US005235669A

United States Patent [19]

Ordentlich et al.

[11] Patent Number: 5,235,669

[45] Date of Patent: Aug. 10, 1993

[54] LOW-DELAY CODE-EXCITED LINEAR-PREDICTIVE CODING OF WIDEBAND SPEECH AT 32 KBITS/SEC

[75] Inventors: **Erik Ordentlich**, Palo Alto, Calif.; **Yair Shoham**, Berkeley Heights, N.J.

[73] Assignee: **AT&T Laboratories**, Murray Hill, N.J.

[21] Appl. No.: **546,627**

[22] Filed: **Jun. 29, 1990**

[51] **Int. Cl.**^{5} **G10L 9/00**

[52] **U.S. Cl.** **395/2**

[58] **Field of Search** 381/29-50; 364/513.5, 724.19

[56] **References Cited**

U.S. PATENT DOCUMENTS

Re. 32,580	1/1988	Atal et al.	381/40
4,133,976	1/1979	Atal et al.	381/47
4,472,832	9/1984	Atal et al.	381/40
4,694,298	9/1987	Milan	364/724.19
4,701,954	10/1987	Atal	381/49
4,811,261	3/1989	Kobayashi et al.	364/724.19
4,827,517	5/1989	Atal et al.	381/41
4,941,178	7/1990	Chuang	381/41

FOREIGN PATENT DOCUMENTS

0331405	2/1989	European Pat. Off. .
2624675	6/1989	France .

OTHER PUBLICATIONS

"Stochastic Coding of Speech Signals at Very Low Bit Rates", *Proc. IEEE Int. Conf. Comm.*, May 1984, B. S. Atal and M. R. Schroeder, pp. 1610-1612.

"Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates", *Proc. IEEE Int. Conf. ASSP.*, 1985, pp. 937-940, M. R. Schroeder and B. S. Atal.

"A Class of Analysis-by-Synthesis Predictive Coders for High Quality Speech Coding at Rates Between 4.8 and 16 kbits/s", *IEEE J. on Sel. Area in Comm.*, SAC-6(2) Feb. 1988, pp. 353-363, P. Kroon and E. F. Deprettere.

"Predictive Coding of Speech Signals and Subjective Error Criteria", *IEEE Tr. ASSP*, vol. ASSP-27, No. 3, Jun. 1979, pp. 247-254, B. S. Atal and M. S. Schroeder.

"Low Delay Code Excited Linear Predictive (LD--CELP) Coding of Wide Band Speech at 32Kbit/sec.," MS Thesis, EE Dept., MIT, Jul. 1990, E. Ordentlich, Abstract only (p. 1).

"Transfor Coding of Audio Signals Using Perceptual Noise Criteria", *IEEE Sel. Areas in Comm.*, vol. 6, No. 2, Feb. 1988, pp. 314-323, J. D. Johnston.

"G.722, A New CCITT Coding Standard for Digital Transmission of Wideband Audio Signals", *IEEE Comm. Mag.*, vol. 26, No. 1, Jan. 1988, pp. 8-15, P. Mermelstein.

"Strategies for improving the performance of CELP coders at low bit rates", ICASSP'88 (1988 International Conf. on Acoustics, Speech, and Signal Processing), vol. 1, pp. 151-154, IEEE, New York; P. Kroon, et al.

"Some experiments of 7 kHz audio coding at 16 kbit/s", ICASSP '89 (1989 International Conference on Acoustics, Speech, and Signal Processing), May 1989, vol. 1, pp. 192-195, IEEE, New York; R. Drogo de Jacovo, et al.

"On different vector predictive coding schemes and their application to low bit rates speech coding", Signal Processing IV: Theories and Applications (Proceedings of EUSIPCO-88, 4th European Signal Processing Conf.) Sep. 1988, vol. II, pp. 871-874, North Holland Publishing Co.; F. Bottau, et al.

Primary Examiner—David D. Knepper
Attorney, Agent, or Firm—William Ryan

[57] **ABSTRACT**

An improved digital communication system, e.g., a CELP code/decoder based system, is improved for use with a wide-band signal such as a high-quality speech signal by modifying the noise weighting filter used in such systems to include a filter section which affects primarily the spectral tilt of the weighting filter in addition to a filter component reflecting formant frequency information in the input signal. Alternatively, the weighting is modified to reflect perceptual transform techniques.

20 Claims, 2 Drawing Sheets

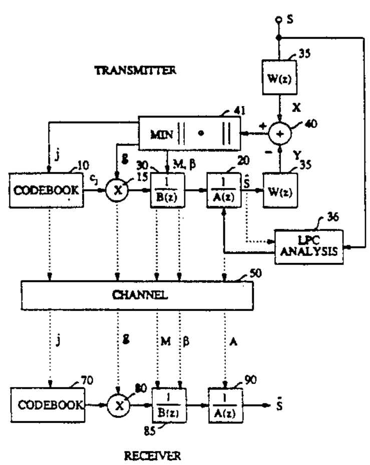


FIG. 1

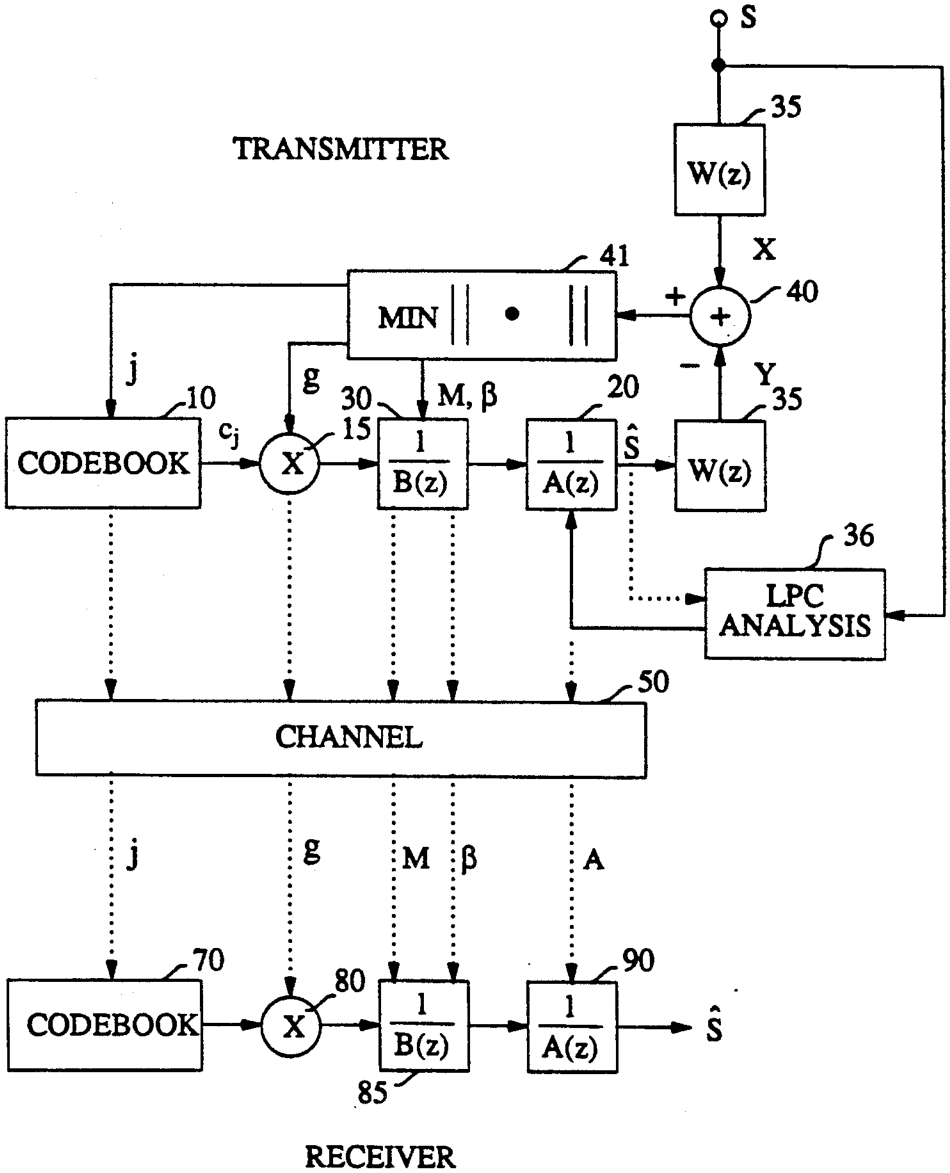

S
35
TRANSMITTER
W(z)
X
41
40
MIN || • ||
+
+
j
g
30
M, β
20
Y
35
10
c_j
15
CODEBOOK
X
1/B(z)
1/A(z)
Ŝ
W(z)
36
LPC ANALYSIS
50
CHANNEL
j
g
M
β
A
70
80
90
CODEBOOK
X
1/B(z)
1/A(z)
Ŝ
85
RECEIVER

FIG. 2

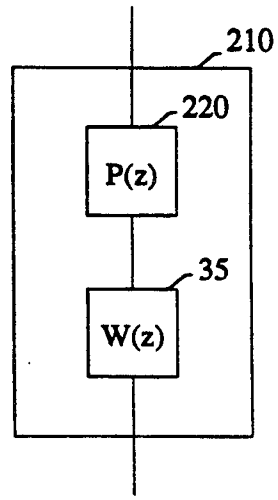
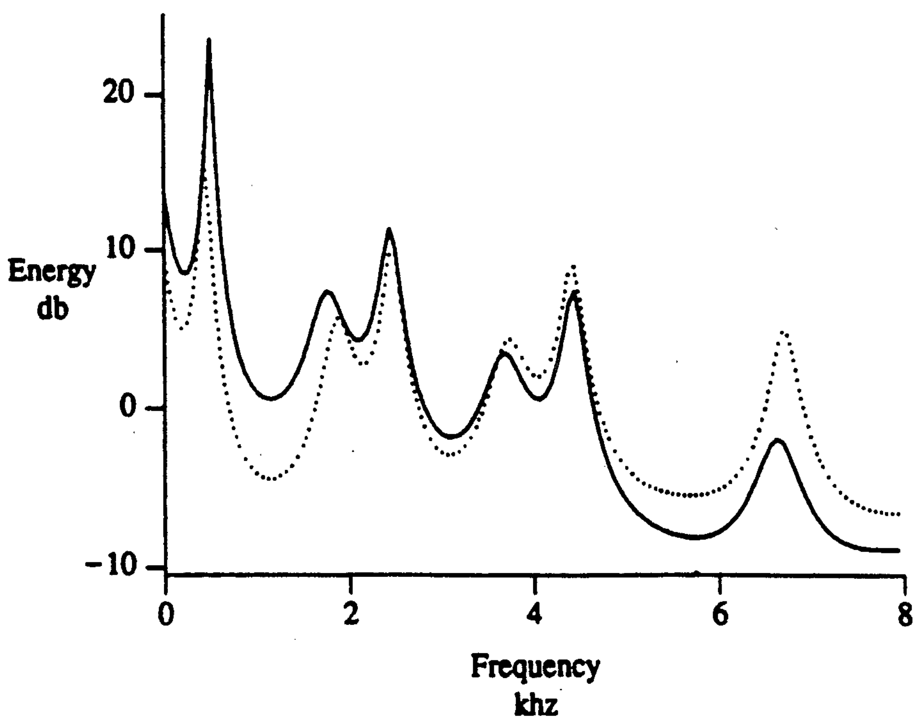


FIG. 3



LOW-DELAY CODE-EXCITED LINEAR-PREDICTIVE CODING OF WIDEBAND SPEECH AT 32 KBITS/SEC

FIELD OF THE INVENTION

The present invention relates to methods and apparatus for efficiently coding and decoding signals, including speech signals. More particularly, this invention relates to methods and apparatus for coding and decoding high quality speech signals. Yet more particularly, this invention relates to digital communication systems, including those offering ISDN services, employing such coders and decoders.

BACKGROUND OF THE INVENTION

Recent years have witnessed many improvements in coding and decoding for digital communications systems. U.S. Pat. No. 4,133,976, issued on Jan. 9, 1979; RE 32,580, issued on Jan. 19, 1988; 4,701,954, issued on Oct. 27, 1987; 4,472,832, issued on Sep. 18, 1984, and 4,827,517, issued on May 2, 1989, to B. S. Atal, et al and assigned to the assignee of the present invention, all present important improvements in this field.

One area of such improvements have came to be called code excited linear predictive (CELP) coders and are, e.g., described B. S. Atal and M. R. Schroeder, "Stochastic Coding of Speech Signals at Very Low Bit Rates," *Proc. IEEE Int. Conf. Comm.*, May 1984, p. 48.1; M. R. Schroeder and B. S. Atal, "Code-Excited Linear Predictive (CELP): High Quality Speech at Very Low Bit Rates," *Proc. IEEE Int. Conf. ASSP.*, 1985, pp. 937–940; P. Kroon and E. F. Deprettere, "A Class of Analysis-by-Synthesis Predictive Coders for High-Quality Speech Coding at Rate Between 4.8 and 16 Kb/s," *IEEE J. on Sel. Area in Comm* SAC-6(2), Feb. 1988, pp. 353–363, and the above-cited U.S. Pat. No. 4,827,517. Such techniques have found application, e.g., in voice grade telephone channels, including mobile telephone channels.

The prospect of high-quality multi-channel/multi-user speech communication via the emerging ISDN has increased interest in advanced coding algorithms for wideband speech. In contrast to the standard telephony band of 200 to 3400 Hz, wideband speech is assigned the band 50 to 7000 Hz and is sampled at a rate of 16000 Hz for subsequent digital processing. The added low frequencies increase the voice naturalness and enhance the sense of closeness whereas the added high frequencies make the speech sound crisper and more intelligible. The overall quality of wideband speech as defined above is sufficient for sustained commentary-grade voice communication as required, for example, in multi-user audio-video teleconferencing. Wideband speech is, however, harder to code since the data is highly unstructured at high frequencies and the spectral dynamic range is very high. In some network applications, there is also a requirement for a short coding delay which limits the size of the processing frame and reduces the efficiency of the coding algorithm. This adds another dimension to the difficulty of this coding problem.

SUMMARY OF THE INVENTION

Many of the advantages of the well-known CELP coders and decoders are not fully realized when applied to the communication of wide-band speech information (e.g., in the frequency range 50 to 7000 Hz). The present invention, in typical embodiments, seeks to adapt existing CELP techniques to extend to communication of such wide-band speech and other such signals.

More particularly, the illustrative embodiments of the present invention provide for modified weighting of input signals to enhance the relative magnitude of signal energy to noise energy as a function of frequency. Additionally, the overall spectral tilt of the weighting filter response characteristic is advantageously decoupled from the determination of the response at particular frequencies corresponding, e.g., to formants.

Thus, whereas prior art CELP coders employ a weighting filter based primarily on the formant content, it proves advantageous in accordance with a teaching of the present invention to use a cascade of prior art weighting filter and an additional filter section for controlling the spectral tilt of the composite weighting filter.

BRIEF DESCRIPTION OF THE DRAWING

FIG. **1** shows a digital communication system using the present invention.

FIG. **2** shows a modification of the system of FIG. **1** in accordance with the embodiment of the present invention.

FIG. **3** shows a modified frequency response resulting from the application of a typical embodiment of the present invention.

DETAILED DESCRIPTION

To simplify the description of the present invention, the above-cited publications by Atal and Schroeder, and the cited U.S. Pat. No. 4,133,976 to Atal and Schroeder are hereby incorporated by reference and should be deemed included in the present disclosure as if set forth in their entirety.

The basic structure of conventional CELP (as described, e.g., in the references cited above) is shown in FIG. **1**.

Shown are the transmitter portion at the top of the figure, the receiver portion at the bottom and the various parameters (j, g, M, β and A) that are transmitted via a communication channel **50**. CELP is based upon the traditional excitation-filter model where an excitation signal, drawn from an excitation codebook **10**, is used as an input to an all-pole filter which is usually a cascade of an LPC-derived filter 1/A(z) (20 in FIG. **1**) and a so-called pitch filter 1/B(z), 30. The LPC polynomial is given by

$$A(z) = \sum_{i=0}^{M} a_i z^{-i}$$

and is obtained by a standard M^{th}-order LPC analysis of the speech signal. The pitch filter is determined by the polynomial

$$B(z) = \sum_{j=0}^{q} b_j z^{-j-P}$$

where P is the current "pitch" lag—a value that best represents the current periodicity of the input and b_j's are the current pitch taps. Most often, the order of the pitch filter is q=1 and it is rarely more than 3. Both polynomial A(z), B(z) are monic.

The CELP algorithm implements a closed-loop (analysis-by-synthesis) search procedure for finding the best excitation and, possibly, the best pitch parameters. In the excitation search loop, each of the excitation vectors is passed through the LPC and pitch filters in an effort to find the best match (as determined by comparator **40** and minimizing circuit **41**) to the output, usually, in a weighted mean-squared error (WMSE) sense. As seen in FIG. **1**, the WMSE matching is accomplished via the use of a noise-weighting filter W(z) **35**. The input speech s(n) is first pre-filtered by W(z) and the resulting signal x(n) ($X(z)=S(z)\ W(z)$) serves as a reference signal in the closed-loop search. The quantized version of x(n), denoted by y(n), is a filtered excitation, closest to x(n) in an MSE sense. The filter used in the search loop is the weighted synthesis filter $H(z)=W(z)/[B(z)\ A(z)]$. Observe, however, that the final quantized signal is obtained at the output of the unweighted synthesis filter $1/[B(z)\ A(z)]$, which means that W(z) is not used by the receiver to synthesize the output. This loop essentially (but not strictly) minimizes the WMSE between the input and output, namely, the MSE of the signal $(S(z)-\hat{S}(z))\ W(z)$.

The filter W(z) is important for achieving a high perceptual quality in CELP systems and it plays a central role in the CELP-based wideband coder presented here, as will become evident.

The closed-loop search for the best pitch parameters is usually done by passing segments of past excitation through the weighted filter and optimizing B(z) for minimum WMSE with respect to the target signal X(z). The search algorithm will be described in more detail.

As shown in FIG. **1**, the codebook entries are scaled by a gain factor g applied to scaling circuit **15**. This gain may either be explicitly optimized and transmitted (forward mode) or may be obtained from previously quantized data (backward mode). A combination of the backward and forward modes is also sometimes used (see, e.g., AT&T Proposal for the CCITT 16 Kb/s speech coding standard, COM N No. 2, STUDY GROUP N, "Description of 16 Kb/s Low-Delay Code-excited Linear Predictive Coding (LD-CELP) Algorithm," March 1989). See also U.S. patent application Ser. No. 07/298451, entitled "A Low-Delay Code-Excited Linear Predictive Coder for Speech or Audio," by J-H. Chen, filed Jan. 17, 1989, and assigned to the assignee of the present invention, which application is hereby incorporated in this disclosure by reference as if set forth in its entirety.

In general, the CELP transmitter codes and transmits the following five entities: the excitation vector (j), the excitation gain (g), the pitch lag (p), the pitch tap(s) (β), and the LPC parameters (A). The overall transmission bit rate is determined by the sum of all the bits required for coding these entities. The transmitted information is used at the receiver in well-known fashion to recover the original input information.

The CELP is a look-ahead coder, it needs to have in its memory a block of "future" samples in order to process the current sample which obviously creates a coding delay. The size of this block depends on the coder's specific structure. In general, different parts of the coding algorithm may need different-size future blocks. The smallest block of immediate future samples is usually required by the codebook search algorithm and is equal to the codevector dimension. The pitch loop may need a longer block size, depending on the update rate of the pitch parameters. In a conventional CELP, the longest block length is determined by the LPC analyzer which usually needs about 20 msec worth of future data. The resulting long coding delay of the conventional CELP is therefore unacceptable in some applications. This has motivated the development of the Low-Delay CELP (LD-CELP) algorithm (see above-cited AT&T Proposal for the CCITT 16 Kb/s speech coding standard).

The Low-Delay CELP derives its name from the fact that it uses the minimum possible block length-the vector dimension. In other words, the pitch and LPC analyzers are not allowed to use any data beyond that limit. So, the basic coding delay unit corresponds to the vector size which only a few samples (between 5 to 10 samples). The LPC analyzer typically needs a much longer data block than the vector dimension. Therefore, in LD-CELP the LPC analysis can be performed on a long enough block of most recent past data plus (possibly) the available new data. Notice, however, that a coded version of the past data is available at both the receiver and the transmitter. This suggests an extremely efficient coding mode called backward-adaptive-coding. In this mode, the receiver duplicates the LPC analysis of the transmitter using the same quantized past data and generates the LPC parameters locally. No LPC information is transmitted and the saved bits are assigned to the excitation. This, in turn, helps in further reducing the coding delay since having more bits for the excitation allows using shorter input blocks. This coding mode is, however, sensitive to the level of the quantization noise. A high-level noise adversely affects the quality of the LPC analysis and reduces the coding efficiency. Therefore, the method is not applicable to low-rate coders. It has been successfully applied in 16 Kb/s LD-CELP systems (see above-cited AT&T Proposal for the CCITT 16 Kb/s speech coding standard) but not as successfully at lower rates.

When backward LPC analysis becomes inefficient due to excessive noise, a forward-mode LPC analysis can be employed within the structure of LD-CELP. In this mode, LPC analysis is performed on a clean past signal and LPC information is sent to the receiver. Forward-mode and combined forward-backward mode LD-CELP systems are currently under study.

The pitch analysis can also be performed in a backward mode using only past quantized data. This analysis, however, was found to be extremely sensitive to channel errors which appear at the receiver only and cause a mismatch between the transmitter and receiver. So, in LD-CELP, the pitch filter B(z) is either completely avoided or is implemented in a combined backward-forward mode where some information about the pitch delay and/or pitch tap is sent to the receiver.

The LD-CELP proposed here for coding wideband speech at 32 Kb/s advantageously employs backward LPC. Two versions of the coder will be described in greater detail below. The first includes forward-mode pitch loop and the second does not use pitch loop at all. The general structure of the coder is that of FIG. **1**, excluding the transmission of the LPC information. Also, if the pitch loop is not used, $B(z)=1$ and the pitch information is not transmitted. The algorithmic details of the coder are given below.

A fundamental result in MSE waveform coding is that the quantization noise has a flat spectrum at the point of minimization, namely, the difference signal between the output and the target is white. On the other hand, the input speech signal is non-white and actually has a wide spectral dynamic range due to the formant structure and the high-frequency roll-off. As a result,

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.