microphone 111 produces a speech signal in real time. The microphone 111 delivers the speech signal to an A/D (analog to digital) converter 115. The A/D converter 115 converts the speech signal to a digital form then delivers the digitized speech signal to a speech encoder 117.

The speech encoder 117 encodes the digitized speech by using a selected one of a plurality of encoding modes. Each of the plurality of encoding modes utilizes particular techniques that attempt to optimize quality of resultant reproduced speech. While operating in any of the plurality of modes, the speech encoder 117 produces a series of modeling and parameter information (hereinafter "speech indices"), and delivers the speech indices to a channel encoder 119.

The channel encoder 119 coordinates with a channel decoder 131 to deliver the speech indices across the communication channel 103. The channel decoder 131 forwards the speech indices to a speech decoder 133. While operating in a mode that corresponds to that of the speech encoder 117, the speech decoder 133 attempts to recreate the original speech from the speech indices as accurately as possible at a speaker 137 via a D/A (digital to analog) converter 135.

The speech encoder 117 adaptively selects one of the pluralities of operating modes based on the data rate restrictions through the communication channel 103. The communication channel 103 comprises a bandwidth allocation between the channel encoder 119 and the channel decoder 131. The allocation is established, for example, by telephone switching networks wherein many such channels are allocated and reallocated as need arises. In one such embodiment, either a 22.8 kbps (kilobits per second) channel bandwidth, i.e., a full rate channel, or a 11.4 kbps channel bandwidth, i.e., a half rate channel, may be allocated.

With the full rate channel bandwidth allocation, the speech encoder 117 may adaptively select an encoding mode that supports a bit rate of 11.0, 8.0, 6.65 or 5.8 kbps. The speech encoder 117 adaptively selects an either 8.0, 6.65, 5.8 or 4.5 kbps encoding bit rate mode when only the half rate channel has been allocated. Of course these encoding bit rates and the aforementioned channel allocations are only representative of the present embodiment. Other variations to meet the goals of alternate embodiments are contemplated.

With either the full or half rate allocation, the speech encoder 117 attempts to communicate using the highest encoding bit rate mode that the allocated channel will support. If the allocated channel is or becomes noisy or

otherwise restrictive to the highest or higher encoding bit rates, the speech encoder 117 adapts by selecting a lower bit rate encoding mode.

Similarly, when the communication channel 103 becomes more favorable, the speech encoder 117 adapts by switching to a higher bit rate encoding mode.

With lower bit rate encoding, the speech encoder 117 incorporates various techniques to generate better low bit rate speech reproduction. Many of the techniques applied are based on characteristics of the speech itself. For example, with lower bit rate encoding, the speech encoder 117 classifies noise, unvoiced speech, and voiced speech so that an appropriate modeling scheme corresponding to a particular classification can be selected and implemented. Thus, the speech encoder 117 adaptively selects from among a plurality of modeling schemes those most suited for the current speech. The speech encoder 117 also applies various other techniques to optimize the modeling as set forth in more detail below.

Fig. 1b is a schematic block diagram illustrating several variations of an exemplary communication device employing the functionality of Fig. 1a.

A communication device 151 comprises both a speech encoder and decoder for simultaneous capture and reproduction of speech. Typically within a single housing, the communication device 151 might, for example, comprise a cellular telephone, portable telephone, computing system, etc. Alternatively, with some modification to include for example a memory element to store encoded speech information the communication device 151 might comprise an answering machine, a recorder, voice mail system, etc.

A microphone 155 and an A/D converter 157 coordinate to deliver a digital voice signal to an encoding system 159. The encoding system 159 performs speech and channel encoding and delivers resultant speech information to the channel. The delivered speech information may be destined for another communication device (not shown) at a remote location

As speech information is received, a decoding system 165 performs channel and speech decoding then coordinates with a D/A converter 167 and a speaker 169 to reproduce something that sounds like the originally captured speech.

The encoding system 159 comprises both a speech processing circuit 185 that performs speech encoding, and a channel processing circuit 187 that performs channel encoding. Similarly, the decoding system 165 comprises a speech processing circuit 189 that performs speech decoding, and a channel processing circuit 191 that performs channel decoding.

Although the speech processing circuit 185 and the channel processing circuit 187 are separately illustrated, they might be combined in part or in total into a single unit. For example, the speech processing circuit 185 and the channel processing circuit 187 might share a single DSP (digital signal processor) and/or other processing circuitry.

Similarly, the speech processing circuit 189 and the channel processing circuit 191 might be entirely separate or combined in part or in whole. Moreover, combinations in whole or in part might be applied to the speech processing circuits 185 and 189, the channel processing circuits 187 and 191, the processing circuits 185, 187, 189 and 191, or otherwise.

The encoding system 159 and the decoding system 165 both utilize a memory 161. The speech processing circuit 185 utilizes a fixed codebook 181 and an adaptive codebook 183 of a speech memory 177 in the source encoding process. The channel processing circuit 187 utilizes a channel memory 175 to perform channel encoding. Similarly, the speech processing circuit 189 utilizes the fixed codebook 181 and the adaptive codebook 183 in the source decoding process. The channel processing circuit 187 utilizes the channel memory 175 to perform channel decoding.

The speech memory 177 is shared as illustrated. Separate copies thereof can be assigned for the processing circuits 185 and 189. Likewise, separate channel memory can be allocated to both the processing circuits 187 and 191. The memory 161 also contains software utilized by the processing circuits 185, 187, 189 and 191 to perform various functionality required in the source and channel encoding and decoding processes.

III. A MULTI-STEP ENCODING

Figs. 2-4 are functional block diagrams illustrating a multi-step encoding approach used by one embodiment of the speech encoder illustrated in Figs. 1a and 1b. In particular, Fig. 2 is a functional block diagram illustrating of a first stage of operations performed by one embodiment of the speech encoder shown in Figs. 1a and 1b. The speech encoder, which comprises encoder processing circuitry, typically operates pursuant to software instruction carrying out the following functionality.

At a block 215, source encoder processing circuitry performs high pass filtering of a speech signal 211. The filter uses a cutoff frequency of around 80 Hz to remove, for example, 60 Hz power line noise and other lower frequency signals. After such filtering, the source encoder processing circuitry applies a perceptual weighting filter as represented by a block 219. The perceptual weighting filter operates to emphasize the valley areas of the filtered speech signal, if the encoder processing circuitry selects operation in a pitch preprocessing (PP) mode as indicated at a control block 245, a pitch preprocessing operation is performed on the weighted speech signal at a block 225. The pitch preprocessing operation involves warping the weighted speech signal to match interpolated pitch values that will be generated by the decoder processing circuitry. When pitch preprocessing is applied, the warped speech signal is designated a first target signal 229. If pitch preprocessing is not selected the control block 245, the weighted speech

signal passes through the block 225 without pitch preprocessing and is designated the first target signal 229.

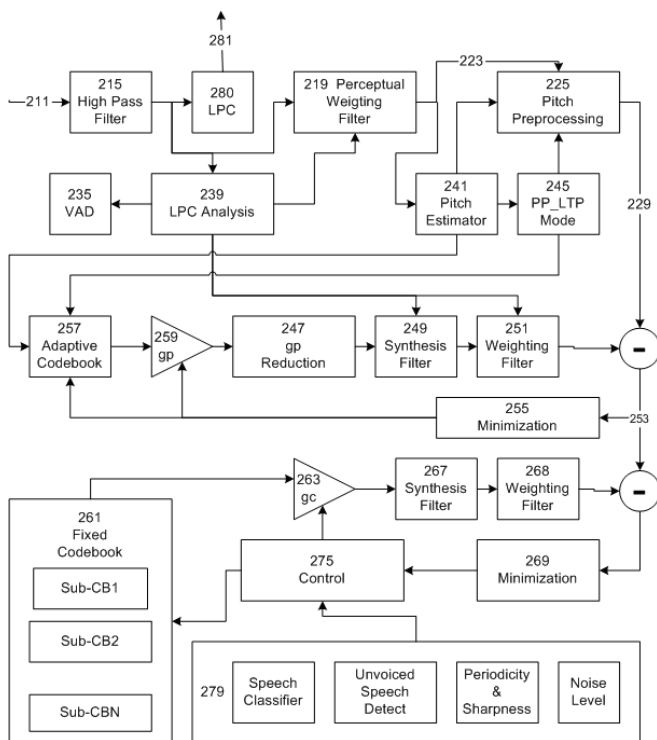


Fig. 2. A first stage of operations is performed by one embodiment of the speech encoder shown in Figs. 1a and 1b

As represented by a block 255, the encoder processing circuitry applies a process wherein a contribution from an adaptive codebook 257 is selected along with a corresponding gain 257 which minimize a first error signal 253. The first error signal 253 comprises the difference between the first target signal 229 and a weighted, synthesized contribution from the adaptive codebook 257.

At blocks 247, 249 and 251, the resultant excitation vector is applied after adaptive gain reduction to both a synthesis and a weighting filter to generate a modeled signal that best matches the first target signal 229. The encoder processing circuitry uses LPC (linear predictive coding) analysis, as indicated by a block 239, to generate filter parameters for the synthesis and weighting filters. The weighting filters 219 and 251 are equivalent in functionality.

Next, the encoder processing circuitry designates the first error signal 253 as a second target signal for matching using contributions from a fixed codebook 261. The encoder processing circuitry searches through at least one of the pluralities of sub codebooks within the fixed codebook 261 in an attempt to select a most appropriate contribution while generally attempting to match the second target signal.

More specifically, the encoder processing circuitry selects an excitation vector, its corresponding sub codebook and gain based on a variety of factors. For example, the

encoding bit rate, the degree of minimization, and characteristics of the speech itself as represented by a block 279 are considered by the encoder processing circuitry at control block 275. Although many other factors may be considered, exemplary characteristics include speech classification, noise level, sharpness, periodicity, etc. Thus, by considering other such factors, a first sub codebook with its best excitation vector may be selected rather than a second sub codebook's best excitation vector even though the second sub codebook's better minimizes the second target signal 265.

Fig. 3 is a functional block diagram depicting of a second stage of operations performed by the embodiment of the speech encoder illustrated in Fig. 2.

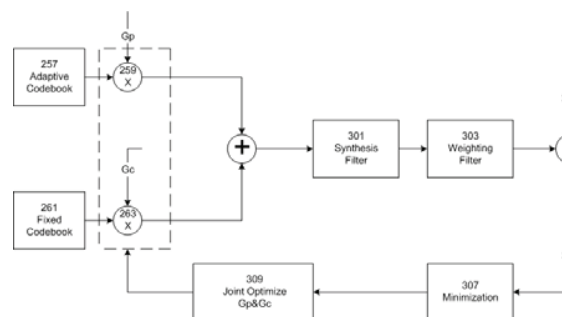


Fig. 3. A functional block diagram depicts of a second stage of operations performed by the embodiment of the speech encoder

In the second stage, the speech encoding circuitry simultaneously uses both the adaptive the fixed codebook vectors found in the first stage of operations to minimize a third error signal 311.

The speech encoding circuitry searches for optimum gain values for the previously identified excitation vectors (in the first stage) from both the adaptive and fixed codebooks 257 and 261. As indicated by blocks 307 and 309, the speech encoding circuitry identifies the optimum gain by generating a synthesized and weighted signal, i.e., via a block 301 and 303, that best matches the first target signal 229 (which minimizes the third error signal 311). Of course if processing capabilities permit, the first and second stages could be combined wherein joint optimization of both gain and adaptive and fixed codebook vector selection could be used.

Fig. 4 is a functional block diagram depicting of a third stage of operations performed by the embodiment of the speech encoder illustrated in Figs. 2 and 3.

The encoder processing circuitry applies gain normalization, smoothing and quantization, as represented by blocks 401, 403 and 405, respectively, to the jointly optimized gains identified in the second stage of encoder processing. Again, the adaptive and fixed codebook vectors used are those identified in the first stage processing

With normalization, smoothing and quantization functionally applied, the encoder processing circuitry has

completed the modeling process. Therefore, the modeling parameters identified are communicated to the decoder. In particular, the encoder processing circuitry delivers an index to the selected adaptive codebook vector to the channel encoder via a multiplexor 419. Similarly, the encoder processing circuitry delivers the index to the selected fixed codebook vector, resultant gains, and synthesis filter parameters etc., to the multiplexor 419. The multiplexor 419 generates a bit stream 421 of such information for delivery to the channel encoder for communication to the channel and speech decoder of receiving device [1].

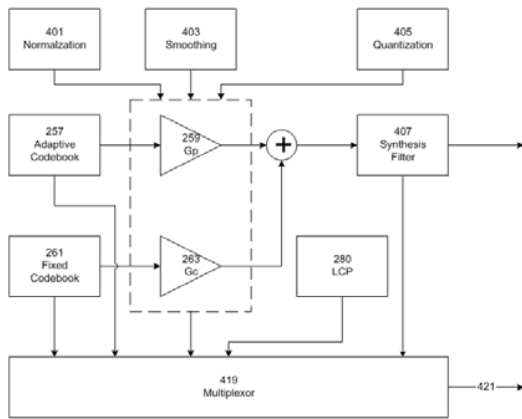


Fig. 4. A functional block diagram depicting a third stage of operations performed by the embodiment of the speech encoder

In many applications it is important to measure how much distortion an operation exerts on the spectrum, that is, to measure how much the spectrum changes, for example, in quantization of parameters. Define the spectral error or difference $V(\omega)$ as [2]

$$V(\omega) = 10 \log_{10}[A(\omega)] - 10 \log_{10}[\hat{A}(\omega)]. \quad (1)$$

The simplest and most used of spectral distortion measures, log spectral distortion (SD), is defined as

$$d^2 = (1/\pi) \int_0^\pi |V(\omega)|^2 d\omega, \quad (2)$$

where $A(\omega)$ is the original spectrum and $\hat{A}(\omega)$ the distorted spectrum [3].

In this paper we describe the measured characteristics of constrained Linear Predictive (LP) models.

IV. LINEAR PREDICTIVE MODELS

A. Hardware

Fig. 5 is a block diagram of an embodiment illustrating functionality of speech decoder having corresponding

functionality to that illustrated in Figs. 2-4.

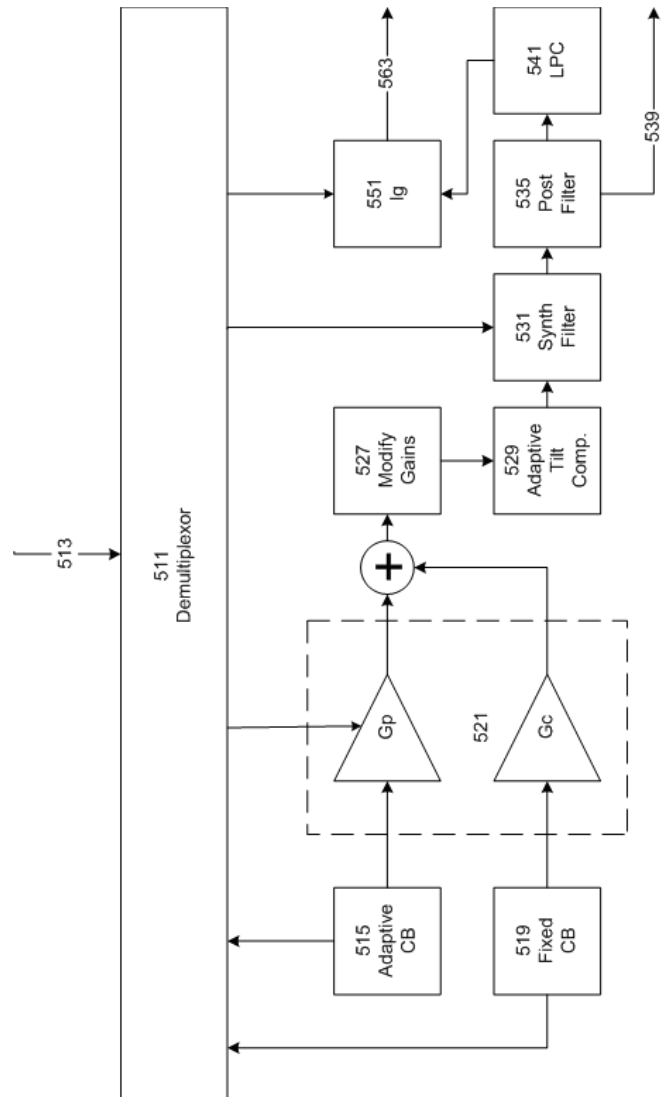


Fig. 5. A functional block diagram depicting a second stage of operations performed by the embodiment of the speech encoder

As with the speech encoder the speech decoder, which comprises decoder processing circuitry, typically operates pursuant to software instruction carrying out the following functionality.

A demultiplexor 511 receives a bit stream 513 of speech modeling indices from an often remote encoder via a channel decoder. As previously discussed, the encoder selected each index value during the multi-stage encoding process described above in reference to Figs. 2-4. The decoder processing circuitry utilizes indices, for example, to select excitation vectors from an adaptive codebook 515 and a fixed codebook 519, set the adaptive and fixed codebook gains at a block 521, and set the parameters for a synthesis filter 531. With such parameters and vectors selected or set, the decoder processing circuitry generates a

reproduced speech signal 539. In particular, the codebooks 515 and 519 generate excitation vectors identified by the indices from the demultiplexor 511. The decoder processing circuitry applies the indexed gains at the block 521 to the vectors which are summed. At a block 527, the decoder processing circuitry modifies the gains to emphasize the contribution of vector from the adaptive codebook 515. At a block 529, adaptive tilt compensation is applied to the combined vectors with a goal of flattening the excitation spectrum. The decoder processing circuitry performs synthesis filtering at the block 531 using the flattened excitation signal

Finally, to generate the reproduced speech signal 539, post filtering is applied at a block 535 deemphasizing the valley areas of the reproduced speech signal 539 to reduce the effect of distortion. The hardware is realized with QuartusII from corporation Altera.

B. Software

A sub design 551 receives a bit stream of Linear Predictive speech modeling indices LPC_j from an often remote encoder via a channel decoder and one LPC_j^{\wedge} via block 541. It have used the following modified nonlinear equation

$$SA^2 = \sum_j 10 \log_{10} |LPC_j^{\wedge} / LPC_j| \quad (3)$$

The software model is realized with MATLAB from corporation Math Works. Text of program is M-file (look appendix).

V. CELLULAR TELEPHONY

In the exemplary cellular telephony embodiment of the present invention, the A/D converter 115 (Fig. 1a) will generally involve analog to uniform digital PCM including: 1) an input level adjustment device; 2) an input anti-aliasing filter; 3) a sample-and-hold device sampling at 8 kHz; and 4) analog to uniform digital conversion to 13-bit representation.

Similarly, the D/A converter 135 will generally involve uniform digital PCM to analog including: 1) conversion from 13-bit/8 kHz uniform PCM to analog; 2) a hold device; 3) reconstruction filter including $x/\sin(x)$ correction; and 4) an output level adjustment device.

In terminal equipment, the A/D function may be achieved by direct conversion to 13-bit uniform PCM format, or by conversion to 8-bit/A-law compounded format. For the D/A operation, the inverse operations take place.

The encoder 117 receives data samples with a resolution of 13 bits left justified in a 16-bit word. The three least significant bits are set to zero. The decoder 133 outputs data in the same format. Outside the speech codec, further processing can be applied to accommodate traffic data having a different representation. A specific embodiment of

an AMR (adaptive multi-rate) codec with the operational functionality illustrated in Figs. 2-5 uses five source codec with bit-rates 11.0, 8.0, 6.65, 5.8 and 4.55 kbps. Four of the highest source coding bit-rates are used in the full rate channel and the four lowest bit-rates in the half rate channel.

All five source codec within the AMR codec are generally based on a code-excited linear predictive (CELP) coding model. A 10th order linear prediction (LP), or short-term, synthesis filter, e.g., used at the blocks 249, 267, 301, 407 and 531 (of Figs. 2-5), is used.

VI. SIMULATION MODEL

Twenty-one encoder input sequences are provided ETSI [4]. Note that for the input sequences TEST0.INP to TEST3.INP, the amplitude figures are given in 13-bit precision. The active speech levels are given in dBov.

TEST0.INP - Synthetic harmonic signal. The pitch delay varies slowly from 18 to 143.5 samples. The minimum and maximum amplitudes are -997 and +971.

TEST1.INP - Synthetic harmonic signal. The pitch delay varies slowly from 144 down to 18.5 samples. Amplitudes at saturation point -4096 and +4095. - TEST2.INP - Sinusoidal sweep varying from 150 Hz to 3400 Hz. Amplitudes ± 1250 .

TEST3.INP - Sinusoidal sweep varying from 150 Hz to 3400 Hz. Amplitudes ± 4000 .

TEST4.INP - Female speech, active speech level: -19.4 dBov, flat frequency response.

TEST5.INP - Male speech, active speech level: -18.7 dBov, flat frequency response.

TEST6.INP - Female speech, ambient noise, active speech level: -35.0 dBov, flat frequency response.

TEST7.INP - Female speech, ambient noise, active speech level: -25.0 dBov, flat frequency response.

TEST8.INP - Female speech, ambient noise, active speech level: -15.6 dBov, flat frequency response.

TEST9.INP - Female speech, car noise, active speech level: -35.5 dBov, flat frequency response.

TEST10.INP - Female speech, car noise, active speech level: -26.1 dBov, flat frequency response.

TEST11.INP - Female speech, car noise, active speech level: -15.8 dBov, flat frequency response.

TEST12.INP - Male speech, ambient noise, active speech level: -34.9 dBov, flat frequency response.

TEST13.INP - Male speech, ambient noise, active speech level: -24.8 dBov, flat frequency response.

TEST14.INP - Male speech, ambient noise, active speech level: -15.0 dBov, flat frequency response.

TEST15.INP - Male speech, babble noise, active speech level: -34.1 dBov, flat frequency response.

TEST16.INP - Male speech, babble noise, active speech level: -24.3 dBov, flat frequency response.

TEST17.INP - Male speech, babble noise, active speech level: -14.4 dBov, flat frequency response.

TEST18.INP - Female speech, ambient noise, active speech level: -26.0 dBov, modified IRS frequency response, with many zero frames.

TEST19.INP - Male speech, ambient noise, active speech level: -36.0 dBov, modified IRS frequency response, with many zero frames.

TEST20.INP - Sequence for exercising the LPC vector quantization codebooks and ROM tables of the codec.

The TEST0.INP and TEST1.INP sequences were designed to test the pitch lag of the GSM enhanced full rate speech encoder. In a correct implementation, the resulting speech encoder output parameters shall be identical to those specified in the TEST0.COD and TEST1.COD sequences, respectively.

The document [5] contains an electronic copy of the ANSI-C code for the Adaptive Multi-Rate codec. The ANSI-C code is necessary for a bit exact implementation of the Adaptive Multi Rate speech transcoder (TS 26.090 [6]).

VII. SIMULATION RESULTS

To begin, we obtain the spectral distortion of the TEST4.INP for the Fast Fourier transform. Table I shows the normalized spectral distortion of Female speech, active speech level: -19.4 dBov, flat frequency response.

TABLE I
TEST4 LAST

BIT RATE, KBPS	QUANTITY	QUANTITY / 8,7797(1)
MR122	2,8397	0,323439
MR102	2,9657	0,337791
MR795	3,5974	0,409741
MR74	3,3838	0,385412
MR67	3,7678	0,429149
MR595	3,6410	0,414707
MR515	4,4173	0,503127
MR475	4,3649	0,497158

Second, we obtain the spectral distortion of the TEST5.INP for the Fast Fourier transform. C, active speech level: -18.7 dBov, flat frequency response.

TABLE II
TEST5 LAST

BIT RATE, KBPS	QUANTITY	QUANTITY / 8,7797(3)
MR122	5,2161	0,594109
MR102	5,4805	0,624224
MR795	6,5916	0,750777
MR74	6,5855	0,750083
MR67	7,5220	0,856749
MR595	7,7835	0,886534
MR515	8,4710	0,964839
MR475	8,7797	1

Third, we obtain the new spectral distortion of the TEST4.INP for the Fast Fourier transform. Table III shows the normalized spectral distortion of Female speech, active speech level: -19.4 dBov, flat frequency response.

TABLE III
TEST4 NEW

BIT RATE, KBPS	QUANTITY	QUANTITY / 2,4408(2)
MR122	1,9257	0,788963
MR102	1,8834	0,771632
MR795	1,9427	0,795928
MR74	1,9843	0,812971
MR67	1,9390	0,794412
MR595	1,9377	0,793879
MR515	2,0095	0,823296
MR475	2,0348	0,833661

Forth, we obtain the new spectral distortion of the TEST5.INP for the Fast Fourier transform. Table IV shows the normalized spectral distortion of male speech, active speech level: -18.7 dBov, flat frequency response.

TABLE IV
TEST5 NEW

BIT RATE, KBPS	QUANTITY	QUANTITY / 2,4408(4)
MR122	2,0334	0,833088
MR102	2,0502	0,839971
MR795	2,0986	0,859800
MR74	2,0925	0,857301
MR67	2,0648	0,845952
MR595	2,0857	0,854515
MR515	2,1277	0,871722
MR475	2,1492	0,880531

Fig. 6 illustrates these results:

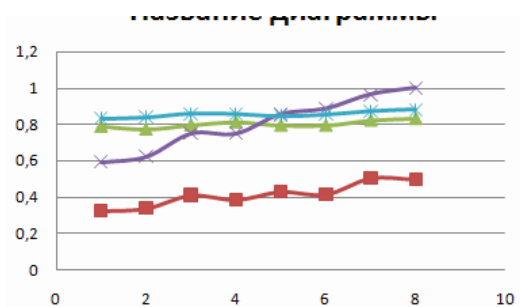


Fig. 6. It is simulation results

Table V shows the error of spectral distortion measurement obtained for female and male speech.

TABLE V
DELTA

BIT RATE, KBPS	LAST DELTA(5)	NEW DELTA(6)
MR122	0,27067	0,044125
MR102	0,286433	0,068338
MR795	0,341037	0,063873
MR74	0,364671	0,044330
MR67	0,427600	0,051540
MR595	0,471827	0,060636
MR515	0,461713	0,048427
MR475	0,502842	0,046870

Fig. 7 shows accuracy new technique.

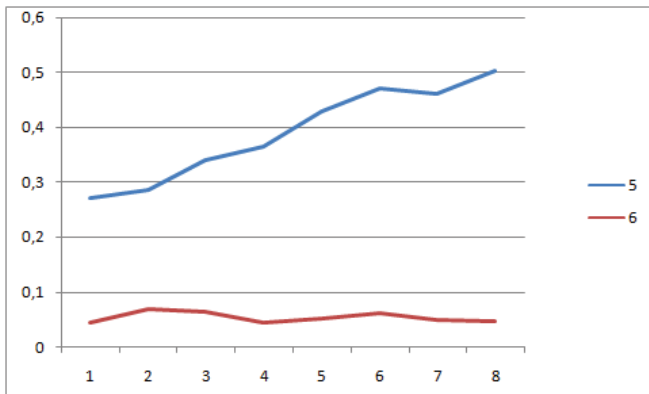


Fig. 7. It is simulation results

VIII. CONCLUSION

The proposed techniques have more high accuracy.

APPENDIX

```

p = 10;
number_of_frame = 64800/320;
max_shift = 5;
kolichestvo_tochek_na_grafike = 9;
yv = wavread('Speech.wav',64800);
old_y = yv;
y(:, 1) = wavread('Speech.wav',64800);
y(:, 2) = wavread('speech_475-1.wav',64800);
y(:, 3) = wavread('speech_475-2.wav',64800);
y(:, 4) = wavread('speech_475-3.wav',64800);
y(:, 5) = wavread('speech_475-4.wav',64800);
y(:, 6) = wavread('speech_475-5.wav',64800);
y(:, 7) = wavread('speech_475-6.wav',64800);
y(:, 8) = wavread('speech_475-7.wav',64800);
y(:, 9) = wavread('speech_475-8.wav',64800);

for index_i = 1:kolichestvo_tochek_na_grafike
    sa_super = 0;
    count = 0;

    for index_ii = 1:(number_of_frame)

        count_i = 0;

        for index_iii = 1:max_shift
            new = y((index_ii*160 -160 + index_iii):(index_ii*160
+ index_iii - 1) , index_i);
            [a,g] = lpc(new,p)
            response_new = real(a);

            old = old_y((index_ii*160 -159):index_ii*160 , 1);
            [aa,gg] = lpc(old,p)
            response_old = real(aa);
            meas_sa = 10 * log10(abs(response_new./response_old));
            f_sa = meas_sa * meas_sa';
            sa_ii = sqrt(f_sa/(p + 1));

```

```

        if (index_iii == 1)
            old_sa_iii = sa_ii;
        end

        if (old_sa_iii > sa_ii)
            old_sa_iii = sa_ii;
        end

        count_i = count_i + 1;

    end
    sa_super = sa_super + old_sa_iii;
    count = count + 1;
end
sa(index_i , 1) = sqrt(sa_super/count);
ayv = fft(yv);
ayvv = fft(y(:, index_i));
eyv = ayv.*conj(ayvv);
eyvv = ayvv.*conj(ayv);
yyv = 10*log10(eyv);
yyvv = 10*log10(eyvv);
meas = minus(yyv,yyvv);
f = meas'*meas;
sd(index_i , 1) = sqrt(f/64800);
end
sd = sd

```

ACKNOWLEDGMENT

One of the authors, S.A. Krivenko, wishes to thank S.S. Krivenko for critical reading of the manuscript and useful discussion.

REFERENCES

- [1] THYSSEN, Jess, "Low complexity random codebook structure," U.S. Patent 99/19135, August 24, 1999.
- [2] Tom Backstrom. Linear predictive modeling of speech -constraints and line spectrum pair decomposition. Dissertation for the degree of Doctor of Science in Technology, Finland, Helsinki University of Technology, 2004. – p.37.
- [3] A. H. Gray and J. D. Markel. Distance measures for speech processing. IEEE Trans. Acoust. Speech Signal Proc., ASSP-24:pp. 380–391, Oct. 1976.
- [4] Digital cellular telecommunications system (Phase 2+); Test sequences for the GSM Enhanced Full Rate (EFR)(3GPP TS 46.054 version 6.0.0 Release 6) Available: <http://www.etsi.com>
- [5] Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech Codec; C-source code (3GPP TS 26.073 version 6.0.0 Release 6) Available: <http://www.etsi.com>
- [6] Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech Codec; Transcoding Functions (3GPP TS 26.090 version 6.0.0 Release 6) Available: <http://www.etsi.com>