Special Article on Mobile Multimedia Signal Processing Technologies

Speech Coding Technology

This article explains high-efficiency and high-quality speech compression/transmission techniques used as core technologies for basic digital mobile communications services. We will briefly explain the principles of speech compression algorithms for mobile communications such as PDC and IMT-2000 as well as speech compression algorithms standardized in the ITU-T for the public switched telephone network (PSTN). We will also describe technologies used to compensate transmission channel errors, which is the most critical problem in mobile communications environments.

Nobuhiko Naka, Tomoyuki Ohya, Masahito Saigusa, and Toyokazu Hama

1. Introduction

Speech coding technologies are used to digitally compress speech signals, and they help to improve signal transmission efficiency. Particularly for mobile communications, which uses a limited bandwidth, speech coding is essential for providing telephony services, hence DoCoMo also has been conducting research and development in this important technological area. Today, technologies developed for mobile communications, such as high-efficiency compression, complexity reduction, and error robustness, are vital in many areas including Internet Protocol (IP) networks.

2. Speech Coding for Mobile Communications

2.1 CELP Algorithm

Speech coding algorithms can be categorized into three types: the waveform coding, the vocoder and the hybrid coding. Waveform coding, such as Pulse Code Modulation (PCM) and Adaptive Differential Pulse Code Modulation (ADPCM), is designed to faithfully encode a signal waveform independent to signal characteristics. Therefore, it can achieve high coding quality as long as the bitrate is sufficiently high. When the bitrate decreases, however, the quality drops sharply. On the other hand, vocoders can achieve low bitrate by employing waveform analysis based on a human vocal model for coding, however as the speech quality depends largely on the vocal model, it is difficult to achieve high quality simply by increasing coding bitrate. Hybrid coding combines the waveform coding and the vocoder. It analyzes and encodes parameters based on a human vocal model, then it applies waveform coding to the residual information that could not be expressed by the model. CELP is one of the most widely used hybrid coding today, which is also used to achieve high-efficiency, high-quality speech coding in mobile applications.

Figure 1 shows a block diagram of a CELP encoder. The CELP encoder has an internal decoder. The CELP decoder consists of a linear prediction synthesis filter and two codebooks (an adaptive codebook and a fixed codebook), which generates excitation signal for the synthesis filter. The linear prediction synthesis filter represents a spectral envelope characteristic of the speech signal, and corresponds to the human vocal tract, while the excitation signal generated by codebooks corresponds to the exhalation from the lungs passing through the human glottis. In other words, CELP was developed using the human vocal system as a model.

Basic technologies used in CELP coding are as follows.

(1) Linear Prediction Analysis

As shown in Figure 2, the linear prediction analysis predicts current signal based on past input by using the temporal correlation of the speech signal. The difference between the predicted signal and the original signal is called linear prediction residual.

The CELP encoder analyses linear prediction coefficients α_i by, for example, Levinson Durvin Itakura method from computed autocorrelations of the speech signal.

Tenth order linear prediction is normally used for telephone band speech coding. Since it is difficult to discriminate the filter stability from linear prediction coefficients, lin-

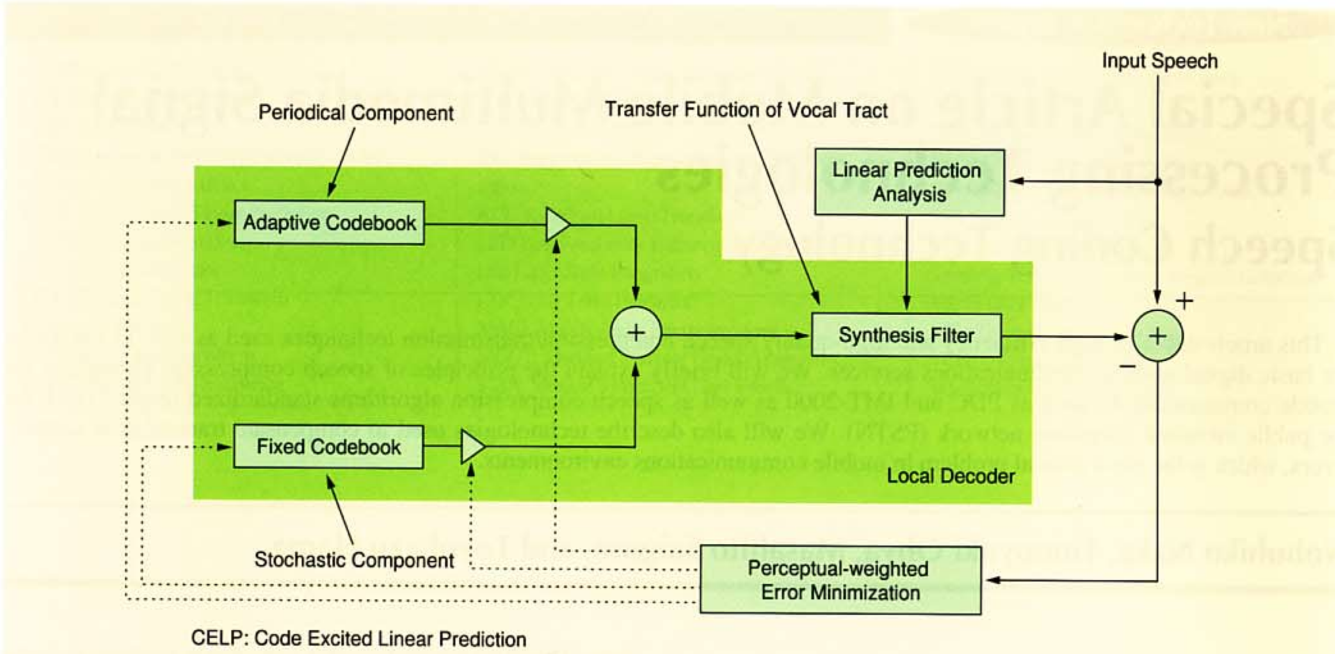


Figure 1 CELP Encoder

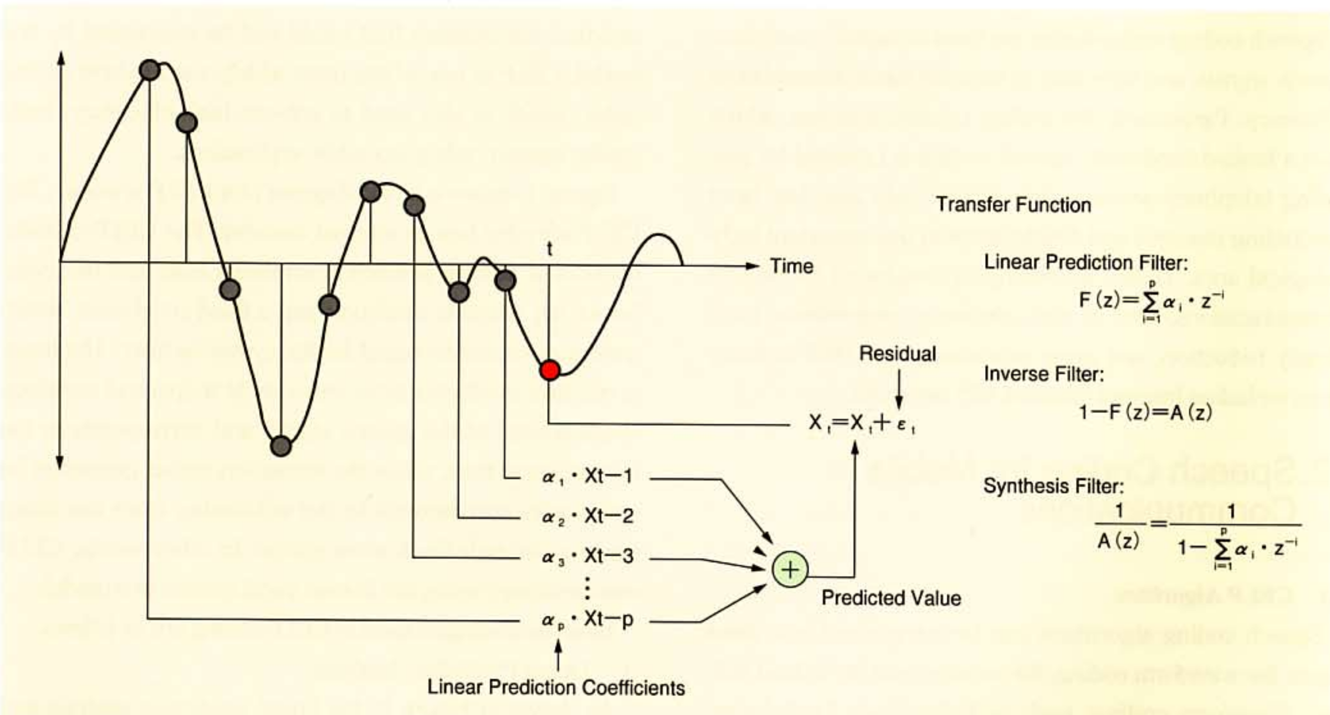


Figure 2 Linear Prediction Analysis

ear prediction coefficients are often converted to reflection coefficients, Line Spectrum Pairs (LSP) or others, which are mathematically equivalent and have guaranteed stability, before quantization and transmission. The decoder reconstructs a synthesis filter with transmitted α_i, and synthesizes the received speech signal from the residual signal as an excitation. The frequency response of the synthesis filter corresponds to the speech spectral envelope.

(2) Perceptual Weighting Filter

The CELP encoder searches the codebook patterns and gain parameters that minimize the quantization error between input speech signal and the synthesized speech signal obtained by the internal decoder. This technique is called Analysis-by-Synthesis (A-b-S) method, and it is one of the most distinctive features of the CELP coding.

The error calculation used in the A-b-S is based on a perceptual weighting filter that models a human auditory mechanism. The perceptual weighting filter is an Auto Regressive Moving Average (ARMA) type filter that uses the coefficients obtained in linear prediction analysis. The filter has vertically inverted frequency responses to the speech spectral envelope, which works to minimizes the quantization error in the section corresponding to the valleys of the spectrum, where the quantization error is relatively noticeable.

Although the use of un-quantized linear prediction coefficients as a perceptual weighting filter improves the quality, it in turn increases the computational complexity. For this reason, quantized linear prediction coefficients have been used in the past implementations to reduce computational complexity through mutual cancellation with a synthesis filter, however, today, the impulse response of the combined synthesis filter and perceptual weighting filter (called an perceptual weighted synthesis filter) is generally used for the computation.

(3) Adaptive Codebook

The adaptive codebook is a memory that stores past excitation, and the content is dynamically changed. In the case of voiced speech that consists of periodic signals, an adaptive codebook can efficiently represent the signal, because it generates the cyclic excitation corresponds to human vocal tone. The pitch lag is selected so as to minimize perceptual weighing error between input speech and the synthesis filter output of the adaptive codebook vector. The normal range for human voice pitch period of 16 to 144 samples is searched. In the case of shorter pitch lag, fractional value of pitch period is quantized by oversampling the speech signal to improve frequency resolution.

As the error minimization requires a huge amount of computation, only the limited range around the estimated pitch frequency derived from autocorrelation of input speech is searched precisely including oversampling resolution for reducing the computational complexity. Furthermore, searching the pitch lag around the past quantized pitch lag and quantizing its difference can also be useful to reduce both bitrates and computation.

(4) Fixed Codebook

The fixed codebook consists of various non-periodic signals, which cannot be represented by an adaptive codebook. In the past, the codebook usually consists of Gaussian noise or trained noise signals, however, today, an algebraic codebook that represents a residual signal with a few pulses is generally used. The new codebook can sharply reduce the memory requirements for fixed codebook storage and the complexity requirements including the orthogonalization to the adaptive code vector and the error minimization calculation.

(5) Post-filter

A post-filter is used in the last stage of the decoding to improve subjective quality by shaping the synthesized speech. A typical post-filter is an ARMA formant emphasis filter with frequency response opposite to perceptual weighting filter for reducing noticeable quantization error by suppressing valleys of the spectrum envelop. In addition, a post-filter is accompanied with a filter that compensates the spectral tilt of the output signal.

2.2 Peripheral Technologies for Mobile Applications

Various peripheral technologies are incorporated for coping with unique conditions of mobile communications, such as the use of wireless channel or the use in outdoor or in traveling. The following outlines these peripheral technologies.

(1) Error Correction

Error correcting codes are used to compensate radio transmission channel errors. A technique to use different level of error correcting capability codes according to the bit error sensitivity of speech bitstream, called Bit Selective-Forward Error Correction (BS-FEC) or Unequal Error Protection (UEP), is used for efficient error correction.

(2) Error Concealment

In case of the error correcting code mentioned above fails to correct errors or the loss of encoded information, there is no way to appropriate decoding with the received information. Then, an error concealment technique is employed to reduce subjective speech quality degradation to the minimum by generating speech signal of corrupted frames based on the interpolation from the parameters of the past-received speech information. The parameters with high temporal correlation, such as linear prediction coefficients, pitch lags and gains, are subject to be interpolated.

(3) Discontinuous Transmission

Discontinuous transmission is a technique to turn off the transmission or to reduce bitrate sharply during silent period. This technique is useful to extend the battery life of mobile terminals, reduce interference and achieve a statistical multiplexing effect. A VAD (Voice Activity Detector) detects the presence of speech signal using speech parameters. During silent period, comfort noise is generated based on the background noise information, which is quite small compared to normal speech information, in order to reduce impairments cased by discontinuous transmission.

(4) Noise Suppression

As the CELP algorithm is based on the model of the human vocal system, it is less efficient on the sounds other than natural voice, such as street noise. Noise suppression can be used to improve speech quality by reducing the amount of the sound, which is not necessary for conversation.

3. History of Speech Coding Technology

(From G.711 Recommendation to Pitch Synchronous Innovation-Code Excited Linear Prediction)

Table 1 shows the years when the major speech codecs for telephony services were standardized, and their bitrates. The International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) has standardized general speech codecs for telecommunications in an effort to ensure the quality internationally. On the other hand, regional standardization organizations, such as the Association of Radio Industries and Businesses (ARIB), the European Telecommunications Standards Institute (ETSI) and the Telecommunications Industry Association (TIA), have also independently developed another speech codecs for their digital mobile communications systems according to regional requirements. Bitrate of these codecs are lower than that of ITU-T codecs at the same period of time, since radio capacity is the most critical requirement. As shown in Table 1, ARIB, for example, adopted for the 6.7 kbit/s (for speech coding bitrate only) full-rate codec as early as 1990, then selected the 3.45 kbit/s half-rate codec in 1993. For above and other speech codecs for mobile communications, not only speech coding algorithm itself but also error correction, discontinuous transmission and other mobile-oriented peripheral technologies were standardized at the same time. As such, technologies for compression and transmission of speech signal have advanced significantly in the mobile communications field. In recent years, ITU-T began considering mobile applications as well as multimedia applications for its standardization and then G.729, which was also adopted to PDC 8 kbit/s enhanced full rate codec, was standardized in 1996.

The following section outlines typical speech codecs for telecommunications applications. Table 2 lists the specifications of speech codecs mentioned in this article.

3.1 G.711 [1]

G.711 is the most basic speech codec for telecommunications applications, and it is adopted in PSTN. G.711 quantizes

Table 1 Major Speech Codecs

10-Year Period	ITU-T (Coding Standard for Fixed Networks)	Coding Standard For PDC (ARIB)	GSM (ETSI)	IMT-2000 (3GPP)	D-AMPS	CDMA-one/3GPP2	ISO/IEC
1970s	1972 G.711 (64 kbit/s PCM)						
1980s	1984 G.726 32 kbit/s ADPCM		1988 RPE-LTP 22.8 kbit/s		1988 VSELP 13 kbit/s		
1990s	1992 G.728 16 kbit/s LD-CELP 1996 G.729 8 kbit/s CS-ACELP 1996 G.723.1 5.3 kbit/s ACELP & 6.3 kbit/s MPC-MLQ	1990 VSELP 11.2 kbit/s 1993 PSI-CELP 5.6 kbit/s 1998 CS-ACELP, ACELP 11.2 kbit/s	1994 EVSELP 11.4 kbit/s 1996 EFR (ACELP) 22.8 kbit/s 1998 AMR (4.75, 5.15, 6.7, 7.4, 7.95, 10.2, 12.2 kbit/s)	1999 AMR	1996 EFR/ ACELP 13 kbit/s	1994 QCELP (8 k, 4 k, 0.8 kbit/s) 1997 EVRC (8 k, 4 k, 0.8 k)	1999 MPEG-4 CELP
2000-	G.4 k (Not Finalized)						

ACELP: Algebraic Code Excited Linear Prediction
AMR: Adaptive Multi Rate
ARIB: Association of Radio Industries and Businesses
CELP: Code Excited Linear Prediction
CS-ACELP: Conjugate Structure-ACELP
D-AMPS: Digital Advanced Mobile Phone Services
EFR: Enhanced Full Rate
ETSI: European Telecommunications Standards Institute
GSM: Global System for Mobile Communications
IEC: International Electrotechnical Commission
IMT-2000: International Mobile Telecommunications-2000
ISO: International Organization for Standardization
ITU-T: International Telecommunication Union-Telecommunication Standardization Sector
MLQ: Maximum Likelihood Quantization
MPC: Multi Pulse Coding
MPEG: Moving Picture Experts Group
PDC: Personal Digital Cellular
PSI-CELP: Pitch Synchronous Innovation-Code Excited Linear Prediction
QCELP: Qualcomm CELP
VSELP: Vector Sum Excited Linear Predictive Coding
3GPP: 3rd Generation Partnership Project

Table 2 Specifications of Speech Codecs

		ITU-T		PDC			IMT-2000
		G.711	G.726	Full Rate	Half Rate	Enhanced Full Rate	AMR
Algorithm		PCM	ADPCM	VSELP	PSI-CELP	CS-ACELP	MR-ACELP
Bitrate	Total	64 kbit/s	32 kbit/s	112.0kbit/s	5.6 kbit/s	11.2 kbit/s	12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15, 4.75 kbit/s
	Speech	64 kbit/s	32 kbit/s	6.7 kbit/s	3.45 kbit/s	8.0 kbit/s	12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15, 4.75 kbit/s
	FEC	—	—	4.5 kbit/s	2.15 kbit/s	3.2 kbit/s	—
Complexity		1 MIPS or Lower	2 MIPS or Lower	7.8 MOPS	18.7 MOPS	Approx. 20 MIPS	Approx. 15 wMOPS
Algorithmic Delay		0.125 ms	0.125 ms	28.125 ms	45 ms	25 ms	25 ms
Frame Length		—	—	20 ms	40 ms	20 ms (10 ms for Speech)	20 ms
Other Features		—	—	VAD/DTX	VAD/DTX Noise Canceller	VAD/DTX G.729 for Speech Coding	Multi Rate Comfort Noise Generation VAD/DTX

ADPCM: Adaptive Differential Pulse Code Modulation
AMR: Adaptive Multi Rate
CS-ACELP: Conjugate Structure-ACELP
DTX: Discontinuous Transmission
IMT-2000: International Mobile Telecommunications-2000
ITU: International Telecommunication Union
MIPS: Million Instructions Per Second
MOPS: Million Operations Per Second
MR-ACELP: Multi Rate-ACELP
PCM: Pulse Code Modulation
PDC: Personal Digital Cellular
PSI-CELP: Pitch Synchronous Innovation-Code Excited Linear Prediction
VAD: Voice Activity Detector
VSELP: Vector Sum Excited Linear Predictive Coding

each speech sample by 8 bits with non-linear quantizaion steps and bitrate is 64 kbit/s. The G.711 specifies two kinds of law for quantizaion steps: A-law and μ-law. In Japan and North America, μ-law is used; hence, μ-law PCM is sometimes used to refer to the G.711.

3.2 G.726 [2]

The speech codec standardized after G.711 is present G.726 (ADPCM). G.726 is a kind of waveform coding and adopts adaptive predictor and adaptive quantizer for high coding efficiency. At 32 kbit/s, which is used generally including PHS, each speech sample (prediction error) is quantized by 4 bits and bitrates of 16, 24 and 40 kbit/s are also specified corresponds to the number of quantization bits per sample. The G.726 at 32 kbit/s is used as a "toll quality" reference standard, which is also a quality requirement of codec for IMT-2000.

3.3 G.729 [3]

The G.729, Conjugate Structure-Algebraic Code Excited Linear Prediction (CS-ACELP), is an 8 kbit/s speech codec standardized by ITU-T after G.728. LSP coefficients are quantized by multi-stage VQ with prediction. The excitation signal is represented by algebraic codebook consists of 4 non-zero pulses. The gain of the algebraic codebook is predicted from past subframes, and its correction factor is quantized. Therefore, there is no need to transmit the frame power. A parity bit for pitch lag and a conjugate-structure gain codebook are adopted for error robustness since the scope of its application includes not only PSTN but also mobile communications from the beginning of the standardization. Because of its toll quality, G.729 is used in speech coding sections of the PDC enhanced full rate codec as well as in VoIP (Voice over IP) applications. G.729 has several annexes including reduced complexity version, VAD, a high-bitrate version, and a low-bitrate version.

3.4 PDC Full-Rate Codec [4]

Following the European and North American digital cellular systems, Japan adopted the PDC full-rate system based on 6.7 kbit/s (for speech coding bitrate only) Vector Sum Excited Linear Predictive Coding (VSELP). VSELP is the algorithm based on CELP, and it utilizes an orthogonal search of the excitation and a simplified perceptual weighted synthesis filter to reduce computational complexity, and vector-sum-structure fixed codebook to achieve error robustness. Therefore, North American Time Division Multiple Access (TDMA) full-rate system and the Global System for Mobile Communications (GSM) half-rate system adopts similar algorithm. In recent years, however, VSELP is being replaced by codecs that have same bitrate but offer higher quality, such as G.729.

3.5 PDC Half-Rate Codec [4]

To cope with the rapid increase of the mobile subscribers, Pitch Synchronous Innovation-Code Excited Linear Prediction (PSI-CELP) was adopted as a PDC half-rate codec. This algorithm was proposed to ARIB by DoCoMo, and competed other candidates in the selection test. PSI-CELP is also the algorithm based on CELP, and achieves a high compression rate of 3.45 kbit/s while maintaining the real-time calculation. It is realized by several advanced technologies such as matrix quantization of the LSP coefficients, pre-selection, fractional pitch analysis, pitch synchronization of the stochastic codebook vectors, conjugate-structure of stochastic codebook, simultaneous optimization of the gains, and delayed decision.

4. PDC Enhanced Full Rate Codec [4]

4.1 Standardization

In addition to the full rate (VSELP) and half rate (PSI-CELP) speech codecs, ARIB standardized the PDC enhanced full rate codec in RCR STD-27H [4] in 1998 as an optional speech codec. The requirement for this optional codec was that the quality of the codec should be higher than or at least equivalent to that of VSELP and PSI-CELP.

4.2 Features of the PDC Enhanced Full Rate Codec

The PDC enhanced full rate codec, which was developed for the purpose of improving speech quality in PDC, employs G.729 for speech coding part as well as DoCoMo's original error correction and concealment technologies. Therefore it achieves better quality at 11.2 kbit/s, which is the same bitrate for speech transmission of former full rate codec (VSELP).

Although speech bitstream from G.729 is produced every 10 ms, channel coding of PDC enhanced full rate is performed every 20 ms in order to fit it into the air interface of PDC and produces the full rate signal of 11.2 kbit/s for transmission. Figure 3 shows the block diagram of the channel coding used for the PDC enhanced full rate codec.

(1) Protected Bit Selection

From an 8 kbit/s G.729 encoded bitstream, the bits to be protected and those that are not to be protected are classified according to error sensitivity. By using 2 frames of the classified G.729 bitstream, a 20ms (160 bits) frame is constructed for channel coding. This is then coded by a convolutional coder with r=1/2 and m=6. Furthermore, the output from the convolutional coder is applied with puncture and 2 slot interleave to produce a 224-bit/20-ms frame.

(2) Channel Decoding

In channel decoding, symbols received are processed by convolutional decoder after de-interleaving. The Viterbi algorithm is used for decoding convolutional codes. It is possible to improve the performance of error correction by using a soft decision decoding with the reliability information of each bit and the envelope information of each symbol for distance metric calculation. After the error correction, CRC (Cyclic Redundancy Check) is computed from the decoded bitstream in the same way at the channel encoder side. If it is different from the received one, the received frame is determined as an erased frame, for which concealment of frame erasures is to be applied.

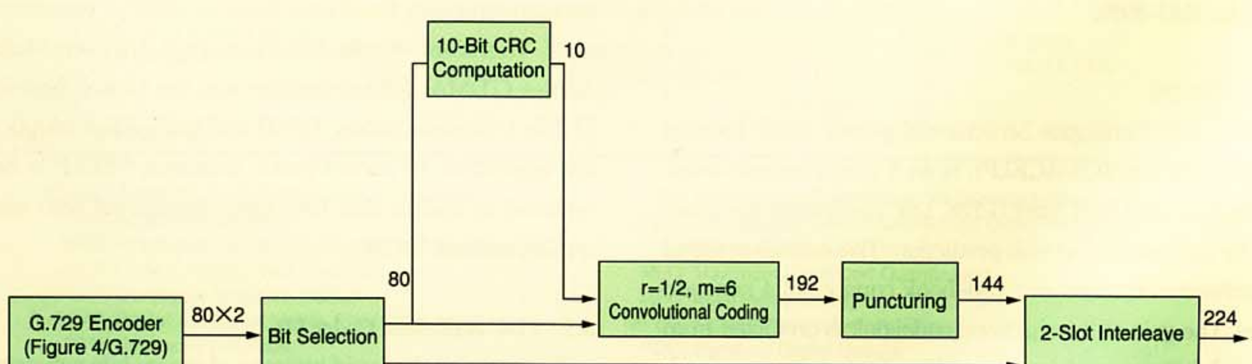
4.3 Concealment of Frame Erasure

In addition to the method defined in G.729, PDC enhanced full rate codec employs following methods optimized for PDC environment for improved speech quality.

(1) Adaptive Pre-filter Control

To prevent the generation of annoying sounds in an erased frame, pitch emphasis processing is omitted for loss compensation in case of burst error.

(2) Reconstruction of the Adaptive Codebook



CRC: Cyclic Redundancy Check

Figure 3 Channel Coding for the PDC Enhanced Full Rate Codec

On an error-recovered frame, the past excitations are reconstructed using pitch lag of the current frame for adaptive codebook reconstruction. Parameters other than pitch lag, such as gains and fixed codebook vectors, used in this procedure are identical to the past ones. This procedure is applied prior to producing adaptive codebook vector of the current frame.

(3) Gain Control of the Excitation

On an error recovered frame, both the adaptive and the fixed codebook gains are controlled to improve the decoded speech quality.

In addition to the higher quality of PDC enhanced full rate codec under error free condition, these error concealment techniques give equivalent or high quality of PDC enhanced full rate codec under error conditions compared with VSELP and PSI-CELP in spite of the fewer redundancy for error correction.

5. IMT-2000 Speech Coding AMR [5]

5.1 Standardization

IMT-2000 Committee established by ARIB in Japan in 1997 has started standardization for 3rd generation mobile communications systems in advance of all over the world. Selection of the speech codec for IMT-2000 had been performed by CODEC (Corder Decorder) Work Group in the IMT-2000 Committee. Since various member companies proposed different speech coding technologies, the Working Group draw up a test plan for assessing performances of the proposed speech codecs and conducted selection tests. Before the selection tests were completed, 3rd Generation Partnership Project (3GPP) was established by ARIB, TTC (Telecommunications Technology Committee), TIA and ETSI at the end of 1998. Then, it was agreed that 3GPP TSG-SA WG4 (CODEC) selected the most appropriate speech codec for IMT-2000 based on ARIB's selection test results. Evaluation test results, to which DoCoMo also contributed, showed superiority of AMR to other candidates and AMR was approved by 3GPP as a mandatory speech codec for IMT-2000 telephony service.

5.2 Algorithm Outline

AMR is a multi-rate speech codec based on ACELP, and was standardized for GSM in 1998. It has eight coding modes ranging from 12.2 kbit/s to 4.75 kbit/s. The 12.2, 7.4 and 6.7 kbit/s modes use common algorithms already standardized as regional standards.

Although the algorithm is similar to that of G.729, innovative technologies were incorporated to achieve the multi rate specification. The frame length was 20 ms in all modes, while the multi rate is accomplished by varying the number of subframes and the number of quantization bit (see Table 3).

In 12.2 kbit/s mode, LPC analysis is preformed twice per frame, and 2 sets of LSP coefficients are divided from the lowest order and 5 sets of 2×2 matrix are constructed for prediction. Then each residual of the matrix is vecoter-quantized. In other modes, LPC analysis is performed once per frame, and, after the prediction in LSP domain, the residual is divided from the lowest order and vector-quantized.

For long-term prediction, 1/6 fractional pitch search is performed in 12.2 kbit/s and 1/3 fractional pitch search in the other modes. Pitch lags are differential-quantized within a frame.

The algebraic codebook consists of 2 to 10 non-zero pulses with amplitude of 1. A pitch pre-filter, which provides the similar effect as PSI, is applied at algebraic codebook search. The adaptive and fixed codebook gains are quantized separately in the 12.2 and 7.95 kbit/s modes, but are vector-quantized in the other modes. At the decoder, decoded speech signal is obtained from the synthesized speech signal through a formant post-filter and tilt compensation filter.

AMR also includes specifications of peripheral technologies necessary for mobile communications. Two VAD algorithms necessary for discontinuous transmission are specified. Background noise information called SID (Silence Insertion Descriptor) consists of short-term prediction coefficients and frame power is quantized by 35 bits, and it is transmitted at equal intervals. Requirements for concealment of lost frames are specified and example solutions described there utilize state machines depends on error conditions for compensation of coding parameters, such as codebook gains and short-term filter coefficients.

In addition to above, classification of AMR bits is specified according to bit error sensitivity so that Unequal Error Protection (UEP) can be applied to AMR bitstreams by IMT-2000 RAN (Radio Access Network) specified as a toolbox for flexible design. ISG (IMT-2000 Steering Group) designates detail radio parameters for transmission of AMR based on this classification.

5.3 Quality

Figure 4 shows the subjective test result of AMR, which was conducted by DoCoMo according to ARIB's selection test plan and was submitted to 3GPP. The result shown in

Table 3 AMR Bit Allocation

(Excerpt from TS 26.090[5])

Mode	Parameter	Subframe #1	Subframe #2	Subframe #3	Subframe #4	Total in a Frame
	LSP×2					38
	Pitch Lag	9	6	9	6	30
12.2 kbit/s	Pitch Gain	4	4	4	4	16
	Algebraic Code	35	35	35	35	140
	Codebook Gain	5	5	5	5	20
	Total					244
	LSP					26
	Pitch Lag	8	5	8	5	26
10.2 kbit/s	Algebraic Code	31	31	31	31	124
	Gain	7	7	7	7	28
	Total					204
	LSP					27
	Pitch Lag	8	6	8	6	28
7.95 kbit/s	Pitch Gain	4	4	4	4	16
	Algebraic Code	17	17	17	17	68
	Codebook Gain	5	5	5	5	20
	Total					159
	LSP					26
	Pitch Lag	8	5	8	5	26
7.40 kbit/s	Algebraic Code	17	17	17	17	68
	Gain	7	7	7	7	28
	Total					148
	LSP					26
	Pitch Lag	8	4	8	4	24
6.70 kbit/s	Algebraic Code	14	14	14	14	56
	Gain	7	7	7	7	28
	Total					134
	LSP					26
	Pitch Lag	8	4	8	4	24
5.90 kbit/s	Algebraic Code	11	11	11	11	44
	Gain	6	6	6	6	24
	Total					118
	LSP					23
	Pitch Lag	8	4	4	4	20
5.15 kbit/s	Algebraic Code	9	9	9	9	36
	Gain	6	6	6	6	24
	Total					103
	LSP					23
	Pitch Lag	8	4	4	4	20
4.75 kbit/s	Algebraic Code	9	9	9	9	36
	Gain	8		8		16
	Total					95

LSP: Line Spectrum Pair

Figure 4 was carried out under the error condition of the Wideband Code Division Multiple Access (W-CDMA) with BER (Bit Error Rate) of 0.1% (Note that the test conditions used here differs from that of current specification). The result shows that the quality of AMR at 12.2 kbit/s is better than that of all the other codecs. In addition, compared with other codecs at the same bitrate quality of AMR is better or at least equivalent.

Furthermore, 3GPP is presently evaluating the quality of AMR over W-CDMA channels using the error patterns provided by DoCoMo and Nortel Networks, besides the quality of AMR under other conditions already reported in 3GPP TR 26.975[6]. AMR also satisfies the 80% clarity level in the articulation test conducted by DoCoMo.

5.4 Non-telephony Applications

Because of its flexibility and high quality, AMR was adopted as a mandatory audio codec for 3G-324M[7], i.e. codecs for circuit switched multimeida telephony services of 3GPP. For IP applications like VoIP, IETF (Internet Engineering Task Force) is working on the standardization of the RTP (Real-time Transport Protocol) payload format for AMR.

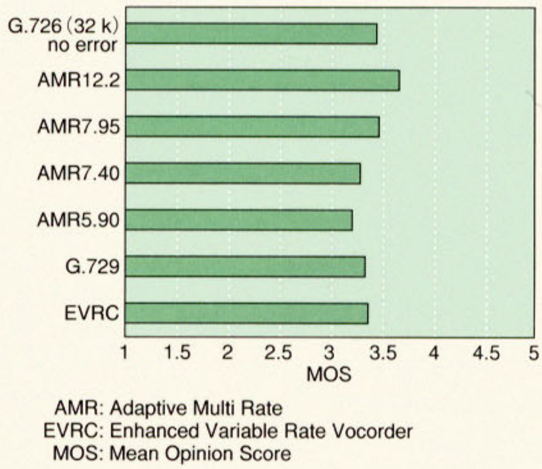


Figure 4 Results of AMR Subjective Evaluation

Consequently, AMR is being considered as a widely-used speech codec beyond the IMT-2000 telephony service.

6. Future Prospects

Presently, 3GPP plans to select AMR-WB, a wideband version (up to 7 kHz) of AMR, in October 2000, and the chosen algorithm is scheduled to be a candidate for ITU-T's low-bitrate wideband speech coding system. ITU-T is also promoting the standardization of a 4 kbit/s toll quality speech coding.

Meanwhile, many organizations are now discussing about speech coding applications for IP-based telecommunications networks, such as VoIP that should provide equivalent telephony services to current circuit switched networks, and streaming services as well.

3GPP is leading the standardization effort for IP-based mobile telecommunications networks in cooperation with those who are standardizing VoIP: ETSI TIPHON (Telecommunication and Internet Protocol Harmonization Over Networks) Project, IETF IPTEL (IP Telephony) and AVT (Audio Video Transport).

7. Conclusion

This article focused on speech coding technologies for mobile communications. Other speech coding algorithms not mentioned in the article include the ITU-T G.723.1 for multimedia systems, EVRC for cdmaOne, and MPEG-4 (Moving Picture Experts Group-4) standardized by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). These speech coding technologies are explained in detail in "Speech Coding" [8] published by the Institute of Electronics, Information and Communication Engineers, which was also used for reference in writing this article.

The speech coding technologies explained in the article are also finding applications in a variety of areas other than the mobile communications field, such as in the Internet. We plan to continue our research and development on wide applications of speech coding technologies capable of offering higher quality and superior flexibility.

References

[1] ITU-T Recommendation G.711, "Pulse code modulation (PCM) of voice frequency," November 1998.
[2] ITU-T Recommendation G.726, "40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)," December 1990.
[3] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction," March 1996.
[4] ARIB STD-27H.
[5] 3GPP 3G TS 26.090, "AMR Speech Codec; Transcoding functions," December 1999.
[6] 3GPP 3G TR 26.975, "Performance charactarization of the AMR speech codec," January 2000.
[7] 3GPP 3G TS 26.111, "Codec for Circuit Switched Multimedia Telephony Services; Modifications to H.324," August 1999.
[8] Takehiro Moriya, "Speech Coding," The Institute of Electronics, Information and Communication Engineers, October 1998.

Glossary

A-b-S: Analysis-by-Synthesis
ACELP: Algebraic Code Excited Linear Prediction
ADPCM: Adaptive Differential Pulse Code Modulation
AMR: Adaptive Multi Rate
ARIB: Association of Radio Industries and Businesses
ARMA: Auto Regressive Moving Average
AVT: Audio Video Transport
BER: Bit Error Rate
BS-FEC: Bit Selective-Forward Error Correction
CELP: Code Excited Linear Prediction
CODEC: Corder Decorder
CRC: Cyclic Redundancy Check
CS-ACELP: Conjugate Structure-ACELP
D-AMPS: Digital Advanced Mobile Phone Services
DTX: Discontinuous Transmission
EFR: Enhanced Full Rate
ETSI: European Telecommunications Standards Institute
EVRC: Enhanced Variable Rate Vocorder
GSM: Global System for Mobile Communications
IEC: International Electrotechnical Commission
IETF: Internet Engineering Task Force
IMT-2000: International Mobile Telecommunications-2000
IP: nternet Protocol
IPTEL: IP Telephony
ISG: IMT-2000 Steering Group
ISO: International Organization for Standardization
ITU: International Telecommunication Union
ITU-T: International Telecommunication Union-Telecommunication Standardization Sector
LSP: Line Spectrum Pair
MIPS: Million Instructions Per Second
MLQ: Maximum Likelihood Quantization
MOPS: Million Operations Per Second
MOS: Mean Opinion Score
MPC: Multi Pulse Coding
MPEG: Moving Picture Experts Group
MR-ACELP: Multi Rate-ACELP
PCM: Pulse Code Modulation
PDC: Personal Digital Cellular
PSI-CELP: Pitch Synchronous Inoveation-Code Excited Linear Prediction
PSTN: Public Switched Telephone Network
RAN: Radio Access Network
RCR: Research and Development Center for Radio System
RTP: Real-time Transport Protocol
QCELP: Qualcomm CELP
SID: Silence Insertion Descriptor
TDMA: Time Division Multiple Access
TIA: Telecommunications Industry Association
TIPHON: Telecommunication and Internet Protocol Harmonization Over Networks
TTC: The Telecommunications Technology Committee
UEP: Unequal Error Protection
VAD: Voice Activity Detector
VoIP: Voice over IP
VSELP: Vector Sum Excited Linear Predictive Coding
W-CDMA: Wideband Code Division Multiple Access
3GPP: 3rd Generation Partnership Project