ADAPTIVE POSTFILTER IN 16KBPS LD-CELP SPEECH CODER

Wang Bingxi He Yinghua
PostBox.1001 Zhengzhou,Henan,China

Abstract

In September 1992,the recommendation G.728, which is a 16kbps LD-CELP speech coder submitted by AT&T,was standarized by CCITT,In the process of ratification test[1],the coder's performances were equivalent to or better than that of 32kbps ADPCM for all conditions tested.This paper,which is based on a G.728 encoding-decoding system simulated in software,studies and tests different parts of the algorithm,espacially that of the postfilter.

1 THE BASIC STRUCTURE OF THE ENCODING-DECODING ALGORITHM

The essense of LD-CELP is CELP[2],the basic idea of which is to quantize and code speech by a closed analysis-by-synthesis book search process.

The input speech signal Si(n) is partitioned with every five continuous samples forming a vector(with an 8k sampling rate).Each candidate vector cj passes a gain control unit and a linear prediction(LP) filter H(z).From the resulting synthesis vectors,one is detected which satisfies a Minimum-Perceptual-Square-Error(MPSE) criterion.In addition to be sent to the decoder,the identified index j will also be fed back to the gain control unit and synthesis filter to determine the excitation gain and filter memory for the next coding ring.The coefficients of H(z) and the gain of gain controllor are extracted backwardly from the restored excitation and synthesis signal So(n) and updated every four vectors which results in low delay and reduced transmitted parameters.At the decoder,the optimum codevector is selected by the received codebook index,then used as excitation to drive gain controllor and synthesis filter.And in the decoder,the excitation gain and LP filter coefficients are extracted backwardly in the same way as in the encoder.To further enhance the output speech quality, a postfilter is added in the output terminal of the decoder.

2 THE CONSTRUCTION OF POSTFILTER AND ITS PERFORMANCES ANALYSIS

The postfilter treats the synthesis signal to enhance the subject effects of the output speech.

2.1 Long-term postfilter

Long-term postfilter,also named pitch-postfilter,has a transform function like:

$$Hl(z)=gl(1+bz^{-p}) \qquad (1)$$

where p is pitch period,ranging from 20 samples to 140 samples,while gl and b are coeffecients deciding the amount of long-term postfiltering.The principle of long-term filter is as following.The partition and quantization of signal will inevitably give rise to some noise.In voiced speech we can see from spectrum that the noise is more serious in the areas between harmonies,while the speech components at the basic frequency(due to the pitch period) and its harmonies have high enegy level themselves and will be less affected by noise.Therefore it is reasonable to belive a adaptive comb filter will be useful in reducing the noise between harmonies while the imformation in the harmonies can pass without much degradation.A zero-typed combfilter as given in (1)is a suitable selection here,which can trace the slow variation of the signal pitch.Fig.1 and Fig.2 show the waveform and the spectrum of signals before and after long-term postfiltering.Althougt there is a slight degradation in the objective average segement SNR(about 0.1db),but the subjective listening test indicates that there is an obvious improvement in the back-ground noise,especially that of

0-7803-2912-0

the high pitched,strongly periodic woman's voice. At the same time, the resulting distortion is almost unnoticeable.All three parameters of gl, b and p are extracted from synthesis signal So(n),and updated periodically every four vectors(a frame).When extracting the pitch,we make use of the stability of speech pitch for the accuracy of pitch predicting.The coefficients gl and b decide the amount of long-term postfiltering,and

$$gl=1/(1+b) \qquad (2)$$

If b is too large,that may cause an unstable signal concurve ,and therefore large distortion.Too small b will have not much effect on the output signal.Therefore,b is restricted in the range of 0 ~ 0.15.Besides, there is no obvious periodicity in unvoiced speech,therefore it is not wise to fixed the b in both the voiced and unvoiced speech. In the method recommended in G.728,the b parameter is determined adaptively and it is larger in periodic voiced speech(not beyond 0. 15) ,and smaller (the minimum of b is 0, which means no long-term postfiltering) in speech without much periodicity(such as unvoiced speech).

2.2 short-term postfilter

Another important component of postfilter is a zero-pole short-term postfilter cascaded by a first-order highpass filter:

$$Hs(z)=\frac{\sum_{i=0}^{10} a_i r_1^i z^{-i}}{\sum_{i=0}^{10} a_i r_2^i z^{-i}}(1+\mu z^{-1}) \qquad (3)$$

where ai is the coefficient of a 10th-order LP filter obtained by backward prediction, r1, r2 are factors controlling the amount of postfiltering.

As we know, LP filter can be used to simulate the formants structure of speech. which matches well at spectral peaks and poorly at spectral valleys.And the energy of frequency components in spectral valleys are usually low,thus human ears are more sensitive to noises there according to mask principle.The idea of re-shaping the noise spectrum is to reduce the noise energy in the area of signal spectral valleys, while although the noise in signal peaks may become larger,it can still be unnoticeable because the high energy of signal has masked the noises.Thus the overall output subjective effects will be better.Suppose

$$H(z)=\frac{1}{\sum_{i=0}^{10} a_i z^{-i}} \qquad (4)$$

is the transform function of the 10th-ordered LP filter.

$$Hr(z)=H(z/r) \qquad (5)$$

is equivalent to remove the poles of H(z) toward origin along radii.Thus,Hr(z) has the same formants positions as H(z),but lower spectral peaks and wider frequency bands,so,r is also called band expanding factor.Fig.4 and Fig.5 are spectrum of a signal and the frequency response of LP filter with different r.Filtering the signal with a pole-typed postfilter as Hr(z) can efficiently reduce the noises in spectral valleys while maintain the imformation of formants. But it is discovered in experiments that there is a phenomenon of muffling after this kind of pole-typed postfiltering.Actually,it is because of a large tilt in the low frequency part of Hr(z),which results in a low pass effects of the speech.So we use zero-pole typed postfilter(as in (3)) to compensate the tilt.As shown in Fig. 5, the frequency response of Hr(z) with r=0.8 has both formants and tilt while the frequency response of Hr(z) with r=0.5 has only tilt.Fig.6 is the frepuency response of Hr1(z)/Hr2(z). where r1=0.9,r2=0.5.We can see that most of tilt has already been compensated with only the tilt near the first formant has not been,which causes the remaining muffling phenomenon.To compensate this tilt, a first-order high pass filter is added:

$$Hh(z)=1+\mu z^{-1} \qquad (6)$$

where μ =0.15k1 and k1 is the first reflect coefficient gained in backward prediction.The frequency response of short-term postfilter with the high-pass filter is shown in Fig. 7.

2.3 gain scaling adapter

After long-term and short-term postfiltering ,the energy of signal has changed,therefore a gain scaling adapter is needed to restore the energy to the pre-postfiltering level.To reduce calculating complexity,the ratio between two signals' short-time-average-amplitude is used and the results are fairly good.To avoid uncontinuity of signal the gain scaling factor which is computed every vector is low-pass filtered to obtain the scaling factor of every sample.Suppose that σ_{-1} is the scaling factor of the last sample in the previous vector, σ' is the ratio computed of current vector,then the scaling factor of every sample in current vector is:

$$\sigma_k = 0.01\sigma' + 0.09\sigma_{k-1} \quad k=0,\ 1,\ 2,\ 3,\ 4; \quad (7)$$

3 THE PROBLEM IN POSTFILTER AND ITS EFFECTS ON THE SELECTION OF PARAMETERS

So far we have given a brife describtion of G.728 LD-CELP speech encoding-decoding system,and discussed in detail the postfilter,for example ,its main components and its main function.With further study,we find out that the principle of postfilter is the same as that of peceptual weight filter in making use of the masking phenomenon.But the postfilter filters the synthesis signal directly,so that,not only noise spectrum is modified but also the spectrum of speech signal.Then why not omit postfilter and use perceptual weight filter alone?On the first hand,the perceptual weight filter does not filter the output noise signal,but try to control the noise by MPSE codebook search.It can only "expect" that the weighted quantization-error will be white noise, but it cannot guarantee that every time.Therefore the noise we actually get may not satisfy the request.The postfilter,which filters the signal directly,can thus have better control of output noise spectrum.Another reason is the tandems of G.728,because the distortion in every coding stage may accumulate to a serious extent after several tandem coding of LD-CELP. Therefore we must adjust the amount of postfiltering and perceptual weight filtering and their relationship. Experiment shows that no postfiltering in one stage can obtain good effect while with three or more stages of encoding-decoding,the output is far from ideal.But when we use the postfiter and adjust the factors in both the postfilter and perceptual weight filter ,we can have excellent performance both in one-stage and in muti-stage (mainly three-stage)coding.The parameters given in G.728 are all tuned to optimize the performance of three tandem coding . What's more,the phase distortion resulted from postfiltering is harmful to some non-speech signals on line such as modem signal.Therefore it is necessary to disable the postfilter when modem signal is detected,that is to set the b,r1,r2 and μ to 0.

4 CONCLUSION

From above theoritic discusses and experimental analysis,it is shown that the LD-CELP technique in G.728 can implement low-delay, high quality speech coding and the postfiltering technique is effective in enhancing the output speech.With further study on CELP and post-filtering we can expect even further advancement in reducing the bitrate and improving the output speech quality.

REFERENCE

[1] Juin-Hwey Chen, Nikil Jayant, and Richard V.cox, "Improving The Performance Of The 16KB/S LD-CELP Speech Coder", ICASSP'92 pp69-72

[2] Recommendation G.728 Coding of Speech At 16kbits/s Using Low-Delay Code Excited Linear Prediction

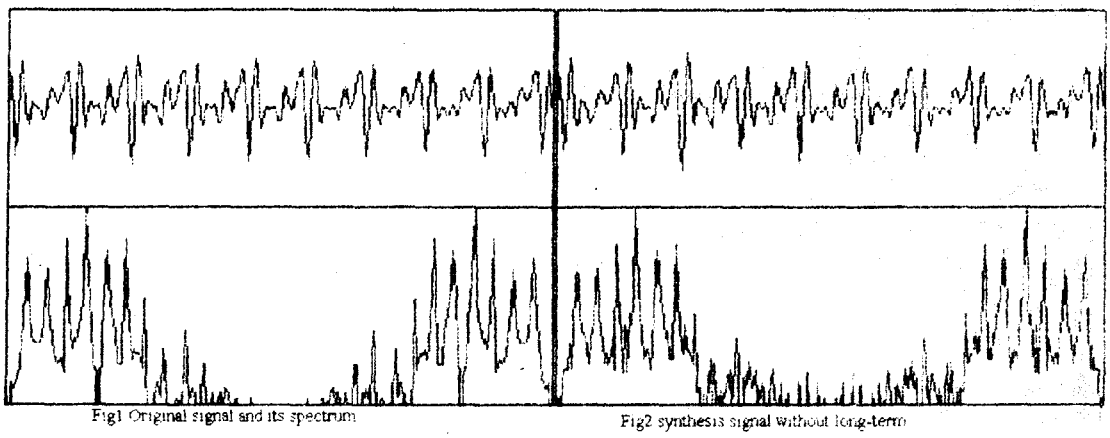

Fig1 Original signal and its spectrum

Fig2 synthesis signal without long-term filtered and its spectrum

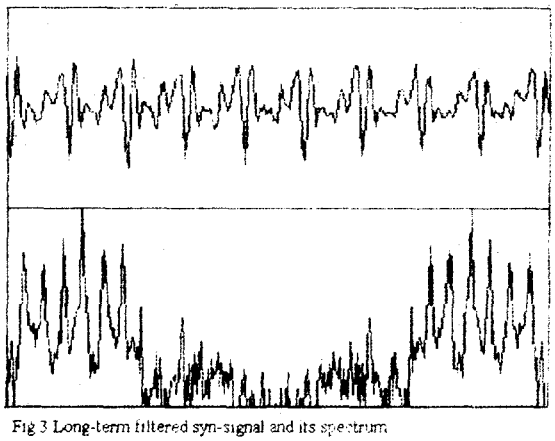

Fig 3 Long-term filtered syn-signal and its spectrum

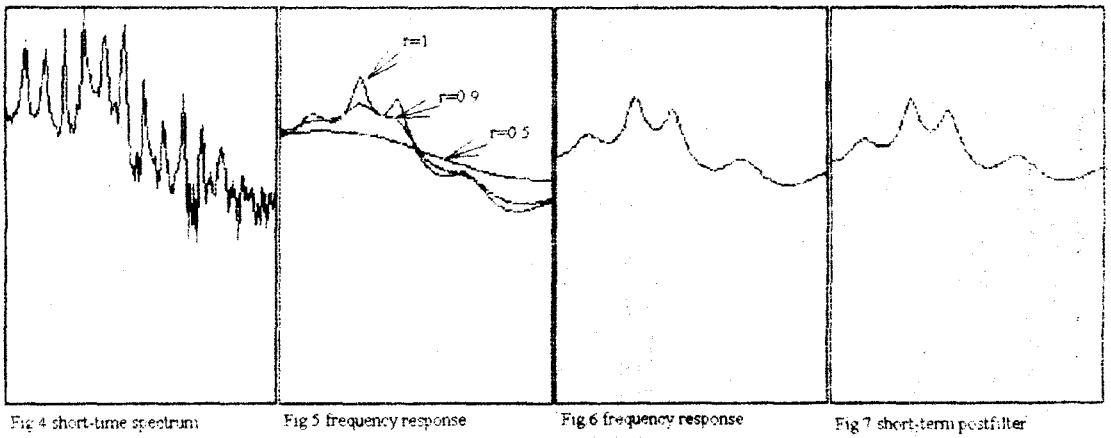

Fig 4 short-time spectrum of signal

Fig 5 frequency response of LP filter with different r

Fig 6 frequency response of Hf1(z)/Hf2(z) with r1=0 5,r2=0 9

Fig 7 short-term postfilter with one-order high pass filter