T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

# 論文 / 著書情報
# Article / Book Information

| | |
|---|---|
| Title | Recent Developments in Audio Coding Technology |
| Author | Sadaoki Furui |
| Journal/Book name | MPEG-4 and Virtual Reality Coding Symposium, , , |
| Issue date | 1995, 7 |

Powered by T2R2 (Science Tokyo Research Repository)

# RECENT DEVELOPMENTS IN AUDIO CODING TECHNOLOGY

*Sadaoki Furui*
NTT Human Interface Laboratories
and
Tokyo Institute of Technology

## ABSTRACT

Speech and audio coding technologies are expected to play important roles in future multimedia services. This paper reviews recent speech and audio coding technologies, focusing on the PSI-CELP, CS-CELP, and Twin VQ methods proposed by the NTT Laboratories for coding at bit-rates of less than 64 kbit/s. This paper also discusses future speech and audio coding technologies, including very low bit-rate speech coding at 200 bit/s and intelligent speech coding using speech recognition techniques.

## INTRODUCTION

Over the last few years, the demand for speech and audio coding has become very high, especially for digital cellular communication systems (mobile and portable telephones). Since there are tradeoffs between coded quality, bit rate, coding delay, and complexity in coding technology, it is necessary to investigate various methods to determine which ones best meet the requirements for specific applications. Speech and audio coding methods can be classified into waveform coding, analysis-synthesis, and hybrid methods. Waveform coding aims at reproducing speech waveforms, whereas analysis-synthesis aims at reproducing the frequency spectrum by taking the human-speech production mechanism into account. Most of the recent speech and audio coding methods are based on hybrid coding, which combines the advantages of waveform coding and analysis-synthesis.

Progress in reducing the transmission rate for speech coding is shown in Fig. 1 [1]. The horizontal axis shows the year in which each coding method was developed or approved as a standard. Figure 2 classifies speech and audio coding techniques according to bandwidth, bit rate, and standardization organization [2]. Figure 3 graphically shows the transmission rate and subjective quality of the coded speech/audio signals of the major coding methods. Table 1 compares their characteristics [4].

## CELP CODERS

Most or the newer low-bit-rate coders are CELP (code-excited linear prediction) coders. In both LPC analysis-synthesis (vocoder) and CELP, the speech waveform is generated by a linear prediction synthesis filter with an excitation source filter. The vocoder models the excitation source either as a pulse sequence of the pitch period or as a fixed noise sequence. In CELP, however, a codebook is used and the best excitation vector is selected so that the perceptually weighted distortion between the input and the synthesized speech is minimized. In addition, a periodic signal is generated by repeating the excitation signal in the previous frame, and a non-periodic signal is selected from a random excitation codebook. Naturalness and robustness are significantly improved by CELP.

Several years ago, extensive efforts to develop new coding algorithms and error correction techniques, as well as progress in LSI technology, enabled the development of 8-kbit/s codecs (coder/decoders) for digital mobile telephones. In 1990, the VSELP (vector sum excitation linear prediction) codec was selected as the full-rate speech codec standard for digital cellular systems in North America and Japan. VSELP is a CELP coding method developed by Motorola; in it the excitation source is generated by summing several fixed-basis vectors.

Digital cellular systems are expected to run out of radio channels in Japan if demand continues to grow at the current rate. NTT has thus been extensively researching ways to expand capacity. This research led to the development of PSI-CELP (pitch synchronous innovation CELP) speech coding; PSI-CELP has quality comparable to that of the current standard (VSELP), while using only half the bit rate. PSI-CELP was chosen as the half-rate codec standard for digital cellular systems in Japan. Table 2 compares the standardized speech coding systems for digital cellular mobile radio in Japan, the U.S. and Europe [3]. Figure 4 shows the synthesizer structures for the LPC vocoder, CELP, VSELP, and PSI-CELP [1].

## PSI-CELP, CS-CELP, AND TWIN VQ METHODS

One of the most important features of the PSI-CELP speech coder is that the random excitation vectors are given pitch periodicity for voiced speech by a pitch synchronization process. This feature is the origin of the name PSI-CELP. It significantly reduces the distortion and improves the quality for voiced speech, especially female voices. In addition to the pitch synchronization, a two-channel conjugate structure and fixed codebook for transient speech signals were newly developed. These two features reduce memory requirements, enhance robustness against channel errors, and improve the quality for transient parts.

The CS-CELP and Twin VQ methods were recently proposed by NTT Laboratories for speech and audio coding at bit rates of less than 64 kbit/s. CS-CELP at 8 kbit/s achieves coded speech quality equivalent to that of 32-kbit/s ADPCM with much shorter coding delay than with PSI-CELP. CS-CELP is being used in NTT's personal handy phone system. It is expected to be adopted as the standard for 8-kbit/s speech coding by the ITU. Twin VQ at 16-64 kbit/s achieves coded audio (music) quality equivalent to that of a compact disc.

## FUTURE PROSPECTS FOR CODING TECHNOLOGY

Very low bit-rate coding technology will be useful not only for digital cellular systems, but also for various applications in multimedia communications. Phono-code, a speech coding method using only 200 bit/s, has been proposed by NTT. This method is based on segment (matrix) quantization, as illustrated in Fig. 5 [4]. As indicated in Fig. 3, while the naturalness of its coded speech is still not satisfactory, it is already fairly intelligible.

The ultimate speech coding method is expected to use not only perceptual constraints but also articulatory constraints and speech recognition technology. Figure 6 illustrates the relationships between the principal speech recognition, synthesis, and coding technologies [4]. The higher the level, the more abstract the information. Some of the technologies indicated in the figure remain to be investigated. Intelligent speech coding using speech recognition techniques will someday reach a processing speed of 30 bit/s, the bit rate of natural language calculated by Shannon. To achieve this level, it is crucial to develop speech coding systems capable of automatically adapting to new speakers and environmental conditions.

## REFERENCES


[1] T. Moriya, T. Kaneko, and N. Kitawaki: "PSI-CELP speech coder for mobile communications", NTT Review, 6, 1, pp. 97-100 (1994).

[2] N. Kitawaki: "Multimedia coding technology", NTT Review, 7, 3, pp. 74-80 (1995).

[3] N. Kitawaki and H. Mishima: "PSI-CELP speech coding doubles the capacity of radio waves", NTT Review, 6, 6, pp. 46-52 (1994).

[4] S. Furui: "Digital speech processing, synthesis, and recognition", Marcel Dekker, Inc., New York (1989).

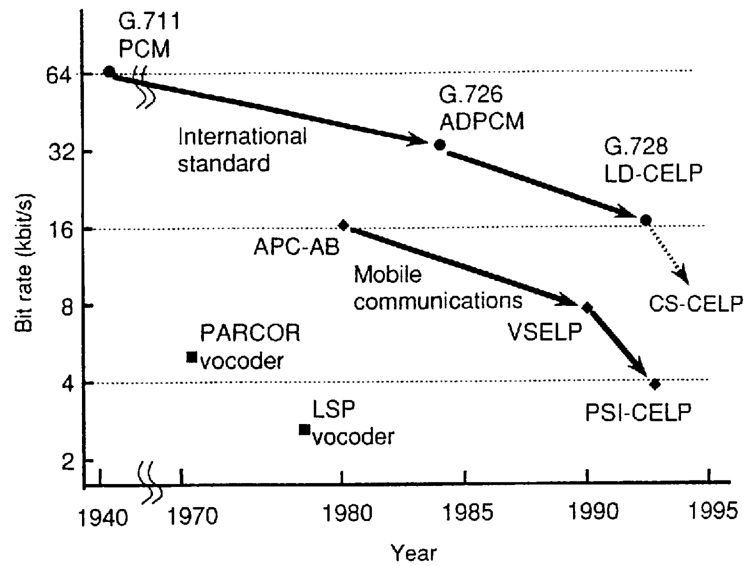



Fig. 1 Progress in reducing the transmission rate for speech coding.

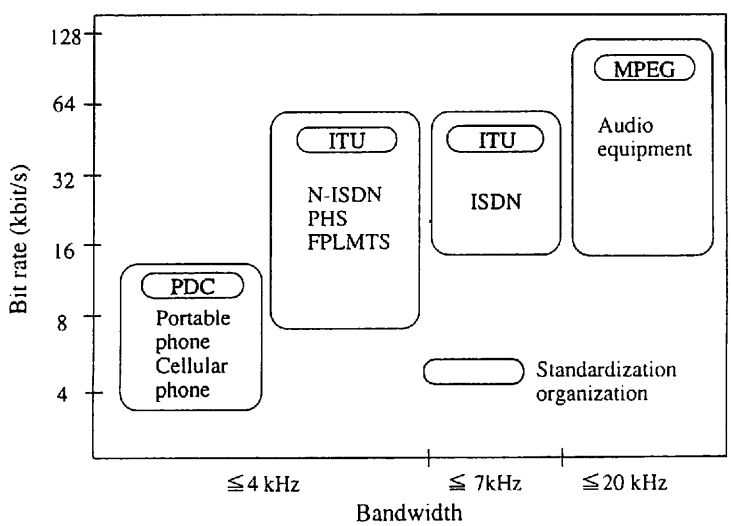



Fig. 2 Speech and audio coding techniques and standardization.

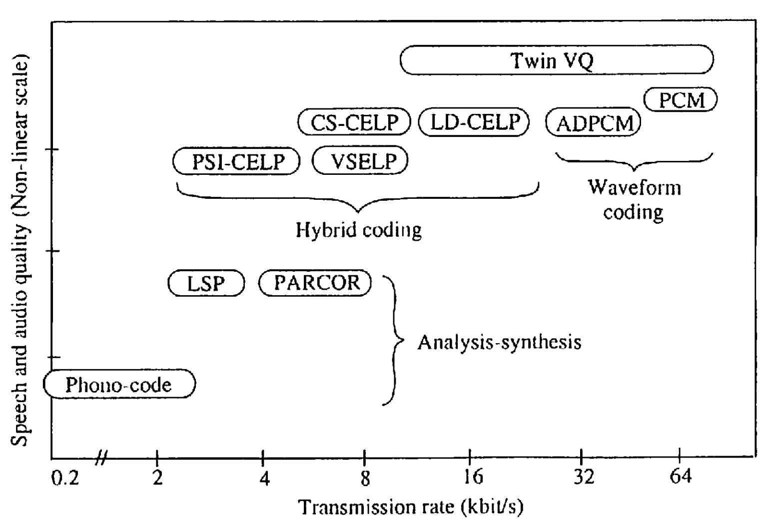


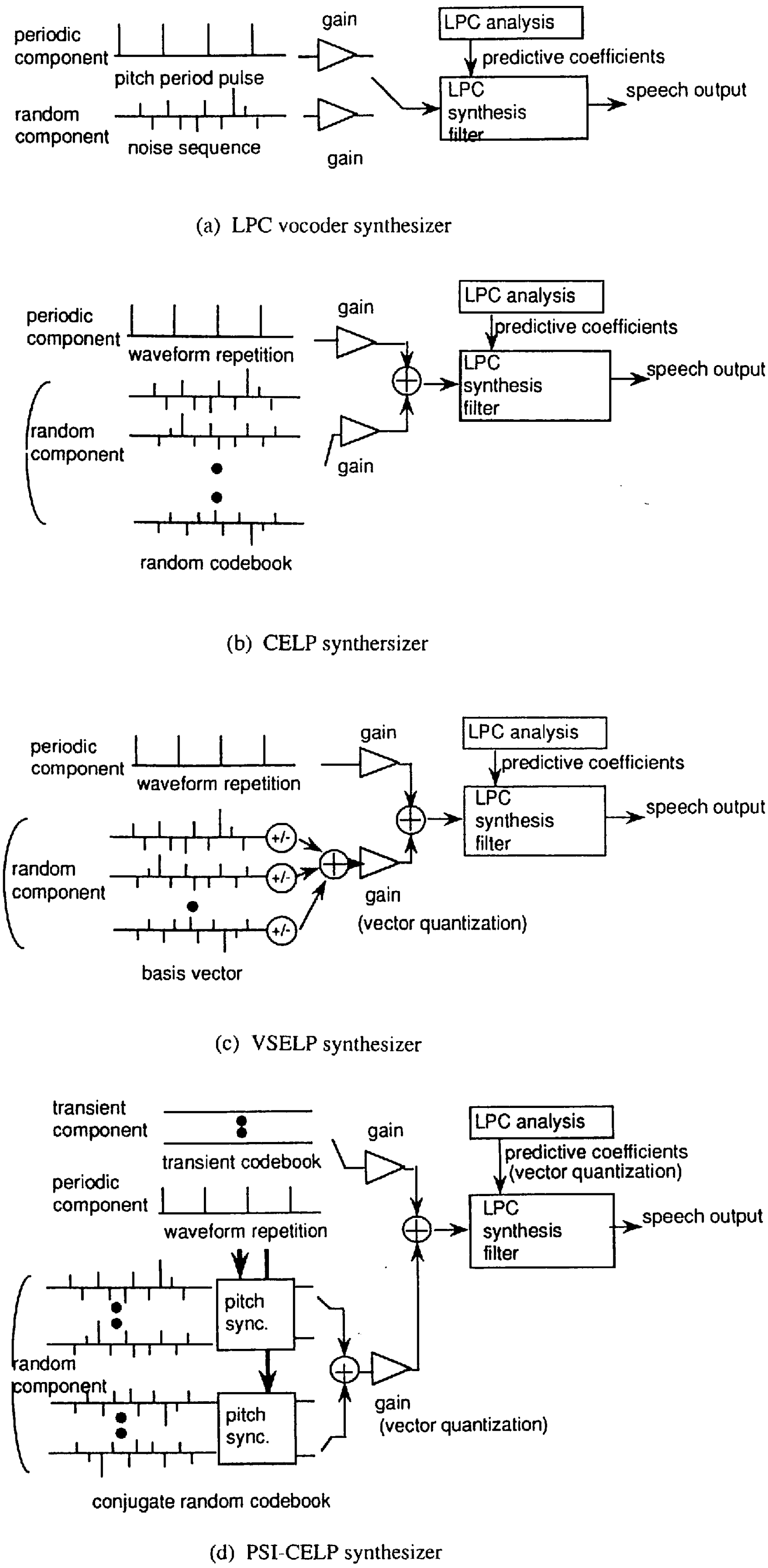

Fig. 3 Coded speech and audio quality versus transmission rate

## Table 1 Comparison of major speech coding systems

| Coding system | Coding technology | Bit rate (kbit/s) | Complexity (MIPS) | Delay (ms) | Speech quality |
|---|---|---|---|---|---|
| PCM (Pulse Code Modulation) | Non-linear quantization | 64 | 0.01 | 0 | High |
| ADPCM (Adaptive Differential PCM) | Adaptive quantization<br>Predictive quantization | 32 | 0.1 | 0 | High |
| APC-AB (Adaptive Predictive Coding with Adaptive Bit Allocation) | Adaptive quantization<br>Predictive quantization<br>Time and frequency division | 16 | 1 | 25 | High |
| MPC (Multipulse LPC) | LPC synthesis filter is excited by multiple pulses | 8 | 10 | 35 | Middle |
| CELP (Code-excited LPC) | LPC synthesis filter is excited by VQ noise | 4-8 | 100 | 35 | Middle |
| LPC vocoder | Analysis-synthesis | 2 | 1 | 35 | Low |

## Table 2 Standardized speech coding for digital cellular mobile radio in Japan, U. S. and Europe

| Speech coding system | | Japan (PDC) | U. S. (IS-54) | Europe (GSM) |
|---|---|---|---|---|
| Full-rate system | Algorithm | VSELP | VSELP | RPE-LTP |
| | Bit-rate<br>Speech<br>Error-correction | Total 11.2 kbit/s<br>6.7 kbit/s<br>4.5 kbit/s | Total 13.00 kbit/s<br>7.95 kbit/s<br>5.05 kbit/s | Total 22.8 kbit/s<br>13.0 kbit/s<br>9.8 kbit/s |
| | System delay<br>Frame length<br>Buffer length | 20 ms<br>6 ms | 20 ms<br>6 ms | 20 ms<br>0 ms |
| | Complexity | 7.8 MOPS | 8.4 MOPS | About 2 MOPS |
| Half-rate system | Algorithm | PSI-CELP | Under study | Under study |
| | Bit-rate<br>Speech<br>Error-correction | Total 5.60 kbit/s<br>3.45 kbit/s<br>2.15 kbit/s | Total 6.5 kbit/s | Total 11.4 kbit/s |
| | System delay<br>Frame length<br>Buffer length | 40 ms<br>5 ms | – | – |
| | Complexity | 18.7 MOPS | – | – |




(a) LPC vocoder synthesizer

(b) CELP synthersizer

(c) VSELP synthesizer

(d) PSI-CELP synthesizer

Fig. 4 Synthesizer structures for LPC vocoder, CELP, VSELP, and PSI-CELP.

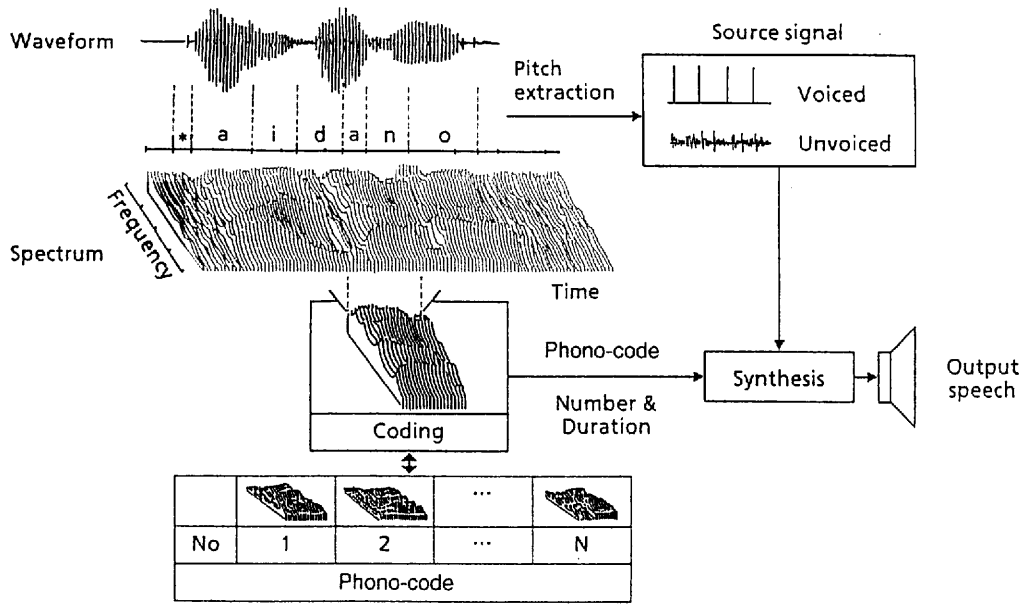



Fig. 5 Very low bit-rate coding by Phono-code method.

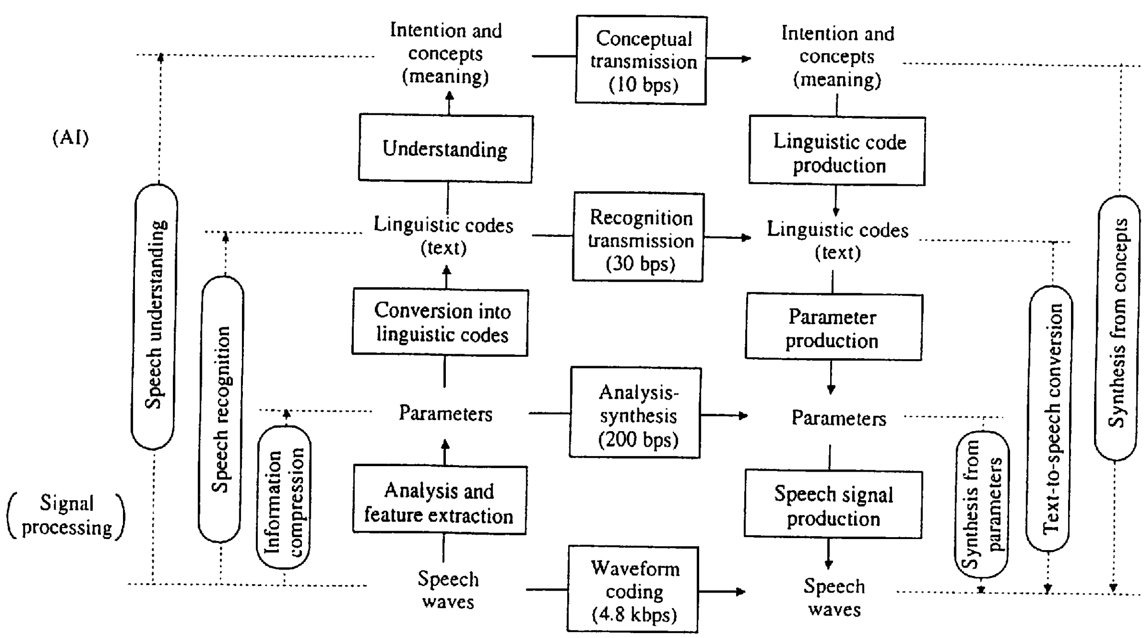



Fig. 6 Relationships between principal speech recognition, synthesis, and coding technologies.