ISCA Archive
http://www.isca-speech.org/archive

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

4.8KBPS SPEECH CODING USING FRAME SYNCHRONOUS TIME DOMAIN COMPRESSION (FS-TDC)

Shinya Takahashi and Kunio Nakajima

Information Systems and Electronics Development Laboratory
Mitsubishi Electric Corporation
Kamakura-shi, Kanagawa, 247 Japan

ABSTRACT

In this paper, we propose a low bit rate (4.0-4.8kbps) speech coding algorithm for digital mobile communication. The coding algorithm is based on a frame synchronous time domain compression (FS-TDC) of LPC residual signal which is optimaly combined with stochastic coding algorithm. Like other TDC schemes, FS-TDC basically uses pitch periodicity of speech, but different from others, its compression and expansion process are both completed within a frame, and the compression ratio is large and fixed. In evaluation tests, the sochastic coders with FS-TDC shows better speech quality than an usual system, and shows good robustness for channel error if parameters are protected by optimized error protection scheme.

I. INTRODUCTION

Recently, a low bit rate speech coder (4.0-4.8kbps) which has almost the same speech quality as that of medium bit rate coder (8.0kbps) is required in mobile communication systems. In the half rate digital cellular systems scheduled in U.S. and Japan for the future, around 6kbps including source and error correction codes will be allocated to speech coder. Then around 4kbps will be available as source rate.

Stochastic codings[1,2] are thought be most promising scheme for obtaining high-quality speech at low/medium bit rates. In these stochastic codings, speech excitation source is needed to be represented correctly. But this is difficult for low bit rate speech codings below 4.8kbps. Then, eliminating redundancy from the excitation is considered to be effective at that low bit rates.

Since speech excitation has pitch periodicity in voiced region, it can be compressed efficiently in time domain by using the pitch periodicity. If compression ratio is large enough, for example 2:1, the number of bits per a compressed excitation sample will be doubled so that noise caused by quantizing excitation should be reduced. On the other hand, it is unavoidable that time domain compression (TDC) scheme itself causes certain degradation on synthesized speech. As a whole, the merit of the TDC would exceed to its demerit below 4.8kbps where very small number of bits can be allocated to quantize excitation.

As one of the promising TDC schemes, we proposed an algorithm based on a frame synchronous time domain compression (FS-TDC) which is optimally combined with stochastic coder[3]. In the algorithm, LPC residual as speech excitation is compressed in time domain by using its pitch periodicity and then vector quantized by stochastic codebook with minimizing a distortion caused by TDC. In this paper, basic algorithm and some recent improvements of FS-TDC are explained, and results of evaluation tests including robustness test for channel error are also described.

II. FRAME SYNCHRONOUS TIME DOMAIN COMPRESSION (FS-TDC)

Conventional TDC scheme[4] has a problem that the extent of compression and expansion depends on the pitch period. So, framing the compressed or expanded signal at fixed interval is very difficult matter. However, to apply TDC scheme to stochastic coders for mobile communication, following 3 conditions are especially important.

(1)Time domain compression and expansion process must be completed within a frame to have small processing delay and good robustness for the channel error.

(2)Compression ratio should be fixed for fixed length vector quantization.

(3)No click sounds must be induced by TDC.

To satisfy these conditions, FS-TDC algorithm is proposed. In the algorithm, compression ratio is set to 2:1 when the pitch periodicity in the frame is high while it is set to 4:3 when the pitch periodicity is low. This algorithm basically depends on a character of residual signal that it can be spliced and combined every one pitch period without causing serious degradation in a output speech when the pitch periodicity is high. Detailed algorithm is described as follows.

(A)Compression (ratio 2:1): At the beginning part of the frame, input residual is comperessed using eq.(1). 2 pitch periods are compressed into 1 period.

$$RC_i = (R_i + R_{i+P})/2 \qquad (i = 0, ..P-1) \tag{1}$$

Where RC_i is compressed residual, Ri is input residual, P is pitch period. This compression, for every 2 pitch periods, is performed M_1 times from the beginning part of the frame. M_1 is defined by eq.(2).

$$M_1 = G[N/2P] \tag{2}$$

Where $G[x]$ is maximum integer value less than x, N is a frame length. At the ending part, $N/2 - M_1P$ samples are simply eliminated. Then input residual is shortened by $N/2$ samples totally.

(B)Compression (ratio 4:3): At the beginning part of the frame, compression using eq.(1) is performed M_2times ($M_2 = G[N/4P]$). And at the ending part, $N/4 - M_2P$ from ending sample is shortened by using eq.(1). Then input residual is shortened by $N/4$ samples totally.

(C)Expansion: Each sample compressed by eq.(1) is placed 2 times at one pitch period interval. And each eliminated sample

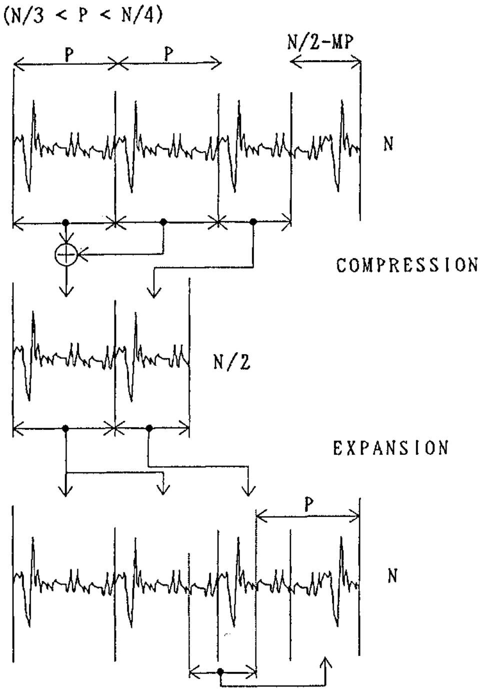

Figure 1. FS-TDC
(compression ratio 2:1)

is substituted by a sample behind one pitch period.

Fig.1 shows compression and expansion manner of FS-TDC in case of compression ratio 2:1.

III. APPLICATION OF FS-TDC TO TC-WVQ

3.1 Basic Structure

Since FS-TDC is performed on residual in a entire frame, TC-WVQ [2], which quantizes residual in a frame together, is considered to be adequate as a stochastic coder to which FS-TDC is applied. While CELP[1] does't treat residual to be quzantized and has sub-frame structure.

Block diagrame of our coding system applying FS-TDC to TC-WVQ is shown in Fig.2. In the encoding part, residual obtained from LPC inverse filter is compressed using FS-TDC. The compression ratio is controlled by compression control flag given by pitch analysis part where pitch periodicity in the frame is estimated. The compressed residual is pitch inverse filtered and then weighted vector quantized. This quantization process includes local decoding part explained in section 3.2.

In the decoding part, output residual from pitch synthesis filter is expanded using FS-TDC and LPC synthesis filtering is parformed on the expanded residual to obtain output speech.

3.2 Optimization of vector quantization for FS-TDC

If residual vector quantization is optimized for FS-TDC, distortion caused by FS-TDC itself will be reduced. For this purpuse, we bring a improved vector quantization process for the compressed residual that minimizes distortion between input and output speech . In the vector quantization process, several code vectors in a high rank are pre-selected as a candidates at the time compressed residual is vector quantized, and local decoding including time domain expansion is perfomed on them (Fig.2). Among the synthesized speeches from the local decoder, a speech which gives minimum distortion against input speech is chosen. And the code vectors that bring this minimum distortion speech are finally selected as outputs of the vector quantization process.

In practice, two code vectors are pre-selected in our coding system for each L sub-vectors to be quantized and 2^L kinds of speeches are synthesized from combination of these code vectors.

3.3 Normalization of residual

Pitch inverse filtering is performed on compressed residual in the encoding part(Fig.2). But pitch gain in the filter tends to be low because FS-TDC already eliminated pitch periodicity from the residual to certain degree. When the pitch gain is low, some sections in the residual from pitch inverse filter sometimes have very high amplitude. This high amplitude sections often appear a pitch synchronously in voiced frame. For vector quantization using gaussian codebook, these sections should be attenuated by proper normalization scheme beforehand.

Then, for a voiced frame, we bring a simple amplitude pattern that synchronizes with pitch pulse position (Fig.3). This pattern consists of two blocks, top and bottom. The pattern of a frame starts from a position S where last pitch pulse of the residual(decoded) in a previous frame exists. Amplitudes of top and bottom blocks, Rt and Rb, are obtained from mean amplitudes at each two blocks. They are normalized by mean amplitude of the frame R0 and quantized using small number of bits. On the other hand, an unvoiced frame is devided into first half block and second half block. At each block, amplitude is obtained by using the same manner as in the voiced frame. Residual is normalized by these simple amplitude patterns before vector quantization.

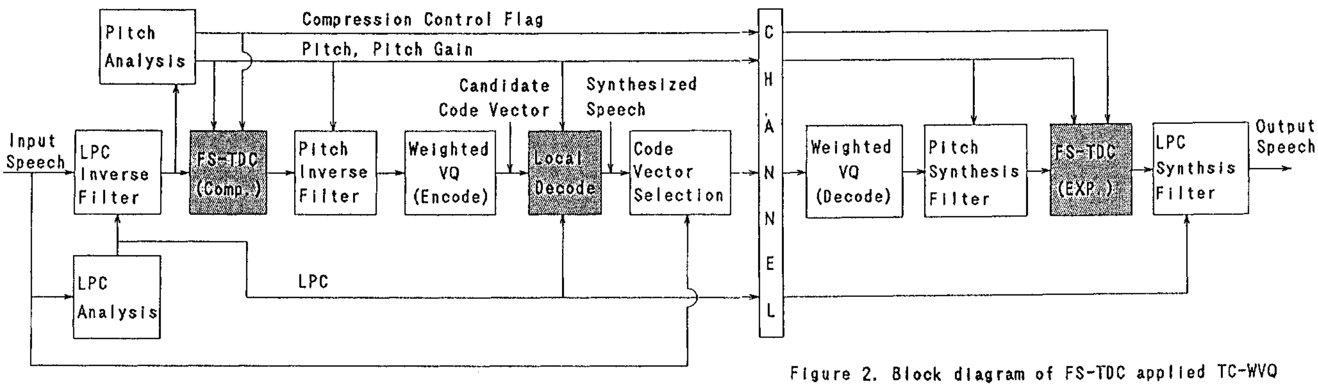

Figure 2. Block diagram of FS-TDC applied TC-WVQ

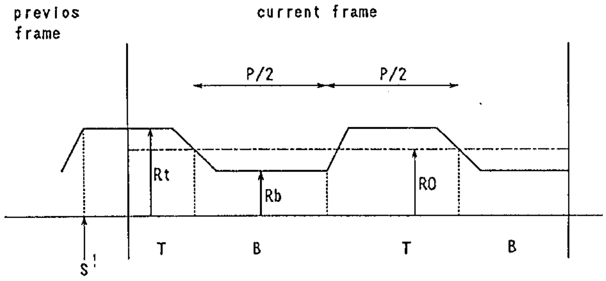

Figure 3. Amplitude pattern for regidual in voiced frame

IV. EVALUATION TESTS

4.1 Test conditions

To evaluate FS-TDC coding system, 3 types of coding systems are prepared as follows.

(1) Without time domain compression (No-TDC)

(2) FS-TDC

(3) FS-TDC with optimized vector quantization

Each system has 2 kinds of bit rate, 4.0kbps and 4.8kbps. All systems include the normalization scheme explained. Bit assignments of parameters are shown in Table 1 and number of bits assigned for one residual sample are shown in Table 2.

Analysis conditions are, sampling rate is 6.4kHz, frame length 20ms, LPC order 8. LPC parameters are transformed to LSPs and are vector-scaler quantized[5]. Pitch prediction is performed by 3-tap filter and a pitch gain is vector quantized [2]. Input speech for testing is spoken by Japanese 5 male and 5 female (14sec total).

4.2 Test results

Table 3. shows SNRseg and CD of 3 types of systems against input speech. TDC speech in Table 3 is a reference obtained by LPC analysis and synthesis process including FS-TDC compression and expansion without any quantizations of parameters. Although SNRseg is degraded, CD of TDC speech is good and its subjective quality is similar to that of input speech. And it does not have any noticable click sounds because the compression ratio is controlled.

SNR of the speech synthesized by FS-TDC against input speech tends to be low because of its large compression ratio. But at 4.0kbps, compared with No-TDC system, improvements on SNR by FS-TDC(with opt.VQ) is notable. In the SNR evaluation, the effect of FS-TDC is clearer at lower bit rate.

Optimization of vector quantization shows good results in Table 3. In practice, SNRseg and subjective quality is much improved by the optimized VQ especially in the frames where FS-TDC(without opt.VQ) gives bad quality. The normalization scheme reduces roughness in the sounds and improves subjective quality very much while its SNR improvement is negligible.

Table 4 shows SNRseg of FS-TDC(with opt.VQ) against TDC speech. SNRseg at both bit rates are high. This explains good speech quality of FS-TDC coding system at low bit rate.

Informal listenning test was done by 10 persons. Preference score of 4.0kbps FS-TDC(with opt.VQ) system is 100% against 4.0kbps No-TDC system and 95% against 4.8kbps No-TDC system.

Table 1. Bit assignment (bit/frame)

Bit rate	4.8kbps	4.0kbps
Power	7	7
LSP	24[25]	22[23]
Pitch Period	6	6
Pitch gain	4	4
Block Amplitude	4	4
C.C. Flag	1[0]	1[0]
Residual (Code Vector)	50 (5vectors)	36 (4vectors)

[ ]No-TDC

Table 2. Bit allocated per a compressed residual sample

TDC	No-TDC	FS-TDC	
Compression ratio	----	2:1	4:3
Number of Samples	128	64	96
Bit/Sample 4.8kbps 4.0kbps	 0.39 0.28	 0.78 0.56	 0.52 0.38

Table 3. Objective quality against input speech, SNRseg[dB], (CD)

TDC	No-TDC	FS-TDC	FS-TDC (Opt. VQ)
4.8kbps	10.6(3.9)	9.9(3.9)	10.9(3.8)
4.0kbps	8.8(4.0)	8.8(4.0)	9.6(3.8)
TDC speech	----	14.9(1.4)	----

Table 4. Objective quality against TDC speech, SNRseg[dB], (CD)

TDC	FS-TDC (Opt. VQ)
4.8kbps	12.5(3.5)
4.0kbps	10.6(3.6)

V. ROBUSTNESS TEST FOR CHANNEL ERROR

First, an error sensitivity of every parameter in the FS-TDC coding system is checked. The error sensitivity is defined by eq.(3)[6]

$$S_e = (D_1 - D_2)/D_0 \tag{3}$$

Where S_e is the error sensitivity of specified parameter bit, D_1 is mean squared error between input speech and output speech synthesized using a parameter whose specified bit always contains channel error, D_2 is mean squared error caused by quantization, D_0 is mean square of input speech.

The error sensitivity of some parameters are shown in Table 5. The error sensitivity of pitch period in FS-TDC system is 2dB greater than that of No-TDC, and compression control flag (C.C.flag) has also very high value. So, together with power and LSP, pitch period and C.C flag should be strongly protected by error correction codes (ECC).

Our error protection scheme is as follows.
(1)Bit errors on significant bits of power and LSP are detected by CRC (cyclic redundancy check). Once error is detected, these parameters are interpolated by parameters of the latest frame where no errors are detected.
(2)Significant bits of power, LSP, and picth period are protected by ECC (coding rate 1/2)
(3)C.C.flag is protected by majority logic (ML).

According to the above scheme, we made 5.6kbps speech coding systems for both FS-TDC and No-TDC which includes 4kbps source rate and 1.6kbps ECC rate. In these systems, BCH is used to protect significant bits of power, LSP, and pitch period.

Error pattern for the channel error test is obtained by fading channel simulator(fading pitch 40Hz). The test results are shown in Fig.3. FS-TDC system has higher SNR value at each error rates than the No-TDC system. On the subjective quality, FS-TDC system is better than No-TDC system at error rates 1% and 0.1%, while their quality is almost the same at 3%.

Table 5. Error sensitivity [dB]

Pitch-A(MSB)	-2.0
Pitch-B(MSB)	-5.4
Power(MSB)	1.7
LSP-VQ	-4.9
Resdual Code	-10.7
C.C.Flag	0.7

Pitch-A: FS-TDC
Pitch-B: No-TDC

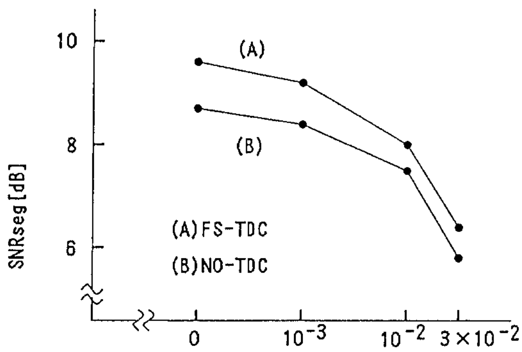

Figure 4. SNR for error channels

VI. APPLICATION OF FS-TDC TO CELP

CELP uses input speech as a reference to find a optimum code vector as speech excitation, and it does not treat residual as quantization object while TC-WVQ does. Then, to apply FS-TDC to CELP, we synthesize compressed input speech S' shown in Fig.5 as a reference. S' is directly synthesized from compressed residual and the optimum code vector for S', instead of for S, is selected by means of CELP coding.

We made 4.8kbps speech coding system applying FS-TDC to CELP. Its sampling rate is 8.0kHz, frame length is 20ms, number of sub-frames is 2. Quality of this coding system is being evaluated now. But in a simple listenning test, it shows better subjective quality than 4.8 kbps CELP coder whose frame length is 30ms and number of sub-frame is 4.

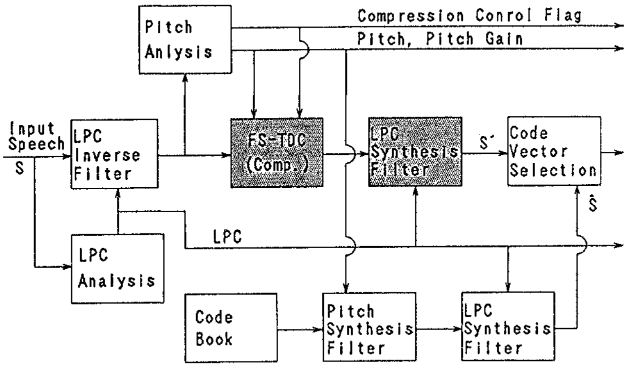

Figure 5. Block diagram of FS-TDC applied to CELP

VII. CONCLUSION

As a low bit rate speech coding algorithm for digital mobile communication, an time domain compression algorithm called FS-TDC which is optimally combined with stochastic coder is proposed.

First, FS-TDC is applied to TC-WVQ, and 4.0/4.8kbps coding systems are made. The optimized vector quantization process for FS-TDC and residual normalization scheme are introduced. In the evaluation test, speech coding system using FS-TDC shows better speech quality than No-TDC system. And in a robustness test for channel error, FS-TDC system also shows good results through the tested error rates (0.1%-3%) if parameters are protected optimimally. Finally, FS-TDC is applied to CELP and 4.8bps system is made. In a simple listenning test, FS-TDC system shows better speech quality than No-TDC system.

References

[1] M. R. Schroerder and B. S. Atal, "Code-excited linear prediction(CELP):high-quality speech at very low bit rates", *Proc. ICASSP*, pp. 937-941, 1985.

[2] T. Moriya and M. Honda, "Transform Coding of Speech Using a Weighted Vector Quantizer" *IEEE J. Selected Areas in Commun.*, vol. 6, pp. 425-431, 1988.

[3] S. Takahashi and K. Nakajima "Low bit rate Speech Coding Method using Time Domain Compression", in *Proc. Spring meeting, ASJ*, pp. 293-294, Mar. 1990.

[4] R. V. Cox et al, "An Implementation of Time Domain Harmonic Scaling with Application to Speech Coding" in *Proc. ICC*, pp. 4G.1.1-4, 1982.

[5] T. Moriya and M. Honda "Adaptive Transform Coding of Speech Using Vector Quantization" in *IEICE Trans.*, vol.J67-A pp. 974-978, 1984.

[6] H. Suda and T. Moriya, "Design and Evaluation of 8kbit/s Transform-coder for Land Mobile Radio" *IEICE Technical Report*, SP88-138, pp. 31-37, 1988.