

Advances in Speech and Audio Compression

ALLEN GERSHO, FELLOW, IEEE

Invited Paper

Speech and audio compression has advanced rapidly in recent years spurred on by cost-effective digital technology and diverse commercial applications. Recent activity in speech compression is dominated by research and development of a family of techniques commonly described as code-excited linear prediction (CELP) coding. These algorithms exploit models of speech production and auditory perception and offer a quality versus bit rate tradeoff that significantly exceeds most prior compression techniques for rates in the range of 4 to 16 kb/s. Techniques have also been emerging in recent years that offer enhanced quality in the neighborhood of 2.4 kb/s over traditional vocoder methods. Wideband audio compression is generally aimed at a quality that is nearly indistinguishable from consumer compact-disc audio. Subband and transform coding methods combined with sophisticated perceptual coding techniques dominate in this arena with nearly transparent quality achieved at bit rates in the neighborhood of 128 kb/s per channel.

I. INTRODUCTION

Compression of telephone-bandwidth speech has been an ongoing area of research for several decades. Nevertheless, in the last several years, there has been an explosion of interest and activity in this area with numerous applications in telecommunications and storage, and several national and international standards have been adopted. High-fidelity audio compression has also advanced rapidly in recent years, accelerated by the commercial success of consumer and professional digital audio products. The surprising growth of activity in the relatively old subject of speech compression is driven by the insatiable demand for voice communication, by the new generation of technology for cost-effective implementation of digital signal processing algorithms, by the need to conserve bandwidth in both wired and wireless telecommunication networks, and the need to conserve disk space in voice storage systems. Most of this effort is focused on the usual telephone bandwidth of roughly 3.2 kHz (200 Hz to 3.4 kHz). There has also

Manuscript received November 1, 1993; revised January 15, 1994. This work was supported in part by the National Science Foundation, Fujitsu Laboratories, Ltd., the UC Micro program, Rockwell International Corporation, Hughes Aircraft Company, Echo Speech Corporation, Signal Technology, Inc., and Qualcomm, Inc.

The author is with the Center for Information Processing Research, Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, USA.

IEEE Log Number 9401177.

been a very large increase in research and development in the coding of audio signals, particularly, wideband audio (typically 20-kHz bandwidth) for transmission and storage of CD-quality music. Interest in wideband (7-kHz) speech for audio in video teleconferencing has also increased in recent years.

Since standards are essential for compatibility of terminals in voice and audio communication systems, standardization of speech and audio coding algorithms has lately become a major activity of central importance to industry and government. As a result, the driving force for much of the research in speech and audio coding has been the challenge of meeting the objectives of standards committees. The most important organization involved in speech coding standardization is the Telecommunication Standardization Sector of the International Telecommunications Union, referred to by the acronym ITU-T (the successor of the International Telephone and Telegraph Consultative Committee, CCITT). Other standards organizations will be mentioned later in this paper.

This paper highlights the state of the art for digital compression of speech and audio signals. The scope is limited to surveying the most important and prevailing methods, approaches, and activities of current interest without attempting to give a tutorial presentation of specific algorithms or a historical perspective of the evolution of speech coding methods. No attempt is made to offer a complete review of the numerous contributions that have been made in recent years, and inevitably some important papers and methods will be overlooked. Nevertheless, the major ideas and trends are covered here and attention is focused on those contributions which have had the most impact on the current state of the art. Many algorithms that are no longer of current importance are not covered at all or only briefly mentioned here, even though they may have been widely studied in the past. We do not attempt to describe the quantitative performance of different coding algorithms as determined from the many subjective evaluations that have taken place in recent years.

For reviews, tutorials, or collections of papers on earlier work in speech compression, see [205], [83], [204], [110],

[67], [115], [85], [46], [79], [253]. A recent survey of audio compression is given in [77]. For a cross section of recent work in speech compression, see [8], [9]. A general perspective of issues, techniques, targets, and standards in signal compression is given in [112]. A comprehensive review of the methods and procedures involved in speech standardization and some recent activity in this area is given in [65].

Virtually all work in speech and audio compression involves *lossy compression* where the numerical representation of the signal samples is never recovered exactly after decoding (decompression). There is a wide range of tradeoffs between bit rate and recovered speech quality that are of practical interest in the coding of telephone speech, where users are accustomed to tolerating various degrees of degradation. On the other hand, for wideband audio compression, consumers have higher expectations today, and quality close to that of the compact disc (CD) is generally needed. Thus research in speech compression includes concurrent studies for different distortion-rate tradeoffs motivated by various applications with different quality objectives. For wideband audio compression, most research aims at the same or similar standard of quality as offered by the CD.

Although the term *compression* is commonly used in the lay press and in the computer science literature, researchers working in speech or audio generally prefer the term *coding*. This avoids ambiguity with the alternative use of *speech compression* that refers to time-scale modification of speech, as in the speeding-up of the speech signal, e.g., in learning aids for the blind. Information theorists refer to signal compression as *source coding*. Henceforth, we shall use the term *coding*.

The ease of real-time implementation of speech-coding algorithms with single-chip digital signal processors has led to widespread implementations of speech algorithms in the laboratory as well as an extension of applications to communication and voice storage systems. The largest potential market for speech coding is in the emerging area of personal communication systems (PCS) where volumes of hundreds of millions are expected in the U.S. alone, and comparable numbers in Western Europe and Japan. In the next decade or so, a significant number (perhaps more than 50%) of telephones are expected to become wireless. Another new area of application is multimedia in personal computing where voice storage is becoming a standard feature. With so many applications already emerging or expected to emerge in the next few years, it is not surprising that speech coding has become such an active field of research in recent years.

Wideband audio coding for high-fidelity reproduction of voice and music has emerged as an important activity in the past decade. Applications of audio coding lie largely with the broadcasting industry, motion picture industry, and consumer audio and multimedia products. A key international standard developed by the Motion Picture Experts Group (MPEG) of the International Standards Organization (ISO) includes an audio coding algorithm [21].

Speech-coding algorithms can be divided into two main categories *waveform coders* and *vocoders*. The term vocoder historically originated as a contraction of *voice coder*. In waveform coders, the data transmitted from encoder to decoder specify a representation of the original speech as a waveform of amplitude versus time, so that the reproduced signal approximates the original waveform and, consequently, provides an approximate recreation of the original sound. In contrast, vocoders do not reproduce an approximation to the original waveform; instead, parameters that characterize individual sound segments are specified and transmitted to the decoder, which then reconstructs a new and different waveform that will have a similar sound. Vocoders are sometimes called *parametric coders* for obvious reasons. Often these parameters characterize the short-term spectrum of a sound. Alternatively, the parameters specify a mathematical model of human speech production suited to a particular sound. In either case, the parameters do not provide sufficient information to regenerate a close approximation to the original waveform but the information is sufficient for the decoder to synthesize a perceptually similar speech sound. Vocoders operate at lower bit rates than waveform coders but the reproduced speech quality, while intelligible, usually suffers from a loss of naturalness and some of the unique characteristics of an individual speaker are often lost.

Most work on speech coding today is based on telephone-bandwidth speech, nominally limited to about 3.2 kHz and sampled at the rate of 8 kHz. Wideband speech coding is of increasing interest today and is intended for speech or audio signals of 7 kHz, sampled at 16 kHz. High-fidelity audio signals of bandwidth 20 kHz are generally sampled at rates of 44.1 or 48 kHz although there is also some interest in 15-kHz bandwidth signals with a 32-kHz sampling rate. Audio coding schemes of interest today include joint coding of multiple audio channels.

Much of the work in waveform speech coding, is dominated by a handful of different algorithmic approaches and most of the developments in recent years have focused on modifications and enhancements of these generic methods. Most notable and most popular for speech coding is *code-excited linear prediction* (CELP). Other methods in commercial use today that continue to receive some attention include *adaptive delta modulation* (ADM), *adaptive differential pulse code modulation* (ADPCM), *adaptive predictive coding* (APC), *multipulse linear predictive coding* (MP-LPC), and *regular pulse excitation* (RPE). MP-LPC, RPE, and CELP belong to a common family of analysis-by-synthesis algorithms to be described later. These algorithms are sometimes viewed as "hybrid" algorithms because they borrow some features of vocoders, but they basically belong to the class of waveform coders.

Although many vocoders were studied several decades ago, the most important survivor is the *linear predictive coding* (LPC) vocoder, which is extensively used in secure voice telephony today and is the starting point of some current vocoder research. Another vocoding approach that has emerged as an effective new direction in the past decade

is *sinusoidal coding*. In particular, *sinusoidal transform coding* (STC) and *multiband excitation* (MBE) coding are both very actively studied versions of sinusoidal coding.

Many waveform coders with other names are closely related to those listed here. Of diminishing interest are RPE, MP-LPC, ADPCM, and ADM although versions of these have become standardized for specific application areas. Perhaps the oldest algorithm to be used in practice is ADM, one well-known version of which is *continuously variable slope delta modulation* (CVSD). Although the performance of ADPCM at 32 kb/s can today be achieved at much lower rates by more "modern" algorithms, ADPCM remains of interest for some commercial applications because of its relatively low complexity.

Subband and transform coding methods have been extensively studied for speech coding a decade ago. Today, they serve as the basis for most wideband audio-coding algorithms and for many image- and video-coding schemes but they are generally not regarded today as competitive techniques for speech coding. Nevertheless, many researchers continue to study subband and transform techniques for speech coding and a few very interesting and effective coding schemes of current interest make use of filter banks or some form of linear transformation. These techniques generally function as building blocks that contribute to an overall algorithm for some effective coding schemes such as IMBE and CELP. One ITU-T standard, Recommendation G.722, for wideband (7-kHz) speech at 64, 56, and 48 kb/s, uses a two-band subband coder [225], [174].

Compression algorithms of current interest for wideband audio are based on signal decompositions via linear transformations or subband filter banks (including wavelet methods) which allow explicit and separate control of the coding of different frequency regions in the auditory spectrum. Efficient coding is achieved with the aid of sophisticated perceptual masking models for dynamically allocating bits to different frequency bands. The quality objectives for audio coding are generally much more demanding than for speech coding. The usual goal is to attain a quality that is nearly indistinguishable from that of the compact disc (CD). In contrast, most speech coding is applied to signals already limited by the telephone bandwidth so that users are not accustomed to high-fidelity reproduction.

This paper is organized as follows. In Section II, we give a brief overview of the most important family of speech coding algorithms that includes CELP and in Section III we review the recent activity in CELP coding, the most widely studied algorithmic approach of current interest. Section IV examines the advances in low-delay speech coding and Section V reviews the area of variable-rate speech coding. In Section VI, we examine recent developments in vocoders. Section VII looks at wideband speech and audio coding. Section VIII summarized the current performance achievable today in speech and audio coding at various bit rates. Finally, in Section IX, some concluding remarks are offered.

II. LPAS SPEECH CODING

The approach to speech coding most widely studied and implemented today is *linear-prediction-based analysis-by-synthesis* (LPAS) coding. An LPAS coder has three basic features:

- Basic decoder structure: The decoder receives data which specify an excitation signal and a synthesis filter; the reproduced speech is generated as the response of the synthesis filter to the excitation signal.
- Synthesis filter: The time-varying linear-prediction-based synthesis filter is periodically updated and is determined by *linear prediction* (LP) analysis of the current segment or *frame* of the speech waveform; the filter functions as a shaping filter which maps a relatively flat spectral-magnitude signal into a signal with an autocorrelation and spectral envelope that are similar to those of the original speech.
- Analysis-by-synthesis excitation coding: The encoder determines the excitation signal one segment at a time, by feeding candidate excitation segments into a replica of the synthesis filter and selecting the one that minimizes a perceptually weighted measure of distortion between the original and reproduced speech segments.

The earliest proposals for LPAS coder configurations appeared in 1981. Schroeder and Atal described a tree-code excitation generator [206] and Stewart proposed a codebook excitation source [222]. The first effective and practical form of LPAS coder to be introduced was *multipulse LPC* (MP-LPC) due to Atal and Remde [11] where in each frame of speech, a *multipulse* excitation is computed as a sparse sequence of amplitudes (pulses) separated by zeros. The locations and amplitudes of the pulses in the frame are transmitted to the decoder. An MP-LPC algorithm at 9.6 kb/s was recently adopted as a standard for aviation satellite communications by the Airlines Electronic Engineering Committee (AECC).

In 1986 *regular pulse excitation* (RPE) coding was introduced by Kroon, Deprettere, and Sluyter [145]. Also an LPAS technique, RPE uses regularly spaced pulse patterns for the excitation with the position of the first pulse and the pulse amplitudes determined in the encoding process. Although inspired by MP-LPC, it is also close in spirit to CELP. A modified version of RPE, called *regular pulse excitation with long-term prediction* (RPE-LTP), was selected as part of the first standard for time-division multiple-access (TDMA) digital cellular telephony by the global system for mobile telecommunications (GSM) subcommittee of the European Telecommunications Standards Institute (ETSI) [93].

Most early LPAS methods were based on a synthesis filter which is a cascade of a short-term or *formant* filter and a long-term or *pitch* filter. The short-delay filter is typically a 10th-order all-pole filter with parameters obtained by conventional LP analysis. The long-term filter is typically

based on a single-tap or three-tap pitch prediction. The properties of these pitch filters were extensively studied by Ramachandran and Kabal [193], [194].

A key element of LPAS coding is the use of *perceptual weighting* of the error signal for selecting the best excitation via analysis-by-synthesis. The error between original and synthesized speech is passed through a time-varying perceptual weighting filter which emphasizes the error in frequency bands where the input speech has valleys and de-emphasizes the error near spectral peaks. The effect is to reduce the resulting quantization noise in the valleys and increase it near the peaks. This is generally done by an all-pole filter obtained from the LP synthesis filter by scaling down the magnitude of the poles by a constant factor. This technique exploits the masking feature of the human hearing system to reduce the audibility of the noise. It is based on the classic work of Atal and Schroeder in 1979 on subjective error criteria [12].

The most important form of LPAS coding today is commonly known as *code-excited linear prediction* (CELP) coding, but has also been called *stochastic coding*, *vector excitation coding* (VXC), or *stochastically excited linear prediction* (SELP). CELP improves on MP-LPC by using vector quantization (VQ) [76], where a predesigned set of excitation vectors is stored in a *codebook*, and for each time segment the encoder searches for that code vector whose set of samples best serves as the excitation signal for the current time segment. The address of the selected code vector is transmitted to the receiver, which has a copy of the codebook, so that the receiver can regenerate the selected excitation segment. For example, a codebook containing 1024 code vectors each of dimension 40 would require a 10-b word to specify each successive 40 samples of the excitation signal. The superior performance capability of CELP compared to MP-LPC and earlier coding methods for bit rates ranging from 4.8 to 16 kb/s has become generally recognized. Today, the terminology "CELP" refers to a family of coding algorithms rather than to one specific technique; all algorithms in this family are based on LPAS with VQ for coding the excitation.

The invention of CELP is generally attributed to Atal and Schroeder [13], [207]. A somewhat similar coding technique was also introduced by Copperi and Sereno [42]. At least one earlier research study contained the key element of CELP, namely, LPAS coding with VQ [222]. In fact, MP-LPC is sometimes viewed as a special form of CELP, in which a multistage VQ structure with a particular set of deterministic codebooks are used [135]. RPE can even more easily be seen as a form of CELP coding. Another coding method, *vector-adaptive predictive coding* (VAPC) has many features of CELP including the use of VQ and analysis-by-synthesis but differs in the encoder search structure and in the ordering of short-term and long-term synthesis filtering [36], [37]. Rose and Barnwell introduced the *self-excited coder* [196], which used prior excitation segments as code vectors for the current excitation. Although MP-LPC perhaps represents a conceptually more fundamental advance in speech coding,

CELP has had a much greater impact in the field. While newer coding techniques have since been developed, none clearly overtakes CELP in the range of bit rates 4–16 kb/s.

III. CELP ALGORITHMS

A. History

Initially viewed as an algorithm of extraordinary complexity, CELP served only as an existence proof (with the help of supercomputers) that it is possible to get very high speech quality at bit rates far below what was previously considered feasible. The first papers on CELP coding by Atal and Schroeder [13], [207] attracted great attention, intrigued researchers, and continue to be widely cited today. In 1986, soon after CELP's introduction, several reduced complexity methods for implementation of the basic CELP algorithm were reported [239], [52], [94], [157]. By circumventing the initial complexity barrier of CELP, these papers indicated that CELP is more than a theoretical curiosity, but rather an algorithm of potential practical importance. It was quickly recognized that real-time implementation of CELP was indeed feasible. The number of studies of CELP coding algorithms has grown steadily since 1986. Numerous techniques for reducing complexity and enhancing the performance of CELP coders emerged in the next seven years, and CELP has found its way into national and international standards for speech coding. Some current speech coding algorithms are hybrids of CELP and other coding approaches. Our definition of CELP encompasses any coding algorithm that combines the features of LPAS with some form of VQ for representing the excitation signal.

Significant landmarks in the history of CELP are the adoptions of several telecommunications standards for speech coding based on the CELP approach. The first of these was the development and adoption of the U.S. Federal Standard 1016, a CELP algorithm operating at 4.8 kb/s, intended primarily for secure voice transmission and incorporating various modifications and refinements of the initial CELP concept. For a description of this standard, see Campbell *et al.* [26]. Another important landmark is the development of a particular CELP algorithm called *vector-sum excited linear prediction* (VSELP) by Gerson and Jasiuk [81] which has been adopted as a standard for North American TDMA digital cellular telephony and, in a modified form, for the Japanese Digital Cellular (JDC) TDMA standard. Very recently, the JDC has adopted a half-rate standard for the Japanese TDMA digital cellular system called *pitch synchronous innovation CELP* (PSI-CELP) [176]. In 1992, the CCITT (now ITU-T) adopted the *low-delay CELP* (LD-CELP) algorithm, developed by Chen *et al.*, [30], [34] as an international standard for 16-kb/s speech coding. Currently, the GSM is establishing a standard for half-rate TDMA digital cellular systems in Europe and the two remaining candidates for the speech-coding component are both CELP algorithms. Also, the Telephone Industry Association (TIA) is now evaluating

eight candidate algorithms for a North American half-rate TDMA digital cellular standard, and most of the candidates are CELP algorithms.

Numerous advances to CELP coding have been developed, to reduce complexity, increase robustness-to-channel errors, and improve quality. Much of this effort is oriented to improving the excitation signal while controlling or reducing the excitation search complexity. Some advances have been made to improve the modeling of the short-term synthesis filter or the quantization of the linear predictor parameters. Below we highlight some of the more important improvements to CELP coding.

B. Closed-Loop Search

In the initial description of CELP [13], [207] only the basic conceptual idea was reported without regard to a practical mechanism for performing the encoder's search operation. Subsequently, some essential details were reported in 1986 for efficiently handling the search operation. In particular, it is efficient to separately compute the *zero-input response* (the ringing) of the synthesis filter after the previously selected optimal excitation vector has passed through it. After accounting for the effect of this ringing, the search for the next excitation vector can be conducted based on a zero initial condition assumption; thus the *zero-state response* of the synthesis filter is computed for each candidate code vector [239], [52]. This use of superposition greatly simplifies the codebook search process.

In [13], the gain scaling factor of the excitation vector was determined from the energy of the original speech prediction error signal, (called the *residual*). The residual is obtained after both short-term prediction and pitch prediction are performed. Subsequently, it was recognized that a *closed-loop gain computation* is easily done so that, in effect, the selection of both gain and code vector is jointly optimized in the analysis-by-synthesis process [7]. This leads to an important quality improvement.

C. Excitation Codebooks

In the *stochastic* excitation codebook initially proposed for CELP, each element of each code vector was an independently generated Gaussian random number. The resulting unstructured character of the codebook is not amenable to efficient search methods, and exhaustive search requires a very high complexity. A variety of structural constraints on the excitation codebook have been introduced to achieve one or more of the following features: reduced search complexity, reduced storage space, reduced sensitivity to channel errors, and increased speech quality. Some of the key innovations are summarized here.

An *overlapped* codebook technique, due to Lin, substantially reduces computation as well as codebook storage [157]. In this method, each code vector of the excitation codebook is a block of samples taken from a larger sequence of random samples by performing a cyclical shift of one or more samples on the sequence. Thus if a one-sample shift is used, a sequence of 1024 Gaussian samples can

generate 1024 distinct code vectors of dimension 40. The effect of filtering each such excitation vector through the synthesis filter is achieved by a single convolution operation on the sequence. The search for the optimal code vector in an overlapped codebook is further accelerated by the use of a modified error weighting criterion introduced by Kleijn *et al.*, allowing a fast recursive computation [136].

A widely used approach to reduce search complexity and storage space is the use of *sparse* excitation codebooks where most of the code vector elements have the value zero. This is usually done in combination with other constraints on the magnitude or location of the nonzero elements. Sparse codebooks for CELP were proposed by Davidson and Gersho [52] and Lin [157]. Sparse codebooks can also be combined with overlapped codebooks. In *ternary* codebooks, proposed by Lin [157], and later Xydeas [254], the nonzero entries of a sparse codebook are forced to be +1 or -1. This can be achieved by hard-limiting the nonzero values of a stochastic codebook or by directly designing specific ternary codebook structures. Salami [200] proposed fixed regularly spaced positions for the nonzero entries so that a short binary word can directly specify the nonzero polarities, eliminating the need for a stored excitation codebook. This technique, called BCELP (for binary CELP), reduces complexity and sensitivity to channel errors while reportedly maintaining good quality. Sparse excitation signals were, of course, central to the technique of MP-LPC and preceded CELP. Attempts to improve MP-LPC by using a codebook of sparse excitation vectors may also be viewed as complexity-reduction methods for CELP. (See in particular Kroon *et al.* [145] and Hernandez-Gomez [71].) Many other sparse codebook schemes have been proposed, for example, Kipper *et al.* [128] and Akamini and Miseki [3].

Another family of excitation codebook methods are based on *lattices*, regularly spaced arrays of points in multiple dimensions. Lattice VQ was proposed in [74] and [75] and extensively studied by various researchers. (See in particular, Gibson and Sayood [84] and Jeong and Gibson [116].) Codebook storage is eliminated since lattices are easily generated and suitable mappings between lattice points (code vectors) and binary words are known. The use of lattice structures for excitation codebooks in CELP has been proposed by Adoul *et al.* who coined the phrase *algebraic codebooks* [1]. In their work, lattice codebooks with all code vectors having the same energy are generated from standard error-correcting codes by replacing the binary symbols 1 and 0 with +1 and -1, respectively. For additional examples of algebraic codebooks for CELP, see [150], [104], [148], and [63]. Le Guyader *et al.* [155] use binary-valued code vectors of unit magnitude so that binary words directly map into excitations without any codebook storage.

An alternative way of generating excitation codebooks is by designing them directly from actual speech files with a suitable *training* algorithm. This is the standard approach to codebook generation in vector quantization (VQ) [76]. However, a closed-loop design method is needed

for CELP, which takes into account the role of the synthesis filter in order to optimize the codebook. A closed-loop design method for vector-predictive coding was reported in [47] and a closed-loop gain-adaptive codebook design was reported in [38]. A similar design specifically applied to CELP coding algorithms was described in [54]. Several codebook design algorithms for CELP were studied by LeBlanc and Mahmoud [152]. Closed-loop codebook design from training data has been used effectively in the LD-CELP algorithm by Chen *et al.* [34] and also in a CELP candidate for the current TIA half-rate standardization due to Serizawa *et al.* [210].

A number of excitation coding methods are based on the use of multistage excitation codebooks, where the excitation is generated as a sum of code vectors, one from each codebook, and the codebooks are sequentially searched. Multistage VQ was introduced by Juang and Gray [121] in 1980 and its application to CELP coding began in 1988. Davidson and Gersho proposed the general use of multiple excitation codebooks with sequential search and separate gain factors for each selected code vector [53]. Kroon and Atal briefly mentioned the idea of multistage excitation codebooks in [141]. Kleijn *et al.* proposed the use of two codebooks, one a stochastic codebook and the other an adaptive codebook [135], [134]. The adaptive codebook, which handles the pitch periodicity, and eliminates the need for a long-term synthesis filter, is now a standard part of most CELP coders and is discussed separately below. Subsequently, many authors have found effective ways to benefit from multiple excitation codebooks for the so-called stochastic (or nonperiodic) excitation component [177], [80], [66], [117], [233], [118], [92], [179], [178], [210]. Multiple codebooks offer reduced search and storage complexity as well as greater robustness to channel errors.

The usual sequential search of multiple codebooks in multistage VQ is suboptimal in comparison with a joint search which, however, would typically have excessive complexity. To approach or attain, with manageable complexity, the jointly optimal excitation as a sum of code vectors with one from each codebook, some form of orthogonalization is needed. Orthogonalization for searching multiple CELP excitation codebooks was proposed by Moreau and Dymarski [177], Gerson and Jasiuk [80], and Johnson and Taniguchi [117]. See also [66] and [179]. Orthogonalization is also used in the PSI-CELP coder [176].

An important and effective CELP coder, the 8-kb/s VSELP coding algorithm of Gerson and Jasiuk, has a multiple codebook structure with two stochastic codebooks and two gain values, and each codebook is itself structured in a manner that can be viewed as a multistage technique [80]. An excitation in VSELP is formed by taking a binary linear combination of N basis vectors, so that each of N codebooks contains two code vectors, a basis vector, and the negative of that vector. With only a small number of basis vectors that need to be passed through the synthesis filter, the excitation search complexity is quite small. With this method, the binary word sent to the receiver directly

specifies the polarities of the linear combination of the basis vectors. Thus a single channel bit error can alter only one term in this sum, causing only a moderate change in the decoded excitation vector. This simple relation between excitation vector and the corresponding binary codeword is very similar to the earlier work of LeGuyader *et al.* [155] and Salami [200].

A number of other complexity-reducing techniques have been proposed which do not fall into the above indicated categories for structuring the excitation codebooks. One interesting example is *CELP with base-band coding* (CELP-BB), due to Kondo and Evans [139], [140], [14]. In this scheme, the short-term linear-prediction residual after low-pass filtering and downsampling is CELP-coded using only long-term (pitch) synthesis. The reduced sampling rate provides a large complexity reduction.

Another approach to reducing search complexity is the *preselection* method, where a simplified but suboptimal search procedure first selects a small subset of candidate code vectors from a codebook and then a second stage search is performed under the desired performance measure to select from the surviving candidates the optimal or near-optimal code vector. Preselection methods were introduced in [52], [94], [43], and have subsequently been applied to other coders, for example [176].

One novel technique for stochastic excitation coding is *pitch synchronous innovation* (PSI) reported by Miki *et al.* as part of their PSI-CELP coder [176]. In this technique, the adaptive codebook is searched in the usual way. If the resulting lag (loosely called the "pitch period") is less than the subframe length, the stochastic codebook is made to have vectors with periodicity based on this lag. This is done by taking the first part of each stored stochastic vector with a number of samples equal to the lag and repeating it till the entire subframe dimension is filled. This technique reportedly gives better performance than the comb filter method in [245] and with lower complexity.

D. Representation of Pitch Periodicity

An important advance in CELP coding came with the introduction of the so-called *adaptive codebook* for representing the periodicity of voiced speech in the excitation signal. In this method, after a search for the optimal time lag, a time-shifted and amplitude-scaled block of prior excitation samples is used as the current excitation; a stochastic codebook is then searched to provide a second vector which is scaled and added to the current excitation. With this technique, only the short-term synthesis filter representing the spectral envelope of the speech is needed—the long-term synthesis filter is eliminated. This method of achieving the needed periodicity in the synthesized speech was introduced by Singhal and Atal for MP-LPC [217] and applied to CELP coding by Kleijn *et al.* who introduced the term *adaptive codebook* [135]. When the pitch period is less than the dimension of the excitation vectors, a modified *virtual search* procedure, proposed in [135] and [122], is generally used. The adaptive codebook has become a standard feature of CELP coders. A somewhat similar

concept to the adaptive codebook was the basis for the *self-excited* coder, introduced by Rose and Barnwell [196], in which no fixed excitation codebook is used but a new excitation segment is obtained (after an initialization) from delayed replicas of the past excitation.

The importance of an accurate reproduction of the periodicity in voiced speech led to the use of high-resolution or *fractional-pitch* methods proposed independently in 1989 by Marques *et al.* [169], [168], and by Kroon and Atal [142]–[144]. In this technique, improved speech quality is achieved by refining the resolution of the pitch period search to a fraction of a sample by means of interpolation. This method increases the size of the adaptive codebook and correspondingly the bit rate for pitch. This leads to more accurate prediction of the current subframe of speech from the filtered past excitation.

With the adaptive codebook, a pitch value is needed for each subframe, leading to a rather high bit rate for pitch information. This can be reduced by differential coding of the pitch within a frame: an average pitch for the frame is first determined, and incremental differences for each subframe are then specified. (See, for example, [265].) A version of this method is used in the FS 1016 CELP coder [26].

An interesting technique called *generalized analysis-by-synthesis* reported by Kleijn *et al.* allows piecewise-linear segments to track the pitch, substantially reducing the bit rate for pitch and eliminating the need for fractional pitch [137].

Further improvements to voiced speech coding were offered by the introduction of the *constrained-excitation* method by Shoham [213]. In this method some of the “gravelly” character of voiced speech in CELP is suppressed, by constraining the gain of the stochastic excitation, based on how good an approximation to the current speech segment is offered by the adaptive codebook. A modified version of constrained excitation, called *pitch sharpening* was proposed by Taniguchi *et al.* [229]. More recently, a smoothing method was proposed by Kleijn which explains the basic impediment to attaining high-pitch periodicity and does not significantly suppress the SNR as do the previous methods [113].

E. Coding of LP Spectral Parameters

In CELP and many other speech coders, the linear prediction parameters are used in modeling the signal and are quantized and transmitted every 20 to 30 ms. These parameters consume a large fraction of the total bit rate for low-rate coders. Hence considerable efforts have been invested in finding efficient ways to represent these parameters, most of them based on the use of VQ. The first application of VQ to quantization of LP parameters is due to Buzo *et al.* [25] in 1980. Most of the recent work on this topic is based on the quantization of *line spectral pairs* (LSP) also known as *line spectral frequencies* (LSF) originally introduced by Itakura in 1979 and first reported by Sugamura and Itakura in [223]. Perhaps the most important study of quantization of the LSF parameters

was presented in the comprehensive report of Kang and Fransen [124] and summarized in their paper [125]. This was the first application of VQ to LSF quantization. They were aiming for an 800-b/s LPC vocoder and designed a 12-b VQ scheme with a weighted distortion measure based on several considerations including auditory perception. Interframe coding of LSF parameters with vector prediction and VQ was later reported in [267] and [248].

Of the various efforts in applying VQ to high-resolution quantization of LSF parameters, the work that is often used as a benchmark for comparing other results is due to Paliwal and Atal [189]. Many authors have studied various alternatives and improvements. Some high-quality work on attaining “transparent” perceptual quality coding of LP parameters at low rates using multistage VQ while controlling complexity was reported by Bhattacharya *et al.* [16]. A multistage VQ technique including a partially adaptive codebook was introduced by Tanaka and Tanichuchi [226]. A computationally efficient algorithm for finding LSP parameters was reported by Kabal and Ramachandran [123]. Other computational methods have also been reported. See, for example, [219] and [203]. Examples of the use of product-code VQ for LSF quantization includes Paksoy *et al.* [187], Chan and Law, [28], Wang *et al.* [248], and Laroia *et al.* [151]. An LSF coding method which adapts to the long-term history of the speech spectral parameters was introduced by Xydeas [255]. Finally, some interesting methods for LSP quantization take into account the effect of channel errors. (See, for example, Secker and Perkis [208] and Hagen and Hedelin [87].) The latter method, based on a linear mapping of codewords in a block code into code vectors, is similar in concept but more general than the prior work of LeGuyader *et al.* [155], Salami [200], and Gerson and Jasuik [80].

F. Multimode Coding and Phonetic Classification

An important advance in CELP algorithms is the use of *multimode* coding or *dynamic bit allocation*, where the bits in each frame are dynamically allocated among the code components (e.g., excitation, pitch, and LP parameters) to adaptively match the local character of the speech. Multimode coding was proposed for low-rate ADPCM coding by Taniguchi *et al.* [236] and subsequently examined for CELP coding by Taniguchi *et al.* [227], [234], [235], Yong and Gersho [264], Kroon and Atal [141], and Jayant and Chen [114]. In one approach to multimode coding, one of several coding modes is selected in each frame by comparing an objective measure of performance offered by each mode. Other criteria have also been used for controlling the mode switching.

Phonetic classification of speech frames in CELP was subsequently introduced as an alternative and more sophisticated means for mode selection. Speech is a highly heterogeneous signal with a time-varying statistical character, ranging from highly predictable quasi-periodic to the almost completely random. The best coding strategies for such extremely diverse classes should adapt to these variations in the waveform and consider both the acoustic

features and the phonetic content of the frame to be coded. With this approach, pattern-classification methods are used to identify each of a small set of phonetically distinct categories, which are defined so that each class is well-suited to a particular coding strategy. Identifying the voiced or unvoiced character of subframes is usually the starting point of such methods. Frames containing *onsets* having a transition from an unvoiced region to a voiced region or from a voiced stop into another voiced phoneme are usually given special attention [244], [246], [247], [188].

G. Interpolative Coding Methods

So much of the effort in CELP coding is devoted to accurately handling the reproduction of periodicity in speech. One promising approach to substantially reduce the rate needed for representing quasi-periodic speech segments is *prototype waveform interpolation* (PWI) introduced by Kleijn [130] and further developed by Kleijn and Granzow [132]. They represent and code a single prototype pitch cycle every 20 to 30 ms and reconstruct the signal by interpolating the sequence of pitch cycles between prototypes either in the time or spectral (i.e., discrete Fourier transform) domain. Linear interpolation of the pitch between prototypes and differential coding of the prototypes in the frequency domain are proposed. For lower rate coding, only the magnitudes of the spectral samples are coded. PWI is used to model the voiced excitation of speech so that synthesis is achieved by passing the reconstructed excitation through an LP synthesis filter. This is combined with conventional CELP coding for unvoiced speech.

By transmitting the coefficients of a Fourier series representation of the prototypes, the coder is in effect specifying a set of harmonic sinusoids which (with the aid of interpolation) could be used to synthesize speech. Although the synthesis does not proceed in this manner, it suggests a strong relation to sinusoidal coding, a parametric coding method to be discussed in the section on vocoders. The PWI approach lies somewhere in a gray region between vocoding and waveform coding.

An effective implementation of a coder based on PWI, called *time-frequency interpolative* (TFI) coding was reported by Shoham [215], [216]. The differential frequency-domain prototype parameters are coded here with a hierarchical VQ technique to obtain a low bit rate while maintaining an accurate representation. Subjective tests indicated an impressive quality at 2.4 and 4 kb/s compared to corresponding conventional schemes for these rates.

H. Postfiltering and Pitch Prefiltering

CELP coders tend to introduce some roughness (a noisy quality or hoarseness) to the reproduced speech. However, it is possible to enhance the speech by a postprocessing operation on the decoder output. Adaptive postfiltering based on the short-term prediction parameters has been proposed earlier to reduce perceived noise in ADPCM speech [195], [111]. This approach was extended to exploit long-term (pitch) prediction parameters in [259]. For LPAS

coders, a particular form of adaptive postfiltering which eliminates most of the muffling effect previously associated with postfiltering methods was introduced for both short-term [29], [37] and long-term postfiltering [29]. By avoiding the spectral tilt that postfiltering tends to introduce in the frequency response, this method has been found effective in enhancing performance for a variety of CELP coders [39]. With minor variations this method has been included in the TIA IS-54 VSELP standard, the JDC digital cellular standard, the U.S. Federal Standard 1016, and the ITU-T G.728 standard.

In the VSELP algorithm [80], the long-term (pitch) postfilter is relocated prior to the short-term synthesis filter, while the short-term postfilter remains after the synthesis filter. This variation, called *pitch prefiltering*, reportedly reduces artifacts sometimes introduced by pitch postfiltering. In another technique, called *adaptive comb filtering*, the pitch filtering is introduced in the encoder search loop as a prefiltering operation prior to the short-term synthesis filter [245]. The idea is to remove some of the annoying noise components from the excitation signal while retaining the various pitch harmonics. To reduce the influence of a random code vectors in one frame on the selection of the code vector in the adaptive codebook for subsequent frames, the comb filter is included in the pitch loop.

I. Other CELP Techniques

The techniques reviewed above constitute only a selected subset of a large variety of techniques and indicate the extensive effort that many researchers and engineers have undertaken to advance the family of CELP coders. Our coverage here represents primarily those methods that have found their way into standards or commercial products or those that have been frequently cited by subsequent papers. It is generally difficult to assess the quality of many individual methods for enhancing coder performance. Reported SNR improvements do not necessarily indicate perceptual quality improvements and self-reported quality assessments of researchers are difficult to calibrate. Thus many other clever or effective contributions may well be overlooked. Some additional examples of contributions to CELP coding are [156], [66], [70], [58], [6], [41], [2], [232], [228], [166].

IV. LOW-DELAY SPEECH CODING

For many applications, the time delay introduced by speech coding into the communications link is a critical factor in overall system performance. While the classical ADPCM algorithm introduces negligible delay, most contemporary coding algorithms must buffer a large block of input speech samples for linear prediction analysis prior to further signal processing. In addition to this buffering delay there is also computational delay in both encoder and decoder as well as delay in unpacking groups of data bits in the decoder. One-way end-to-end coder/decoder (codec) delays of 60 to 100 ms and occasionally even higher are common in speech coders. Algorithms which include error

correcting codes and bit interleaving to combat high channel error rates can incur a substantial additional delay. In 1988, the CCITT (now ITU-T) established a maximum delay requirement of 5 ms with a desired objective of only 2 ms for a 16-kb/s standard algorithm. This culminated in the adoption of the LD-CELP algorithm as CCITT standard G.728 in 1992. The 16-kb/s G.728 algorithm was developed by Chen *et al.* [30]–[32], [34] as an international standard for 16-kb/s speech coding.

Several important ideas for modifying the CELP algorithm to achieve very low coding delay with high quality for both 16- and 8-kb/s rates emerged in a sequence of papers from 1987 to 1992. The key idea of gradient-based backward adaptation of the LP synthesis was well known prior to its application to LPAS (see Gibson [83]). Its application to LPAS coding was reported by Taniguchi *et al.* [230] and by Watts and Cuperman [250]. A high-quality low-delay tree coder with lattice-structured backward prediction filters and backward pitch prediction was reported by Iyengar and Kabal [109]. Backward adaptation of the pitch predictor in a CELP coder was reported by Pettigrew and Cuperman [191]. The use of lattice predictors for backward adaptation of the LP synthesis filter in a CELP configuration was studied by Peng and Cuperman [190]. A low-delay CELP coder called *low-delay vector excitation coding* (LD-VXC) for 16 kb/s was reported in [49].

The LD-CELP algorithm of Chen *et al.* [30]–[34] has the unusual feature of using a 50th-order LP block backward-adaptive synthesis filter, whereas prior CELP coders were almost invariably of 10th order or thereabouts. Furthermore, no pitch predictor or adaptive codebook is used at all. This coder also includes backward gain-adaptive VQ [38], pseudo-Gray coding [268], [61], [269], a novel hybrid of block and recursive windowing for LP analysis, white-noise correction, bandwidth expansion, and spectral smoothing, resulting in excellent robustness to channel errors.

Several studies of low-delay 8-ms coders have shown that in spite of the severe constraint on coding delay (e.g., a 5-ms buffering constraint), a fairly high quality is achievable [49], [40], [256], [258], [102], [266], [103], and [218]. However, these coders did not attain the quality of the G.728 LD-CELP coder at 16 kb/s.

Recently, the ITU-T has been conducting a standardization study of *medium-delay* coders where the delay requirement allows a frame size of up to 16 ms and total codec delay of at most 32 ms [90]. A medium-delay 8-kb/s coder with a 12-ms frame size was recently developed by Adoul *et al.*, and was submitted to the ITU-T as a candidate [154], [202]. A technique for combining forward and backward pitch prediction for medium-delay coding at 8 kb/s was introduced by Kataoka and Moriya at NTT in Japan [126]. More recently, a high-quality medium-delay coder with a 10-ms frame length was reported by Kataoka *et al.* [127]; this is also a candidate for the ITU-T 8-kb/s standardization program. Their method makes use of the conjugate VQ technique [178].

A general review of the principal methods and techniques for low-delay coding is given in [48]. Finally, we note

that an excellent historical review of the development of the G.728 LD-CELP algorithm is given by Chen and Cox in [33]. This fascinating paper offers a unique insider's perspective on speech coding by describing the four-year effort that led to the final algorithm.

V. VARIABLE-RATE SPEECH CODING

For digital transmission, a constant bit-rate data stream at the output of a speech encoder is usually needed. However, for digital storage and for some applications in telecommunications a variable bit-rate output is advantageous. *Variable bit-rate* (VBR) speech coders can exploit the pauses and silent intervals which occur in conversational speech and may also be designed to take advantage of the fact that different speech segments may be encoded at different rates while maintaining a given reproduction quality. Consequently, the *average* bit rate for a given reproduced speech quality can be substantially reduced if the rate is allowed to vary with time. Typically, VBR coders switch from one rate to another at intervals as short as 10 ms. The rate may be controlled *internally* by the statistical character of the incoming speech signal and/or *externally* by the current traffic level in a multi-user communication network. For a recent review of variable rate coding, see [78].

Traditional applications that have motivated the study of variable-rate speech coding include speech storage, packetized voice, and digital speech interpolation (DSI) for digital circuit multiplication equipment (DCME). Multiple-access schemes for wireless communication, particularly code-division multiple access (CDMA) systems have lately become an important application for VBR coding. Recently, the TIA has adopted a CDMA digital cellular telephony standard, known as IS-95, as an alternative to the earlier time-division multiple-access (TDMA) standard IS-54. A variable rate coding algorithm known as QCELP has been evaluated by the TIA as a speech-coding standard for use with IS-95. For a description of the QCELP algorithm, see [73], [60].

Most of the literature in VBR coding report on older methods such as ADPCM [260], [182] or subband coding [252], [162]. Recently, several interesting studies on variable-rate coding based on CELP have been reported [62], [64], [243], [251], [59], [188], [27].

An important component in variable rate speech coding is *voice activity detection* (VAD) which is needed to distinguish active speech segments from pauses, when the speaker is silent and only background acoustical noise is present. An effective VAD algorithm is critical for achieving low average rate without degrading speech quality in variable rate coders. An example of a good VAD scheme for asynchronous transfer mode (ATM) digital transmission is the work of Nakada and Sato (Japan) [182]. The design of a VAD algorithm is particularly challenging for mobile or portable telephones due to vehicle noise or other environmental noise. An important contribution in this category is due to Freeman *et al.* [68] whose VAD technique

has been adopted as a part of the ETSI/GSM digital mobile telephony standard [220], [24]. More recently, a new VAD algorithm was reported which improves on the GSM algorithm in high-background-noise environments considering both vehicle and babble noise [221].

VI. VOCODERS

Speech quality obtained by waveform coding methods including CELP coding is generally found to degrade rapidly as the bit rate drops below 4 kb/s. This is usually explained by the fact that the sparsity of bits (less than 1 b for every two amplitude samples of speech) makes it impossible to adequately approximate the original waveform. Even though LPAS coders attempt to pay more attention to accuracy in reproducing the short-term spectral magnitude in the perceptually important regions, they still devote precious bits to reproducing the general shape of the waveform. Vocoders, on the other hand, make no attempt to reproduce a waveform similar to the original. They generally abandon any attempt to encode the phase of the short-term spectrum and provide only information about the spectral magnitude for the decoder to synthesize speech. Thus vocoders have greater potential for reproducing a signal perceptually similar to the original, while operating at bit rates in the region of 2 kb/s and below, where effective waveform reproduction is virtually impossible.

A. LPC Vocoders

Among the various vocoders, the most widely studied in the past was the classical *linear-prediction coding* (LPC) vocoder due to Itakura and Saito [106] and Atal and Hanauer [10]. A version of the LPC vocoder has been used for many years as a U.S. Government standard, Federal Standard 1015, for secure voice communication. This particular coder, known informally as *LPC-10* because it uses 10th-order linear prediction, is based on a simple model of speech production [241]. The decoder synthesizes speech by passing an excitation signal through an LP synthesis filter. However, unlike LPAS coders, the excitation is generated in the receiver only from a relatively crude specification of the general character of the current speech frame and without actually sending bits that specify the excitation waveform. Each frame is characterized as *voiced* or *unvoiced* and for voiced frames the pitch period is specified. The gain (or, equivalently, the energy) of the excitation is coded and transmitted. The receiver generates a random-noise excitation for unvoiced frames and a train of impulses with the given periodicity for voiced frames.

While vocoders have been studied for many years [205], most of the classical methods are of little current interest because of their poor quality. Often the reproduced speech sounds artificial or “unnatural” with a “buzzy” character and the identity of the speaker is hard to recognize. These coders tend to degrade even further if the original speech contains acoustical background noise of various kinds. Recently, several new vocoder algorithms have emerged which appear

to be competitive with CELP coders at 4 kb/s and superior to CELP at 2.4 kb/s.

McCree and Barnwell have developed a very effective vocoder called the *mixed-excitation vocoder*, based on a number of substantive enhancements to the LPC vocoder concept [172], [173]. A mixed pulse and noise excitation signal is generated and applied to a synthesis filter. The excitation provides a frequency-dependent voicing strength which removes much of the buzzy quality of the standard LPC-10. Separate voicing decisions for different subbands of the speech band are made, similar to that of the MBE coder [86] discussed below. An adaptive spectral enhancement technique, similar to the adaptive postfiltering method of [37], is included as the first (rather than last) stage of the synthesis filter. Subjective test results reported in [173] indicate that the quality at 2.4-kb/s approaches that of the 4.8-kb/s Federal Standard 1016 CELP at 4.8 kb/s under clean speech conditions, and exceeds the 4.8-kb/s standard for noisy speech.

B. Sinusoidal Coders

An important class of vocoders, generically called *sinusoidal coders*, has emerged in recent years as a viable alternative to CELP, particularly for rates of 2–4 kb/s. These coders characterize the evolving short-term spectra of the speech by extracting and quantizing certain parameters which specify the spectra, giving particular attention to the pitch harmonics present in voiced speech. The key feature of sinusoidal coders is that voiced speech is synthesized in the decoder by generating a sum of sinusoids whose frequencies and phases are carefully modified in successive frames to represent and track the evolving short-term spectral character of the original speech. Three main variants of sinusoidal coding have been studied: harmonic coding, sinusoidal transform coding (STC), and multiband excitation coding (MBE). In some versions, the phase information (as well as magnitude) of the sinusoids is obtained from the input speech spectrum and transmitted to the receiver. Other versions, operating at lower rates, do not transmit phase information.

The conceptual introduction of this approach is due to Hedelin [91]. Later Almeida and Tribolet [5] developed *harmonic coding* algorithms and in subsequent papers reported very high quality at 6 to 8 kb/s. (See for example, [167].) More recently, Marques, Almeida, and Tribolet studied critical issues needed to achieve high quality with harmonic coding at lower rates and they presented a 4.8-kb/s version of the algorithm [167].

Another version of sinusoidal coding, called *sinusoidal transform coding* (STC) was developed and extensively refined by McAuley and Quatieri [170], [171]. A third version of sinusoidal coding called *multiband excitation coding* (MBE) was developed by Griffin and Lim [86] and one version called *improved MBE* (IMBE), [23], was subsequently adopted by Inmarsat as a standard for satellite voice communications. A coder based on MBE is currently one of the finalists for the TIA half-rate TDMA digital cellular standardization [183].

Both STC and MBE identify spectral peaks in each successive frame of speech and encode and transmit the amplitude (and in some cases the phase) of these peaks. The receiver synthesizes speech with similar time-varying spectra by controlling the magnitude and phase of a set of sine waves. In MBE, the selected spectral samples are harmonics of the pitch.

Several studies have shown that more efficient quantization can allow the MBE coder to operate with little drop in quality at rates of 2.4 kb/s or below. Brandstein [22], Yeldener *et al.* [197], [261]–[263], Garcia-Mateo *et al.* [72], and Rowe and Secker [198], have shown that the bit rate may be substantially reduced by replacing the spectral modeling with an LP modeling technique. Vector quantization of the spectral magnitudes without the use of LP models has been reported in [175] for 2.4 kb/s and in [183] for 3 kb/s. A different approach was taken by Hassenein *et al.* [89] for 2.4 kb/s, where the MBE analysis is followed by a postprocessor which selects three fixed bandwidth windows and sends spectral information only for these regions. They report comparable quality with full-band MBE for noise-free speech. Recently, a high-quality variable rate was reported, which combines MBE with phonetic classification and a novel spectral VQ technique [270].

Although the amount of work on sinusoidal coders has been very small compared to CELP, indications are that this is a promising approach and will lead to more efforts in future. The sinusoidal coders still retain some remnants of the traditional vocoder type of imperfections in the reproduced speech but generally give a cleaner, crisper reproduction than is available with CELP coders at comparable rates, i.e., 2.4–4 kb/s. An interesting comparison between CELP and sinusoidal coding by Trancoso *et al.* [240] suggests that these techniques are complementary and future work might lead to some merging of these two approaches. In fact, the PWI approach described earlier already performs a similar merger.

VII. AUDIO AND WIDEBAND SPEECH COMPRESSION

Audio coding usually refers to the compression of high-fidelity audio signals, i.e., with 15- or 20-kHz bandwidth for consumer hi-fi, professional audio including motion picture and HDTV audio, and various multimedia systems. Sometimes the term audio coding is also used to refer to wideband speech coding, the compression of 7-kHz bandwidth speech audio for videoteleconferencing and for future integrated subscriber digital network (ISDN) voice communication, where higher quality speech is feasible and desirable.

Digital coding of audio probably began in the early 1970's. Initial efforts simply used uniform or nonuniform (e.g., logarithmic) quantization of audio samples for digital transmission and storage. The British Broadcasting Corporation developed an audio compression scheme called *nearly instantaneous companding audio multiplex* (NICAM) for digital audio transmission. NICAM uses a

block adaptive-gain amplitude scale, where one of five scale factors is specified for every block of 32 samples represented with 10 b/sample. Including overhead bits, the NICAM standard carries a stereo audio signal of 15-kHz bandwidth at a rate of 728 kb/s. For an overview of NICAM and its application to digital transmission, see, for example, [199].

Virtually all the current work in hi-fi audio coding relies on either subband or transform coding to achieve a spectral decomposition of the signal. A transform coding technique with fully overlapping windows called *time-domain alias cancellation* (TDAC) was introduced by Princen and Bradley in 1986 [192] and combines features of both subband and transform coding. Scalar quantization and entropy coding are generally performed on the transformed signal components. Perceptual masking models determine adaptive bit allocations across the spectral components. Important contributions to transform-based audio coding with perceptual masking techniques include Brandenburg [18], Johnston [119], [120], Brandenburg *et al.* [19], [20], and the AC-2 coder of Davidson *et al.* [50], [51]. A collaborative effort led to the ASPEC (adaptive spectral perceptual entropy coding) of high-quality music signals, a transform coding scheme [20].

Recently, a transform coding scheme called AC-3, developed by Todd *et al.* [238] at Dolby Laboratories, was adopted for the multichannel audio portion of the forthcoming high-definition television (HDTV) terrestrial broadcasting standard of the U.S. Federal Communications Commission (FCC). The algorithm operates at a range of bit rates as low as 32 kb/s per channel with up to 5.1 channels. (The 0.1 channel is a low-frequency effects (subwoofer) channel.) The coder uses TDAC filter banks and perceptual masking. Other features include the transmission of a variable-frequency resolution spectral envelope and hybrid backward/forward adaptive bit allocation.

Subband coding has also been the basis of effective audio coding methods. An early example of subband coding is the ITU-T G.722 standard for 7-kHz audio which employs ADPCM to code each of two subbands. For wideband audio compression, a subband coding scheme called *masking pattern adapted subband coding and multiplexing* (MASCAM) was developed by Theile *et al.* [237]. Subsequently, a closely related algorithm called *masking pattern adapted universal subband integrated coding and multiplexing* (MUSICAM) was adopted in Europe for use in digital audio broadcasting. See Dehery *et al.* [56].

Most of the current international interest in audio compression algorithms is centered around the recently completed ISO/MPEG audio standardization. For an outline of the MPEG audio algorithm, see Brandenburg and Stoll [21] or Noll [77]. The standard supports sampling rates of 32, 44.1, and 48 kHz and bit rates ranging from 32 to 448 kb/s per monophonic or stereo channel. The ISO/MPEG audio algorithm has three layers of coding, each of increasing complexity and quality, which offer different versions suited to distinct application needs. A polyphase filter bank of 32 equal-size bands is used. Layer 1 has the lowest com-

plexity: it performs a relatively simple perceptual weighting for bit allocation and is less adaptive to transitory material. Layer 2 is more flexible in sending gains for blocks of samples in one band or shared for two or more adjacent bands. Layer 2 is based on MUSICAM and layer 1 is a simplified version of layer 2. Block companding is used in each subband to quantize blocks of 12 samples each. Quantization resolution is determined by a masking model. Layer 2 differs from Layer 1 only in the joint quantization and coding of each triplet of scaling factors from three consecutive companding blocks in each subband. Layer 3 has the highest complexity and best quality versus rate tradeoff and consists of a combination of the ASPEC transform coding algorithm and the MUSICAM filterbank: an overlap-DCT transform is performed in each subband to provide increased frequency resolution. An adaptive block size for transform coding, inspired by the work of Sugiyama *et al.* [224], [108], mitigates "pre-echos," the audible noise preceding the onset of a sound. Thus both temporal masking as well as the usual frequency-domain masking give an improved performance in Layer 3. Each layer of the MPEG standard also includes a technique for joint coding of two stereo channels. For examples of interesting recent research in stereo coding see [120] and [95].

Two new consumer hi-fi audio products both use audio coding. They are the DCC (digital compact cassette) and the MiniDisc. Both use compression based on perceptual masking methods. The DCC scheme is called *precision adaptive subband coding* or PASC, the MiniDisc system is called *adaptive transform acoustic coding* (ATRAC). The PASC algorithm is essentially the same as the ISO/MPEG Layer 1 algorithm. The ATRAC coder is reported by Tsutsui *et al.* [242] and PASC is described by Lockhoff [161].

There has been considerable interest in wideband coding of speech for ISDN and teleconferencing applications. Effective coding methods here are often based on CELP. (See for example, Laflamme *et al.* [149], Salami *et al.* [201], Ordentlich and Shoham [184], Fuldseth *et al.* [69], and Paulus *et al.* [186].) In some cases, the perceptual weighting used in CELP is modified and low-delay constraints are imposed. Much of this work is quite similar in methodology to the coding of telephone speech.

VIII. STATE OF THE ART

The current state of speech and audio coding from the users perspective is best summarized by describing the performance achievable with established algorithms. We give here only a qualitative description of quality.

The quality of a good connection in a normal wired network telephone call, i.e., *wireline* or *toll* quality, is achievable at 16 kb/s with the LD-CELP G.728 algorithm. The performance at this rate offers low delay and is suited to a large variety of applications. The coder is robust to moderate bit errors, moderate background acoustical noise, a reasonable range of input power levels, and to tandemed network connections where up to three separate encoding/decoding stages may arise.

Speech coding at 8 kb/s with medium delay is now under study for ITU-T standardization. Based on the reported results obtained by two candidate algorithms [127], [154], it is likely that the emerging standard will achieve wireline quality for most operating conditions.

At 4–6 kb/s, with the best current CELP algorithms, the speech exhibits noticeable coding noise, but the features of intelligibility, naturalness, and identifiability of the speaker's voice are retained. This quality is sometimes described as *digital cellular* quality since it is typical of the performance of current standards for digital cellular telephony. Sinusoidal vocoders at this rate region have a different type of distortion but still belong to the same general category of quality. These coders generally suffer from a lack of robustness to nonspeech sounds, such as the noise in a moving vehicle or babble noise. They also introduce a codec delay in the region of 60 to 100 ms.

At 2–3 kb/s, the quality of CELP is further degraded with noisy "hoarse" speech quality and the new generation of vocoders (sinusoidal and mixed-excitation), and PWI or TFI coders appear to offer better quality. The quality at this rate, described as *communications* quality, is intelligible with speaker intonation and identity preserved, but there is a slight loss of naturalness with some slight degree of buzziness in most coders. These coders generally have poor robustness to nonspeech sounds and they also introduce a relatively large coding delay.

At rates below 1 kb/s, the available speech coders are generally vocoders that operate on large segments of speech. They range from barely intelligible to reasonably intelligible but distortion is substantial and speaker identity and naturalness are lost. The coders introduce a delay of hundreds of milliseconds.

Wideband speech coding with 7 kHz based on CELP algorithms can achieve at 32 kb/s the same quality as the ITU-T G.722 algorithm at 64 kb/s. This provides roughly the quality of an FM radio announcer, with a richness notably greater than telephone bandwidth speech and with very high intelligibility, naturalness, and free from any noticeable distortion.

Wideband audio coding in the range of 96 to 128 kb/s per channel for 15- to 20-kHz bandwidth music achieves nearly transparent quality. In other words, most casual listeners will find the music indistinguishable from the original CD audio output for most source materials. Some particularly difficult music segments can reveal an audible distortion when coded at rates of 128 kb/s at least with a more discriminating than average listener. The ISO/MPEG standard with its choice of bit rates represents the current state of the art in audio coding.

IX. CONCLUDING REMARKS

Speech coding in the last decade has been dominated by the extensive studies and advances in the LPAS approach and more specifically with CELP algorithms. As the research frontier moves towards 2.4 kb/s and below, waveform coding with the best CELP techniques available

today appears to be inadequate to meet the increasing quality objectives. Consequently, vocoder studies are experiencing a resurgence today as the focus of research on speech compression is gradually moving to lower bit rates. Sinusoidal and mixed-excitation coders appear to offer the potential for meeting the needs of future standards for 2.4-kb/s coding of speech.

Nevertheless, there are considerable difficulties to surmount before these low-rate coders can become telecommunication standards. In particular, adequate robustness of these coders in the presence of background noise and non-speech sounds or transmission errors is not easy to achieve. For wireless schemes, a very large increase in bit rate is essential to handle the high error rates that arise. Mobile environments also have high levels of background noise and sophisticated adaptive noise cancellation schemes may be needed to achieve adequate performance. There are indeed many challenges ahead for researchers in speech coding.

Since ADPCM was standardized in 1984 (ITU-T G.721 standard), research at 32 kb/s diminished rapidly. With the new 16-kb/s ITU-T G.728 standard, there is relatively little remaining research interest at this bit rate. The recently developed first generation of standards for TDMA digital cellular telephony concentrated on rates ranging from a high of 13 kb/s for the ETSI/GSM RPE-LTP standard, to the (North American) Telephone Industry Association (TIA) IS-54 standard with VSELP at 8 kb/s, and the JDC standard at 6.7 kb/s. All of these cellular applications also provide channel error protection that add to the data rates needed for transmission, but which are not included in above bit-rate specifications. Most of the mobile applications of speech coding currently focus on the second-generation "half-rate" digital speech compression, where the rate is half of the first generation rates. In Japan, the JDC has already adopted a half-rate standard. The U.S. CDMA standard, IS-95, provides for alternative service options and a new variable-rate coding option is currently under study. Following the imminent "half-rate" standards for speech coding in Europe and in the U.S., future efforts in the next five years are likely to concentrate on coding algorithms for 2.4 kb/s with some continued activity at 4.8 and 8 kb/s for some specialized applications.

Audio coding activities have been dominated by the work developed for the MPEG audio standardization. New research in wideband audio coding at lower rates is now in progress, stimulated by plans for future MPEG standards.

There are many topics of importance in speech coding that have not been discussed in this paper due to length limitations. Progress in pitch and voicing detection (e.g., [98]), very-low bit-rate coding at 200 to 600 b/s (see, for example, [100]), coding for robustness to channel errors (e.g., [268] [35], [82] [44], [138]), improved perceptual error criteria for codebook searching (e.g., [249], [209]), objective measures of perceptual quality for evaluating coding algorithms [146], [129], [105] [185], [88], [45], [96], [55], [249], and the development of effective subjective testing methods are some of the additional topics actively being pursued but not covered here.

Speech and audio compression is indeed a very active area of research and development and generally requires a high level of specialization which combines strength in digital signal processing with a good understanding of human psychophysics and modern quantization methods.

ACKNOWLEDGMENT

Several colleagues have offered very helpful corrections and suggestions based on an early draft of this manuscript. In particular, the author wishes to thank S. Dimolitsas, J.-H. Chen, V. Cuperman, B. Kleijn, P. Kroon, M. Iwadare, E. Paksoy, A. Sekey, E. Shlomot, and T. Taniguchi.

REFERENCES

- [1] J.-P. Adoul, P. Mabilieu, M. Delprat, and S. Morissette, "Fast CELP coding based on algebraic codes," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 1957-1960, Apr. 1987.
- [2] M. Akamine and K. Maseki, "CELP coding with an adaptive density pulse excitation model," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 29-32, vol. 1, 1990.
- [3] —, "Efficient excitation model for low bit rate speech coding," in *Proc. 1991 IEEE Int. Symp. on Circuits and Systems*, vol. 1, pp. 586-589, 1991.
- [4] —, "Adaptive bit-allocation between the pole-zero synthesis filter and excitation in CELP," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, 1991), vol. 1, pp. 229-232.
- [5] L. B. Almeida and J. M. Tribolet, "Harmonic coding: a low bit-rate good-quality speech coding technique," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Paris, France, 1982), pp. 1664-1667.
- [6] F. G. Andreotti, V. Maiorano, and L. Vetrano, "A 6.3 kb/s CELP codec suitable for half-rate system," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, 1991), vol. 1, pp. 621-624.
- [7] B. S. Atal, "High-quality speech at low bit rates: multi-pulse and stochastically excited linear predictive coders," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, 1986), pp. 1681-1684.
- [8] B. S. Atal, V. Cuperman, and A. Gersho, *Advances in Speech Coding*. Norwell, MA: Kluwer, 1991.
- [9] —, *Speech and Audio coding for Wireless and Network Applications*. Norwell, MA: Kluwer, 1993.
- [10] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, vol. 50, pp. 637-655, 1971.
- [11] B. S. Atal and J. R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Paris, France, May 1982), vol. 1, pp. 614-617.
- [12] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 3, pp. 247-254, 1979.
- [13] —, "Stochastic coding of speech signals at very low bit rates," in *Proc. Int. Conf. on Communications*, pp. 1610-1613, May 1984.
- [14] S. A. Atungsiri, A. M. Kondoz, and B. G. Evans, "Robust 4.8 kbit/s CELB-BB speech coder for satellite-land mobile communications," *Space Commun.*, vol. 7, no. 4-6, pp. 589-595, Nov. 1990.
- [15] T. P. Bamwell, III, "Recursive windowing for generating autocorrelation coefficients for LPC analysis," *IEEE Trans. Acoust., Speech., Signal Process.*, vol. ASSP-29, pp. 1062-1066, Oct. 1981.
- [16] B. Bhattacharya, W. LeBlanc, S. Mahmoud, and V. Cuperman, "Tree searched multi-stage vector quantization of LPC parameters for b/s speech coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Mar. 1992), vol. 1, pp. 105-108.

- [17] R. Boite, H. Leich, and G. Yang, "Simplification and improvement of the binary coded excited linear prediction (BCELP) for speech coding," in *Signal Processing V. Theories and Applications. (Proc. 5th European Signal Processing Conf.)*, vol. 2, pp. 1211-1214, 1990.
- [18] K. Brandenburg, "OCF—A new coding algorithm for high quality sound signals," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Dallas, TX, Mar. 1992), vol. 1, pp. 141-145.
- [19] K. Brandenburg, H. Gerhauser, D. Seitzer, and T. Sporer, "Transform coding of high quality digital audio at low bit rates—algorithms and implementations," in *Proc. IEEE Int. Conf. on Communications*, vol. 3, pp. 932-936, 1990.
- [20] K. Brandenburg, J. Herre, J. D. Johnston, Y. Mahieux, and E. F. Schroeder, "ASPEC: Adaptive spectral perceptual entropy coding of high quality music signals," presented at the 90th Audio Engineering Soc. Conv., Paris, France, 1991, Reprint 3011.
- [21] K. Brandenburg and G. Stoll, "The ISO/MPEG-audio codec: A generic standard for coding of high quality digital audio," presented at the 92nd Audio Engineering Soc. Conv., Vienna, Austria, Mar. 1992, Preprint 3336.
- [22] M. S. Brandstein, "A 1.5 kbps multi-band excitation speech coder," S.M. thesis, EECS Dept., Mass. Inst. Technol., 1990.
- [23] M. S. Brandstein, P. A. Monta, J. C. Hardwick, and J. S. Lim, "A real-time implementation of the improved MBE speech coder," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), vol. 1, pp. 5-8.
- [24] H. J. Braun, G. Cosier, D. Freeman, A. Gilloire, D. Sereno, C. B. Southcott, and A. Van der Krogt, "Voice control of the pan-European digital mobile radio system," *CSELT Tech. Repts.*, vol. 18, no. 3, pp. 183-187, June 1990.
- [25] A. Buzo, A. H. Gray, Jr., R. M. Gray, and J. D. Markel, "Speech coding based upon vector quantization," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 5, pp. 562-574, Oct. 1980.
- [26] J. P. Campbell, Jr., T. E. Tremain, and V. C. Welch, "The DOD 4.8 KBPS Standard (Proposed Federal Standard 1016)," in *Advances in Speech Coding*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Norwell, MA: Kluwer, 1991, pp. 121-133.
- [27] L. Cellario and D. Sereno, "Variable rate speech coding for UMTS," in *Proc. IEEE Workshop on Speech Coding for Telecommunications* (Ste. Adele, Que., Canada, 1993), pp. 1-2.
- [28] C.-F. Chan and K.-W. Law, "New multistage scheme for vector quantization of PARCOR coefficients," *Electron. Lett.*, vol. 28, pp. 1267-1268, June 18, 1992.
- [29] J.-H. Chen, "Low-bit-rate predictive coding of speech waveforms based on vector quantization," Ph.D. dissertation, Univ. of Calif., Santa Barbara, Mar. 1987.
- [30] ———, "A robust low-delay CELP speech coder at 16 kbits/s," in *Conf. Rec. IEEE Global Telecomm. Conf.* (Dallas, TX, Nov. 1989), vol. 2, pp. 1237-1241.
- [31] "A robust low-delay CELP speech coder at 16 kb/s," in *Advances in Speech Coding*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Dordrecht, The Netherlands: Kluwer, 1991, pp. 25-36.
- [32] "LD-CELP: A high quality 16 kb/s speech coder with low delay," in *Conf. Rec. IEEE Global Telecomm. Conf.* (San Diego, CA, Dec. 1990), vol. 1, pp. 528-532.
- [33] J.-H. Chen, "The creation and evolution of 16 kbit/s LD-CELP: From concept to standard," *Speech Commun.*, vol. 12, no. 2, pp. 103-111, June 1993.
- [34] J.-H. Chen, R. V. Cox, Y.-C. Lin, N. Jayant, and M. J. Melchner, "A low delay CELP coder for the CCITT 16 kb/s speech coding standard," *IEEE J. Sel. Areas Commun.*, vol. 10, pp. 830-849, June 1992.
- [35] J.-H. Chen, G. Davidson, A. Gersho, and K. Zeger, "Speech coding for the mobile satellite experiment," in *Proc. IEEE Int. Conf. on Communications*, pp. 756-763, 1987.
- [36] J.-H. Chen and A. Gersho, "Vector adaptive predictive coding of speech at 9.6 kb/s," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, Apr. 1986), vol. 3, pp. 1693-1696.
- [37] ———, "Real-time vector APC speech coding at 4800 bps with adaptive postfiltering," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 2185-2188, Apr. 1987.
- [38] ———, "Gain-adaptive vector quantization with application to speech coding," *IEEE Trans. on Commun.*, vol. COM-35, no. 9, pp. 918-930, Sept. 1987.
- [39] ———, "Adaptive postfiltering for quality enhancement of coded speech," submitted for publication, 1994.
- [40] J.-H. Chen and M. S. Rauchwerk, "An 8 kb/s low-delay CELP speech coder," in *Conf. Rec. IEEE Global Telecommunications Conf.* (Phoenix, AZ, Dec. 1991), vol. 3, pp. 1894-1898.
- [41] M. Copperi, "Efficient excitation modeling in a low bit-rate CELP coder," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 233-236, May 1991.
- [42] M. Copperi and D. Sereno, "Improved LPC excitation based on pattern classification and perceptual criteria," in *Proc. 7th Int. Conf. on Pattern Recognition* (Montreal, Que., Canada, 1984), pp. 860-862.
- [43] ———, "CELP coding for high-quality speech at 8 kbit/s," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, Apr. 1986), vol. 3, pp. 1685-1689.
- [44] R. V. Cox, W. B. Kleijn, and P. Kroon, "Robust CELP coders for noisy backgrounds and noisy channels," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, May 1989), pp. 739-742.
- [45] D. P. Crowe, "Objective quality assessment," in *Dig. IEE Coll. on Speech Coding-Techniques and Applications* (London, England, Apr. 1992), pp. 5/1-5/4.
- [46] V. Cuperman, "Speech coding," *Adv. Electron. Electron Phys.*, vol. 82, pp. 97-196, 1991.
- [47] V. Cuperman and A. Gersho, "Vector predictive coding of speech at 16 kb/s," *IEEE Trans. Commun.*, vol. COM-33, pp. 685-696, July 1985.
- [48] ———, "Low delay speech coding," *Speech Commun.*, vol. 12, no. 2, pp. 193-204, June 1993.
- [49] V. Cuperman, A. Gersho, R. Pettigrew, J. S. Shynk, and J.-H. Yao, "Backward adaptation techniques for low delay vector excitation coding of speech," in *Conf. Rec. IEEE Global Telecomm. Conf.*, pp. 1242-1246, Nov. 1989.
- [50] G. Davidson, W. Anderson and A. Lovrich, "A low cost adaptive transform decoder implementation for high-quality audio," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Mar. 1992), vol. 2, pp. 193-196.
- [51] G. Davidson, L. Fielder, and M. Antill, "High-quality audio transform coding at 128 kbit/s," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), vol. 2, pp. 1117-1120.
- [52] G. Davidson and A. Gersho, "Complexity reduction methods for vector excitation coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, 1986), pp. 3055-2058.
- [53] ———, "Multiple-stage vector excitation coding of speech waveforms," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (New York, Apr. 1988), pp. 163-166.
- [54] G. Davidson, M. Yong, and A. Gersho, "Real-time vector excitation coding of speech at 4800 bps," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Dallas, TX, Apr. 1987), pp. 2189-2192.
- [55] A. De and P. Kabal, "Rate-distortion function for speech coding based on perceptual distortion measure," in *Conf. Rec., IEEE Global Telecomm. Conf.*, pp. 452-456, 1992.
- [56] Y. F. Dehery, M. Lever, and P. Urcun, "A MUSICAM source codec for digital audio broadcasting and storage," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* vol. 1, pp. 3605-3609, 1991.
- [57] R. Drogo de Jacovo, R. Montagna, F. Perosino, and D. Sereno, "Some experiments of 7 kHz audio coding at 16 kbit/s," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, May 1989), vol. 1, pp. 192-195.
- [58] R. Drogo de Jacovo and D. Sereno, "6.55 kbit/s speech coding for application in the pan-European digital mobile radio cellular system," in *Proc. 5th European Signal Processing Conf.*, vol. 2, pp. 1231-1234, 1990.
- [59] ———, "Embedded CELP coding for variable bit-rate between 6.4 and 9.6 kb/s," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, May 1991), pp. 681-683.
- [60] A. DeJaco, W. Gardner, P. Jacobs, and C. Lee, "QCELP: The North American CDMA digital cellular variable rate speech coding standard," in *Proc. IEEE Workshop on Speech Coding*

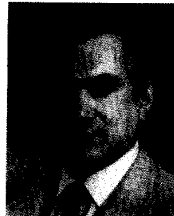
- for *Telecommunications* (Ste. Adele, Que., Canada, 1993), pp. 5-6.
- [61] J. R. B. De Marca and N. S. Jayant, "An algorithm for assigning binary indices to the codevectors of a multi-dimensional quantizer," in *Proc. IEEE Int. Conf. on Communications* (Seattle, WA, June 1987), vol. 2, pp. 1128-1132.
- [62] R. J. Di Francesco, "Real-time speech segmentation using pitch and convexity jump models: application to variable rate speech coding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 741-748, May 1990.
- [63] —, "Algebraic speech coding: ternary code excited linear prediction," *Annal. Telecommun.*, vol. 47, no. 5-6, pp. 214-226, May-June 1992.
- [64] R. Di Francesco, C. Lamblin, A. Leguyader, and D. Massaloux, "Variable rate speech coding with online segmentation and fast algebraic codes," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 233-236, 1990.
- [65] S. Dimolitsas, "Standardizing speech-coding technology for network applications," *IEEE Commun. Mag.*, vol. 31, no. 11, pp. 26-33, Nov. 1993.
- [66] P. Dymarski, N. Moreau and A. Vigier, "Optimal and sub-optimal algorithms for selecting the excitation in linear predictive coders," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, April 1990), pp. 485-488.
- [67] J. L. Flanagan, M. R. Schroeder, B. S. Atal, R. E. Crochiere, N. S. Jayant, and J. M. Tribolet, "Speech coding," *IEEE Trans. Commun.*, vol. COM-27, no. 4, pp. 710-737, Apr. 1979.
- [68] D. K. Freeman, G. Cosier, C. B. Southcott, and I. Boyd, "The voice activity detector for the pan-European digital cellular mobile telephone service," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, May 1989), vol. 1, pp. 369-372.
- [69] A. Fuldseth, E. Harborg, F. T. Johansen, and J. E. Knudsen, "Wideband speech coding at 16 kbit/s for a videophone application," *Speech Commun.*, vol. 11, no. 2-3, pp. 139-148, June 1992.
- [70] C. Galand, J. Menez, and M. Rosso, "Adaptive code excited predictive coding," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1317-1327, June 1992.
- [71] R. Garcia-Gomez, F. J. Casajus-Quiros, and L. Hernandez-Gomez, "Vector quantized multipulse LPC," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Paris, France, Apr. 1987), vol. 4, pp. 2197-2200.
- [72] C. Garcia Mateo, E. Rodriguez Banga, J. L. Alba, and L. A. Hernandez Gomez, "Analysis, synthesis and quantization procedures for a 2.5 kbps voice coder obtained by combining LP and harmonic coding," in *Proc. European Signal Processing Conf.* (Brussels, Belgium, Aug. 1992), vol. 1, pp. 471-474.
- [73] W. Gardner, P. Jacobs, and C. Lee, "QCELP: A variable rate speech coder for CDMA digital cellular," in *Speech and Audio Coding for Wireless and Network Applications*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Norwell, MA: Kluwer, 1993, pp. 77-84.
- [74] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Informat. Theory*, vol. IT-25, no. 4, pp. 373-380, July 1979.
- [75] —, "On the structure of vector quantizers," *IEEE Trans. Informat. Theory*, vol. IT-28, no. 2, pp. 157-166, Mar. 1982.
- [76] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer 1991.
- [77] P. Noll, "Wideband speech and audio coding," *IEEE Commun. Mag.*, vol. 31, no. 11, pp. 34-44, Nov. 1993.
- [78] A. Gersho and E. Paksoy, "Variable rate speech coding for cellular networks," in *Speech and Audio Coding for Wireless and Network Applications*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Norwell, MA: Kluwer, 1993, pp. 77-84.
- [79] A. Gersho and S. Wang, "Recent trends and techniques in speech coding," in *Conf. Rec. 24th Asilomar Conf. on Signals, Systems, Computers* (Pacific Grove, CA, Nov. 1990), vol. 2, pp. 634-638.
- [80] I. Gerson and M. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 8 kb/s," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), vol. 1, pp. 461-464.
- [81] —, "Vector sum excited linear prediction (VSELP)," in *Advances in Speech Coding*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Norwell, MA: Kluwer, 1991, pp. 69-79.
- [82] I. A. Gerson, M. A. Jasiuk, M. J. McLaughlin, and E. H. Winter, "Combined speech and channel coding at 11.2 kbps," in *European Signal Processing Conf.* (Barcelona, Spain, Sept. 1990), vol. 2, pp. 1339-1342.
- [83] J. D. Gibson, "Adaptive prediction for speech differential encoding systems," *Proc. IEEE*, vol. 68, pp. 1789-1797, Nov. 1974.
- [84] J. Gibson and K. Sayood, "Lattice quantization," *Adv. Electron. Electron Phys.*, vol. 72, pp. 259-330, 1988.
- [85] N. Gouvanakis and C. Xydeas, "Advances in analysis by synthesis LPC speech coders," *J. Inst. ERE*, vol. 57, no. 6 (suppl.), pp. S272-S286, Nov./Dec. 1987.
- [86] D. W. Griffin and J. S. Lim, "Multi-band excitation vocoder," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 8, pp. 1223-1235, Aug. 1988.
- [87] R. Hagen and P. Hedelin, "Robust vector quantization in spectral coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 13-16, 1993.
- [88] U. Halka and U. Heute, "A new approach to objective quality-measures based on attribute-matching," *Speech Commun.*, vol. 11, no. 1, pp. 15-30, Mar. 1992.
- [89] H. Hassanein, A. Brind'Amour, and K. Bryden, "A hybrid multiband excitation coder for low bit rates," in *Proc. IEEE Int. Conf. on Wireless Communications* (Vancouver, BC, Canada, 1992), pp. 184-187.
- [90] S. Hayashi and M. Taka, "Standardization activities on 8-kbit/s speech coding in CCITT SGXV," in *Proc. IEEE Int. Conf. on Wireless Communication* (Vancouver, BC, Canada, June 1992), pp. 188-191.
- [91] P. Hedelin, "A tone-oriented voice-excited vocoder," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 205-208, 1981.
- [92] P. Hedelin and A. Bergstrom, "Amplitude quantization for CELP excitation signals," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, 1991), pp. 225-228.
- [93] K. Hellwig, P. Vary, D. Massaloux, and J. P. Petit, "Speech codec for the European mobile radio system," in *Conf. Rec., IEEE Global Telecomm. Conf.* (Dallas, TX, Nov. 1989), vol. 2, pp. 1065-1069.
- [94] L. A. Hernandez-Gomez, F. Casajus-Quiros, A. R. Figueiras-Vidal, and R. Garcia-Gomez, "On the behavior of reduced complexity code-excited linear prediction (CELP)," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, 1986), pp. 469-472.
- [95] J. Herre, E. Eberlein, and K. Brandenburg, "Combined stereo coding," presented at the 93rd Audio Engineering Society Conv., San Francisco, CA, Oct. 1992, Preprint 3369.
- [96] J. Herre, E. Eberlein, H. Schott, and K. Brandenburg, "Advanced audio measurement system using psychoacoustic properties," presented at the 92nd Audio Engineering Society Conv., Vienna, Austria, Mar. 1992, Preprint 3321.
- [97] W. Hess, *Pitch Determination of Speech Signals: Algorithms and Devices*. New York: Springer-Verlag, 1983.
- [98] W. Hess, "Pitch and Voicing Determination," in *Advances in Speech Signal Processing*, S. Furui and M. M. Sondhi, Eds. New York, Basel, Hong Kong: Marcel Dekker, 1992, pp. 3-48.
- [99] M. Honda, "Speech coding using waveform matching based on LPC residual phase equalization," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), vol. 2, pp. 213-216.
- [100] M. Honda and Y. Shiraki, "Very low-bit-rate speech coding," in *Speech Signal Processing*, S. Furui and M. M. Sondhi, Eds. New York, Basel, Hong Kong: Marcel Dekker, 1992, pp. 209-230.
- [101] G. Hult, "Some remarks on a halting criterion for iterative low-pass filtering in a recently proposed pitch detection algorithm," *Speech Commun.*, vol. 10, no. 3, pp. 223-226, Aug. 1991.
- [102] A. Husain and V. Cuperman, "Low delay vector excitation speech coding at 8 kbits/s," in *1992 IEEE Int. Workshop on Intelligent Signal Processing Communication Systems* (Taipei, Taiwan, ROC, Mar. 1992), pp. 148-155.
- [103] —, "Lattice low delay vector excitation for 8 kb/s speech coding," in *Speech and Audio Coding for Wireless and Network Applications*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Norwell, MA: Kluwer, 1993.
- [104] M. A. Ireton and C. S. Xydeas, "On improving vector excitation coders through the use of spherical lattice codebooks

- (SLCs)," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 57-60, May 1989.
- [105] H. Irii, K. Kurashima, N. Kitawaki, and K. Itoh, "Objective measurement method for estimating speech quality of low-bit-rate speech coding," *NTT Rev.*, vol. 3, no. 5, pp. 79-87, Sept. 1991.
- [106] F. I. Itakura and S. Saito, "Analysis-synthesis telephony based on the maximum likelihood method," in *Proc. 6th Int. Congr. on Acoustics* (Tokyo, Japan, 1968), pp. C17-20.
- [107] K. Itoh and N. Kitawaki, "Real and artificial speech signals for objective quality evaluation of speech coding systems," *Electron. Commun. in Japan*, pt. 3 (*Fund. Electron. Sci.*), vol. 72, no. 11, pp. 1-9, Nov. 1989.
- [108] M. Iwadare, A. Sugiyama, F. Hazu, A. Hirano, and T. Nishitani, "A 128 kb/hi-fi audio codec based on adaptive transform coding with adaptive block size MDDCT," *J. Selected Areas Commun.*, vol. 10, no. 1, pp. 138-144, 1992.
- [109] V. Iyengar and P. Kabal, "A low delay 16 kb/s speech coder," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 243-246, 1988.
- [110] N. S. Jayant, Ed., *Waveform Quantization and Coding*. New York: IEEE Press, 1976.
- [111] N. S. Jayant and V. Ramamoorthy, "Adaptive postfiltering of 16 kb/s-ADPCM speech," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, Apr. 1986), pp. 829-832.
- [112] N. S. Jayant, "Signal compression: technology targets and research directions," *IEEE Trans. Selected Areas Commun.*, vol. 10, no. 5, pp. 795-818, June 1992.
- [113] N. S. Jayant, J. Johnston, and R. Sofranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, no. 10, pp. 1385-1422, Oct. 1993.
- [114] N. S. Jayant and J. H. Chen, "Speech coding with time-varying bit allocations to excitation and LPC parameters," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, May 1989), vol. 1, pp. 65-69.
- [115] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [116] D. G. Jeong and J. D. Gibson, "Uniform and piecewise uniform lattice vector quantization for memoryless Gaussian and Laplacian sources," *IEEE Trans. Informat. Theory*, vol. 39, no. 3, pp. 786-804, May 1993.
- [117] M. Johnson and T. Taniguchi, "Pitch-orthogonal code-excited LPC," in *Conf. Rec. IEEE Global Telecomm. Conf.*, vol. 1, pp. 542-546, Dec. 1990.
- [118] —, "Low-complexity multi-mode VXC using multi-stage optimization and mode selection," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, May 1991), vol. 1, pp. 221-224.
- [119] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Selected Areas Commun.*, vol. 6, pp. 314-323, Feb. 1988.
- [120] J. D. Johnston and A. J. Ferreira, "Sum-difference stereo transform coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Apr. 1992), vol. 2, pp. 569-572.
- [121] B.-H. Juang and A. H. Gray, Jr., "Multiple stage vector quantization for speech coding," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Paris, France, Apr. 1982), vol. 1, pp. 597-600.
- [122] P. Kabal, J. L. Moncet, and C. C. Chu, "Synthesis filter optimization and coding: applications to CELP," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (New York, Apr. 1988), vol. 2, pp. 569-572.
- [123] P. Kabal and R. P. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 6, pp. 1419-1426, Dec. 1986.
- [124] G. S. Kang and L. J. Fransen, "Low-bit-rate speech encoders based on line-spectrum frequencies (LSFs)," *Naval Res. Lab., Rep.* 8857, Nov. 1984.
- [125] —, "Application of line-spectrum pairs to low-bit-rate speech encoders," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Tampa, FL, Mar. 1985), pp. 244-247.
- [126] A. Kataoka and T. Moriya, "A backward adaptive 8 kb/s speech coder using conditional pitch prediction," in *Conf. Rec. IEEE Global Telecommunication Conf.*, pp. 1889-1893, Dec. 1991.
- [127] A. Kataoka, T. Moriya, and S. Hayashi, "An 8-kbit/s speech coder based on conjugate structure CELP," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Minneapolis, MN, 1993), vol. 2, pp. 592-595.
- [128] U. Kipper, H. Reininger, and D. Wolf, "Improved CELP coding using adaptive excitation codebooks," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, May 1991), vol. 1, pp. 237-240.
- [129] N. Kitawaki, "Research of objective speech quality assessment," *NTT Rev.*, vol. 3, no. 5, pp. 65-70, Sept. 1991.
- [130] W. B. Kleijn, "Continuous representations in linear predictive coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, May 1991), vol. 1, pp. 201-204.
- [131] —, "Improved pitch prediction," in *Proc. IEEE Workshop on Speech Coding for Telecommunications* (Ste. Adele, Que., Canada, 1993), pp. 19-20.
- [132] W. B. Kleijn and W. Granzow, "Methods for waveform interpolation in speech coding," *Dig. Signal Process.*, vol. 1, no. 4, pp. 215-230, Oct. 1991.
- [133] W. B. Kleijn, "Encoding speech using prototype waveforms," *Proc. IEEE Trans. Acoust., Speech, Signal Process.*, vol. 1, no. 4, pp. 386-399, Oct. 1993.
- [134] W. B. Kleijn, D. J. Krasinski, and R. H. Ketchum, "An efficient stochastically excited linear predictive coding algorithm for high quality low bit rate transmission of speech," *Speech Commun.*, vol. 7, no. 3, pp. 305-316, Oct. 1988.
- [135] W. B. Kleijn, D. J. Krasinski, and R. H. Ketchum, "Improved speech quality and efficient vector quantization in SELP," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (New York, 1988), pp. 155-158.
- [136] —, "Fast methods for the CELP speech coding algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 8, pp. 1330-1342, Aug. 1990.
- [137] W. B. Kleijn, R. P. Ramachandran, and P. Kroon, "Generalized analysis-by-synthesis coding and its application to pitch prediction," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Mar. 1992), vol. 1, pp. 337-40.
- [138] W. B. Kleijn and R. A. Sukkar, "Efficient channel coding for CELP using source information," *Speech Commun.*, vol. 11, pp. 547-566, 1992.
- [139] A. M. Kondoz and B. G. Evans, "CELP base-band coder for high quality speech coding at 9.6 to 2.4 kbps," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 159-162, 1988.
- [140] A. M. Kondoz, K. Y. Lee, and B. G. Evans, "Improved quality CELP base-band coding of speech at low-bit rates," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, May 1989), vol. 1, pp. 128-131.
- [141] P. Kroon and B. S. Atal, "Strategies for improving the performance of CELP coders at low bit rates," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (New York, 1988), vol. 1, pp. 151-154.
- [142] —, "On improving the performance of pitch predictors in speech coding systems," in *Proc. IEEE Workshop on Speech Coding for Telecommunications* (Vancouver, BC, Canada, Sept. 1989), pp. 49-50.
- [143] —, "Pitch predictors with high temporal resolution," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), vol. 2, pp. 661-664.
- [144] —, "On the use of pitch predictors with high temporal resolution," *IEEE Trans. Signal Process.*, vol. 39, no. 3, pp. 733-735, 1991.
- [145] P. Kroon, E. F. Deprettere, and R. J. Sluyter, "Regular-pulse excitation: A novel approach to effective and efficient multiple-pulse coding of speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 1054-1063, Oct. 1986.
- [146] R. F. Kubichek, E. A. Quinicy, and K. L. Kiser, "Speech quality assessment using expert pattern recognition techniques," in *Proc. IEEE Pacific Rim Conf. on Computers, Communications, and Signal Processing*, Jun. 1989.
- [147] G. Kubin, B. S. Atal, and W. B. Kleijn, "Performance of noise excitation for unvoiced speech," in *Proc. IEEE Workshop on Speech Coding for Telecommunications* (Ste. Adele, Que., Canada), pp. 1-2.
- [148] C. Laflamme, J.-P. Adoul, H. Y. Su, and S. Morissette, "On reducing computational complexity of codebook search in CELP coders through the use of algebraic codes," *Proc. Int. Conf.*

- on *Acoustics, Speech, and Signal Processing*, pp. 177–180, Apr 1990.
- [149] C. Laflamme, J.-P. Adoul, R. Salami, S. Morissette, and P. Mabileau, "16 kbps wideband speech coding technique based on algebraic CELP," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, 1991), pp. 13–16.
- [150] C. Lamblin, J. P. Adoul, D. Massaloux, and S. Morissette, "Fast CELP coding based on the Barnes-Wall lattice in 16 dimensions," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, 1989), vol. 1, pp. 61–64.
- [151] R. Laroia, N. Phamdo, and N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, May 1991), vol. 5, pp. 641–644.
- [152] W. P. LeBlanc and S. A. Mahmoud, "Structured codebook design in CELP," in *Proc. 2nd Int. Mobile Satellite Conf.* (Ottawa, Ont., Canada, June 1990), pp. 667–672.
- [153] J. I. Lee and C. K. Un, "Multistage self-excited linear predictive speech coder," *Electron. Lett.*, vol. 25, no. 18, pp. 1249–1251, Aug. 1989.
- [154] R. Lefebvre, R. Salami, C. Laflamme, and J.-P. Adoul, "8 kbit/s coding of speech with 6 ms frame-length," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Minneapolis, MN, Apr. 1993), vol. 2, pp. 612–615.
- [155] A. Le Guyader, D. Massaloux, and F. Zucher, "A robust and fast CELP coder at 16 kbit/s," *Speech Commun.*, vol. 7, no. 2, pp. 217–226, July 1988.
- [156] A. Le Guyader, D. Massaloux, and J. P. Petit, "Robust and fast code-excited linear predictive coding of speech signals," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, May 1989), vol. 1, pp. 120–123.
- [157] D. Lin, "New approaches to stochastic coding of speech sources at very low bit rates," in *Signal Processing III: Theories and Applications*, I. T. Young *et al.*, Eds. Amsterdam, The Netherlands: Elsevier, North-Holland, 1986 pp. 445–447.
- [158] —, "Speech coding using efficient pseudo-stochastic block codes," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Dallas, TX, Apr. 1987), pp. 1354–1357.
- [159] X. Lin, R. A. Salami, and R. Steele, "High quality audio coding using analysis-by-synthesis technique," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, May 1991), vol. 5, pp. 3617–2620.
- [160] T. M. Liu and H. Hoegge "Phonetically-based LPC vector quantization of high quality speech," in *Proc. European Conf. on Speech Communication and Technology* (Paris, France, Sept. 1989), vol. 2, pp. 356–359.
- [161] G. C. P. Lohhoff, "Precision adaptive sub-band coding (PASC) for the digital compact cassette (DCC)," *IEEE Trans. Consumer Electron.*, vol. 38, no. 4, pp. 784–789, Nov. 1992.
- [162] L. M. Lundheim and T. A. Ramstad, "Variable rate coding for speech storage," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, Apr. 1986), vol. 4, pp. 3079–3082.
- [163] J. Makhoul and M. Berouti, "Adaptive noise spectral shaping and entropy coding in predictive coding of speech," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 63–73, Feb. 1979.
- [164] J. Makhoul, S. Roucos and Gish, "Vector quantization in speech coding," *Proc. IEEE*, vol. 73, no. 11, pp. 1551–1588, Nov. 1985.
- [165] K. Mano and T. Moriya, "4.8 kbit/s delayed decision CELP coder using tree coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 21–24, 1990.
- [166] —, "Delay decision CELP coding using tree coding," *Trans. Inst. Electron., Informat. Commun. Eng.—A*, vol. J74A, no. 4, pp. 619–627, Apr. 1991.
- [167] J. S. Marques, L. B. Almeida, and J. M. Tribolet, "Harmonic coding at 4.8 kb/s," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 17–20, 1990.
- [168] J. S. Marques, J. M. Tribolet, I. M. Trancoso, and L. B. Almeida, "Pitch prediction with fractional delays in CELP coding," in *European Conf. on Speech Communication and Technology* (Paris, France, Sept. 1989), vol. 2, pp. 509–512.
- [169] —, "Pitch prediction with fractional delays in CELP coding," in *European Conf. on Speech Communication and Technology* (Paris, France, Sept. 1989), vol. 2, pp. 509–512.
- [170] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech Signal Processing*, vol. ASSP-34, pp. 744–754, 1986.
- [171] —, "Low-rate speech coding based on the sinusoidal model," in *Advances in Acoustics and Speech Processing*, M. Sondhi and S. Furui, Eds. New York: Marcel Dekker, 1992, pp. 165–207.
- [172] A. V. McCree and T. P. Barnwell, III, "Improving the performance of a mixed excitation LPC vocoder in acoustic noise," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Mar. 1992), vol. 2, pp. 163–166.
- [173] —, "Implementation and evaluation of a 2400 bps mixed excitation LPC vocoder," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Minneapolis, MN, Apr. 1993), vol. 2, pp. 159–162.
- [174] P. Mermelstein "G. 722, A new CCITT coding standard for digital transmission of wideband audio signals," *IEEE Comm. Mag.*, vol. 26, no. 1, pp 8–15, Jan. 1988.
- [175] P. C. Meuse, "A 2400 bps multi-band excitation vocoder," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), vol. 1, pp. 9–12.
- [176] S. Miki, K. Mano, H. Ohmuro, and T. Moriya, "Pitch synchronous innovation CELP (PSI-CELP)," in *Proc. European Conf. on Speech Communication and Technology* (Berlin, Germany, Sept. 1993), pp. 261–264.
- [177] N. Moreau and P. Dymarksi, "Mixed excitation CELP coder," in *Proc. European Conf. on Speech Communication and Technology* (Paris, France, Sept. 1989), pp. 322–325.
- [178] T. Moriya, "Two-channel conjugate vector quantizer for noisy channel speech coding," *IEEE J. Selected Areas Commun.*, vol. 10, no. 5, pp. 866–874, June 1992.
- [179] T. J. Mousley and P. W. Elliot, "Fast vector quantisation using orthogonal codebooks," in *6th Int. Conf. on Digital Processing of Signals in Communications* (Loughborough, England, Sept. 1991), pp. 294–299.
- [180] T. J. Mousley and P. R. Holmes, "An adaptive voiced/unvoiced speech classifier," in *European Conf. on Speech Communication and Technology* (Paris, France, Sept. 1989), vol. 1, pp. 466–469.
- [181] J.-M. Muller, "Improving performance of code-excited LPC-coders by joint optimization," *Speech Commun.*, vol. 8, no. 4, pp. 363–369, Dec. 1989.
- [182] H. Nakada and K.-I. Sato, "Variable rate speech coding for asynchronous transfer mode," *IEEE Trans. Commun.*, vol. 38, pp. 277–284, Mar. 1990.
- [183] M. Nishiguchi, J. Matsumoto, R. Wakatsuki, and S. Ono "Vector quantized MBE with simplified V/UV division at 3.0 kbps," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Minneapolis, MN, Apr. 1993), vol. 2, pp. 151–154.
- [184] E. Ordentlich and Y. Shoham, "Low-delay code-excited linear predictive coding of wideband speech at 32 kbps," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, May 1991), pp. 9–12.
- [185] B. Paillard, P. Mabilieu, S. Morissette, and J. Soumagne, "PERCEVAL: Perceptual evaluation of the quality of audio signals," *J. Audio Eng. Soc.*, vol. 40, no. 1–2 pp. 21–31, Jan.–Feb. 1992.
- [186] J. Paulus, C. Antweiler, and C. Gerlach, "High quality coding of wideband speech at 24 kbit/s," in *Proc. European Conf. on Speech Communication and Technology* (Berlin, Germany, Sept. 1993), vol. 2, pp. 1107–1110.
- [187] E. Paksoy, W.-Y. Chan, and A. Gersho, "Vector quantization of speech LSF parameters with generalized product codes," in *Proc. Int. Conf. Spoken Language Processing* (Banff, Alta, Canada, Nov. 1992), pp. 33–36.
- [188] E. Paksoy, K. Srinivasan, and A. Gersho, "Variable rate speech coding with phonetic segmentation," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processin* (Minneapolis, MN, Apr. 1993), vol. 2, pp. 155–158.
- [189] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Toronto, Ont., Canada, May 1991), pp. 661–664.
- [190] R. Peng and V. Cuperman, "Variable-rate low-delay analysis-by-synthesis speech coding at 8–16 kbit/s," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 29–32 1991.
- [191] R. Pettigrew and V. Cuperman, "Backward pitch prediction for low-delay speech coding," in *Conf. Rec., IEEE Global Telecommunications Conf.*, pp. 34.3.1–34.3.6, Nov. 1989
- [192] J. Princen and A. Bradley, "Analysis/synthesis filter bank design based on time-domain aliasing cancellation," *IEEE Trans.*

- Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1153–1161, 1986.
- [193] R. P. Ramachandran and P. Kabal, "Stability and performance analysis of pitch filters in speech coders" *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 7, pp. 937–946, 1987.
- [194] —, "Pitch prediction filters in speech coding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 4, pp. 467–477, 1989.
- [195] V. Ramamoorthy and N. S. Jayant, "Enhancement of ADPCM speech by adaptive postfiltering," *Bell Syst. Tech. J.*, vol. 63, no. 8, pp. 1465–1475, Oct. 1984.
- [196] R. C. Rose and T. P. Barnwell III, "The self-excited vocoder—an alternative approach to toll quality at 4.8 kbs." in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, Apr. 1986), vol. 1, pp. 453–456.
- [197] D. Rowe, W. Cowley, and P. Secker, "A multiband excitation linear predictive hybrid speech coder," in *Proc. European Conf. on Speech Communication and Technology (Eurospeech 91)* (Genova, Italy, Sept. 1991), vol. 1, pp. 239–242.
- [198] D. Rowe and P. Secker, "A robust 2400 bit/s MBE-LPC speech coder incorporating joint source and channel coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Mar. 1992), vol. 2, pp. 141–144.
- [199] F. Rumsey, "Hearing both sides-stereo sound for TV in the UK," *IEE Rev.*, vol. 36, no. 5, pp. 173–176, May 10, 1990.
- [200] R. A. Salami, "Binary code excited linear prediction (BCELP): new approach to CELP coding of speech without codebooks," *Electron. Lett.*, vol. 25, no. 6, pp. 401–403, Mar. 1989.
- [201] R. Salami, C. LaFlamme, and J.-P. Adoul, "Real-time implementation of a 9.6 kbit/s ACELP wideband speech coder," in *Conf. Rec., IEEE Global Telecomm. Conf.*, pp. 447–451, 1992.
- [202] —, "ACELP speech coding at 8 kbit/s with a 10 ms frame: a candidate for CCITT standardization," presented at the IEEE Workshop on Speech Coding for Telecommunications, Oct. 1993.
- [203] S. Saoudi, J. M. Boucher and A. Le Guyader, "A new efficient algorithm to compute the LSP parameters for speech coding," *Signal Process.*, vol. 28, pp. 201–212, 1992.
- [204] R. W. Schafer and L. R. Rabiner, "Digital representation of speech signals," *Proc. IEEE*, vol. 63, pp. 662–677, Apr. 1975.
- [205] M. R. Schroeder, "Vocoders: Analysis and synthesis of speech," *Proc. IEEE*, vol. 54, no. 3, pp. 720–734, May 1966.
- [206] M. Schroeder and B. S. Atal, "Rate distortion theory and predictive coding," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Atlanta, GA, Mar. 1981), vol. 1, pp. 201–204.
- [207] —, "Code-excited linear prediction (CELP) high quality speech at very low bit rates," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 937–940, Mar. 1985.
- [208] P. Secker and A. Perkis, "Joint source and channel trellis coding of line spectrum pair parameters," *Speech Commun.*, vol. 11, pp. 149–158, 1992.
- [209] D. Sen, D. H. Irving, and W. H. Holmes, "Use of an auditory model to improve speech coders," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 411–414, 1993.
- [210] M. Serizawa, K. Ozawa, T. Miyano, and T. Nomura, "M-LPC/LPC speech coding at bit-rates below 4 kbps," in *Proc. IEEE Workshop on Speech Coding for Telecommunications*, pp. 45–46, 1993.
- [211] Y. Shiraki and M. Honda, "LPC speech coding based on variable-length segment quantization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 9, pp. 1437–1444, Sept. 1988.
- [212] Y. Shoham, "Constrained excitation coding of speech at 4.8 kb/s," in *Proc. IEEE Workshop on Speech Coding for Telecommunications* (Vancouver, BC, Canada, Sept. 1989), p. 65.
- [213] —, "Constrained-stochastic excitation coding of speech at 4.8 kb/s," in *Proc. Int. Conf. on Spoken Language Process* (Kobe, Japan, Nov. 1990).
- [214] —, "Constrained-excitation coding of speech at 4.8 kb/s," in *Advances in Speech Coding*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Dordrecht, The Netherlands: Kluwer, 1991, pp. 339–348.
- [215] —, "High-quality speech coding at 2.4 to 4.0 kbps based on time-frequency interpolation," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Minneapolis, MN, Apr. 1993), vol. 2, pp. 167–170.
- [216] —, "High-quality speech coding at 2.4 kbps based on time-frequency interpolation," in *Proc. European Conf. on Speech Communication and Technology* (Berlin, Germany, Sept. 1993), vol. 2, pp. 741–744.
- [217] Y. Shoham, S. Singhal, and B. S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 1.3.1–1.3.4, 1984.
- [218] R. Soheili, A. M. Kondoz, and B. G. Evans, "An 8 kb/s LD-CELP with improved excitation and perceptual modelling," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Minneapolis, MN, 1993), vol. 2, pp. 616–619.
- [219] F. K. Soong and B.-H. Juang, "Optimal quantization of LSP parameters," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 1, pp. 15–24, Jan. 1993.
- [220] C. B. Southcott, D. Freeman, G. Cosier, D. Sereno, A. van der Krogt, A. Gilloire, and H. J. Braun, "Voice control of the pan-European digital mobile radio system," in *Conf. Rec. IEEE Global Telecomm. Conf.*, vol. 2, pp. 1070–1074, Nov. 1989.
- [221] K. Srinivasan and A. Gersho, "Voice activity detection for digital cellular networks," in *Proc. IEEE Workshop on Speech Coding for Telecommunications*, pp. 85–86, Oct. 1993.
- [222] L. C. Stewart, "Trellis data compression," Information Systems Lab., Tech. Rep. L905-1, Stanford Univ., July 1981. See also L. C. Stewart, R. M. Gray, and Y. Linde, "The design of trellis waveform coders," *IEEE Trans. Commun.*, vol. 30, pp. 702–710, Apr. 1982.
- [223] N. Sugamura and F. Itakura, "Line spectrum representation of linear predictor coefficients of speech signal and its statistical properties," *Trans. Inst. Electron., Commun. Eng. Japan*, vol. J64-A, pp. 323–340, 1981.
- [224] A. Sugiyama, F. Hazu, M. Iwadare, and T. Nishitani, "Adaptive transform coding with an adaptive block size (ATC-ABS)," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), vol. 2, pp. 1093–1096.
- [225] M. Taka, P. Combescure, P. Mermelstein, and F. Westall, "Overview of the 64 kb/s (7 kHz) audio coding standard," in *Conf. Rec. IEEE Global Telecomm. Conf.* (Houston, TX, 1986), pp. 17.1.1–17.1.6.
- [226] Y. Tanaka and T. Taniguchi, "Efficient coding of LPC parameters using adaptive prefiltering and MSVQ with partially adaptive codebook," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 5–8, 1993.
- [227] T. Taniguchi, S. Unagami, and R. M. Gray, "Multimode coding: a novel approach to narrow- and medium-band coding," *J. Acoust. Soc. Amer.*, suppl. 1, vol. 84, p. S12, Nov. 1988.
- [228] T. Taniguchi, F. Amano, and M. Johnson, "Improving the performance of CELP-based speech coding at low bit rates," in *Proc. IEEE Int. Symp. on Circuits and Systems* (Singapore, June 1991), vol. 1, pp. 590–593.
- [229] T. Taniguchi, M. Johnson, and Y. Ohta, "Pitch sharpening for perceptually improved CELP, and the sparse-delta codebook for reduced computation," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 241–244, May 1991.
- [230] T. Taniguchi, K. Okazaki, F. Amano, and S. Unagami, "4.8 kbps CELP coding with backward prediction," in *IEIC-Nat. Conv. Rec.* (in Japanese), p. 1346, Mar. 1987.
- [231] T. Taniguchi, Y. Tanaka, and Y. Ohta, "Tree-structured delta codebook for an efficient implementation of CELP," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Mar. 1992), vol. 1, pp. 325–328.
- [232] T. Taniguchi, Y. Tanaka, Y. Ohta, and F. Amano, "Improved CELP speech coding at 4 kb/s and below," in *Proc. Int. Conf. on Spoken Language Processing* (Banff, Alta, Canada, Nov. 1992), pp. 41–44.
- [233] T. Taniguchi, Y. Tanaka, A. Sasama and Y. Ohta, "Principal axis extracting vector excitation coding: high quality speech at 8 kb/s," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), vol. 1, pp. 241–244.
- [234] T. Taniguchi, S. Unagami, and R. M. Gray, "Multimode coding: application to CELP," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, Scotland, May 1989), vol. 1, pp. 156–159.
- [235] T. Taniguchi, Y. Tanaka, and R. M. Gray, "Speech coding with dynamic bit allocation (multimode coding)," in *Advances in*

- Speech Coding*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Dordrecht, The Netherlands: Kluwer, 1991.
- [236] T. Taniguchi, S. Unagami, K. Iseda, and S. Tominaga, "ADPCM with a multiquantizer for speech coding," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 410-424, Feb. 1988.
- [237] G. Theile, G. Stoll, and M. Link, "Low bit-rate coding of high-quality audio signals. An introduction to the MASCAM system," *EBU Rev. Tech.*, no. 230, pp. 158-181, Aug. 1988.
- [238] C. Todd, G. Davidson, M. Davis, L. Fielder, B. Link, and S. Vernon, "AC-3: Flexible perceptual coding for audio transmission and storage," presented at the 96th Audio Eng. Soc. Conv., Amsterdam, The Netherlands, Feb. 26-Mar. 1 1994, Preprint 3796.
- [239] I. M. Trancoso and B. S. Atal, "Efficient procedures for finding the optimum innovation in stochastic coders," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, 1986), pp. 2379-2382.
- [240] I. M. Trancoso, J. S. Marques, C. M. Ribeiro, "CELP and sinusoidal coders: two solutions for speech coding at 4.8-9.6 kbps," *Speech Commun.*, vol. 9, no. 5-6, pp. 389-400, Dec. 1990.
- [241] T. E. Tremain, "The government standard linear predictive coding algorithm: LPC-10," *Speech Technol.*, pp. 40-49, Apr. 1982.
- [242] K. Tsutsui, H. Suzuki, O. Shimoyoshi, M. Sonohara, K. Agagiri, and R. M. Heddl, "ATRAC: Adaptive transform acoustic coding for MiniDisc," in *Conf. Rec., Audio Eng. Soc. Conv.* (San Francisco, CA, Oct. 1992).
- [243] S. V. Vaseghi, "Finite state CELP for variable rate speech coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, Apr. 1990), pp. 37-40.
- [244] S. Wang and A. Gersho, "Phonetically-based vector excitation coding of speech at 3.6 kbps," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 49-52, May 1989.
- [245] —, "Improving the excitation for phonetically-segmented VXC speech coding below 4 KBPS," in *Conf. Rec., IEEE Global Telecomm. Conf.* (San Diego, CA, 1990), vol. 2, pp. 946-950.
- [246] —, "Phonetic segmentation for low rate speech coding," in *Advances in Speech Coding*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Dordrecht, The Netherlands: Kluwer, 1991, pp. 225-234.
- [247] —, "Improved phonetically-segmented vector excitation coding at 3.4 kb/s," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Mar. 1992), vol. 1, pp. 349-352.
- [248] S. Wang, E. Paksoy, and A. Gersho, "Product code vector quantization of LPC parameters," in *Speech and Audio Coding for Wireless and Network Applications*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Dordrecht, The Netherlands: Kluwer, 1993, pp. 251-258.
- [249] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE J. Selected Areas Commun.*, vol. 10, pp. 819-829, June 1992.
- [250] L. Watts and V. Cuperman, "A vector ADPCM analysis-by-synthesis configuration for 16 kb/s speech coding," in *Conf. Rec. IEEE Global Telecomm. Conf.*, pp. 275-279, 1988.
- [251] D. Y. K. Wong, "Issues on speech storage," in *IEE Coll. on Speech Coding—Techniques and Applications* (London, England, Apr. 1992), pp. 711-714.
- [252] G. Wu and J. W. Mark, "Multiuser rate subband coding incorporating DSI and buffer control," *IEEE Trans. Commun.*, vol. 38, no. 12, p. 2159-2165, Dec. 1990.
- [253] C. Xydeas, "An overview of speech coding techniques," in *Dig. IEE Colloq. Speech Coding—Techniques and Applications* (London, Apr. 1992), pp. 111-125.
- [254] C. S. Xydeas, M. A. Ireton, and D. K. Baghadrani, "Theory and real time implementation of a CELP coder at 4.8 and 6.0 Kbit/sec using ternary code excitation," in *Proc. IERE 4th Int. Conf. on Digital Processing of Signals in Communications* (Univ of Loughborough, Sept. 20-23, 1988), pp. 167-174.
- [255] C. S. Xydeas and K. K. M. So, "A long history quantization approach to scalar and vector quantization of LSP coefficients," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1-4, 1993.
- [256] J.-H. Yao, J. J. Shynk, and A. Gersho, "Low delay vector excitation coding of speech at 8 kbit/sec," in *Conf. Rec., IEEE Global Telecomm. Conf.* (Phoenix, AZ, Dec. 1991), vol. 3, pp. 695-699.
- [257] —, "Low-delay VXC at 8 Kb/s with interframe coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* vol. 1, (San Francisco, CA, Mar. 1992), pp. 45-48.
- [258] —, "Low-delay speech coding with adaptive interframe pitch tracking," in *Proc. IEEE Int. Conf. on Communication* (Geneva, Switzerland, May 1993).
- [259] Y. Yatsuzuka, S. Iizuka, and A. T. Yamazaki, "A variable rate coding by APC with maximum likelihood quantization from 4.8 kbit/s to 16 kbit/s," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Tokyo, Japan, Apr. 1986), pp. 3071-3074.
- [260] Y. Yatsuzuka, "Highly sensitive speech detector and high-speed voiceband data discriminator in DSI-ADPCM systems," *IEEE Trans. Commun.*, vol. COM-30, pp. 739-750, Apr. 1982.
- [261] S. Yeldener, A. M. Kondo, and B. G. Evans, "High quality multiband LPC coding of speech at 2.4 kb/s," *Electron. Lett.*, vol. 27, no. 14, pp. 1287-1289, July 4, 1991.
- [262] —, "Natural sounding speech coder operation at 2.4 kb/s and below," in *Proc. IEEE Inter. Conf. Wireless Communication* (Vancouver, BC, Canada, 1992), pp. 176-179.
- [263] S. Yeldener, W. Ma, A. M. Kondo, and B. G. Evans, "Low bit rate speech coding at 1.2 and 2.4 kb/s," in *IEE Coll. on Speech Coding—Techniques and Applications* (London, England, Apr. 1992), pp. 611-614.
- [264] M. Yong and A. Gersho, "Vector excitation coding with dynamic bit allocation," in *Conf. Rec. IEEE Global Telecomm. Conf.* (Phoenix, AZ, Dec. 1991), vol. 3, pp. 695-699.
- [265] "Efficient encoding of the long-term predictor in vector excitation coders," in *Advances in Speech Coding*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Dordrecht, The Netherlands: Kluwer, 1991, pp. 29-338.
- [266] J.-H. Yao, J. J. Shynk, and A. Gersho, "Low-delay VXC at 8 Kb/s with interframe coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (San Francisco, CA, Nov. 1988), vol. 1, pp. 290-294.
- [267] M. Yong, G. Davidson, and A. Gersho, "Encoding of LPC spectral parameters using switched-adaptive interframe vector prediction," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (New York, Apr. 1988), pp. 402-405.
- [268] K. Zeger and A. Gersho, "Zero redundancy channel coding in vector quantisation," *Electron. Lett.*, vol. 23, no. 12, pp. 654-656, 1987.
- [269] —, "Pseudo-Gray coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147-2158, Dec. 1990.
- [270] A. Das and A. Gersho, "A variable rate natural quality parametric coder at 1.5 kb/s," in *IEEE Int. Conf. on Communications*, vol. 1, pp. 216-220, May 1994.



Allen Gersho (Fellow, IEEE) received the B.S. degree from the Massachusetts Institute of Technology, Cambridge, in 1960, and the Ph.D. degree from Cornell University, Ithaca, NY, in 1963.

He was at Bell Laboratories from 1963 to 1980. He is now Professor of Electrical and Computer Engineering at the University of California, Santa Barbara (UCSB). His current research activities are in signal compression methodologies and algorithm development for speech, audio, image, and video coding. He holds patents on speech coding quantization, adaptive equalization, digital filtering, and modulation and coding for voiceband data modems. He is co-author with R. M. Gray of the book *Vector Quantization and Signal Compression* (Dordrecht, The Netherlands: Kluwer Academic Publishers, 1992), and co-editor of two books on speech coding. He received NASA "Tech Brief" awards for technical innovation in 1987, 1988, and 1992. In 1980, he was co-recipient of the Guillemin-Cauer Prize Paper Award from the IEEE Circuits and Systems Society. He received the Donald McClennan Meritorious Service Award from the IEEE Communications Society in 1983, and in 1984 he was awarded the IEEE Centennial Medal. In 1992, he was co-recipient of the 1992 Video Technology Transactions Best Paper Award from the IEEE Circuits and Systems Society. He served as a member of the Board of Governors of the IEEE Communications Society from 1982 to 1985, and is a member of various IEEE technical, award, and conference management committees. He has served as Editor of the *IEEE COMMUNICATIONS MAGAZINE* and Associate Editor of the *IEEE TRANSACTIONS ON COMMUNICATIONS*.