IEEE JOURNAL ON

SELECTED AREAS IN COMMUNICATIONS

JUNE 1992 VOLUME 10 NUMBER 5 ISACEM (ISSN 0733-8716)

A PUBLICATION OF THE IEEE COMMUNICATIONS SOCIETY

SPEECH AND IMAGE CODING
Guest Editor—N. Hubing

IEEE COMMUNICATIONS SOCIETY

The field of interest of the IEEE Communications Society consists of all telecommunications including telephone, telegraphy, facsimile, and point-to-point television, by electromagnetic propagation including radio; wire; aerial, underground, coaxial, and submarine cables; waveguides, communication satellites, and lasers; in marine, aeronautical, space, and fixed station services; repeaters, radio relaying, signal storage, and regeneration; telecommunication error detection and correction; multiplexing and carrier techniques; communication switching systems; data communications; and communication theory.

In addition to the above, this JOURNAL or the IEEE TRANSACTIONS ON COMMUNICATIONS contains papers pertaining to analog and digital signal processing and modulation, audio and video encoding techniques, the theory and design of transmitters, receivers, and repeaters for communications via optical and sonic media, the design and analysis of computer communication systems, and the development of communication software. Contributions of theory enhancing the understanding of communication systems and techniques are included, as are discussions of the social implications of the development of communication technology. All members of the IEEE are eligible for membership in the Society upon payment of the annual Society membership fee of $15.00. Members may receive this JOURNAL or the IEEE TRANSACTIONS ON COMMUNICATIONS upon payment of an additional $15.00 ($30.00 total), or both publications upon payment of an additional $30.00 ($45.00 total). For information on joining, write to the IEEE at the address below. *Member copies of Transactions/Journals are for personal use only.*

IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS is published nine times a year in January, February, April, May, June, August, September, October, and December by The Institute of Electrical and Electronics Engineers, Inc. Responsibility for the contents rests upon the authors and not upon the IEEE, the Society/Council, or its members. **IEEE Headquarters:** 345 East 47 Street, New York, NY 10017. **NY Telephone:** 212-705 + extension: Information -7900; General Manager -7910; Public Information -7867; Publishing Services -7560; Spectrum -7556. **Telecopiers:** NY (Headquarters) 212-752-4929, NY (Publications) 212-705-7682; **NY Telex:** 236-411 (international messages only). **IEEE Service Center** (for orders, subscriptions, address changes, Educational Activities, Region/Section/Student Services, Standards): 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. **NJ Telephone:** Information: 908-981-0060; 908-562 + extension: Controller -5365; Technical Activities -3900. **IEEE Washington Office** (for U.S. professional activities): 1828 L Street, NW, Suite 1202, Washington, DC 20036-5104. **Washington Telephone:** 202-785-0017. **Price/Publication Information:** Individual copies: IEEE members $10.00 (first copy only), nonmembers $20.00 per copy. (Note: Add $4.00 postage and handling charge to any order from $1.00 to $50.00, including prepaid orders.) Member and nonmember subscription prices available on request. Available in microfiche and microfilm.

A Low-Delay CELP Coder for the CCITT 16 kb/s Speech Coding Standard

Juin-Hwey Chen, *Senior Member, IEEE,* Richard V. Cox, *Fellow, IEEE,* Yen-Chun Lin, Nikil Jayant, *Fellow, IEEE,* and Melvin J. Melchner

(Invited Paper)

***Abstract*—This paper presents a low-delay code-excited linear prediction (LD-CELP) speech coder, which is expected to be standardized in 1992 as a CCITT G Series Recommendation for universal applications of speech coding at 16 kb/s. The coder achieves a one-way coding delay of less than 2 ms by making both the LPC predictor and the excitation gain backward-adaptive, and by using a small excitation vector size of five samples. The pitch predictor in conventional CELP coders is not used due to its sensitivity to channel errors, and the resulting performance loss is compensated for by increasing the LPC predictor order from 10 to 50. The excitation gain is updated by a 10th-order adaptive linear predictor based on the logarithmic gains of previously quantized and scaled excitation vectors. The LPC predictor and the gain predictor are updated by performing LPC analysis on previously coded speech and previous log-gain sequence, respectively, with the autocorrelation coefficients calculated by a novel hybrid windowing method. The excitation codebook is closed-loop optimized and its index is Gray-coded for better robustness to channel errors. The perceptual weighting filter is designed to optimize speech quality for three asynchronous tandem encodings without significantly degrading the performance for a single encoding. An adaptive postfilter is used at the decoder to improve coder performance, the tandeming performance in particular. This coder has been implemented in real-time using the AT&T WE® DSP32C digital signal processor. The official CCITT laboratory tests revealed that the speech quality of this 16 kb/s LD-CELP coder is either equivalent to or better than that of the CCITT G.721 standard 32 kb/s ADPCM coder for almost all conditions tested. This paper describes the LD-CELP algorithm, its implementation on the DSP32C for CCITT testing, and some performance results from these tests.**

I. Introduction

In the past, for telephone-bandwidth speech transmission, high-quality speech was attainable only by high-bit-rate coders such as 64 kb/s log PCM or 32 kb/s ADPCM. Currently, several coding techniques can produce high-quality speech at 16 kb/s. These techniques include code-excited linear prediction (CELP) [1], multipulse linear predictive coding (MPLPC) [2], adaptive predictive coding (APC) [3], adaptive transform coding (ATC) [4], subband coding (SBC) combined with ADPCM [5], etc. However, all of these techniques require a fairly large coding delay—typically in the range of 40 to 60 ms—to achieve high-quality speech. While a large coding delay is necessary for these coders to buffer enough speech to exploit the redundancy, the large delay is undesirable or unacceptable in many applications, especially when echo cancellation is involved. Achieving high-quality or "toll-quality" speech at 16 kb/s with a coding delay less than 1 or 2 ms has been a major challenge to speech coding researchers.

The most widely used speech coding standards are 64 kb/s μ-law and A-law PCM, which were standardized in the 1960's, and 32 kb/s adaptive differential PCM (ADPCM), which was standardized in 1984 and later revised in 1986. Both standards were established by the CCITT, which assigned the code names of G.711 and G.721 to 64 kb/s PCM and 32 kb/s ADPCM, respectively. These two standards can reproduce high-quality telephone-bandwidth speech with essentially negligible coding distortion. Furthermore, both standards have negligible coding delay because the encoding is performed sample-by-sample.

At the other end of the spectrum, several other speech coders have been standardized at much lower bit rates by other organizations. Prominent examples (in an order of increasing bit rates) include the U.S. Government standards 2.4 kb/s LPC-10E vocoder [6], [7] and 4.8 kb/s CELP coder (FS1016) [8], the Japanese digital cellular radio standard 6.7 kb/s VSELP coder, the North American digital cellular radio standard 8 kb/s VSELP coder (IS54) [9], [10], and the Pan-European digital cellular radio standard 13 kb/s RPE-LTP coder (GSM) [11]. However, all these coders have large coding delays and, due to their lower bit rates, none of them are considered a "toll-quality" coder—a term usually reserved only for such high bit rate, high-quality coders as 64 kb/s G.711 PCM or 32 kb/s G.721 ADPCM.

There is a gap between the two groups of speech coding standards mentioned above. At 16 kb/s, there has not been any predominant speech coding standard, although there is an increasing need for such a standard (hopefully toll-quality and low-delay). Some localized 16 kb/s

Manuscript received September 30, 1991; revised January 15, 1992. This paper has been presented in part at the IEEE Global Communications Conferences in 1989 and 1990 and IEEE International Conferences on Acoustic, Speech, and Signal Processing in 1990, 1991, and 1992.

J.-H. Chen, R. V. Cox, N. Jayant, and M. J. Melchner are with AT&T Bell Laboratories, Murray Hill, NJ 07974.

Y.-C. Lin is with Macronix International Co., Taiwan.

IEEE Log Number 9107595.

speech coding standards now already exist for specific applications. To avoid proliferation of different 16 kb/s standards and the potential difficulty of interworking between different standards, in June 1988 the CCITT decided to investigate the possibility of establishing a single 16 kb/s speech coding standard for universal applications. The CCITT's intended applications include videophone, cordless telephone, digital satellite systems, Digital Circuit Multiplication Equipment (DCME), Public Switched Telephone Network (PSTN), Integrated Service Digital Network (ISDN), digital leased lines, voice store and forward systems, voice messages for recorded announcements, land-digital mobile radio, packetized speech, etc. [12].

Because of the variety of applications this 16 kb/s standard has to serve, the CCITT determined a stringent set of performance requirements and objectives for the candidate coding algorithms. Not every application would need every requirement to be met. Yet, to be accepted by the CCITT as a universal 16 kb/s standard, a candidate algorithm must meet all of the requirements. The major requirement was that the speech quality should be roughly comparable to that of G.721 while the one-way encoder/decoder delay should not exceed 5 ms (the objective was ≤ 2 ms) [12]. In other words, the CCITT was looking for a toll-quality low-delay speech coder at 16 kb/s.

The CCITT specifies the speech quality requirements in terms of qdu, or the quantization distortion unit. By definition, one encoding with a 64 kb/s G.711 PCM codec introduces 1 qdu of distortion (CCITT Recommendation G.113). For asynchronous tandem connections, the qdu of individual speech coders is supposed to be additive. For example, N asynchronous tandeming stages of G.711 codecs should result in N qdu. Another example is that a single encoding of a 32 kb/s G.721 ADPCM codec is rated at 3.5 qdu, and therefore two ADPCM encodings should be rated at 7 qdu and four encodings at 14 qdu. The CCITT performance requirements for the 16 kb/s standard specify that, for a clear channel (i.e., no bit errors), a candidate coder should produce 4 qdu or less for a single encoding and 14 qdu or less for three asynchronous tandeming stages. In effect, this says that a candidate coder can be slightly worse than G.721 ADPCM for a single encoding, and three asynchronous encodings of the candidate coder should match the speech quality of four asynchronous encodings of G.721 ADPCM. For noisy channels, the CCITT required that, for random bit errors at a bit error rate (BER) of 10^{-3} or 10^{-2}, a candidate coder should produce decoded speech quality not worse than that of G.721 ADPCM under the same conditions. In addition, the coder should pass network signaling tones such as DTMF and CCITT Signaling Systems No. 5, 6, and 7. It was not so difficult to meet each one of these quality requirements individually. However, in 1988, it was a major challenge to create a 16 kb/s coder that would meet all of these requirements simultaneously. Furthermore, the

Because of the low-delay requirement, none of the 16 kb/s coders mentioned above (CELP, MPLPC, APC, ATC, and SBC-ADPCM) could be used in their current form. With all these well-established coders ruled out, the only hope seemed to be *backward-adaptive* predictive coders which derive their predictor coefficients from previously quantized speech and thus do not need to buffer a large frame of speech samples. (The G.721 ADPCM coder belongs to this category.)

Prior to the CCITT's recent standardization effort at 16 kb/s, several researchers had previously reported their work on low-delay speech coding at 16 kb/s. Jayant and Ramamoorthy [13], [14] used an adaptive postfilter to enhance 16 kb/s ADPCM speech and achieved a mean opinion score (MOS) [15] of 3.5 at nearly zero coding delay. Cox *et al.* [16] combined SBC and vector quantization (VQ) and achieved an MOS of roughly 3.5–3.7 at a coding delay of 15 ms. Berouti *et al.* [17] reduced the coding delay of an MPLPC coder to 2 ms by reducing the frame size to 1 ms. However, the speech quality was equivalent to 5.5-bit log PCM—a significant degradation. Taniguchi *et al.* [18] developed an ADPCM coder with multiquantizers where the best quantizer was selected once every 2.5 ms. The coding delay of their real-time prototype codec was 8.3 ms. With the help of postfiltering, the coder produced speech quality "nearly equivalent to a 7-bit μ-law PCM" [18], but this was achieved with a nonstandard 6.4 kHz sampling rate and a resulting nonstandard bandwidth of the speech signal. Gibson *et al.* [19], [20] studied backward-adaptive predictive tree coders and predictive trellis coders which should have low coding delays. Unfortunately, they did not report the exact coding delay or the subjective speech quality. Iyengar and Kabal [21] also developed a backward-adaptive predictive tree coder with a 1 ms coding delay and a level of speech quality equivalent to 7-bit log PCM. Watts and Cuperman [22] proposed a vector generalization of ADPCM with a delay between 1 and 1.5 ms. They did not report the subjective speech quality of the coder.

Since 1988, when the CCITT announced its intention to standardize a 16 kb/s low-delay speech coder, there has been a great deal of research activity in the area of low-delay speech coding at 16 kb/s [23]–[38]. In response to the CCITT's standardization effort, we have created a 16 kb/s coder called low-delay CELP, or LD-CELP, which achieves high speech quality with a one-way coding delay less than 2 ms [24], [29], [32], [33], [35], [38].

The LD-CELP coder is a predictive coder that combines: 1) high-order backward-adaptive linear prediction; 2) backward gain-adaptive vector quantization [39], [40] for excitation; 3) the analysis-by-synthesis excitation codebook search of CELP; and 4) adaptive postfiltering [14], [41]. The low coding delay is achieved by using backward-adaptive prediction to avoid the long speech buffer required by forward-adaptive prediction, and by

With the processing delay and transmission delay also included, the total one-way coding delay can be less than 2 ms. This not only surpasses the CCITT delay requirement of 5 ms but actually meets the objective of 2 ms.

This LD-CELP coder was submitted by AT&T to the CCITT and has been the only candidate coder since 1989. This coder has been implemented in real-time hardware using the AT&T WE® DSP32C floating-point digital signal processor, and the resulting hardware prototype LD-CELP coder has been used in the official CCITT laboratory tests.

In the standardization process, there were two phases of laboratory testing. The first phase of testing was conducted in late 1989 and early 1990, while the second phase was in early 1991. The LD-CELP coder submitted for the first phase of testing (called the *Phase 1 coder* from here on) met all of the CCITT's performance requirements except for the requirement of three asynchronous tandems. Based on the Phase 1 test results, the Speech Quality Experts Group (SQEG) of the CCITT indicated that the LD-CELP coder could be standardized for point-to-point applications but not for networking applications where tandeming may occur, unless the coder could be improved to meet the tandeming performance requirement in the Phase 2 test.

In late 1990 to early 1991, we improved the LD-CELP coder's tandeming performance significantly and produced what we called the *Phase 2 coder*. The hardware prototype Phase 2 coder was then tested in the second phase of laboratory testing in 1991. From the Phase 2 test results, the SQEG concluded that the 16 kb/s LD-CELP coder ''*has a performance equivalent to or better than G.721*,'' and ''*meets all the speech quality requirements set by Study Group XV and tested by Study Group XII*'' [42]. Therefore, ''*the SQEG recommends that the 16 kb/s LD-CELP codec can be standardized as a CCITT G Series Recommendation as regards to its speech quality*'' [42]. According to the current standardization schedule, this 16 kb/s LD-CELP coder is expected to be standardized by the first half of 1992.

In this paper, we will describe the 16 kb/s LD-CELP coding algorithm, its implementation, and its performance. Section II introduces system concepts and provides an overview of the LD-CELP coder. Section III describes the LD-CELP coding algorithm. Section IV discusses the implementation issues. Section V describes the subjective and objective performance, and Section VI gives some concluding remarks.

II. System Concepts and Overview

In this section, we review the conventional CELP algorithm [1] and then give an overview of the LD-CELP algorithm and point out the differences between conventional CELP and LD-CELP. Along the way, we also discuss the issue of coding delay.

A. Review of Conventional CELP

''source-filter'' speech production model [43], with the short-term synthesis filter modeling the vocal tract and the excitation VQ, together with the long-term synthesis filter, modeling the glottal excitation. The CELP coder synthesizes speech by passing a gain-scaled excitation sequence through long-term and short-term synthesis filters. Both synthesis filters are all-pole filters containing either a long-term or a short-term predictor in a feedback loop. Basically, the CELP coder encodes speech frame-by-frame, and within each frame it attempts to find the best predictors, gain, and excitation such that a perceptually weighted mean-squared error (MSE) between the input speech and the synthesized speech is minimized.

The long-term predictor is often referred to as the *pitch predictor*, because its main function is to exploit the pitch periodicity in voiced speech. Typically, a one-tap pitch predictor is used, in which case the predictor transfer function is:

$$P_1(z) = \beta z^{-p} \tag{1}$$

where p is the bulk delay or *pitch period*, and β is the predictor tap. The short-term predictor is sometimes referred to as the *LPC predictor*, because it is also used in the well-known LPC (linear predictive coding) vocoders which operate at 2.4 kb/s or below. The LPC predictor is typically a 10th-order predictor with a transfer function of:

$$P_2(z) = \sum_{i=1}^{10} a_i z^{-i} \tag{2}$$

where a_1 through a_{10} are the predictor coefficients. The excitation VQ codebook contains a table of codebook vectors (or *codevectors*) of equal length. The codevectors are typically populated by Gaussian random numbers with possible center clipping.

In the actual encoding process, the encoder first buffers an input speech frame of about 20 ms or so, and then performs linear predictive analysis [43] (or *LPC analysis*) on the buffered speech. The resulting LPC parameters are then quantized. The pitch predictor parameters, including the pitch period and the predictor tap, are then determined either in an open-loop fashion [1] or in a closed-loop fashion [44]. The quantized LPC parameters and pitch predictor parameters are both sent as side information to the decoder. This scheme is called *forward-adaptive prediction*.

The input speech frame is further subdivided into several equal-length *subframes*, or *vectors*, typically of size 4 to 8 ms. Then, for each vector, the encoder passes each candidate codevector in the excitation VQ codebook through the gain scaling unit and the two synthesis filters, and then compares the corresponding filtered output vector with the input speech vector and computes the associated perceptually weighted MSE distortion. The encoder repeats this process for all candidate excitation

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE