UNIVERSITY OF CALIFORNIA

Los Angeles

Adaptive Wireless Voice Communications with Embedded Source and Channel Coding

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Electrical Engineering

by

Benjamim Tang

1995

UMI Number: 9610861

UMI

300 North Zeeb Road
Ann Arbor, MI 48103

The dissertation of Benjamim Tang is approved.

Abeer A. H. Alwan

Patricia A. Keating

Kung Yao

Gregory J. Pottie, Committee Chair

University of California, Los Angeles

1995

To my wife, Judy,

and my parents, Jong and Kay,

for valuing who I am,

much more than what I do.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to extend my deep-felt gratitude to all those who have helped, encouraged, supported, and stood by me these past 4 years. It has been a privilege to do graduate research at UCLA, and I thank you all for the experience.

I thank my wife Judy, who is the only reason I need to try to be the best person I can be. If I could learn just one thing in my lifetime, it would be how to be a good husband to her.

I thank my parents, Jong and Kay Tang, for their love, sacrifice, encouragement, affirmation, and discipline. I also thank my sister, Nancy Imamoto, her husband Daniel, her daughters Alyson and Melissa, my in-laws, Kanji and Jane Sahara, and my brother-in-law, Richard Sahara. No one could ask for a more loving family.

I thank Professor Greg Pottie, my graduate advisor, and Professor Abeer Alwan for their instruction, guidance, and support. They have guided my research activities, provided me with much insight and direction, and encouraged my development as an independent researcher.

I thank Professor Kung Yao and Professor Patricia Keating for their enthusiastic participation in my doctoral committee, and their support as I brought my research efforts to a conclusion.

I thank Albert Shen for his contributions to this work. Albert's development of the perceptual model, conceptualization, implementation and optimization of the speech coding algorithm, and conduction of the listening tests were significant milestones in this research activity, and I gladly share the accomplishments in this work with him.

I thank Victor Lin, Charles Wang, Chris Hansen, Rei Chen, Jonathan Min, Ben Belzer, and the rest of the gang at the UCLA Communication Systems & Optimization Laboratory, the UCLA Integrated Circuits & Systems Laboratory, the UCLA Speech Processing & Auditory Perception Laboratory, and the UCLA Image Processing Laboratory. I've greatly enjoyed going through classes, homework, exams, discussion groups, seminars, and especially lunch breaks together.

I thank Cindy Torsney, Brian Wong, Bill Ashley, Steve Mishima, and the rest of my management at TRW for their support. I have appreciated the flexibility in work schedule and assignments, the financial support from the fellowship, and the access to computing and research resources. I hope their investment will yield many future contributions from me.

I thank Sue, Jay, and Joe Hosoda, and the Print 'n Copy Center for printing this manuscript, and for being such great relatives.

I thank the financial supporters of this work, NSF grant IRI-9309418, ARPA/CSTO Contract J-FBI-93-112, and TRW Space & Electronics Group Fellowship.

I would like to acknowledge the re-publishing of previously copyrighted material. Portions of sections 3.1 and 3.2 were previously published in part at the 1995

International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95, Detroit MI, May 9-12, 1995, pp. 249-252, and also in Albert Shen's 1994 UCLA M.S. Thesis. Section 3.1.2.1 was previously published in part at the 1995 International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95, Detroit MI, May 9-12, 1995, pp. 1324-1327. Section 3.1.4 was previously published in part at the 5th International Conference on Signal Processing Applications and Technology, ICSPAT '94, Dallas TX, October 18-21, 1994, pp. 456-460.

I lift up my eyes to the hills —
where does my help come from?
My help comes from the Lord,
the Maker of heaven and earth.

Psalm 121:1-2

VITA

November 3, 1964	Born, São Paulo, Brazil
1985	B.S., Electrical Engineering UCLA Los Angeles, California
1986	M.S., Electrical Engineering UCLA Los Angeles, California
1985 - 1988 1989 - 1991 1991 - 1995	Member of Technical Staff Section Head Department Staff Digital Products Center TRW Redondo Beach, California

PUBLICATIONS AND PRESENTATIONS

Shen, A., Tang, B., Alwan, A., and Pottie, G., "A Robust Variable Rate Speech Coder," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95, Detroit MI, May 1995, pp. 249-252.

Tang, B., and Pottie, G., "Embedded Non-Uniform Scalar Quantizer for Variable Rate Applications," International Conference on Signal Processing Applications and Technology, October 1994, pp. 456-460.

Tang, B., Shen, A., Pottie, G., and Alwan, A., "Spectral Analysis of Subband Filtered Signals," IEEE International Conference on Acoustics, Speech, and Signal Processing, May 1995, pp. 1324-1327.

ABSTRACT OF THE DISSERTATION

Adaptive Wireless Voice Communications
with Embedded Source and Channel Coding

by

Benjamim Tang
Doctor of Philosophy in Electrical Engineering
University of California, Los Angeles, 1995
Professor Gregory J. Pottie, Chair

Embedded source and channel coding offer improvements in quality and robustness over traditional non-embedded methods and make low complexity adaptive wireless voice communication possible. Embedded coding allows source rate reduction through truncation of the information bitstream, and channel rate increase through puncturing of the channel symbols. The source and channel coders always operate at a base rate, and rate adaptation does not require interaction with the coders.

This research introduces methodologies for embedded source coding, embedded channel coding, and adaptive transmission that significantly improve the performance, flexibility, robustness, and complexity of wireless voice communication systems.

The embedded speech coding methodology is based on subband decomposition and perceptually optimized bit allocation and prioritization. Dynamic bit allocation and prioritization is combined with embedded quantization to implement robust, high quality encoding with little performance degradation relative to non-embedded implementations. The coder output is scalable from high quality at higher bit rates to lower quality at lower rates, supporting a wide range of service and resource utilization.

The embedded channel coding methodology is based on block based tailbiting rate compatible punctured convolutional codes. The use of tailbiting in short block lengths results in a scheme with little degradation relative to traditional convolutional coding, but with much greater flexibility. Novel reduced complexity, fixed computational workload decoders are presented for both traditional maximum likelihood and soft output decoding of tailbiting codes.

Adaptive transmission capability is demonstrated in two schemes making use of the flexibility of embedded source and channel coding. A rate adaptive scheme maximizes the perceptual quality of the voice transmission by determining the optimal source and channel rates given the existing channel conditions and bandwidth availability. A progressive transmission scheme uses ARQ to adapt the source and channel rates to improve transmission reliability and maximize throughput. Compared to traditional fixed rate, variable rate, and unequal error protection schemes, adaptive transmission provides improved error rate, robustness to poor transmission and congested network conditions, and better utilization of channel resources.

1. Introduction

Communication is the process of exchanging information. With electronics, communication over free space, wires, and fiber-optics is possible. Voice communications, the most widespread use of electronic communications, has been in place for nearly a century in the form of the public wireline telephone network. It has become an indispensable part of people's everyday life throughout the world. Now, wireless communications using cellular telephones allow nearly ubiquitous access to this huge infrastructure, making it possible to "reach out and touch" someone just about anywhere, anytime.

Electronic communications may be used to transmit information that is either analog or digital in nature. Analog information is represented by a continuous one-dimensional or multi-dimensional waveform, such as voice, audio, and video. Digital information is represented by a string of numbers or symbols, such as text, numerical lists, or even analog information that has been transformed so that it is represented digitally.

Analog information may be transmitted by superimposing the waveform over a carrier, using analog modulation schemes such as Amplitude Modulation (AM), Frequency Modulation (FM) or Phase Modulation (PM). Noise is introduced by the electronics and the transmission channel, so some degradation of the reconstructed signal always occurs in analog transmissions, and the degradation is dependent on the channel quality.

On the other hand, digital information may be transmitted without errors, where the channel quality only degrades the amount of information that may be reliably transmitted.

The capability for reliable transmission over poor channels has always motivated the use and evolution of digital communications: telegraph communications in the 1800's, space communications in the 1950's, long distance telecommunications point-to-point trunk links in the 1970's, and wireless communications in the 1990's [113]. Digital systems have also benefited from the invention of the transistor in 1947 and the integrated circuit in 1961, which led to explosive growth in computing power, and resulted in the availability of significant digital processing capability.

Due to these advantages, communication systems have evolved to take advantage of digital communications even for sources that are analog in nature. Signals that are analog in nature must be transformed into digital information in order to be transmitted over a digital communication system. This transformation requires two basic operations: sampling and quantization.

Sampling is the process by which discrete values or samples are used to represent a continuous waveform. The sampling theorem, based on Nyquist's celebrated work in 1928 [83], states that perfect reconstruction of an analog waveform is possible through interpolation of periodic samples, if the sampling is done at higher than twice the highest frequency component of the waveform. Thus, for bandlimited analog signals, sampling can be done without loss of information, as long as the sampling is done at high enough frequency.

Quantization is the process by which the discrete samples are assigned a representation from a finite set or codebook. Unlike the sampling operation, quantization destroys information, since a representation from a finite set cannot represent all possible analog values. However, if some reconstruction error or distortion is acceptable, it is possible to represent the analog value with sufficiently high fidelity given a large enough representation set. Thus, an analog source may be digitized through sampling and quantization, provided the codebook is sufficiently large to meet Nyquist sampling and fidelity criteria.

Communications theory provides an additional framework for evaluating both digitization and transmission of analog sources over a digital communication system. Digital communication systems are typically modeled and described as shown in Figure 1.1. The operations performed by the communication system are separated into three functions: source coding, channel coding, and modulation. Source coding attempts to describe the source with the smallest possible digital representation, and the smallest possible error or distortion. The source signal may be analog or digital in nature. Channel coding attempts to correct errors occurring during transmission by introducing redundancy in the digital representation transmitted. Modulation represents the digital information with waveforms for transmission over a physical transmission channel. The channel introduces degradation in the form of noise and interference, affecting the ability of the receiver to properly detect the transmitted waveform. The communications system is designed such that the effect of this degradation is handled in an efficient manner.

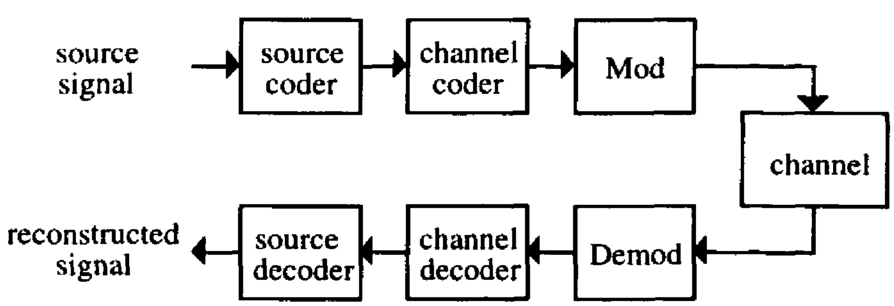


Figure 1.1 - Digital Communication System Model

Source coding studies information and its conversion into digital form. The information content of the source is related to the uncertainty in source outcome. For example, a deterministic source does not convey any information. Shannon [93] introduced a measure of information content, called entropy, which is essentially the minimum number of symbols required to represent the source eliminating all its redundancy. For analog sources, however, some distortion must be introduced to obtain a digital representation due to quantization. Shannon also introduced rate-distortion theory [94] to obtain theoretical limits on the minimum distortion achievable with a number of symbols given the characteristics of the source, although infinite decoding delay and complexity may be required to achieve this level of distortion. Motivated by the theoretical limit, the goal of source coding is to obtain a digital representation of the source that can be represented most efficiently, with the fewest number of bits, and with the least amount of distortion [1][15][75].

Channel coding studies the limitations of transmitting information over a non-ideal channel. Shannon [93] showed that the bandwidth and signal-to-noise ratio (SNR) limit the channel capacity, the amount of information that may be reliably transmitted over the channel. Error-free communication is possible below the channel capacity by introducing redundancy in the transmitted symbols, although infinite decoding delay and complexity is required to achieve error-free performance. In practical implementations, increasingly small error probabilities may be attained as decoding delay and complexity are allowed to increase. The goal of channel coding is to introduce just enough redundancy so that reliable transmission can be achieved with the highest throughput at a given decoding complexity.

Modulation is the process of converting discrete symbols to be transmitted into the actual waveforms used for transmission over the physical channel. The waveform representation may be a simple symbol-by-symbol mapping, or it may be more complicated, grouping a set of symbols into a set of waveforms.

The theoretical limit of the digital communication system is set by the channel capacity given the transmission channel, and the rate distortion of the source given the channel capacity. In principle, the three basic processes, source coding, channel coding, and modulation, can be designed and optimized independently to achieve the theoretical performance. In practice, there may be implementation limitations, and joint design of the three processes is sometimes required to maximize the achievable performance.

Optimization of the digital communication system performance is particularly important for wireless voice communications. Wireless communication systems must cope with a channel that is non-Gaussian, non-stationary, subject to deep fading, and highly congested. Large cells, fast moving cars, low RF frequency, multipath propagation, and high signal interference contribute to a hostile transmission environment for wireless communications. Consumers are accustomed to the level of quality provided by the wireline telephone system, which provides a much more stable transmission environment. Economic considerations and the scarcity of bandwidth drives the system to maximize the number of users that can be supported without significantly increasing resources. These factors contribute to significant technical challenges in delivering wireless voice communications at the required level of performance.

Analog cellular systems suffer from several shortcomings, among which are the inefficient utilization of available bandwidth, high output power, and the inability to support applications other than voice due to the low data rate available. Due to these shortcomings, the cellular network is currently shifting from analog to digital transmission. Digital transmission offers the potential to lower costs, increase the number of users, increase the quality of service, improve privacy, and introduce new applications. As of yet, however, this potential has not been fulfilled. Digital cellular systems are only slowly gaining acceptance, and the majority of consumers are still served by the older analog technology.

Current digital cellular systems have been driven to use low bit rate speech coders in order to increase system capacity, resulting in poorer quality and robustness than offered by the existing wireline telephone network (toll-quality). Recent advances in speech coding research have focused on lowering the data rate while trying to maintain near-toll quality and reasonable robustness in the wireless environment. Next generation wireless systems will significantly expand the range of applications from those available today, applications that require higher quality voice and audio support, including video conferencing, broadcasting, and multimedia. Flexible implementations will be required to support varied network topologies, transmission environments, and service requirements.

The goal of this research is to add capability, address deficiencies, and improve performance by investigating the major issues associated with the implementation of wireless voice communications, particularly the framework for optimization of the source and channel coding given the large variability in transmission channel quality. The result of this work was the development of a methodology for adaptive transmission using embedded source and channel coding, which allows higher quality, robustness, and flexibility to be achieved, without significant increase in complexity compared to fixed transmission schemes.

Embedded source coding is a methodology that allows partial reconstruction of the source at higher distortion, given only a portion of the bitstream is available to the decoder, as shown in Figure 1.2. Embedded source coding is advantageous in that it allows the source rate to be modified simply by truncating the bitstream. Thus it is attractive for variable rate applications, where interaction with the source coder to modify the rate may not be desirable, and robust applications, where transmission of the entire bitstream cannot be guaranteed.

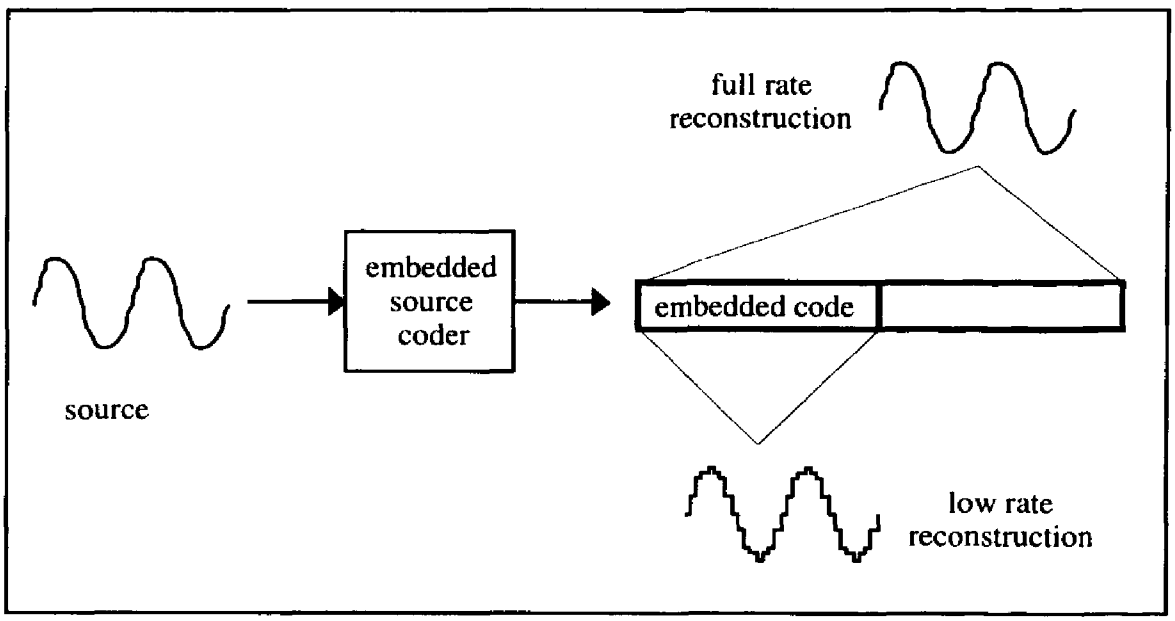


Figure 1.2 - Embedded Source Coding

Embedded channel coding is an error protection methodology that allows a higher rate code to be embedded into the bitstream of a lower rate base code. The codewords at the higher rate are obtained simply by not transmitting, or puncturing, channel symbols according to a puncturing pattern, as represented by the darkened areas in Figure 1.3. The base code and puncturing pattern are selected such that the transmitted information is still decodable with partial transmissions. Only redundant information is lost through puncturing, although the error protection capability is reduced. The partial transmission then corresponds only to an increase in channel rate and decrease in coding gain, not loss of information. Embedded channel coding is advantageous in that it allows rate adaptation and progressive transmission to be easily implemented by simply selecting which channel symbols are to be transmitted.

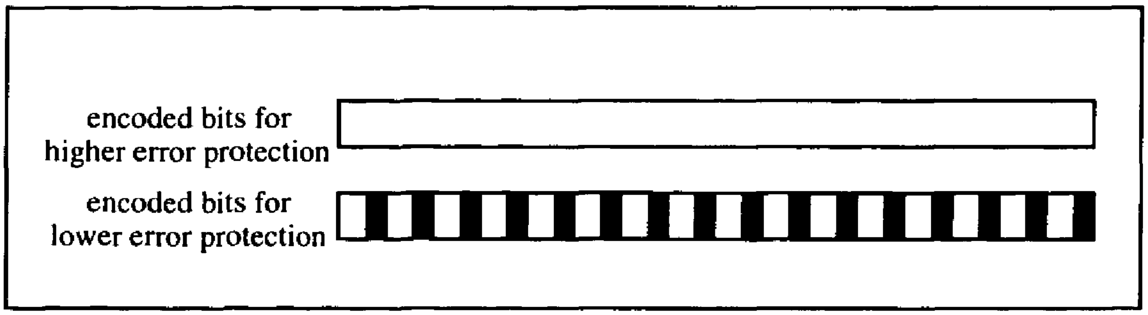


Figure 1.3 - Embedded Channel Coding

For transmission channels that are highly variable, such as the wireless radio channel, adaptive transmission offers a significant improvement over fixed transmission schemes. Adaptive transmission allows the source and channel coding to be dynamically optimized such that the overall quality of the reconstructed signal is maximized. The flexibility of adaptive transmission systems makes robust, high quality wireless voice communication possible over a large range of channel conditions. Embedded source and channel coding are ideally suited to adaptive transmission since they allow rate adaptation to be implemented with little complexity increase over fixed transmission schemes.

The statement of the problem addressed in this research is formally provided in Section 2.4, following a brief presentation of technical background material on wireless communications, speech coding, and channel coding in Sections 2.1 - 2.3.

Section 3 describes methodologies for improved implementation of embedded source coders and an embedded subband speech coder design developed jointly with Albert Shen [96][97][104]. The main contribution of the work in this area is the demonstration that embedded source coding is possible with very little performance degradation and complexity increase relative to comparable non-embedded designs. Since the performance penalty is small, embedded source coding is very attractive for wireless communications because of the added flexibility it provides. The embedded speech coder employs several novel techniques:

- Use of an infinite impulse response (IIR) quadrature-mirror filterbank (QMF) instead of a finite impulse response (FIR) QMF to reduce filterbank implementation complexity, reduce delay, and improve stopband attenuation, using the design technique developed by Zhong Jiang [41].
- Dynamic bit allocation and prioritization based on perceptual modeling and a novel optimization cost measure, the interpolated mask threshold.
- Subband spectral analysis, a technique for performing reduced complexity transform-based spectral analysis on subband samples with an aliasing error cancellation technique.
- Embedded quantization with optimal non-uniform embedded scalar quantization.

Section 4 describes methodologies for improved implementation of embedded channel coders. The main contribution of the work in this area is the demonstration that embedded channel coding is possible with very little performance degradation and complexity increase relative to traditional non-embedded fixed rate channel coding methods. The embedded channel coder demonstrates several novel techniques:

- Block based application of tailbiting rate compatible punctured convolutional coding as a flexible approach to obtain adaptive unequal error protection.
- Asymptotically optimal fixed computational workload decoding of tailbiting convolutional codes with the Tailbiting Viterbi Decoder (TVD).
- Soft output decoding of tailbiting convolutional codes with the Soft Output bi-directional Viterbi Decoder (SOBVD).

Section 5 describes development of adaptive transmission with embedded source and channel coding. The main contribution here is the demonstration that adaptive transmission with embedded source and channel coding offers substantial improvements in the quality and robustness, without significantly added complexity, compared to non-adaptive implementation of wireless voice communications. The adaptive transmission schemes demonstrate several novel techniques:

- Adaptive source and channel rate selection with unequal error protection as a flexible, robust, and high performance scheme for wireless voice communications.
- Progressive transmission with automatic-repeat request (ARQ) protocol for better robustness and performance than forward error correction (FEC) alone when a reverse channel is available.

- Suitability of adaptive transmission to multi-hop networks, since adaptation does not require interaction with the source and channel coders.

Section 6 summarizes this work, discusses its accomplishments and relevance, and presents possible topics for further research.

The methodology presented as a result of this research shows that adaptive transmission with embedded source and channel coding is an attractive methodology for high quality, robust, flexible, low complexity wireless voice communications. The main accomplishment of this work is the added motivation for development of adaptive transmission and embedded source and channel coding techniques, demonstrating that embedded coding can make wireless voice communication systems not only simpler, but also better.

2. Technical Background

The basic theories, issues, approaches, and implementation techniques for wireless communications, speech coding, and channel coding are discussed in the following sections to familiarize the reader with the research subjects addressed in this work. Section 2.1 describes wireless communications and applications, providing a context for understanding the main issues related to the implementation of wireless voice communications. Section 2.2 describes the digitization of speech from source coding, speech and auditory modeling, and coder implementation points of view. Section 2.3 describes protection from errors in transmission with channel coding. Section 2.4 describes the problems in implementation of wireless voice communications addressed in this work.

2.1 Wireless Communications

The goal of wireless communications is to provide ubiquitous access to a global information network capable of carrying all conceivable types of data traffic. Consumers will demand that access be not merely available; it must be convenient, reliable, and affordable.

The implementation of wireless communications is primarily driven by the characteristics of the radio channel. The radio channel is a dispersive, time-varying medium. The physical location of the user and base station in a metropolitan environment guarantees the presence of obstructions. Communication is then possible only through diffractions, reflections, and attenuated transmissions through the obstructions.

The presence of obstructions leads to transmission degradations such as shadowing and fading. Shadowing occurs when there is no direct line of sight between the transmitter and receiver, causing the received power to be dependent on many different transmission paths. Fading occurs because very small changes in the location of the obstructions can lead to large phase changes in the many transmission paths. The received power will then vary greatly for locations varying by as little as one wavelength, with the variations governed by Rayleigh statistics. If a single dominant path is available, the variations are not as severe, and are governed by Rician statistics.

An additional difficulty with the wireless environment is the degradation in signal strength compared to free space propagation. While the transmitted power in free space only drops as the square of the distance, the two main paths in a wireless transmission, a direct path and a ground reflected path, nearly cancel each other out due to the 180° phase change in the reflection. The cancellation of the two main paths results in much lower received power, inversely proportional to the fourth power of the distance.

The radio channel in a wireless environment is thus susceptible to large variations in transmission quality. In addition, the radio spectrum must be shared with numerous other users and applications. The high number of users leads to increased noise and interference, which degrade the quality of the radio channel, and ultimately limit the available resources to each user. The scarcity of bandwidth and the poor transmission environment requires that wireless communications be developed in an efficient and effective manner.

The cellular telephone network is an early example of both the capability that is achievable and the hurdles that need to be overcome with wireless communications.

Cellular telephones provide access to the wireline telephone network from most populated areas via wireless communication with a network of base stations. The main feature of cellular networks is the high capacity achieved by frequency reuse through spatial separation of users sharing the same spectrum. Capacity can be increased by using more cells, but this eventually reaches practical limitations due to added cost of network hardware, and increased interference due to the smaller cell spacings.

The first generation cellular systems, introduced in the late 1970's and early 1980's and still in use today, AMPS in North America, TACS and NMT in Europe, and NTT in Japan, use analog frequency modulation (FM) for speech transmission [84]. Each user is assigned a channel under frequency division multiple access (FDMA), consisting of a frequency band in his cell. These systems have been hugely successful, with over 20 million subscribers in the United States alone [84].

Second generation cellular systems, such as IS-54 and IS-95 in North America, GSM in Europe, and PDC in Japan, introduced digital transmission to increase the capacity of the cellular system [84]. Digital transmission also offers the potential to support new digital services, reduce RF transmission power to improve handset battery life, improve communication privacy through encryption, and reduce system complexity through continued increase in digital hardware integration capability. However, the second generation wireless systems support only low data rate communications, at high error rates, with voice quality noticeably inferior to wireline systems, and in most cases, even analog cellular [88].

Next generation microcellular wireless communication systems currently under development, such as Personal Communication Services (PCS), should offer improved capabilities [71]. Higher frequencies will result in lower out-of-cell interference. Antenna placement and cell size will lead to fewer obstructions and Rician instead of Rayleigh fading. Multipath lengths will not differ as greatly, leading to less dispersion and inter-symbol interference. Higher data rates will be available to the user, who will need them to support additional applications such as networking, high rate data transfer, and video and multimedia. Even with these advantages, digital microcellular will still be a highly dynamic transmission environment, and applications must be designed with suitable robustness or the service provided will be of poor quality.

From a digital communications perspective, wireless voice communications presents a great research challenge and opportunity. There is great motivation to optimize the system, to offer the most reliable, highest quality, most affordable service to the greatest number of users. Optimization of the source and channel coding, and efficient methodologies for transmission of digitized speech are desirable if they offer the potential to improve system performance without significantly increasing system complexity.

Speech coders designed for wireless communications must be robust because transmission errors cannot be completely avoided. The extent to which errors affect the coder performance, the types of errors likely to occur, and the frequency at which they occur are all factors which should be taken into account in the design of the coder.

Channel coders designed for the wireless channel must be flexible and efficient. The channel is highly variable, outages must be very infrequent, and capacity must be maximized.

Handling of different traffic sources and support for different network topologies are additional considerations in developing the next generation wireless communication systems. Voice traffic will require different handling than data traffic, since the delay and error requirements differ. Base-stationless network topologies will better serve applications such as crisis management, military communications, and low cost peer-to-peer networks.

A flexible approach to accommodate localized, distributed allocation of resources is desirable to make optimal use of system resources. The need for flexible, robust, high performance, and low complexity solutions continues to be a major technical challenge, demonstrating the need for improvement in design methodologies for applications such as wireless voice communications.

2.2 Speech Coding

The goal of speech coding is to find a digital representation such that reconstruction is possible at the highest possible quality, given constraints in the coding rate, delay, and complexity of the coding algorithm. Various techniques have been developed which allow the source rate to be reduced or compressed, with minimal loss of perceived quality based on both standard source coding techniques for analog waveforms and speech production and auditory system models.

An overview of speech coding theory is presented in the following sections. Section 2.2.1 describes models for speech production, which gives insight into the characteristics of speech signals. Section 2.2.2 describes models for the human auditory system, which gives insight into how the quality of a reconstructed speech or audio signal is perceived. Section 2.2.3 describes the main compression techniques for both speech and audio. Section 2.2.4 describes objective and subjective measures of how speech coder quality is evaluated.

2.2.1 Models for Speech Production

The mechanism for speech production in the human vocal system can be modeled to improve the coding efficiency of the speech coder by allowing the redundancy in voice waveforms to be exploited.

The human vocal system is typically modeled as the response of an excitation and filter, the source-system model [86]. The excitation is provided by the lungs and vocal cords. The filter response is determined by the vocal tract, nasal tract, mouth, and lips. The output of such a system constitutes the speech waveform.

The excitation and filter thus characterize the speech signal. For example, a speech segment is classified as voiced or unvoiced depending on whether the excitation is tonal or white (random). Peaks in the filter response, called formants, characterize the resonances in the vocal and nasal tract.

Speech can be efficiently coded using parametric representations of the excitation and filter characteristics. The speech production model has been used to obtain efficient parametrization of the source and high coding efficiency in many coding techniques, from early vocoders to today's most advanced low rate coders, as will be described in section 2.2.3.

2.2.2 Human Auditory System

Modeling of the human auditory system plays a key role in the design of speech and audio coders. Properties such as the frequency response, frequency sensitivity and selectivity, and spectral and temporal masking affect the perceived quality of the reconstructed signal. An efficient coder design avoids what the human ear perceives as annoying noise and exploits what the human ear ignores. This leads to coder implementations that do not optimize a classical source coding optimization criterion, such as signal to noise ratio (SNR) or minimum square error (mse), but try to use a perceptually significant cost measure instead.

A completely accurate model of human auditory perception does not yet exist [59]. Although several properties are difficult to model, and there are variations in perception from one human to another, models that approximately incorporate several key characteristics have been successful in improving perceived quality of speech and audio coders. Coders that incorporate such features are referred to as perceptual coders [58].

One of the key characteristics of the human auditory system is its frequency response. The frequency response of the auditory system is a band pass response that attenuates the responses at low and high frequencies. The auditory system also exhibits frequency selectivity in that the ability to resolve neighboring frequencies decreases nonlinearly above 500 Hz. The frequency resolution of the auditory system is modeled by critical-band filters, where sound is preprocessed by a bank of such filters with center frequency spacings and bandwidths increasing in frequency. The frequency scale corresponding to the critical band center freqencies is the Bark scale [115].

The auditory system also exhibits masking effects, where the presence of a strong signal is capable of making a weaker, otherwise audible signal, inaudible. In spectral masking, the presence of a tone results in noise masking in a band of frequencies around the tone. The global masking threshold, the Just Noticeable Distortion (JND) [59], is evaluated by considering all the spectral components in the source signal and their masking contributions. The JND can be used in coder design so that the distortion introduced by the coding process is shaped such that its perceptual effect is minimized.

2.2.3 Coding Techniques

Speech and audio coding techniques fall into several broad categories: waveform coders, sinusoidal analysis coders, vocoders, and hybrid coders [41][100].

Waveform coding is a classical method of source compression, where the coding algorithm attempts to reconstruct the source signal time domain representation. Waveform coding methods can be further categorized into scalar quantization, vector quantization, and transform domain techniques.

Scalar quantization uses a digital word representation for each sample or parameter. It is the simplest coding method, but also the least efficient. Pulse Code Modulation (PCM) is based on scalar quantization. PCM incorporates companding, a non-linear mapping of the input using either μ-Law or A-Law logarithmic compression, to better accommodate the large amplitude fluctuations of the speech signal. Adaptive Pulse Code Modulation (APCM) is an alternative to PCM using adaptive quantizer step sizes to handle the non-stationary amplitude fluctuations of the speech signal. Differential Pulse Code Modulation (DPCM) and Adaptive Differential Pulse Code Modulation (ADPCM) are also based on scalar quantization, but incorporate a linear predictor to improve coding efficiency [38].

Vector quantization (VQ) uses a digital word representation for a block of samples or parameters. The digital word represents an index to a codebook containing possible reconstruction blocks or vectors. The VQ allows unconstrained optimal placement of the reconstruction vectors, resulting in substantial quantization performance improvements over scalar quantization. The optimal VQ codebook can be found using the Linde-Buzo-Gray (LBG) algorithm.

Transform or frequency domain coders use a transform to allow quantization of transform coefficients instead of waveform samples. Transforms such as the Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Karhunen Loeve Decomposition (KL), or Discrete Wavelet Transform (DWT) are performed to decorrelate the coefficients, and high coding efficiency can be achieved with coefficient quantization based on perceptual significance [75]. The subband coder is a variation of the transform coder that decomposes a source into several spectral bands using a filterbank.

Vocoders utilize the speech model to parametrize a representation of the speech source. Homomorphic (cepstral) and Linear Predictive Coding (LPC) methods can be used to deconvolve the source into an excitation and filter characteristic [86]. The source is modeled as either tonal or white noise. The filter is modeled as an all-pole filter. These techniques achieve very low coding rates but result in synthetic sounding speech.

Sinusoidal analysis-synthesis or harmonic sinusoidal modeling coders improve on vocoder techniques by representing the speech signal as a sum of sinusoids and bandlimited noise. Spectral analysis characterizes each frame by the excitation period and spectral envelope. The additional characterization of the excitation parameters results in more natural sounding speech than those achieved with vocoders. This technique is used to achieve very low coding rate in Sinusoidal Transform Coders (STC) [117] and Multi-band Excitation (MBE) coders [55].

Hybrid Coders or Analysis by Synthesis Linear Predictive Coders [89] are similar to vocoders in that they perform an efficient parametrization of the speech source based on an excitation and filter model, but differ in that they also combine waveform coding methods by searching for the optimal excitation to minimize a time-domain mean squared error criterion. Analysis by synthesis refers to a closed loop selection of excitation, where all possible excitation representations are tried and the one corresponding to the smallest error is selected as the optimal excitation. Hybrid coding techniques include Multi-pulse Excited Linear Predictive (MPLP) [100], Regular Pulse Excited, Long Term Prediction (RPE-LTP) [69], Code Excited Linear Prediction (CELP) [12][14], and Vector Sum Excited Linear Prediction (VSELP) [43]. These differ mainly in how the excitations available in the codebook are constrained to obtain an efficient representation. MPLP models the excitation as a number of pulses. RPE-LTP models the excitation as a periodic train of pulses. CELP uses training sequences to optimize the values in the codebook. VSELP uses multiple orthogonal codebooks to simplify the codebook search.

These coding methods have all found application in speech and audio coding standards. PCM and DPCM were the early standards in digital telephony. The telephone network incorporates 64 kbps μ-Law PCM and 32 kbps DPCM, so these coders are "toll-quality" by definition [45][67][68][90]. 64 kbps PCM is typically referred to as "uncompressed" speech. DPCM is capable of good quality and reasonable scalability, and has a strong base of existing standards such as The International Telephone and Telegraph Consultative Committee (CCITT) standards G.721, G.723, G.726, and G.727 [95][101]. The main drawbacks in DPCM are the need for high data rates (32 kbps to achieve toll quality) and sensitivity to error propagation. Nevertheless, DPCM is ubiquitous in existing networks and has been incorporated into some early wireless communication standards [95][101].

Hybrid coders are incorporated into existing and emerging digital cellular standards such as FS1016 (CELP) [12], IS-54 (VSELP) [43], IS-96 (QCELP) [124], GSM full-rate (RPE-LTP) [69], and GSM half-rate (VSELP) [110]. The main benefit of hybrid coders is that low bit rate is achievable with reasonable quality. Near toll quality can be achieved at bit rates from 6.5 to 9.6 kbps, but the robustness to background noise and channel errors is poor. A backward adaptive CELP algorithm, LD-CELP, achieves toll quality and low delay at 16 kbps. It has been adopted by CCITT as standard G.728 for networks and videoconferencing applications and is expected to come into wide use [14]. The main drawbacks to hybrid coders are the lack of robustness, high implementation complexity, and sub-toll quality. The quality is particularly poor for female speech, non-speech signals such as music, and for speech in the presence of background noise.

An alternative to hybrid coders for low rate coding is IMBE. IMBE is also capable of achieving sub-toll quality at low rates (2.4 to 6.4 kbps). An IMBE coder was selected by the maritime satellite communications system INMARSAT-M over several competing CELP and VSELP implementations based on superior quality over a range of test conditions [55][117].

Subband coders are attractive for wideband and high quality speech coding at medium bit rates. CCITT G.722 is a 64 kbps, 7 kHz wideband speech standard based on a two band subband decomposition of the speech source [57]. With the 2 band decomposition, the lower band is equivalent to a narrowband source that is then coded with standard narrowband speech coding methodology, in this case 48 kbps ADPCM. The upper band also utilizes ADPCM. Although there is less prediction gain available, the higher band is perceptually less significant, so it can be coded at a lower rate, 16 kbps.

Extensive research of subband coding has been done by Cox et. al. [17][18][48]. Bit rates from 12 kbps to 32 kbps have been studied for mobile and indoor wireless applications. Their system achieves significant robustness improvements over existing standards and provides good speech quality.

The main advantages of subband coding are its robustness to both source variation and channel errors, scalability to high quality at high source rate, and relatively low computational complexity.

Transform coding and subband coding are the predominant schemes used in audio coding. Typically, transform and subband coders do not incorporate a source model, but take advantage of auditory system modeling, which is facilitated by the frequency domain representation available in transform and subband coding [9][21][64][82].

The ISO-MPEG audio coding standard [9] for 20 kHz audio is a high quality standard based on 32 band subband decomposition and the auditory perceptual model. Subband coding is well suited for audio applications, where a source production model does not adequately represent the wide variety of sources, and high fidelity is an important requirement.

Audio coders based on transform coding, such as those proposed by Dolby [21][22] and Johnston [64], offer performance similar to those achievable with the subband based ISO-MPEG coder, but at higher complexity and less flexibility.

Schemes for embedded source coding with varying degrees of optimality have been devised for many speech coding architectures such as DPCM [31][46][95], VQ [54], Delta Modulation [116], subband coding [17][18], CELP [25], and RPE-LTP [123]. The viability of embedded source coding concepts and motivation to design systems exploiting embedded source and channel coding for voice transmissions will grow as the performance penalty from embeddability constraints diminishes with further progress in this research area.

2.2.4 Quality Measures

The fundamental goal of a speech quality measure is to predict how a typical individual would rate the perceived quality, the naturalness, intelligibility, and pleasantness of the reconstructed speech signal. Quality measures may be objective [28], a mathematical criterion of how well the speech was reconstructed, or subjective [29], an average of how individual listeners rate the reconstructed speech.

Objective quality measures attempt to obtain a quantitative measure of quality based on a mathematical relationship between the source speech signal and the reconstructed coded speech signal. There is no single predominant measure in use. Selection of an appropriate measure is highly dependent on the source coding technique used. Within related techniques, objective measures tend to be highly correlated with perceived quality, but this does not necessarily hold if different techniques are used or if the measure is used to evaluate coders implemented using different techniques.

Some commonly used objective measures are signal to noise ratio (SNR), segmental signal to noise ratio (seg SNR), spectral distortion (SD), cepstral distance (CD), noise to mask ratio (NMR), and coherence function (CF) [28].

SNR is the ratio of signal energy to error energy. It is the classical mathematical method of specifying the difference between two waveforms. Segmental SNR is a logarithmic average of SNR measured over short (10-20 ms) frames of speech.

Spectral distortion is a measure of the difference in spectral magnitude between the source and the reconstructed signal. Ignoring the signal's phase has some benefit since the human listener is less sensitive to phase errors.

Cepstral Distance is also a spectral distortion measure, but it is based only on the envelope of the spectrum. Thus it has the added benefit of de-emphasizing frequency resolution in the reconstructed signal.

Noise to Mask ratio is a different method of evaluating quality in the frequency domain. Instead of comparing the source and reconstructed signals, the Just Noticeable Distortion (JND) masking threshold is computed. The NMR is the ratio of the error energy to the masking threshold [8][58].

The coherence function is also a spectral measure between the source and reconstructed signal, but it separates the comparison between coherent and noncoherent components. Through selective thresholds and weighting functions, less significant components can be deemphasized.

Basically, all objective measures compute the error between the source and reconstructed signal and try to separate the error into perceptually significant and insignificant components. Unfortunately, no measure has been shown to be highly accurate and reliable, so subjective measures must be used to verify the quality of all coder designs.

In subjective measures, a qualitative measurement is made by asking different listeners to rate the perceived quality of the reconstructed speech. The Mean Opinion Score (MOS) is the most common rating scale used [29], a 1-5 scale representing bad, poor, fair, good, and excellent quality.

The simplicity of the MOS scale belies the difficulty of incorporating it in the design process. Clearly, the MOS is the most reliable measurement method of speech quality and the only viable means of validating other measures of speech quality. However, because the MOS scale is subjective, results from separate experiments are difficult to compare without some calibration.

The most common way to incorporate calibration is to include other coders for comparison. Including appropriate coders, with similar bit rates, compression techniques, or target applications, increase the relevance and reliability of the subjective testing results. However, such a process becomes very cumbersome and costly, as the listening tests must include a fairly large set of test sentences.

2.3 Channel Coding

Shannon demonstrated that the limitation of a noisy channel is not the inability to transmit digital data without errors, but rather the capacity, or amount of data, that the channel can transmit without errors [93]. The key to achieving error-free performance is to introduce redundancy so that error correction may be performed on the received data. However, the decoding complexity grows infinitely to achieve lower and lower error rates. Channel codes are designed to introduce redundancy in an effective manner, so that error protection is maximized while minimizing the amount of redundancy, decoding delay, and complexity required to achieve a certain error rate.

Embedded channel coding is attractive for wireless applications because it provides flexibility in selecting the amount of error protection and the channel rate. In a varying transmission channel, the channel code must adapt to the existing conditions to optimize the data throughput and reliability. Traditional non-embedded channel codes are not well suited for channel rate adaptation. Both the encoder and decoder structures may be greatly complicated by the need to support different block sizes and channel rates. Rate adaptation thus results in a much more complex system. Embedded channel coding is advantageous over non-embedded approaches in that it greatly simplifies adaptive transmission.

This section provides an overview of channel coding concepts, with an emphasis towards implementation of embedded channel coding and adaptive transmission. Section 2.3.1 discusses the two general categories of error correction codes: block codes and convolutional codes. Section 2.3.2 describes rate compatible punctured convolutional

(RCPC) codes, a practical approach for embedded channel coding. Section 2.3.3 describes tailbiting convolutional codes, which allow efficient encoding of short blocks of data. Section 2.3.4 discusses adaptive transmission and unequal error protection, two important aspects of channel protection for wireless voice communications. Section 2.3.5 discusses Automatic Response Request (ARQ) and Hybrid-ARQ systems, which allow further improvements in communications performance when a reverse channel is available.

2.3.1 Error Correction Codes

Forward error correction (FEC) channel codes introduce redundancy to data prior to transmission so that a decoder can use the redundancy to correct errors that occur during transmission. FEC codes generally fall into 2 categories: block codes and convolutional codes. Block codes generate codewords from a fixed block of information and convolutional codes generate codewords from a sliding block in a continuous stream of data.

With block codes, the information is grouped into blocks of data, and the data is encoded by selecting a codeword from a fixed codebook. In an (n,k) block code, a sequence of k information symbols is encoded with a codeword containing n information symbols.

Block codes can be efficiently described if they are constructed with certain algebraic or geometric structures. Linear codes have algebraic structure such that the sum of two codewords is also a codeword. Cyclic codes are linear codes with the additional property that a cyclic shift of a codeword is also a codeword. Linearity facilitates analysis of the code's error protection capabilities. In a linear code, every codeword has similar distance properties, so the error protection capability need only be evaluated for a single codeword. Cyclic codes provide a compact representation of the codewords using polynomial multiplication, and provide some structure for the decoder, using shift registers and polynomial division to determine the location of errors in the codeword.

Most useful block codes, such as Hamming, Bose-Chaudhuri-Hocquenghem (BCH), and Reed-Solomon (RS) codes, are cyclic codes. Hamming codes are simple, single error correcting, codes. BCH codes are multiple error correcting codes with long block length, good distance properties, and efficient decoder structures. The BCH codes are generated such that the generating polynomials' zeros in the extension field may be used to set up a set of non-linear equations to determine the locations of errors in the codeword. This results in a more efficient decoder structure than possible with just the cyclic decoder structures. RS codes are a non-binary subclass of BCH codes with optimal distance properties; no other codes offer more error protection for the same block length and rate. The Cyclic-Redundancy-Check (CRC) code is another important type of cyclic code. CRC codes are high rate cyclic codes offering little error correction capability, but they are efficient for use in error detection applications.

Unlike block codes, convolutional codes are not restricted to transmitting code words in blocks. In a rate k/n convolutional code, k information symbols are input into a finite state machine, generating an output of n symbols. The number of previous inputs on which the present state and output are dependent is the constraint length v of the code, the size of the sliding block used in encoding. The number of states in the convolutional code is q^v, where q is the size of the symbol alphabet, usually 2 (binary).

The encoder for a simple convolutional code, a 4 state rate 1/2 code, is shown in Figure 2.1. For each binary input x, two outputs y_1 and y_0 are generated. The encoder is a finite state machine with 2 memory elements, where the state transitions are given in Figure 2.2. The 4 nodes are labeled with the state variables, and the transitions are labeled with the input x causing the transition and the outputs y_1 and y_0.

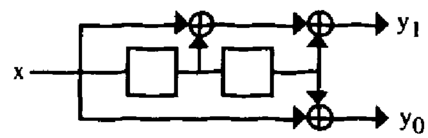


Figure 2.1 - Convolutional Code Encoder

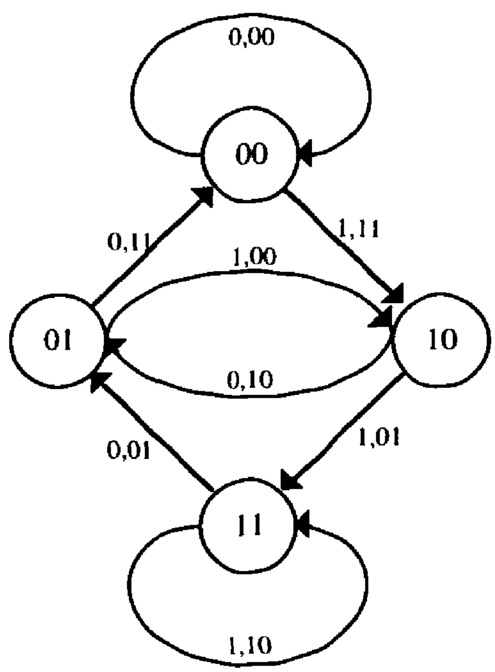


Figure 2.2 - Convolutional Code State Transition Diagram

The sequence of possible state transitions defines the trellis for the convolutional code. The trellis is a representation of the possible state transitions over time, where each of the paths through the trellis corresponds to one of the possible state sequences. The trellis for the 4 state, rate 1/2 convolutional code over 5 state transitions is shown in Figure 2.3. The nodes in the trellis correspond to the states being traversed, and the branches correspond to the possible inputs and outputs corresponding to those transitions.

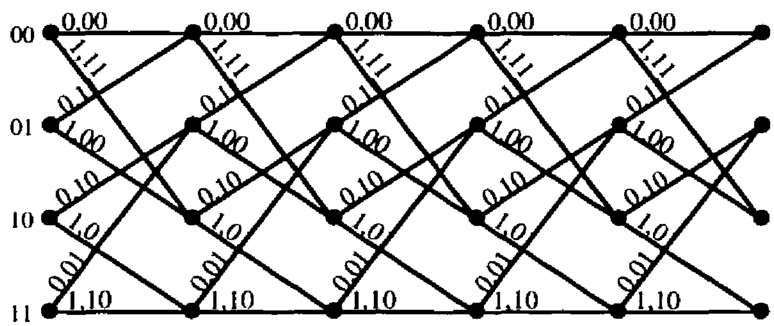


Figure 2.3 - Convolutional Code Trellis

The Viterbi Algorithm (VA) [35][112] is a maximum likelihood (ML) decoder that uses the branching in the trellis to compare and eliminate paths until the only path remaining is the maximum likelihood path through the trellis. The VA is a very structured decoder that easily incorporates soft decisions, a probabilistic measure for the likelihood of traversing each branch in the trellis. A path length or metric is assigned to each branch in the trellis, where the difference in length corresponds to the log-likelihood of the transition given the observation of the symbols transmitted through the channel. Finding the ML path then becomes simply a search for the shortest path through the trellis.

Block codes and convolutional codes provide error correction capability by maximizing the distance between codewords. Greater redundancy and greater codeword distance are available as the channel code rate, the ratio of information to codeword symbols, is decreased. The effectiveness of the error protection is given by the coding gain, the attainable savings in energy per information bit required to achieve a given error probability compared to uncoded transmission. Block codes generally provide the most coding gain for a given code rate and encoder memory, but convolutional codes have the advantage of structured decoding with soft decisions using the Viterbi Algorithm.

Channel codes can also be used in a concatenated manner to provide even greater protection against channel errors. For example, a concatenated code may use a Reed-Solomon outer code and a convolutional inner code, as shown in Figure 2.4. For some applications, such as deep space communications, concatenated coding is often the only solution to the poor channel [1]. The main drawbacks to concatenated coding are the decrease in coding rate, increase in decoding complexity, and the need for additional interleaving to improve the outer code's efficiency, which further increases the decoding delay.

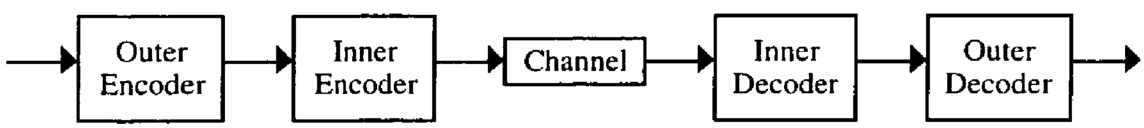


Figure 2.4 - Concatenated Coding Example

Joint source and channel coding approaches combine compression of the source and protection to channel errors in a single codebook to minimize the total encoding delay and complexity. The codebook design attempts to optimize noisy channel performance for a given codebook complexity by maximizing the codebook robustness to transmission errors. The codebook may be obtained using either vector quantization [34] or trellis coded [2][49] schemes. The main drawback with these types of schemes is that a model of the channel is required for optimization of the codebook design. The scheme also lacks flexibility for source and channel coder adaptation, which is important when the channel variability is very high.

A scheme related to joint source and channel coding is robust indexing of codebooks. This scheme improves the robustness of the source code to transmission errors without combining the source and channel coder designs. The source coder is modified, without degrading its performance, by considering the reconstruction error introduced by transmission errors [87]. For example, the pseudo-Gray indexing of VQ codebooks, introduced by Zeger and Gersho [122], improves the robustness of VQ codebook indexing by assigning the indices such that the sensitivity to bit errors is minimized.

Block based channel coding with interleaving is the predominant method of incorporating channel coding for wireless voice communications. Interleaving is beneficial in randomizing bursty noise in a fading environment so that the channel code is more effective. The need for interleaving also eliminates the decoding delay advantage of continuous coding approaches over block based coding approaches.

Both block codes and convolutional codes can be used to encode blocks of data, although modifications to the traditional convolutional coding methodology must be incorporated to handle blocks. Block based coding is advantageous for embedded channel coding because it allows the source information to be grouped such that error protection may be varied from block to block, and block-by-block adaptation can be performed in a flexible manner.

Unfortunately, block codes do not allow easy incorporation of erasures and soft decision information in the decoder. Thus, convolutional codes are preferable for embedded channel coding since they allow structured soft decision Viterbi decoders to be used and provide a larger range of usable rates.

2.3.2 Rate Compatible Punctured Convolutional Codes

Rate Compatible Punctured Convolutional (RCPC) codes are an attractive methodology for embedded channel coding. Embedded channel coding requires a higher rate code to be embedded into the bitstream of a lower base rate code. Most block codes and convolutional codes are compatible with a set of puncturing patterns and decoding rules that allow an embedded code to be defined. However, block codes require hard decision decoders, where the punctures correspond to an "erasure" symbol. On the other hand, convolutional codes allow soft decision decoders to be used, where the branch metric calculation is modified to account for the punctured symbols. RCPC codes are attractive in that they allow the implementation of embedded channel coding with a simple decoder structure incorporating soft decisions, with little performance penalty relative to non-embedded channel coding schemes of similar complexity.

Cain et al. [11] and Yasuda et al. [119], introduced punctured convolutional coding as a lower complexity alternative to high rate convolutional coding. Punctured convolutional codes reduce the complexity of the decoder for high rate codes because it allows the use of the relatively simpler trellis of the base rate code to be used.

Hagenauer introduced RCPC codes [51][52][53] by imposing a rate compatibility restriction on punctured convolutional codes The main advantage of RCPC codes is that they allow the use of the same decoder structure for multiple code rates, since the trellis remains unchanged through puncturing. Only the branch metric calculations need to be changed depending on the puncturing pattern. Furthermore, Hagenauer showed the RCPC codes to be almost as good as the best known convolutional codes of the same constraint length for rates between 1/4 to 8/9. Thus, the performance degradation of the RCPC codes is minimal relative to traditional convolutional codes.

Shi, Ho, and Cuperman [98] introduced rate compatible punctured Reed-Solomon (RS) codes as an alternative to RCPC codes. The advantages of the RS code are its good distance properties and burst error correction capabilities. The RS code provides for the maximum distance for a given block length, and its burst error correction reduces the amount of interleaving required. However, the RS codes have several disadvantages. Large block lengths are required to implement the RS codes. Hard decision decoding must be used since incorporation of soft decision greatly increases the complexity of the decoder. At low SNR, the hard decision decoding of RS codes leads to substantially lower performance than RCPC codes.

Complementary punctured convolutional (CPC) codes were introduced by Kallel [65] for diversity transmission schemes. CPC codes are a set of codes derived from a base low rate such that each code is individually decodable and when combined yield the original low rate code. The main advantage of CPC codes is that the code can be used in progressive transmission schemes where each transmission is individually decodable. The main drawback is that the use of CPC codes reduces the number of available rates that can be used, although a combination of RCPC and CPC rates may be attractive.

2.3.3 Tailbiting Convolutional Codes

Convolutional codes normally operate on a continuous stream of information bits. In order to encode blocks, convolutional codes must be initialized and terminated in each block. There are four basic methods to initialize and terminate blocks: direct truncation, coding across block boundaries, coding with tail bits, and tailbiting.

With direct truncation, the encoder is initialized to a known state, usually the all-zero state, and the encoded sequence is simply terminated after the last information bit arrives at the encoder. This truncation reduces the amount of error protection to the information bits at the end of the block.

In coding across block boundaries, the encoder saves the final state and uses it as the initial state for the next block, so the error protection at the end of the block is preserved. This requires the next block to be received before completing the decoding of the present block, so it is unattractive for packet networks where delay and packet loss can be of serious consequence.

Coding with tail bits is the most common method to encode blocks with convolutional codes. The encoder is initialized to a known state, usually the all-zero state, at the beginning of the block. A known sequence, usually all zeros, of length equal to the code's constraint length is added to the end of the block, such that the final state of the trellis is now known. The disadvantage of tail bits is that coding efficiency is lost since information bits are replaced by the termination sequence.

Tail bits can be eliminated by using tailbiting [76]. Tailbiting convolutional codes force the initial state and the final state to be the same for any given block by initializing the convolutional coder with the last few bits of the block, using a circular buffer. This allows the entire block to be encoded without the overhead of tail bits, and without incurring decoding error degradation at the end of the block. This can be of significant importance for systems utilizing short blocks, where additional overhead and increased error rate are unacceptable. Tailbiting has not been widely used because the decoding structures designed to this point are not very efficient.

The Viterbi algorithm must be modified to decode tailbiting convolutional codes. Since all valid paths must begin and end at the same state, the Viterbi algorithm must provide some means of eliminating invalid paths and comparing valid paths beginning at different initial states. The decoding complexity is increased due to the uncertainty in initial state, since the Viterbi algorithm is usually initialized with a known initial state. Several decoding strategies are possible, either through multiple iterations assuming different initial states or through some method of first estimating or decoding the initial state.

The maximum likelihood (ML) decoder uses the straightforward approach of comparing all possible valid paths. This requires an exhaustive search comparing the distance metric from the best surviving paths from each of the possible initial states [76]. The complexity is very high since the Viterbi algorithm over the entire block must be run once for each state.

Bar-David's probabilistic algorithm [76] uses a suboptimal iterative approach that guesses the initial state until a path that ends in the same state as the guessed initial state is found. However, this does not guarantee that the maximum likelihood path is found, so the decoding is suboptimal. The number of iterations required may also become prohibitively high, especially under poor SNR.

The two-step algorithm from Ma [76] first orders all possible initial states using an algebraic method then attempts to find a good path. This reduces the number of iterations relative to the Bar-David algorithm, but still results in a suboptimal search.

The Circular Viterbi Algorithm (CVA) introduced by Cox [19] utilizes continuous Viterbi decoding over a number of repeated blocks. The Viterbi algorithm eventually converges so that the maximum likelihood path is found. An adaptive rule is used to determine when the metrics have converged, and the traceback is modified to consider only valid paths. The CVA significantly reduces the computational complexity of optimal tailbiting decoding, and is the best of the previously published decoding methodologies. However, the computational workload for the CVA is variable, and the adaptive rules increase the decoding complexity.

2.3.4 Adaptive Transmission and Unequal Error Protection

Channel coding for wireless voice communications should take into account the unique characteristics of the source and its error protection requirements. The information bits do not have equal significance, so they differ in their error tolerance. High quality reconstruction does not require an extremely low error rate, especially in the less significant bits. On the other hand, the transmission channel is highly variable, making it difficult, if not impossible, to devise a fixed channel coding scheme that performs well under all conditions. Adaptive transmission schemes allow improvements in performance and robustness by matching the channel rate more appropriately to the existing transmission conditions. Unequal error protection improves both the throughput and robustness of the transmission, by more closely matching the requirements to the effect of errors in perceived quality.

Modestino and Daut [79] introduced source and channel rate adaptation with unequal error protection for transmission of images encoded with two-dimensional DPCM. The reconstructed image quality degraded rapidly when the error rate increased, so decreasing the source rate and increasing the channel protection when the channel was poor resulted in significant improvement in the reconstructed image quality. Their scheme demonstrated that selective error control based on the channel conditions and bit significance greatly improved reconstructed image quality without additional transmission bandwidth. The main drawback with their scheme is that neither the source nor channel coding schemes were embedded, so the rate adaptation scheme was not very flexible.

Goodman and Sundberg [47] extended the concept of source and channel rate adaptation to speech transmission scheme using embedded DPCM and punctured convolutional coding. When the channel was good, no error correction was necessary, but when the channel was poor, rate 1/2 and rate 3/4 punctured convolutional codes could be selected by dropping the source rate. Embedded DPCM allowed the source rate to be easily modified, demonstrating the advantages of embedded coding. However, their system did not provide for a mechanism for adapting channel code rates and providing unequal error protection, which are made possible by the use of short block lengths in embedded channel coding.

Hagenauer [51] demonstrated the flexibility of RCPC codes for both rate adaptation and unequal error protection. RCPC codes are attractive both for their excellent performance over a large range of rates and their simple decoder structures. Unequal error protection was obtained by simply grouping information bits according to their significance. Cox et al. [18] demonstrated the effectiveness of source and channel rate adaptation using an embedded subband coder and RCPC codes.

Masnick and Wolf [77] introduced unequal error protection (UEP) codes that provided different error correction capability for different positions in the codeword. UEP codes have remained of general research interest, but have not found wide application since unequal error protection can be more easily and flexibly generated from block-based methods.

2.3.5 Automatic Response Request (ARQ) and Hybrid-ARQ Systems

Protection from channel noise is possible without forward error correction (FEC) if a return channel is available. Automatic Response Request (ARQ) systems use detection of errors in decoded blocks to request retransmission [1]. Error detection is typically done with the use of a Cyclic-Redundancy-Check (CRC) code, a high rate cyclic block code. ARQ systems provide a flexible method of accommodating channel degradation. ARQ ensures a more appropriate allocation of system resources by concentrating effort when it is needed.

The retransmission protocol may be of several types. "Stop-and-Wait" ARQ requires acknowledgment before subsequent blocks are transmitted. "Go Back N" ARQ allows the receiver to request the previous N blocks to be retransmitted when an error is detected. "Selective-Repeat" ARQ allows the receiver to ask for specific blocks to be retransmitted.

Hybrid ARQ schemes combine FEC with ARQ to improve the overall performance and throughput of the system, by combining the best features of the two error protection schemes. FEC provides some protection from channel noise so that the number of retransmissions are not excessively high. ARQ ensures reliable transmission by requesting retransmissions when the protection from the FEC code is not sufficient for error-free transmission.

Type I hybrid ARQ schemes attempt to match the error correction code to the channel conditions, so that throughput will not suffer as a result of either too many retransmission requests or too much wasted redundancy for error correction. This requires a fairly stable transmission environment, so that the code selection can be optimal. Deng [26] proposed a modified type I hybrid ARQ system using punctured convolutional coding and adaptive rate selection based on BER estimation from channel state information. The adaptive nature of this modified scheme improved the system's performance over a large range of channel conditions.

Type II hybrid ARQ uses incremental redundancy in the form of progressive transmission for improved performance and flexibility compared to type I schemes. Combined decoding is made possible through the use of systematic block codes or punctured convolutional codes [51].

Kallel [65] proposed another variation which he called Type III Hybrid ARQ. It is similar to Type II hybrid ARQ in that it uses progressive transmission packets for combined decoding, but replaces the error correction code with complementary punctured convolutional (CPC) codes. With CPC codes, every packet is self decodable, so the decoder does not have to rely on previously received sequences as is the case for typical incremental redundancy ARQ schemes. This is desirable when the original packet can be lost or severely damaged as a result of interference.

Darnell et al [20] used array codes with multiple levels of protection in an ARQ scheme. Repeat requests were used to indicate the lowest level at which the decoding was successful, and unsuccessful blocks were retransmitted.

The main drawback with ARQ is the delay required for the acknowledgments to be transmitted and potential network overload due to increased retransmission requests under poor channel conditions. Although ARQ schemes are well suited for adaptive transmission, some modifications are necessary when the data is voice traffic. An important consideration is the short tolerable delay and need for synchronization. With wireless voice, it is best to eventually give up attempts at transmission and go on to the next packet. Intelligent packet dropping is an important feature for efficient implementation of wireless voice networks.

For example, IS-54, the North American digital Time Division Multiple Access (TDMA) standard for cellular radio, classifies the coded data into 2 levels, and uses rate 1/2 convolutional coding for the most significant data. It generates a CRC check on the most perceptually significant bits, so that error detection may be used to discard bad packets.

Asynchronous Transfer Mode (ATM) networks also incorporate data prioritization to more effectively handle data traffic. Chang and Rubin [13] showed that statistical packet dropping based on packet prioritization yielded significant improvements in voice transmission over ATM networks.

In addition to the stringent delay requirement, most of encoded speech is relatively tolerant of transmission errors. Improvements in throughput can be obtained by controlling the error rate rather than requiring error-free transmission.

2.4 Statement of the Problem

As wireless communications evolve from the present cellular systems to future microcellular systems, several aspects will change. Users will demand new applications to improve personal productivity, such as networking, high rate data transfers, video, and multimedia. Greater traffic variety will lead to more complex systems, designed to optimally handle the different types of data. Network topologies will be less constraining, accommodating more flexible configurations. The increased system complexity will require more decentralized control to make optimal use of all resources. At the same time, expectations for high performance will demand good quality of service under even the most difficult conditions. Along with all the new applications and capabilities, consumers will demand better quality of service for digital wireless voice communications than what is provided by the current digital cellular network.

Adaptive transmission is the best method for efficiently and effectively dealing with the large variations in the transmission environment. Over a large range of diverse conditions, adaptive transmission will outperform any fixed transmission scheme. However, adaptive transmission schemes are more complex and more difficult to implement than fixed transmission methods. Traditional fixed rate source and channel coding schemes cannot be integrated into adaptive transmission schemes without great compromise in either performance or complexity. They also are not compatible with decentralized, "on-the-fly" adaptation, as may be needed in advanced topologies such as multi-hop peer-to-peer wireless networks.

Current speech coders have several shortcomings that make them poorly suited to the needs of future applications. Fixed rate implementations make it difficult to scale coders to high quality, and are not sufficiently flexible to be incorporated in an adaptive transmission system. The speech coders must also be robust to transmission errors, but the low rate coders typically are very sensitive to bit errors. Improvements in coding rate, efficiency, scalability, robustness, complexity, and ease of integration, are all highly desirable.

Flexible schemes for channel coding are also needed to deal with the variation in channel conditions. Outages must be infrequent, the quality of service must be acceptable, and system capacity must be maximized. It is important that the channel coding scheme be efficient, taking advantage of the voice traffic characteristics through unequal error protection, source rate adaptation, and throughput and error rate optimization.

Speech and channel coding methodologies need to be developed to support robust, high quality, and low complexity wireless voice communications. Current methodologies lack the flexibility to efficiently handle different traffic sources, accommodate localized allocation of resources, and support different network topologies. Embedded source and channel coding not only allow reduction in system complexity, but are in some cases essential for the implementation of adaptive transmission. The flexibility of embedded source and channel coding and the substantial performance advantage of adaptive transmission systems makes it a potential solution, but design methodology is not in place demonstrating that this can be done effectively. New methodologies for adaptive transmission with embedded source and channel coding improving the performance, flexibility, robustness, and complexity are desirable to allow the full advantage of digital communications to be exploited for wireless voice communications.

3. Embedded Source Coding

In this section, a methodology for embedded source coding is described that allows efficient digital representation of speech signals and facilitates wireless voice communications. The objective of source coding is to remove redundancy in the source signal and provide a digital representation that minimizes distortion at a given coding rate. Embedded source coding provides for an incremental description of the source, such that partial reconstruction is possible if only a portion of the bitstream is available. In an embedded source coder, the embedded bitstream constraint causes the partial reconstruction at lower bit rates to produce higher distortion than a non-embedded fixed-rate source coder at the same rate. However, it is possible to design an embedded coder such that the degradation in distortion performance and increase in implementation complexity relative to a non-embedded implementation is minimized, as will be shown with the methodology developed in this work.

For variable rate applications, an embedded coder is preferable to multiple non-embedded designs, because the system design can be simplified if the rate selection does not have to be done before the source is encoded. Embedded source coding is also desirable if the system allows reconstruction with partial transmissions in order to improve robustness to transmission degradation and network congestion. Thus, unless the performance penalty is severe or the complexity increase is significant, there is significant motivation to use embedded source coders in applications such as wireless voice communications, where the added flexibility may be useful in designing systems with improved performance.

Techniques for effective embedded source coding of speech were investigated in this research through the design and evaluation of an embedded speech coder. High perceptual quality, robustness, and rate scalability were the main goals of the design. Quality is important because current algorithms do not adequately address the issue of extending performance to higher quality. Robustness to both source and channel variation is important to ensure the quality of service is maintainable across a large range of users, applications, and channel conditions. Rate scalability is important to allow efficient use of system resources through rate adaptation. The embedded speech coder, designed jointly with Albert Shen [96][97][104], incorporates several new techniques to achieve these goals.

The embedded speech coder provides a flexible implementation with prioritized bitstream and scalable quality. Rate adaptation allows efficient utilization of system resources, providing consistent service under varying conditions. Channel coding and interleaving of the encoded bitstream can be easily implemented to take advantage of the scalability of the coder and provide the required robustness to transmission errors under various transmission and traffic conditions, as will be shown in section 4 and section 5.

The basic approach is to use a subband decomposition of speech, spectrally shape the quantization noise so the coding distortion is perceptually minimized for the human listener, and use bit prioritization and embedded quantization to assemble an embedded bitstream. The coder performance is scalable so that as the bit rate is changed, the quality of encoded speech is not significantly degraded compared to non-embedded designs at the same rate. The implementation of the coder and description of novel features are discussed in section 3.1.

The coder was designed and evaluated for narrowband (3.5 kHz) speech and audio sampled at 8 kHz, although the techniques applied are also suitable for wideband speech (7 kHz) and broadcast audio (20 kHz). The listening results are summarized in section 3.2.

The benefits of embedded source coding, and the particularly novel techniques employed in this design are discussed in section 3.3. The main accomplishment of this work is the demonstration that embedded source coding is possible with very little performance degradation relative to non-embedded designs, motivating their use in applications that benefit from the added flexibility, such as wireless voice communications.

3.1 Embedded Speech Coder Implementation

A perceptually based embedded subband coder incorporating several novel features was implemented to investigate techniques for effective embedded coding of speech. The techniques used for encoding and embedding ensure that the performance over a large range of rates is only slightly degraded compared to fixed-rate coders optimized for each rate.

Subband coding is based on separating the spectral content of the source speech or audio signal into several frequency bands and spectrally shaping the quantization noise so that the distortion introduced by the coding process is minimized. The distortion can be optimized in a perceptually significant manner, using a model of the perceptual masking properties of the human auditory system to determine the optimal shaping of the quantization noise. The coder is made embedded by assembling the bitstream in a prioritized manner, and using embedded quantization to represent the subband samples, such that truncation of the bitstream corresponds to encoding at a lower source rate.

The coder consists of five main components: analysis QMF filterbank, perceptual quality model, subband bit allocation, quantizer, and bitstream formatter. The decoder consists of a bitstream decoder, inverse quantizer, and synthesis QMF bank. These components are highlighted in the coder block diagram shown in Figure 3.1.

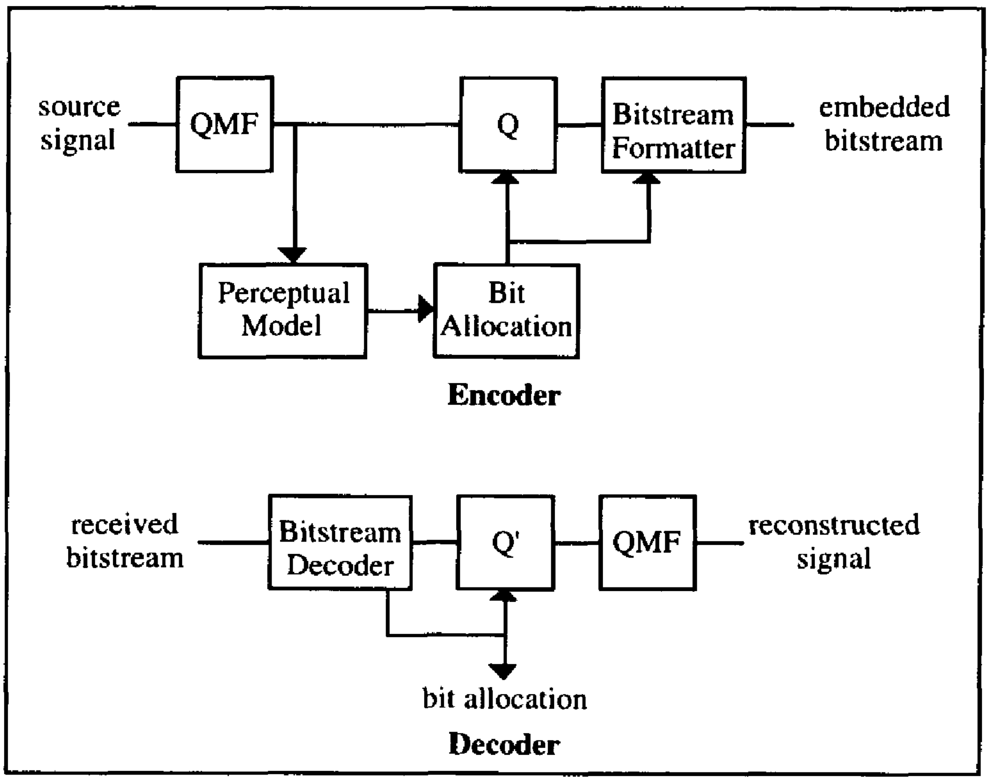


Figure 3.1 - Embedded Subband Coder

The source signal is 3.5 kHz filtered speech or audio, sampled at 8 kHz. Non-overlapping frames of 20 ms duration are used for processing and transmission.

The analysis and synthesis filterbanks, described in Section 3.1.1, are 8-channel, tree structured Infinite Impulse Response (IIR) Quadrature Mirror Filterbanks (QMF). The filterbanks are designed to provide good out-of-band rejection, alias cancellation, no amplitude distortion, minimal delay, and minimal phase distortion.

The perceptual model, described in Section 3.1.2, estimates the Signal-to-Noise Ratio (SNR) required to mask quantization noise for transparent coding based on the Just Noticeable Distortion (JND) for each frame of the speech source signal. The complexity of computing the JND is reduced by using a novel subband spectral analysis technique, described in Section 3.1.2.1.

The dynamic bit allocation and prioritization algorithm, described in Section 3.1.3, determines the optimal bit assignment for each frame, minimizing the interpolated masking threshold (IMT) cost measure.

The subband samples are normalized and quantized using an optimal embedded scalar quantizer, described in Section 3.1.4. The bitstream formatting is described in Section 3.1.6. The overall bit rate is made adaptive by specifying the bit allocation required to achieve perceptual transparency. The bits are arranged such that bits are truncated in a perceptually optimized manner by placing the less important bits at the end of the bitstream. Bit truncation of the quantization indices is allowed on a frame by frame basis to reduce the bit rate. Various source-rate constraints, such as fixed bit-rate, variable bit-rate, and network controlled adaptive source rate, can be met by simply specifying how the bitstream is to be truncated.

3.1.1 Filterbank

The filterbank performs subband decomposition of the source signal. Quadrature Mirror Filterbanks (QMF) are a specific class of subband filters designed such that no aliasing is introduced even though the subband samples are critically decimated [107]. The synthesis filterbank is designed such that the aliasing introduced by the decimation process is canceled when the source signal is reconstructed.

A 2 band (halfband) QMF is shown in Figure 3.2. The transfer function realized by such a system is given by:

$$\begin{aligned} Y(z) &= \tfrac{1}{2}[H_0(z)F_0(z) + H_1(z)F_1(z)]X(z) + \\ &\quad \tfrac{1}{2}[H_0(-z)F_0(z) + H_1(-z)F_1(z)]X(-z) \end{aligned} \tag{3.1}$$

where $X(z)$ and $Y(z)$ are the z transforms of the input $x(n)$ and output $y(n)$, $H_0(z)$ and $H_1(z)$ are the analysis filter transfer functions, and $F_0(z)$ and $F_1(z)$ are the synthesis filter transfer functions.

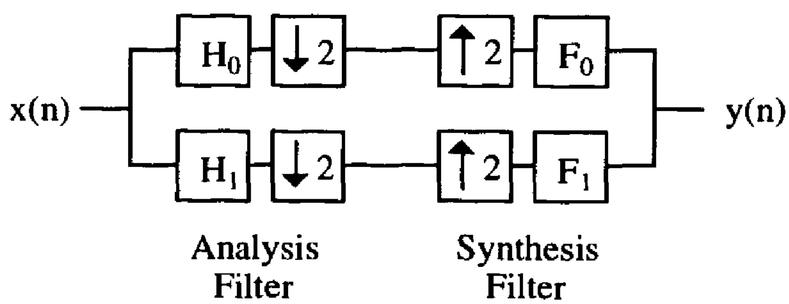


Figure 3.2 - Halfband Quadrature Mirror Filterbank

The second term in (3.1) represents the aliasing term. By choosing $F_0(z)=H_1(-z)$ and $F_1(z)=-H_0(-z)$, the second term becomes zero and there is no aliasing introduced by the filterbank. By also choosing $H_1(z)=H_0(-z)$, the filter becomes a lowpass/highpass pair. Given these conditions, the filterbank response becomes:

$$T(z) \equiv \frac{Y(z)}{X(z)} = \tfrac{1}{2}[H_0^2(z) - H_0^2(-z)] \tag{3.2}$$

If $|T(z)| = 1$, then the filterbank introduces no amplitude distortion. If $T(z)$ has linear phase, then the filterbank introduces no phase distortion.

Infinite Impulse Response (IIR) QMF filterbanks for speech and audio coding developed by Zhong Jiang [62] were incorporated into the design. Finite Impulse Response (FIR) QMF filterbanks have traditionally been used in subband speech and audio coders [16][17][63][108]. IIR QMF filterbanks have been avoided because of their non-linear phase response, but they have several advantages over FIR filterbanks: lower complexity, lower delay, sharp transition regions, high stopband attenuation, and low sensitivity to coefficient truncation [62][107][108].

The human auditory system is somewhat insensitive to phase distortion [24]. Thus it was reasonable to expect that an IIR QMF filterbank might be suitable for speech coding. Some phase distortion in the filterbank may actually be desirable, since it allows a lower delay, sharper transition, and higher stopband attenuation implementation than linear phase filterbanks. Using informal listening tests, Shen [96] was able to confirm the IIR QMF filterbanks used in this design did not produce noticeable phase distortion and were suitable for high quality speech and audio coding.

Subband coders for speech coding applications typically utilize about 40 dB of stopband attenuation from FIR QMF filterbanks [17]. This is a limitation due to both the complexity of higher order FIR filters, and the additional delay required to achieve higher stopband attenuation. However, since quantization introduces noise, aliased noise can become a source of perceptual distortion in high quality coders. Since the IIR QMF filterbanks offered an improvement in stopband attenuation, with no perceptible phase distortion, and at lower complexity, they were shown to be a better choice for performing the subband decomposition.

A 7th order elliptic IIR QMF halfband filter with 60 dB stopband attenuation was designed and implemented. The 8-channel tree-structured IIR QMF filterbank, shown in Figure 3.3, was designed with the methodology given in [62]. Decimation of the high frequency subband leads to spectral inversion, so the crossovers are necessary to maintain a low frequency to high frequency ordering of the subbands.

The analysis and synthesis filterbanks have identical responses, shown in Figure 3.4. The filterbank has no aliasing distortion, no amplitude distortion and non-perceptible phase distortion.

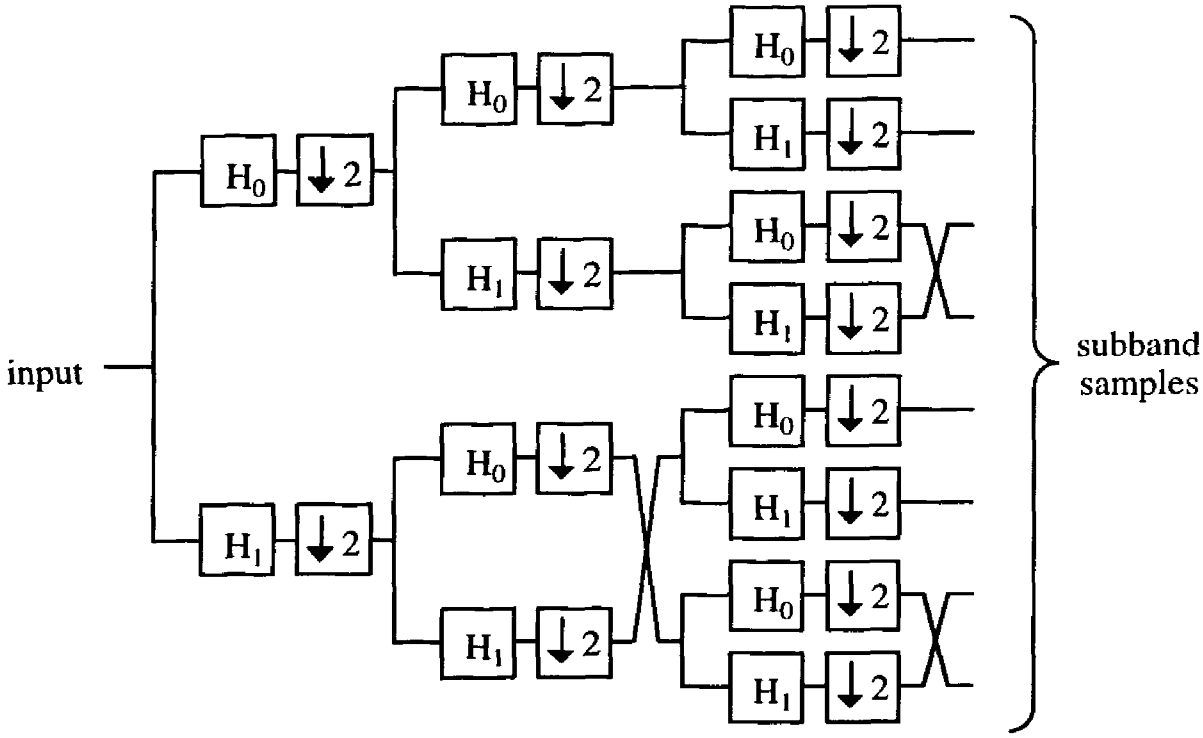


Figure 3.3 - 8 Band Tree Structure QMF

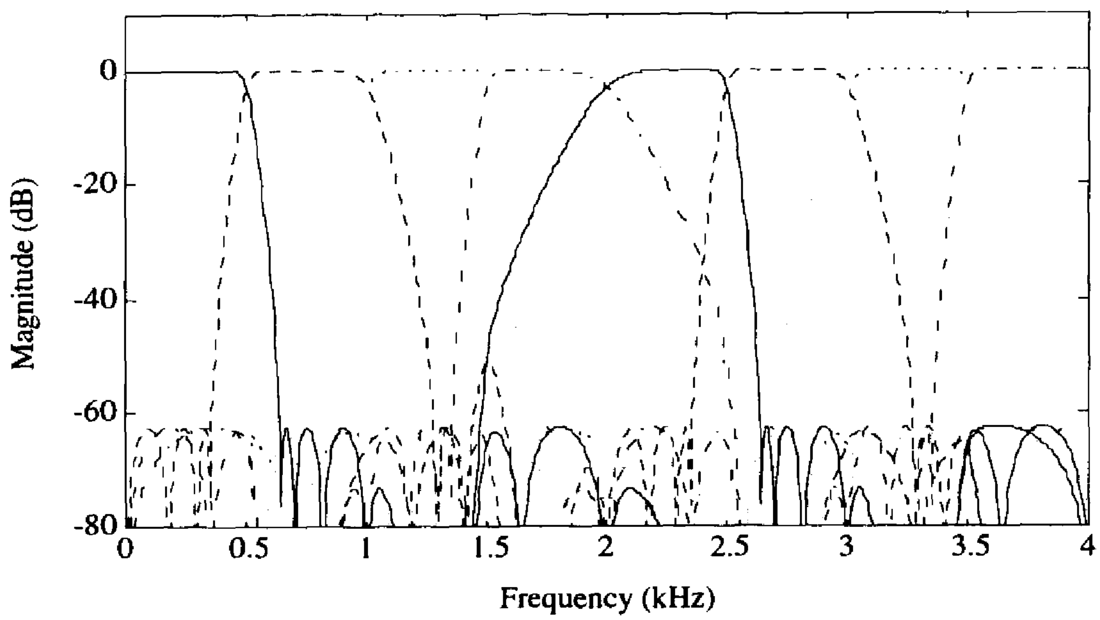


Figure 3.4 - 8 Band, 7th Order IIR QMF Response

The imperceptibility of phase distortion in this filter design was verified through informal listening tests [96][97]. Processing tones and sentences, both 7th and 9th order elliptic filters provided reconstructed signals free from audible distortion. The 7th order design was selected based on its good performance and low complexity.

The complexity of the IIR QMF filtering operation is approximately 300 K operations per second and it introduces a delay of only 5 ms. In contrast, an 8 band 64 tap FIR QMF filterbank with a delay of 8 ms and complexity of 900 K operations per second provided only 40 dB of stopband attenuation [17].

3.1.2 Perceptual Model

A perceptual model is used to determine the optimal shaping of the quantization noise that minimizes the perceptual distortion of the reconstructed speech. The perceptual model takes into account the spectral masking properties of the human auditory system to compute the minimum level of an additive signal (the distortion) that is perceptible in the presence of masking from another signal. The minimum perceptible distortion level is defined as the Just Noticeable Distortion (JND) [59]. Perceptual models are an integral part of several high-quality audio coding schemes [9][21][64].

The JND is usually referenced relative to the signal spectrum using the Signal-to-Mask Ratio (SMR), rather than in absolute levels. Thus, if the distortion introduced by the coder is below the JND from the speech signal, the Signal-to-Noise Ratio (SNR) exceeds the SMR, and the listener is not able to detect the distortion.

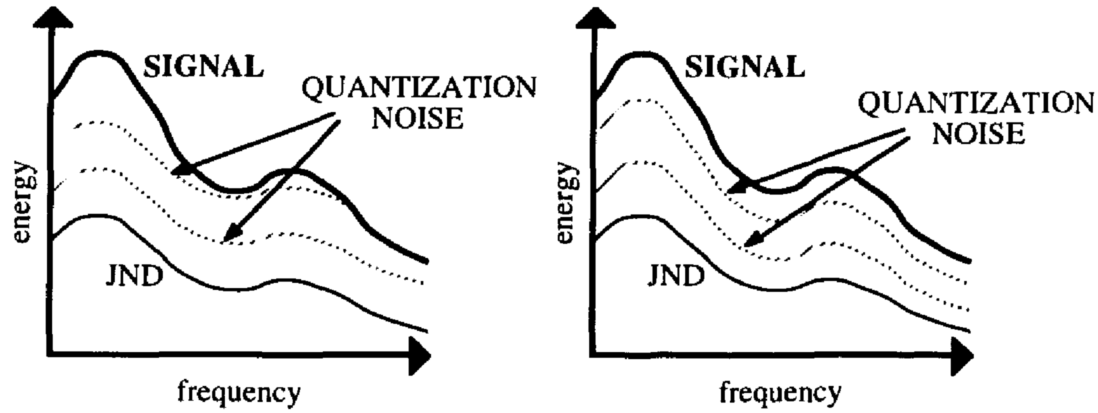
An estimate of the JND is made by computing the signal spectrum in each frame using subband spectral analysis. Subband spectral analysis is a computationally efficient method of performing the analysis on subband samples directly instead of on the source signal. The QMF properties are utilized to cancel aliasing and amplitude errors in the spectral analysis, such that similar spectral resolution characteristics are achieved at significantly reduced complexity. Subband spectral analysis is discussed in more detail in section 3.1.2.1.

Once the source signal spectrum is obtained, both frequency and magnitude axes of the spectrum are transformed into dimensions that are more closely related to the characteristics of the human auditory system. The magnitudes of the spectrum are converted into Sound Pressure Level (SPL) using calibration curves, and the frequencies are converted into critical bands, using the Bark scale [115]. The local noise masking curve is calculated from the energy in each of the 17 Bark bands. Next, an overall noise-masking curve is estimated by power summing the masked power curves due to individual Bark bands. The power of the spectral components is then integrated over each of the 17 Bark bands in the 4 kHz signal bandwidth, and masking curves are calculated from the signal energy. The overall noise curve is obtained by power summing all the masking curves, giving the noise masking level in each of the bands. Bark bands are then grouped corresponding to the eight 500 Hz subbands of the QMF bank. The maximum Bark signal level and minimum Bark noise masking level in each subband is the SMR required to achieve the JND in each subband [96].

Once the JND and SMR have been obtained, the perceptual quality metric usually optimized in perceptual coders is the noise-to-mask ratio (NMR), defined as the ratio of distortion (noise) power to the JND masking threshold [8]. The assumption inherent in this metric is that a signal with lower NMR should be perceived as higher quality than one with higher NMR. Informal listening experiments, however, indicated that this was not always true. Shaping noise based on the NMR often resulted in zero bits being allocated to some subbands at low to medium bit rates, resulting in noticeable distortion. Ad hoc schemes that limited the maximum and minimum number of bits allocated to each subband sometimes resulted in improved perceptual quality at those rates [96].

An interpolated masking threshold (IMT) was devised as an alternative to NMR noise shaping. The IMT is a log-linear interpolation of the SMR, assumed to be the optimal spectrum for noise shaping for a given bit allocation. The perceptual quality of the coded speech is then maximized by achieving the highest possible IMT. Listening tests showed that IMT based noise shaping, compared to NMR, resulted in improved perceptual quality as the bit rate was reduced [96].

A comparison of NMR and IMT noise shaping is shown in Figure 3.5. In both cases, if sufficient bits are available, the JND is the desired noise shape. If fewer bits are available, NMR attempts to distribute the additional quantization noise power evenly among the subbands. This leads to the classical reverse water filling solution for bit allocation. With IMT noise shaping, the desired noise shape is a log-linear interpolation between the source signal spectrum and the JND. Thus, the additional quantization noise power should be distributed in proportion to the required SNR in each band.



(a) NMR noise shaping (b) IMT noise shaping

Figure 3.5 - Comparison of NMR (a) and IMT (b) Noise Shaping

The implementation of the perceptual model is similar to the implementation in the ISO-MPEG Audio Standard [9] in that it computes the JND. However, the JND by itself is not a perceptual metric; it simply indicates a threshold below which noise is not perceptible. The interpolated masking threshold, while not necessarily the optimal perceptual metric, is a significant innovation in that it outperforms the NMR and results in a bit allocation and prioritization algorithm that is very simple to implement.

Optimizing the IMT leads to a bit allocation algorithm based on assignment of bits in proportion to the amount required to reduce the quantization noise below the JND. The allocation and prioritization algorithm is discussed in section 3.1.3.

3.1.2.1 Subband Spectral Analysis

A novel subband spectral analysis methodology was introduced in this work to allow reduced complexity implementation of the perceptual model computation. In order to evaluate the perceptual model, the spectrum of the source signal must be evaluated over each frame. The typical implementation consists of performing a Discrete Fourier Transform (DFT) on the input samples, and computing the power sum to determine the spectral content. However, the filterbank itself can be thought of performing a coarse spectral analysis, and suggests a possible advantage in performing analysis on the subband samples instead of on the input samples directly.

The methodology for transform-based spectral analysis of subband filtered signals is shown in Figure 3.6. Aliasing due to decimation is eliminated by including the effects of the adjacent subband in the analysis of frequencies near the filterbank transition regions. The frequency resolution and spectral leakage are nearly the same as if the transform had been performed on the input directly. In an M band filterbank, the analysis block length is reduced by a factor of M. This reduces the complexity of source compression techniques based on subband decomposition and spectral analysis

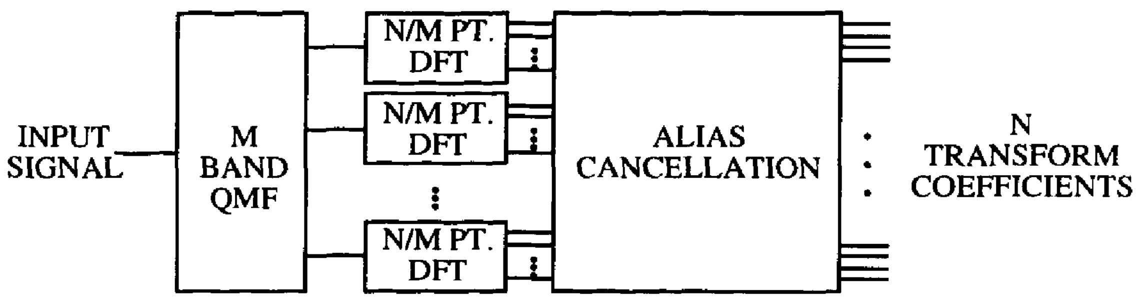


Figure 3.6 - Subband Spectral Analysis Methodology

Spectral analysis can be viewed as the correlation of the sampled input source signal with an analysis sequence to determine coefficients as follows:

$$Y_k = \sum_{n=-\infty}^{\infty} x(n)a_k^*(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})A_k^*(e^{j\omega})d\omega \qquad (3.3)$$

where x(n) is the input signal, $a_k(n)$ is the k-th analysis sequence, $X(e^{j\omega})$ and $A_k(e^{j\omega})$ are their Fourier transforms respectively, and Y_k is the k-th analysis coefficient.

The properties of $A_k(e^{j\omega})$ determine the spectral analysis characteristics. Ideally, as in a DFT analysis, $A_k(e^{j\omega})$ should be a function of the form $\delta(\omega-\omega_k)$, so that for $\omega \neq \omega_k$, Y_k is independent of $X(e^{j\omega})$. In practice, however, $a_k(n)$ must be a finite length (windowed) sequence so that $A_k(e^{j\omega})$ approximates the delta function by having a narrow main lobe with attenuated side lobes.

In subband spectral analysis, the analysis is done on the decimated subband samples. Consider first the 2 band case, as depicted graphically in Figure 3.7.

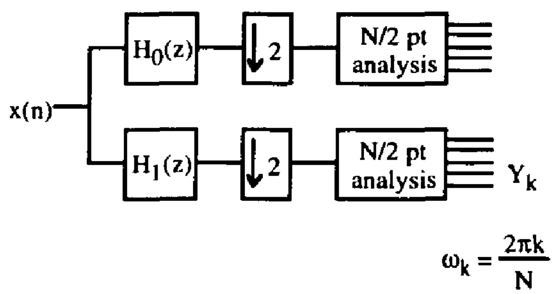


Figure 3.7 - Subband Spectral Analysis Model

The analysis coefficient Y_k is given by:

$$\begin{aligned} Y_k &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} x(m) h_i(n-m) w_k^*(n) \\ &= \sum_{m=-\infty}^{\infty} x(m) \sum_{n=-\infty}^{\infty} h_i(n-m) w_k^*(n) \end{aligned} \tag{3.4}$$

so the analysis sequence is now given by

$$a_k^*(n) = w_k^*(n) * h_i(-n) \tag{3.5}$$

where $w_k(n)$ is the decimated k-th analysis sequence and $h_i(n)$ is the subband impulse response of the band.

The analysis response $A_k(e^{j\omega})$ is the product of the decimated analysis transfer function consisting of two aliased "main" lobes located at ω_k and $\omega_{k'}=\pi+\omega_k$, and the subband conjugate response. One of the lobes corresponds to the analysis frequency and the other corresponds to the aliased frequency. Ideally, the subband filter response would leave the analysis lobe unaltered and would reject the aliased lobe.

However, given the finite transition regions and out-of-band rejection, errors are introduced in the analysis. If the frequency to be analyzed is near a transition region, amplitude error is introduced due to attenuation in the passband edge and aliasing error is introduced due to the limited attenuation in the stopband edge. For this reason, subband spectral analysis had not been considered to produce acceptable results [9].

However, the subband spectral analysis methodology can be modified such that the results are acceptable. For analysis frequencies not near the transition region, the errors may be acceptable since there is less attenuation of the analysis frequency in the passband and more attenuation of the aliasing frequency in the stopband. Near the transition region, the properties of the QMF filterbank can be used to cancel out the errors as follows.

Let $h_i(n)$ and $f_i(n)$ be the impulse responses and $H_i(e^{j\omega})$ and $F_i(e^{j\omega})$ the frequency responses of the analysis and synthesis filterbank, respectively, for subband i. Consider now the impulse response

$$g_k(n) = F_i(e^{j\omega_k})h_i(n) + F_j(e^{j\omega_k})h_j(n) \tag{3.6}$$

with frequency response

$$G_k(e^{j\omega}) = F_i(e^{j\omega_k})H_i(e^{j\omega}) + F_j(e^{j\omega_k})H_j(e^{j\omega}) \tag{3.7}$$

If the filterbank is designed to cancel aliasing and has no amplitude distortion, then $|G_k(e^{j\omega})| = 1$ for $\omega=\omega_k$, and $|G_k(e^{j\omega})| = 0$ for $\omega=\omega_{k'}$, where ω_k is the analysis frequency in subband i and $\omega_{k'}$ is the aliasing frequency in subband j.

Note that even in the presence of amplitude distortion and aliasing in the subband filterbank design, the error in the subband spectral analysis is minimized by minimizing the amplitude and aliasing errors in the filterbank. This observation is useful in extending this methodology to the M band filterbank case. Rather than considering all M bands, only the adjacent band needs to be considered. For non-adjacent bands, the stopband attenuation ensures that the attenuation and aliasing error contribution from that band is minor.

For subband spectral analysis, the aliasing and amplitude distortion introduced at $\omega_{k'}$ can be canceled by modifying the analysis sequence to

$$\begin{aligned} a_k^*(n) &= w_k^*(n) * g_k(-n) \\ &= F_i(e^{j\omega_k})(w_k^*(n) * h_i(-n)) + F_j(e^{j\omega_k})(w_k^*(n) * h_j(-n)) \end{aligned} \tag{3.8}$$

This method of modified analysis is denoted alias-canceled subband spectral analysis. Figure 3.8 depicts the operations to evaluate Y_k, the k-th coefficient of an N point analysis in an M band filterbank example.

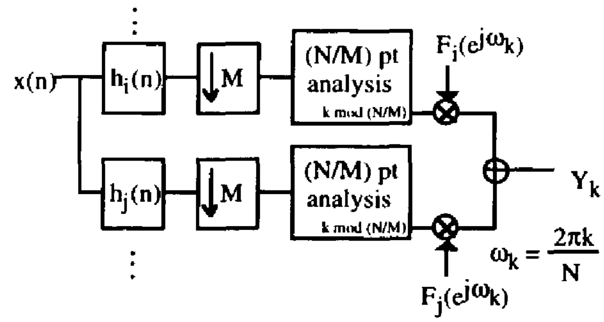


Figure 3.8 - Alias-Canceled Subband Spectral Analysis Model

To summarize, in order to perform subband spectral analysis with no aliasing errors, the only additional operation required is a linear combination of the coefficients at ω_k and ω_k' with the non-modified analysis sequences. In this manner, the analysis coefficients can be obtained by performing a shorter block analysis on the subband samples rather than on the input signal directly. Windowing and various transform based spectral analysis methods such as the DFT and the Discrete Cosine Transform (DCT) are easily adaptable to subband analysis with this method.

To illustrate the effectiveness of this method, we compared the analysis response of a 256 point rectangular window DFT, the response of a 32 point subband DFT based on an 8 band, zero amplitude distortion, tree structured IIR QMF filterbank, and the response of the alias-canceled subband DFT for the same filterbank structure. The response of the filterbank is shown in Figure 3.3.

The spectral analysis response was computed as a power sum of the analysis at ω_k and 2π-ω_k (i.e., $|A_k(e^{j\omega})|^2 + |A_k(e^{j(2\pi-\omega)})|^2$). This makes the alias frequency more easily noticeable in the frequency response since it then simply appears as the image of the analysis frequency. In Figure 3.9, the analysis frequency is $\pm\ 63\pi/128$ and alias frequency is $\pm 65\pi/128$ (k=63, analysis subband i=3, aliasing subband j=4, N=256).

The desired analysis response is given by the 256 point DFT analysis with a rectangular window, shown in Figure 3.9 (a). The non alias-canceled subband DFT analysis response, shown in Figure 3.9 (b), results in both amplitude distortion (the main lobe at $63\pi/128$ is at -3 dB) and considerable leakage (the adjacent lobe at $65\pi/128$ is at -6 dB). Note that there is very little leakage from the other bands due to the stopband response of the subband filter.

The alias-canceled subband DFT analysis response, shown in Figure 3.9 (c), corrects both amplitude and aliasing distortion. The main lobe is at 0 dB and the adjacent lobe is at -13 dB, with a null at $\omega = 65\pi/128$. Compared to the direct DFT analysis response, as shown in Figures 3.9 (d), there is little difference in amplitude and aliasing errors. In addition, there is rejection of spectral components in the other subbands, since they have been "filtered" out of the subband samples. In practice, this is only a minor advantage, since the spectral leakage is dominated by frequencies closer to the analysis frequency.

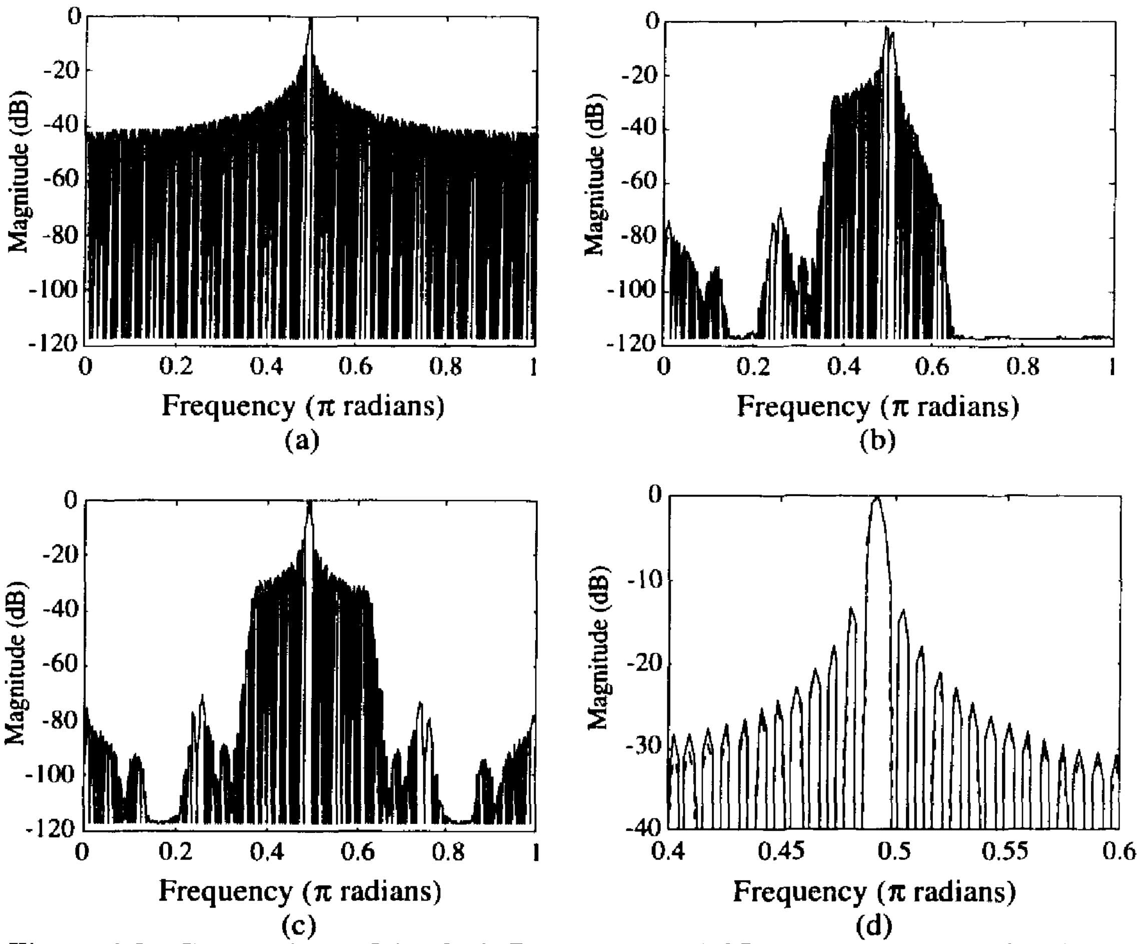


Figure 3.9 - Comparison of Analysis Responses: (a) 256 pt. Direct DFT, (b) 32 pt. Non Alias-Canceled Subband DFT, (c) 32 pt. Alias-Canceled Subband DFT, (d) Direct DFT (dashed) and Alias-Canceled Subband DFT (solid)

This example demonstrates how a simple methodology involving a linear combination of coefficients obtained by a DFT of subband samples achieves essentially the same results as obtained from a DFT of the input samples directly. If this method is not used, the recourse is to either perform the spectral analysis on the source sequence directly or perform the analysis on the subband samples and accept the distortion introduced in the transition region. The former requires higher computational complexity, while the latter leads to inadequate results due to the errors introduced in the analysis.

Subband spectral analysis allows the desired accuracy to be achieved at reduced complexity. With a fast transform such as the Fast Fourier Transform (FFT), the complexity is reduced from N/2 log N to N/2M log N/M for an M band coder. When compared to performing the spectral analysis on the input signal directly, the analysis block length for subband spectral analysis with an 8-band filterbank is reduced by a factor of 8 and the complexity of a 256 point FFT computed every frame is reduced by 37.5%. This reduces the overall complexity of computing the perceptual model from 200 K to 125 K operations per second. The reduction in complexity is even greater if the perceptual model is simplified so that analysis is required only over a limited interesting region, ignoring one or more of the subbands.

3.1.3 Bit Allocation and Prioritization

The bit allocation determines how many bits are used to quantize the samples in each subband, thus determining the shaping of the quantization noise in a subband coder. In an embedded coder, bit prioritization determines the significance of the bits, so the embedded representation can more effectively scale over the entire range of coding rates. The perceptual model discussed in section 3.1.2, provides a cost measure for allocating bits that may be used to optimize the quality of the subband coder.

The dynamic bit allocation and prioritization algorithm determines how many bits are allocated to quantize the samples in each subband for the given frame. The bit allocation determines the quantization noise power in each subband, and therefore the overall quantization noise spectrum. The bit prioritization algorithm orders each bit allocated in priority, so that the encoded bitstream can be truncated in a perceptually prioritized manner.

The Interpolated Masking Threshold (IMT) leads to a simple bit allocation and prioritization algorithm. The quantization noise is determined by the number of bits allocated to each subband. The signal to noise ratio (SNR) is approximately given by

$$SNR_j = 6.02\ dB \cdot n_j \tag{3.9}$$

where n_j is the number of bits per subband sample used to quantize subband j. This approximation holds fairly well since the subband samples are normalized in each frame such that the statistical distribution is nearly Gaussian, and the rate distortion limit for quantizing a memoryless Gaussian source is given by (3.9).

If sufficient bits are available, then each subband is allocated the number of bits required for the SNR to exceed the Signal to Mask Ratio (SMR) given by the Just Noticeable Distortion (JND) masking threshold. Thus, the bit-allocation algorithm to achieve perceptual transparency is

$$n_j > SMR / 6.02 \tag{3.10}$$

If the number of bits available is not sufficient, then the bit allocation should be based on the IMT. Thus the bit allocation is

$$n_j > \alpha\, SMR / 6.02 \tag{3.11}$$

where α is the IMT scale such that IMT = α SMR. If $\alpha = 1$, then the JND is met and the quantization noise is not perceptible. For $0 < \alpha < 1$, the quantization noise is perceptible, but is shaped to optimize the IMT.

Since the number of bits allocated to each subband is proportional to the ideal bit allocation (the number of bits required to achieve the JND), we denote this algorithm proportional bit allocation.

The proportional bit allocation algorithm also implies a prioritization of each bit allocated. The bit prioritization corresponds to allocating fewer bits for the frame when truncation occurs, such that the allocation after truncation still maximizes the IMT for the given bit rate.

The implementation of bit prioritization can be simplified by using the ideal bit allocation directly instead of the IMT to compute the actual bit allocation. Thus the IMT need not be calculated explicitly. Table 3.1 illustrates an example of the bit allocation and prioritization computation for a frame. First, the JND is computed by the perceptual model for the frame, giving the required SMR. The ideal bit allocation n_j is then determined from the SMR. The bits are prioritized one at a time based on maximizing the proportion of the assigned bits versus the ideal. If subbands have already had an equal proportion of bits assigned, priority is given first to subbands with higher ideal bit allocations, then to lower frequency subbands. The bit allocation at each step N (shown in bold) corresponds to the effective bit allocation if the bitstream is truncated at that point. The bit prioritization follows from the allocation at each step. At each step N, a bit is allocated (bold), resulting in the bit allocation and prioritization to the subbands in the following order: 1 3 4 2 6 1 1 3 4 1 (10 bits/subband sample total).

Table 3.1 - Bit Allocation and Prioritization Example

subband	1	2	3	4	5	6	7	8
SMR (dB)	22.0	4.0	12.0	8.0	0.0	4.0	0.0	0.0
n_i	4	1	2	2	0	1	0	0
N=1	**1**	0	0	0	0	0	0	0
N=2	1	0	**1**	0	0	0	0	0
N=3	1	0	1	**1**	0	0	0	0
N=4	1	**1**	1	1	0	0	0	0
N=5	1	1	1	1	0	**1**	0	0
N=6	**2**	1	1	1	0	1	0	0
N=7	**3**	1	1	1	0	1	0	0
N=8	3	1	**2**	1	0	1	0	0
N=9	3	1	2	**2**	0	1	0	0
N=10	**4**	1	2	2	0	1	0	0

The bit allocation and prioritization scheme in this coder is a significant improvement over uniform bit allocation and water-filling bit allocation, the bit allocation schemes prevalent in subband speech and audio coders. Uniform bit allocation is the simplest method to perform bit allocation, but has poor correspondence to perceptual quality. Water filling bit allocation is used to maximize either the SNR or the NMR, both of which also have poorer correspondence to perceptual quality than IMT.

3.1.4 Quantization

In quantizing the subband samples, we attempt to minimize the quantization noise power for a given bit allocation. The subband samples in each 20 ms frame are scaled so that the distribution of subband samples closely matches a Gaussian distribution. The scale is quantized with a logarithmic quantizer and the subband samples with an embedded non-uniform scalar quantizer. The scale and bit allocation are transmitted as overhead, and the embedded quantization indices of the subband samples are transmitted according to the bit allocation and prioritization algorithm.

This scheme is based on the assumption that the subbands have sufficiently narrow bandwidth to decorrelate the subband samples. This assumption neglects that speech and audio signals are non-stationary, and that significant intraband correlation may exist. It is common, however, to assume a memoryless source model for subband coders [17].

Source statistics were obtained by analyzing male and female speech segments from the DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus (NTIS PB91-505065). The probability distribution for these source signals were more peaked (high probability density near zero and very slow tail decay) than Laplacian and Gaussian distributions. Even when the silence portions of speech are removed (by removing segments where the energy falls below a minimum threshold), the distribution of the samples is still highly peaked. The highly peaked distribution results in very poor quantization performance when a simple quantization algorithm such as uniform quantization is used. However, if the samples over a short frame are normalized (i.e., scaled by the variance), then the distribution of the normalized samples is nearly Gaussian, as shown in Figure 3.10. Since the normalized subband samples are nearly Gaussian, the Signal to Noise Ratio (SNR) closely follows the 6.02 dB per bit assumption in the bit allocation and prioritization scheme.

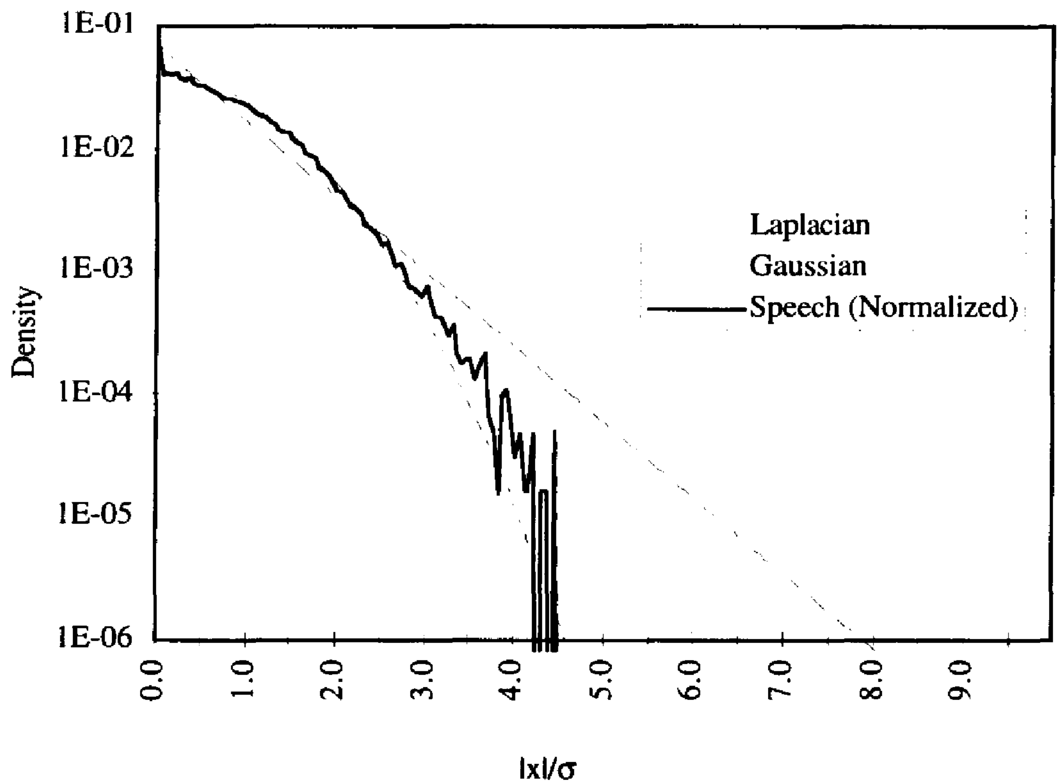


Figure 3.10 - Statistical Distribution of Normalized subband samples

Although the rate distortion bound implies the more peaked the distribution the lower the achievable distortion should be at the given rate, this is not necessarily true at a fixed complexity. For example, scalar quantizers have better performance when quantizing Gaussian variables than when quantizing Laplacian variables. This is true for both uniform and non-uniform (Lloyd-Max) quantizers. The performance for scalar quantizers degrades even more for more peaked distributions, even though the rate distortion function decreases. This apparent paradox results from the fact that more gain is achievable with multi-dimensional shaping when the distribution is highly peaked.

The highly peaked distribution results in very poor quantization performance when a simple quantization algorithm such as uniform quantization is used. Companding provides a simple way of improving the performance, so u-law and A-law PCM have traditionally been used to provide improved quantization performance with low complexity.

A more complex approach is to frame the data and use a scaling parameter for the frame. Given a short enough frame length, the scaling of the frame results in a nearly Gaussian distribution of the samples. This is due both to the "bounding" of the samples by the scale factor and the correlation of the magnitude of adjacent samples.

When the source signal is split into subbands, the statistical analysis showed a similar peaked distribution for the subband samples. Therefore, it is reasonable to expect that the same quantization methodologies (companding, frame scaling, and ADPCM) can be used to quantize the subband samples. However, the available gain to ADPCM has been reduced, since some of the redundant information has actually been removed, resulting in non-equal energies in each of the subbands. For this reason, most subband based compression algorithms assume the subband samples are uncorrelated.

Since the subband coder approach assumes no correlation (i.e., memoryless source), we selected a memoryless embedded scalar quantizer matched to the Gaussian statistical distribution of the samples. Since the perceptual model is based on the SNR for each subband, the appropriate error measure is the squared error or Euclidean distance.

The embedded scalar quantization approach was chosen based on its low complexity in implementation. The principle behind embedded quantizers is that even a partial index should provide sufficient information to determine a suitable reconstruction value. This is a desirable property if the encoded pattern is not necessarily received in its entirety at the decoder. However, an embedded quantizer incurs a performance penalty relative to the optimal quantizer at any one rate. Some distortion tradeoff at the different effective bit rates must occur in the embedded quantizer design. The quantizer performance can be made optimal at one of the target bit rates, but the embeddability constraint implies that a suboptimal choice must be made at other rates. However, the performance penalty can be minimized through proper design of the embedded quantizer, as is shown in Section 3.1.4.1.

An additional issue considered in the quantizer design is robustness to source statistics. Robustness is the property of the quantizer being able to perform adequately if the statistics are different from those assumed. Most commonly, this is observed if the variance of the source does not match that assumed in the design of the quantizer. A robust quantizer implementation makes the error performance relatively insensitive to non-optimal scaling. A non-uniform scalar quantizer offers significant improvement in robustness compared to uniform scalar quantizers for Gaussian sources [4].

The embedded scalar quantizer offers several advantages over uniform scalar quantization. It is superior in minimum square error and robustness to a uniform quantizer. Even at the lower bit rates, the degradation of the embedded quantizer compared to the optimal Lloyd-Max quantizer is minimal [102].

Embedded quantization with bit prioritization allows an embedded bitstream to be assembled as described in the section 3.1.5. The bit prioritization ensures that the truncated bitstream's effective bit allocation is the same as would have been generated at the lower bit rate. The embedded quantizer design ensures that the quantizer performance with a truncated bitstream is also approximately the same as if a non-embedded quantizer had been used at the same bit rate. Thus, the design methodology ensures that only a slight performance penalty has been paid in generating a scalable, embedded output.

3.1.4.1 Optimal Embedded Scalar Quantization

The principle behind embedded quantizers is that even a partial index should provide sufficient information to determine a suitable reconstruction value. In variable rate applications, this is a desirable property if the encoded pattern is not necessarily received in its entirety at the decoder. Using embedded quantizers results in greater quantization noise then non-embedded quantizers, since embedding imposes an additional constraint on the threshold of quantizers, whereas non-embedded quantizers are free to select the thresholds for each quantization range independently. However, there are similarities in the characteristics of optimal quantizers and embedded quantizers that allows optimal embedded quantizers to be defined such that their performance is very close to optimal non-embedded quantizers over the entire range of rates.

The simplest and most commonly used scalar quantizer is the uniform quantizer. The uniform quantizer is constrained to having uniform step sizes, and reconstruction values at the midpoint of each quantization level.

The optimal scalar quantizer for non-uniform source probability distribution functions (pdf) requires non-uniform step sizes. The optimal scalar quantizer under minimum-squared-error (mse) criteria for memoryless sources of a given distribution can be found using the Lloyd-Max algorithm [38][73][78]. Quantization noise can be reduced at the same bit rate only by resorting to non-scalar methods, either through variable length (entropy-coded) quantizer or multi-dimensional (vector) quantizer approaches [39][118]. For many applications, a fixed-length scalar quantizer is desirable for its simplicity, so optimal scalar quantization is a very practical approach.

The Lloyd-Max quantizer is characterized by thresholds and reconstruction values that meet two criteria:

- The optimal reconstruction values are the centroids of the source pdf over the corresponding quantization region.
- The thresholds between adjacent regions are the midpoints between the corresponding reconstruction values.

Lloyd-Max quantizers can be easily designed through an iterative process using the rules described above. Design values for Gaussian and Laplacian distributions at various bit rates can be found in [61].

The Lloyd-Max quantizer can also be approximated for high bit rates by Bennett's formula [6], the companding function

$$f(x) = c\int_0^x p(u)^{\frac{1}{3}}\,du \tag{3.12}$$

followed by a uniform quantizer [38]. Quantizers designed with the companding approximation have performance close to the Lloyd-Max quantizers.

Any scalar quantizers can be embedded by using sign-magnitude or 2's complement binary notation indices and imposing an embeddability constraint on the thresholds of the lower bit rate representations. The rate of the quantizer can then be varied by simply truncating the Least-Significant-Bits (LSBs) and assigning new reconstruction values.

The embeddability constraint for a scalar quantizer can be described [102] as follows:

- The indexing scheme must be tree-structured such that the branch at each level is described by part of the index. A partial index allows decoding to a partial level of the tree. Truncation of the index occurs for the lowest level first. A binary tree is the most straightforward implementation, allowing truncation one bit at a time.
- The quantization region corresponding to each branch should be the union of the quantization regions of its daughter branches. At the full depth of the tree, the quantization regions should be non-overlapping and their union consists of the entire input range. The regions should be arranged such that adjacent regions share the same parent at all levels of the tree. This facilitates the tree search and ensures that the mse is minimized at all levels of the tree. The optimal reconstruction value is the centroid of the quantization region.

The embedded scalar quantizer is fully described by the tree structure of the quantization indices and the quantization regions or thresholds at the highest bit rate. The indexing scheme must be tree-structured such that the branch at each level is described by part of the index. A partial index allows decoding to a partial level of the tree. At the lowest level of the tree, the quantization regions are non-overlapping and their union consists of the entire input range. The quantization region corresponding to each branch is the union of the quantization regions of its daughter branches. The optimal reconstruction value is the centroid of the quantization region. Under these constraints, the only design values to be determined for an embedded scalar quantizer are the tree structure and the quantization regions or thresholds. The assignment to each branch (indexing) and reconstruction values are given by the above definition.

Although it is not possible for the embedded quantizer to meet the Lloyd-Max criteria at all levels, it is possible to meet them at a single level. Doing so defines the thresholds for all higher levels of the trees and constrains the thresholds for all the lower levels. The lower level thresholds can be defined by placing the additional constraint of minimum mse. In this case the undefined thresholds at lower levels are selected to be the midpoint between the centroids and can be found with an iterative procedure similar to the Lloyd-Max criteria, but constrained to the quantization region boundaries.

The design procedure for an optimal embedded non-uniform scalar quantizer with binary indexing is to meet the Lloyd-Max criteria at the full bit rate (lowest level) and to use the embeddability constraint to define the higher levels. The embedded quantizer is also optimal (for symmetric densities) at a bit rate of 1, since it automatically assigns the centroid as the reconstruction value. The embedded quantizer will therefore be optimal at the full rate and at the lowest rate, and somewhat sub-optimal at the other rates, relative to Lloyd-Max quantizers. But the sub-optimality is very small, and the design is in fact the optimal embedded scalar design.

Embedded uniform quantizers are far from optimal. The reconstruction values are constrained to be the midpoint of the quantization region instead of the centroid. This is not a problem if the probability density is relatively uniform over the quantization region, but this is especially not true for the outlying regions, the tail of the distributions. This results in reconstruction values that are too far from the center of the distribution.

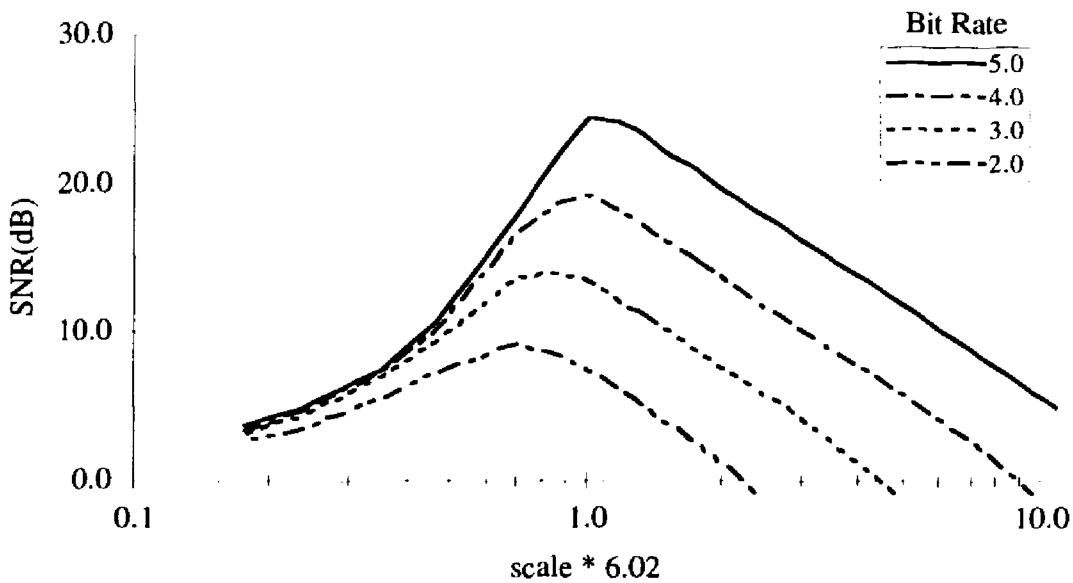
A second reason that embedded uniform scalar quantizers are far from optimal is the difference in optimal scale required at each bit rate. Table 3.2 shows the required scale (2^N * step size / σ, where N is the quantizer bit rate and σ is the square root of the variance) for Gaussian and Laplacian source densities. The change in scale required is more drastic for heavily tailed densities such as the Laplacian.

Table 3.2 - Optimal Uniform Quantizer Scale for Gaussian and Laplacian Densities

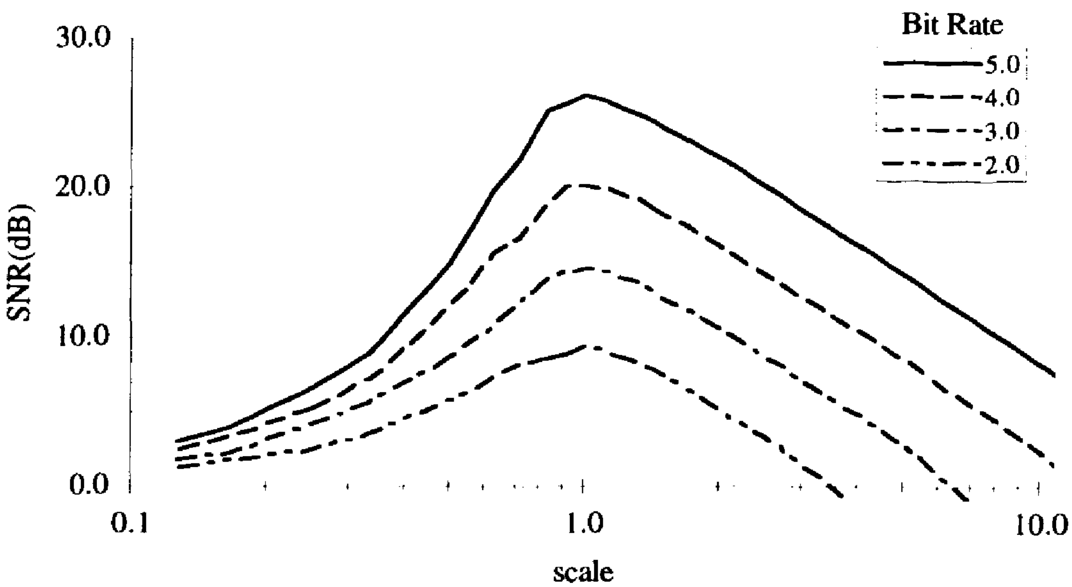
N	Gaussian	Laplacian
1	3.19	2.83
2	3.98	4.35
3	4.69	5.85
4	5.36	7.38
5	6.02	8.96

Embedding the quantizer forces a single scale to be used for all bit rates. If the scale is too large, then the quantization regions are too large, and the Signal to Noise Ratio (SNR) degrades because of loss of granularity. If the scale is too small, then the quantization regions are too small, and the SNR degrades because of clipping or overloading. Since the effect of clipping is more drastic than loss of granularity, the scale should be optimized for the highest bit rate. However, using that scale results in poor SNR at the lower bit rates. The net result is that an embedded uniform scalar quantizer has poor performance at low bit rates.

The scaling problem is illustrated in Figures 3.11 and 3.12 Uniform quantizers require different scales to achieve optimal SNR performance at different rates, as observed by the location of peaks in the SNR curve. Uniform quantizers are also more sensitive to scaling errors, as observed by the sharper SNR curves compared to non-uniform quantizers. Non-uniform quantizers are more suitable for embedding and result in more robust quantization.

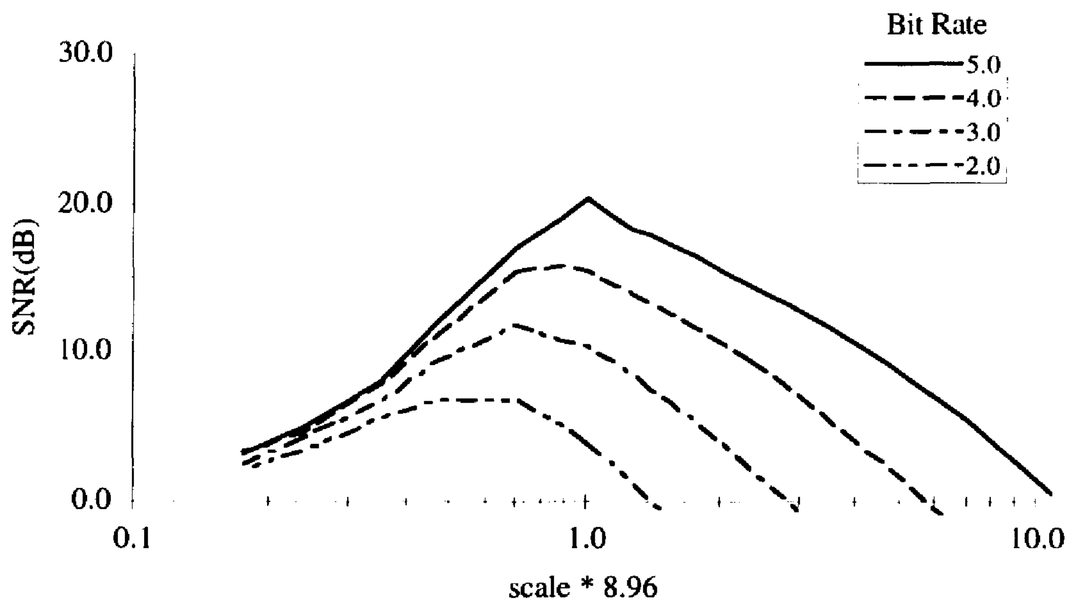


(a)

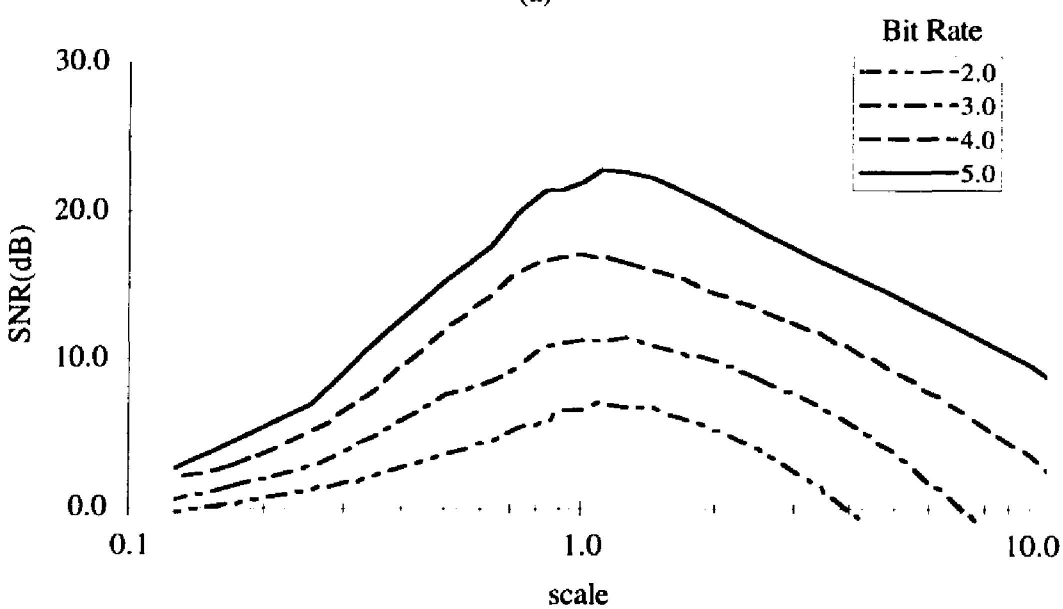


(b)

Figure 3.11 - Uniform (a) and Non-Uniform (b) Scalar Quantizer Performance for Gaussian Memoryless Sources as a function of scale



(a)



(b)

Figure 3.12 - Uniform (a) and Non-Uniform (b) Scalar Quantizer Performance for Laplacian Memoryless Sources as a function of scale

The non-uniform embedded quantizer does not suffer from the scaling problems. The key is the fact that the optimal non-uniform quantizer can be approximated by Bennett's companding function. The companding function is independent of bit rate, so the same function is used at all bit rates to approximate the optimal scalar quantizer. Since the companding function is followed by a scalar quantizer with a fixed quantization range and uniform step size, the quantizers generated by such a procedure over various bit rates result in an embedded quantizer.

The relationship between Bennett's companding function and optimal embedded quantization is of significant importance. Embedded uniform scalar quantizers will always result in performance degradation for non-uniformly distributed sources. Ad-hoc schemes for non-uniform embedded scalar quantizers, such as those proposed in [17], allow performance to be optimized at a single rate, but are sub-optimal at other rates.

The methodology employed here allows the quantizer to be optimal at full rate, and nearly optimal quantizer at the lower rates. The performance of the optimal embedded quantizer at the full bit rate and at the bit rate of 1 is the same as the performance of the Lloyd-Max quantizer. The performance of the embedded quantizer at other bit rates is only slightly degraded from the Lloyd-Max quantizer at that bit rate, the degradation due to the companding approximation.

Optimal 5 bit embedded quantizers for normalized Gaussian and Laplacian distributions were designed and evaluated. The thresholds and reconstruction values are given in Table 3.3 for Gaussian sources and Table 3.4 for Laplacian sources. The performance was evaluated by integrating the quantization mean square error over the probability distribution, and comparing to optimal non-embedded uniform and non-uniform scalar quantization, as summarized in Figures 3.13 and 3.14. The embedded quantizer offers performance nearly as good as the Lloyd-Max quantizer. The performance is better than the non-embedded uniform quantizer and substantially better than an embedded uniform quantizer. The performance advantage is greater for Laplacian sources than Gaussian sources, due to the degradation of the uniform quantizer for heavily tailed densities.

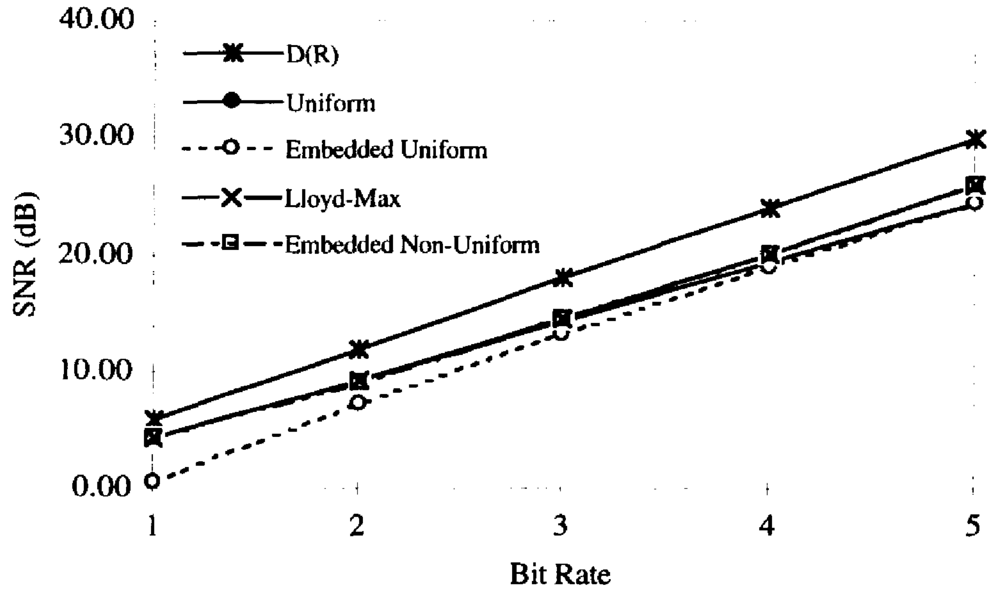
The embedded quantizer design is particularly attractive for variable rate applications. In contrast to the embedded uniform quantizer, the performance over the entire range of possible rates does not degrade significantly, allowing higher performance and more robust system design. Other non-uniform embedded scalar quantizers have been described in the literature [17], but were not optimally implemented. The performance was substantially inferior to the Lloyd-Max quantizer and degraded significantly at the higher bit rates.

Table 3.3 - Thresholds and Reconstruction Values for 5 Bit Embedded Gaussian Quantizer

index	T_k	$R_k(5)$	$R_k(4)$	$R_k(3)$	$R_k(2)$	$R_k(1)$
00000	0.0000	0.0660	0.1318	0.2618	0.5082	0.7979
00001	0.1322	0.1983				
00010	0.2651	0.3318	0.3983			
00011	0.3995	0.4672				
00100	0.5364	0.6056	0.6745	0.8096		
00101	0.6768	0.7479				
00110	0.8217	0.8954	0.9687			
00111	0.9726	1.0497				
01000	1.1313	1.2128	1.2935	1.4467	1.6312	
01001	1.3001	1.3874				
01010	1.4824	1.5773	1.6704			
01011	1.6828	1.7883				
01100	1.9091	2.0298	2.1443	2.2929		
01101	2.1743	2.3188				
01110	2.5050	2.6912	2.8271			
01111	2.9764	3.2616				

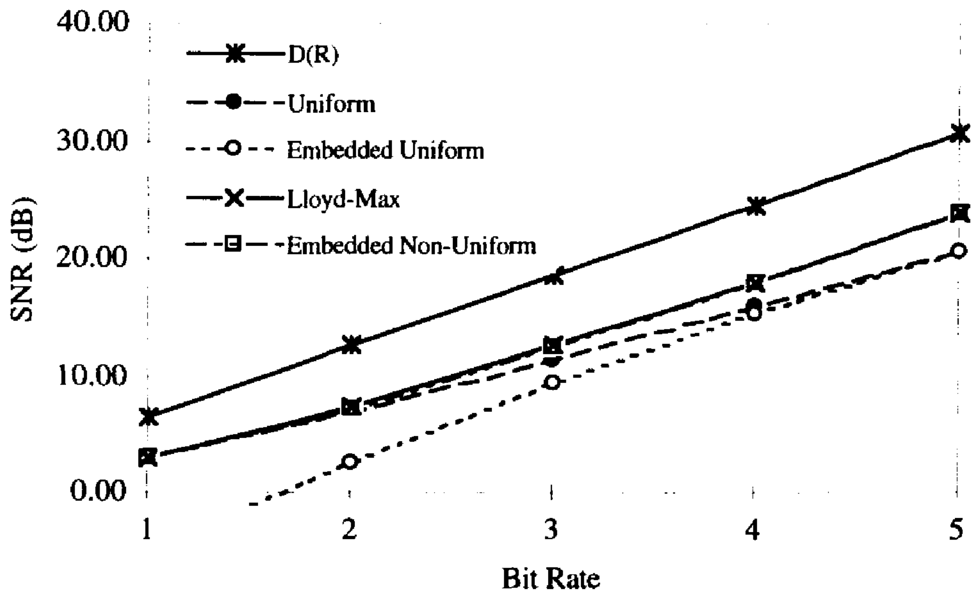
Table 3.4 - Thresholds and Reconstruction Values for 5 Bit Embedded Laplacian Quantizer

index	T_k	$R_k(5)$	$R_k(4)$	$R_k(3)$	$R_k(2)$	$R_k(1)$
00000	0.0000	0.0641	0.1279	0.2535	0.4834	0.7071
00001	0.1323	0.2005				
00010	0.2734	0.3463	0.4188			
00011	0.4246	0.5028				
00100	0.5873	0.6718	0.7557	0.9183		
00101	0.7637	0.8555				
00110	0.9560	1.0565	1.1560			
00111	1.1676	1.2786				
01000	1.4026	1.5266	1.6486	1.8737	2.1097	
01001	1.6671	1.8076				
01010	1.9695	2.1314	2.2885			
01011	2.3226	2.5138				
01100	2.7473	2.9807	3.1977	3.4544		
01101	3.2805	3.5803				
01110	4.0001	4.4199	4.7072			
01111	5.1270	5.8341				



Bit Rate	D(R)	Uniform	Lloyd-Max	Embedded-Uniform	Embedded-Non-Uniform
1	6.02	4.40	4.40	0.65	4.40
2	12.04	9.25	9.30	7.19	9.14
3	18.06	14.27	14.62	13.16	14.51
4	24.08	19.38	20.22	19.02	20.20
5	30.10	24.57	26.01	24.57	26.01
source	Noll[81]	Max[78]	Jain[56]	simulation	simulation

Figure 3.13 - Optimal Embedded Quantizer SNR vs. Non-embedded Uniform and Lloyd-Max Quantizers for quantization of Gaussian sources



Bit Rate	D(R)	Uniform	Lloyd-Max	Embedded-Uniform	Embedded-Non-Uniform
1	6.62	3.01	3.01	-4.55	3.01
2	12.66	7.07	7.54	2.70	7.32
3	18.66	11.44	12.64	9.35	12.52
4	24.69	15.96	18.13	15.38	18.10
5	30.73	20.60	23.87	20.62	23.89
source	Noll[81]	Jeong [61]	Jain [56]	simulation	simulation

Figure 3.14 - Optimal Embedded Quantizer SNR vs. Non-embedded Uniform and Lloyd-Max Quantizers for quantization of Laplacian sources

3.1.5 Bitstream Formatting

Bitstream formatting arranges the coded bitstream for each 20 ms frame by perceptual significance. The overhead consisting of the subband bit allocation and scaling is the most important information in each frame. The bit allocation information consists of 24 bits, 3 per subband. The scaling consists of 36 bits, 4 for scaling the frame, and 4 for scaling each of the 8 subbands. The subband sample quantization indices are grouped into 20 bit blocks (1 per subband sample), and arranged based on the bit allocation and prioritization scheme. The variable length embedded bitstream frame structure is shown in Figure 3.15.

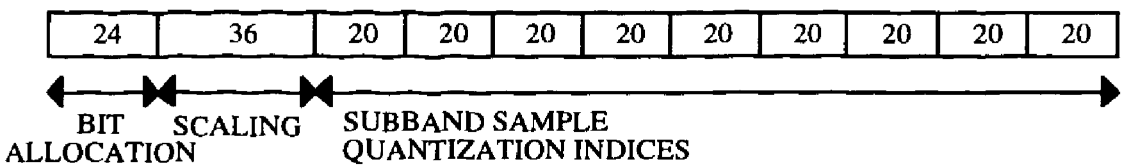


Figure 3.15 - Bitstream Frame Structure

Each frame can be truncated as required to meet various source-rate requirements. The coder can thus be operated in several configurations as desired by the system, such as fixed bit-rate, variable bit-rate, or network controlled adaptive source rate configurations.

In a fixed bit-rate configuration, the coder output is truncated to a particular rate in every frame, resulting in a fixed frame size. The bit allocation corresponds to a dynamic-frequency, fixed-time allocation minimizing the Interpolated Masking Threshold (IMT) on a frame-by-frame basis. This is the typical configuration in which most speech coders currently operate, minimizing a cost measure over each frame while constrained to a fixed use of system bandwidth.

In a variable bit-rate configuration, the coder operates in a perceptually transparent mode, producing varying frame sizes such that the number of bits is the minimum required to mask out all the quantization noise in each frame. Alternatively, a fixed proportion of the bitstream is truncated, such that the coder operates in a variable rate, fixed frame quality mode. This is similar to variable rate configurations that maintain a fixed quality while minimizing the average use of system bandwidth, but with the added capability of adjusting the quality as desired.

In a network controlled adaptive source rate configuration, the source rate is selected according to the status of the network. The source-rate adaptation provides the network with the capability to effectively deal with the transmission environment, which may be limited by interference, fading, congestion, or lack of resources. The coder provides an estimate of the signal quality, allowing the system to optimize the source and channel rate, transmission power and bandwidth, in a perceptually significant manner.

In all these configurations, no overhead is required to describe each frame. Since the bitstream is embedded, truncation of frames due to any rate control algorithm affects only the frame size. The decoder treats each frame as a truncated frame and no information other than the received frame size is required.

3.1.6 Algorithm Complexity and Delay

The algorithmic complexity is dominated by the filterbanks and perceptual model computations. The analysis and synthesis filterbanks require 150 K operations per second each, and the perceptual model requires an additional 125 K operations per second. The total speech coder algorithm requires approximately 325 K operations per second for the encoder and 200 K operations for the decoder. This is a significant improvement over subband coders such as the one proposed by Cox et al. [17], which required 900 K operations per second just for the filterbank.

The algorithmic delay of 25 ms is due to buffering of one frame (20 ms) and the delay of the filterbank (5 ms). This does not include additional decoder implementation delay, channel coder delay, and possible interleaving of frames needed to combat fading in the transmission channel.

3.2 Performance Evaluation

Traditionally, the performance of waveform coding systems has been evaluated using the SNR criterion. For encoded speech and audio signals, however, the SNR is a poor indicator of distortion in the coded signals. Signals with high SNR may contain significant levels of audible distortion, whereas signals with moderate to low SNR may contain noise levels that are not perceptible.

Other quality measures have improved correlation to subjective quality, but are still somewhat unreliable as a measure of speech coder design. Therefore, evaluation of the coder design was based on subjective listening tests.

Listening tests using four subjects were conducted. The signal sources consisted of speech, speech with background noise, and music segments. Training sets were used to familiarize the subjects with the types and degrees of coding distortion in the listening tests. Signals were presented monaurally to simulate audition with a telephone handset or portable radio unit, using the Ariel ProPort model 656 16 bit Digital to Analog Converter and calibrated TDH-49 headphones in a double-walled sound chamber.

The speech sources consisted of 6 phonetically balanced sentences, 2 male and 2 female speakers each, from the TIMIT database. The sentences were approximately 4 seconds in length. Two sequences were generated for the listening experiments. The first consisted of the sentences coded without background noise, and the second consisted of the same sentences with added background road noise. The road noise, provided by Qualcomm Inc, was recorded in an automobile traveling at highway speeds with the windows rolled up. The sentences were presented at levels ranging between 80-90 dB SPL and the additive road noise measured at 85 dB SPL.

A five point Mean Opinion Score (MOS) scale was used, with the scale representing (1) very annoying, (2) annoying, (3) slightly annoying, (4) perceptible but not annoying, and (5) imperceptible.

The listening test results are shown in Figure 3.16. The coder operating at 32 kbps received the highest average score for both conditions. As the coder bit rate decreased from 4 to 1 bit per sample (32-8 kbps), the scores decreased as well. The highly distorted speech at 4 kbps scored poorly and consistently with or without road noise. In the presence of road noise, the clear distinction between 32 and 24 kbps diminishes (score difference drops an average of 0.31 points). At the same time, the score for the 8-24 kbps coders increased in the presence of road noise. It appears that, at the SPL levels we used, distortions in these coders were masked by road noise and became more difficult to detect. These results indicate that at 12 kbps the coder can provide acceptable speech quality (score above 3), and high quality at the higher rates.

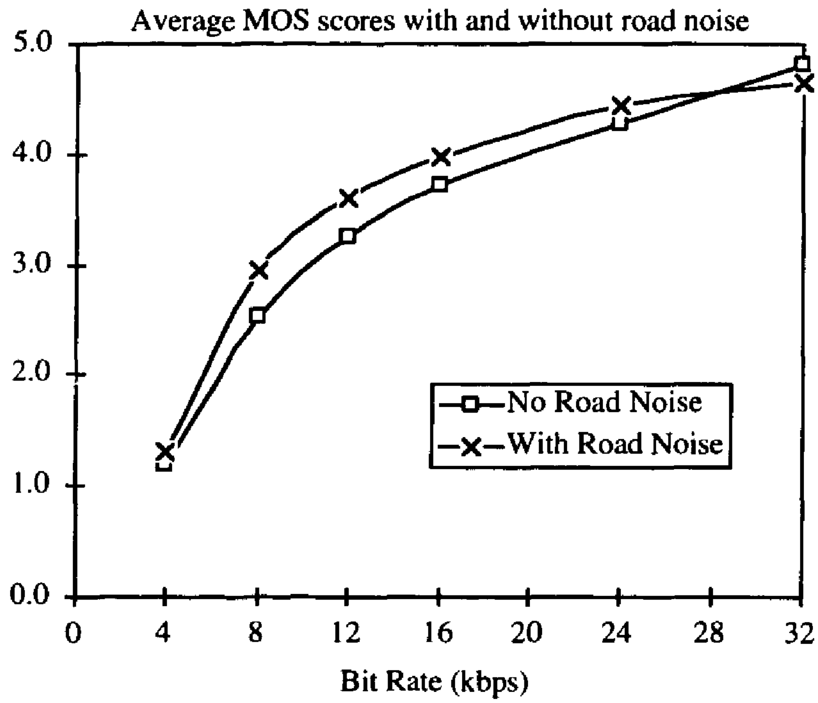


Figure 3.16 - Listening Test Results

Listening tests with non-speech sources were also conducted to demonstrate robustness to non-speech sources. Two 4 second segments of baroque music (Water Music by Handel, Autumn Concerto by Vivaldi) were coded at rates of 12 kbps and 16 kbps. The average scores were 2.7 and 4.0 respectively.

These results indicate both the high quality and scalability achievable by this coder design. The results also demonstrate robustness to non-speech sources such as background noise and music.

3.3 Discussion

Embedded source coding offers significant advantages over non-embedded schemes for applications such as wireless voice communications. The methodology described in this work for perceptually-based embedded subband coding of speech demonstrates that it is possible to achieve performance and complexity similar to non-embedded schemes.

The embedded speech coder developed in this work offers several advantages over the current speech and audio coders. Current speech and audio coding standards have mostly focused on coding efficiency as the main design criterion. As a result, low coding rates have been achieved, but designs suffer from high complexity, poor flexibility, and poor robustness.

The embedded speech coder features perceptually based bit prioritization and optimal quantization, which allow the implementation to have similar performance over a large range of rates compared to fixed rate, non-embedded perceptually based subband speech coder with little added complexity. The coder is scalable and robust to source statistics, so the same design can support medium quality to high quality sources, various source rates, narrowband and wideband sources, speech and audio. In addition, the design was made flexible and robust to channel conditions, so it can accommodate high channel error rate, varying transmission environment, and varying network congestion without severe degradation of quality.

The embedded speech coder implementation is similar to the subband coder proposed by Cox et al. [17][18], which also features embedded coding with a prioritized bitstream.

However, it differs in that an IIR QMF is used to reduce the filterbank complexity, a perceptual model is used to improve the scalability of the bit allocation and prioritization, and optimal embedded scalar quantization is implemented.

The main drawback of the embedded speech coder implementation is that subband coding does not perform well at low bit rates compared to hybrid and IMBE coders. However, the methodology developed in this work, namely the techniques for bit prioritization and embedded quantization, are significant in that they should be applicable to most other coding techniques. Thus, this methodology may also be fruitful in development of good low rate embedded speech coders.

Embedded source coding may also find application in image and video coding. Image and video coding make use of transform, wavelet, subband, or predictive coding, essentially multi-dimensional versions of the techniques used in speech and audio coding. The significance of the encoded bits also vary in a similar manner, so bit prioritization and embedded quantization are also applicable in generating an embedded bitstream.

The requirements for image and video transmission differ from voice communications, but they also benefit from embedded source coding. For image transmission, delay is not a significant factor, but embedded source coding is beneficial in allowing progressive transmission and scalable quality to be implemented in a simple manner. For video transmission, delay is also be an important consideration, so embedded source coding is beneficial in allowing rate adaptation as well as progressive transmission and scalable quality.

The scalability and flexibility of embedded source coding is a significant advantage in integrating the coder into communication systems. Embedded source coders support variable source rates, making it suitable for a variety of transmission conditions, network resource availability, and grades of service, without significantly increasing the complexity of the system. The main contribution from this work in embedded source coding is that it demonstrates that embedding does not necessarily lead to a non-efficient coder implementation; good performance and flexibility are consistent within an embedded source coding methodology.

4. Embedded Channel Coding

In section 3, a methodology for embedded source coding was described that allows efficient digital representation of speech signals and facilitates wireless voice communications. In this section, an embedded channel coding methodology is described that allows reliable wireless voice communications over a noisy channel in a flexible and efficient manner, taking advantage of the properties of digitized speech.

Embedded channel coding allows incremental redundancy to be provided such that the amount of error protection may be varied by selecting how many of the symbols generated by the base rate channel code to transmit. Embedded channel coding cannot provide the same amount of error protection as fixed-rate non-embedded channel coding schemes over the entire range of rates, but the performance penalty and complexity increase can be minimized, as the methodology in this work will show.

The flexibility of embedded channel codes is beneficial in matching the error requirements for the type of data being transmitted. In applications such as digital voice, audio, image, and video transmission, the significance of individual bits in encoded bitstreams naturally vary so that the error tolerance of the less significant bits may be exploited to achieve more efficient transmission in noisy channels. Unequal error protection (UEP) can be easily obtained with embedded channel coding by appropriately selecting which channel symbols need to be transmitted.

The ease with which the channel code rate may be varied with embedded channel codes is also advantageous in adaptive transmission schemes. Adaptive transmission makes it possible to match the channel code to the existing channel conditions by varying the code rate and the amount of error protection provided. Adaptive transmission results in substantial performance improvement when the channel is highly variable, as is the case for wireless communications.

Embedded channel coding also makes progressive transmission possible because of the incremental nature of the redundancy in the code. Progressive transmission allows efficient utilization of the channel, maximizing throughput while providing improved robustness to channel variations.

The system advantages of embedded channel coding make it preferable to non-embedded schemes as long as the performance penalty and added implementation complexity are not significant. Techniques for efficient and effective implementation of embedded channel coding for wireless voice communications were investigated in this research through the development of an embedded channel coding methodology based on tailbiting rate compatible punctured convolutional (RCPC) codes. This methodology is attractive in that it allows flexible implementation of channel codes over a large range of rates, supporting adaptive and variable rates over short blocks of data, with a single efficient decoder structure allowing incorporation of both soft decisions and soft outputs, with comparable performance and implementation complexity to continuous implementation of convolutional codes.

Section 4.1 describes the embedded channel coding methodology and the implementation of the channel coder. The design issues associated with using tailbiting convolutional codes are discussed in Section 4.1.1. The analysis of the distance properties of tailbiting convolutional codes led to the development of two novel decoding schemes. Section

4.1.2 describes the Tailbiting Viterbi Decoder (TVD), which efficiently decodes tailbiting convolutional codes with a fixed complexity algorithm by using the distance properties of the code to determine the number of recursions required to achieve asymptotically optimal decoding. Section 4.1.3 describes the Soft Output Bi-directional Viterbi Decoder (SOBVD), which uses path distance computations with bi-directional recursions of the Viterbi algorithm to produce soft outputs with reliability measures corresponding to the likelihoods for all state transitions. The two decoders are compatible with the same encoded bitstream, allowing the choice at the receiver of the appropriate decoder depending on the particular content of the bitstream.

The benefits of embedded channel coding and the novel techniques introduced in this work are discussed in section 4.2. The main accomplishment of this work is the demonstration that embedded channel coding is possible with very little performance degradation and complexity increase relative to non-embedded designs, motivating their use in applications that benefit from the added flexibility, such as wireless voice communications.

4.1 Embedded Channel Coder Implementation

The embedded channel coding methodology presented in this work is based on tailbiting rate compatible punctured convolutional (RCPC) codes. Tailbiting RCPC codes allow an efficient embedded channel coding methodology for wireless voice communications, providing performance comparable to traditional convolutional codes, but with much greater flexibility.

Any channel code can be embedded simply by defining decoding rules in the presence of missing or punctured channel symbols. However, these decoding rules do not ensure that the channel code has good performance over the entire range of channel code rates. The goal in embedded channel coding is to design a code, puncturing patterns, and decoding rules, such that the channel code has performance close to a fixed rate code of the same rate and decoding complexity.

Channel codes generally fall into 2 categories, block codes and convolutional codes. Block codes generally provide the most coding gain for a given channel code rate and encoder memory, but convolutional codes have the advantage of incorporation of soft decisions in simpler decoder structures with the Viterbi Algorithm (VA).

RCPC codes are a family of convolutional codes generated from the same low rate base code with puncturing patterns structured such that the higher rate codes are embedded into the lower rate codes [11][51][52][53][119]. Efficient structures for decoding embedded codes are possible since using a single decoder structure is sufficient to decode the entire range of rates. RCPC codes are also attractive because they offer performance close to the best available convolutional codes at similar complexity.

Block based coding is advantageous for embedded channel coding because it allows the source information to be grouped such that error protection may be varied from block to block in a flexible manner. However, block codes do not allow easy incorporation of erasures and soft decision information in the decoder. Thus, convolutional codes are preferable since they allow structured soft decision Viterbi decoders to be used and provide a larger range of usable rates.

Tailbiting convolutional codes are an efficient methodology to encode blocks of data with convolutional codes without incurring the rate degradation due to termination of blocks with tail bits [76]. Tailbiting does not decrease the minimum distance nor increase the multiplicity of the convolutional code, given sufficiently long block length, so it does not incur a performance penalty relative to continuous convolutional coding. Tailbiting can be combined with RCPC codes [19] to essentially produce a block code that is puncturable, has error performance comparable to the best known convolutional codes over a large range of code rates, has moderate block length, and may be efficiently decoded with soft or hard decisions and soft output capability.

The embedded channel coder implementation is shown in Figure 4.1. The source is grouped into data blocks. Each block is encoded with a block based embedded channel code. This allows independent rate adaptation and unequal error protection of the data blocks, so it is well suited for protecting the embedded speech coder with its prioritized bitstream. Each block is individually decodable and independent of errors occurring in other blocks.

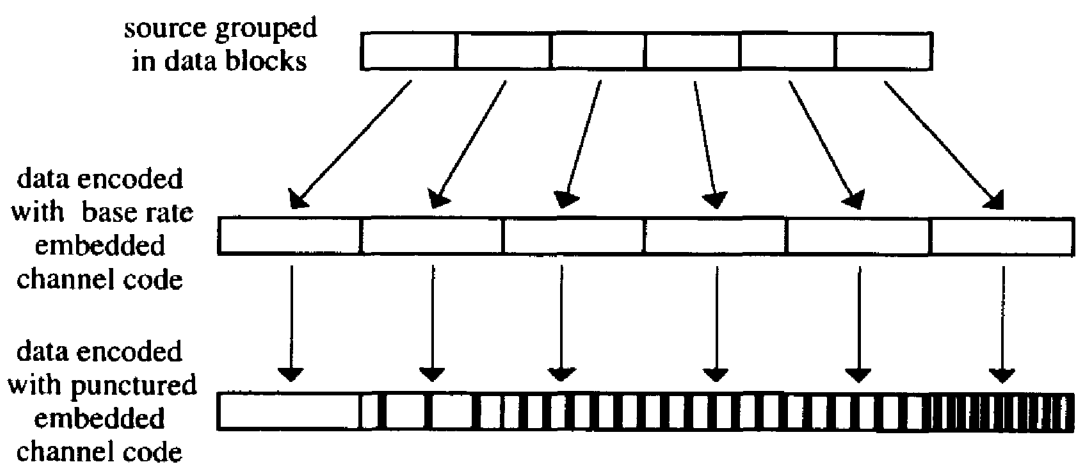


Figure 4.1 - Embedded Channel Coder Implementation

Short block tailbiting RCPC codes, as proposed by Cox [19], are used to implement the channel coding. The generating polynomials and puncturing patterns for rates from 8/9 to 1/3 for 4, 8, 16, 32, and 64 state codes are given in [51] and [53]. The innovative aspects of this embedded coding methodology are discussed in the following sections. Section 4.1.1 discusses distance properties of tailbiting convolutional codes and their effect on decoder performance. Section 4.1.2 and 4.1.3 introduce novel decoding structures that allow embedded channel coding with short block tailbiting RCPC codes to be implemented with little performance degradation and complexity increase relative to non-embedded channel coding methods.

4.1.1 Tailbiting Convolutional Codes

The widespread use of tailbiting convolutional codes has been limited due to the lack of efficient decoding algorithms [19]. In this section, a detailed discussion of the distance properties of tailbiting codes and the shortcoming of existing decoding approaches is presented. This presents a framework for the analysis methodology that led to the development of two novel asymptotically optimal decoder implementations, the Tailbiting Viterbi Decoder (TVD), described in section 4.1.2, and the Soft Output Bi-directional Viterbi Decoder (SOBVD), described in section 4.1.3.

The distance properties of tailbiting convolutional codes were analyzed by Ma [76] using generalized transfer functions that included the analysis of the codeword weights for 3 types of paths: paths starting and ending at the zero state, merged paths starting at non-zero states, and unmerged paths starting at a non-zero state.

Minimum distance paths starting at the zero state are the same as for the tail bit terminated case; thus they are trivial to analyze using standard transfer function techniques. Merged paths starting at non-zero states are essentially wrap-around versions of the paths that start and end at the zero state. The distance for unmerged paths, the paths that never merge with state zero, increase with the length of the block. For sufficiently long blocks, unmerged paths in non-catastrophic codes will exceed the minimum distance, so the additional minimum distance paths introduced by tailbiting are merged paths. Thus, the minimum distance and multiplicity of the tailbiting convolutional code are equal to the continuous convolutional code case, for sufficiently long blocks.

Ma did not consider the effects of invalid paths on the performance of decoders for tailbiting convolutional codes. Invalid paths are those which begin and end at different states. They are not valid codewords and thus do not affect the distance properties of the code. However, the invalid paths present a serious difficulty in implementation of the decoder. If the path distances of the invalid paths are less than the code minimum distance, the decoder needs an efficient method of eliminating them. The performance and complexity of tailbiting decoders are largely influenced by how invalid paths are handled.

Most, if not all, decoders proposed for tailbiting convolutional codes are based on the Viterbi Algorithm (VA), taking advantage of its efficient decoding of convolutional codes with soft decisions. However, the VA does not differentiate between valid and invalid paths in tailbiting codes. For traditional convolutional codes, the Viterbi decoder relies on the fact that the maximum likelihood (ML) path differs from all other paths by at least the minimum distance of the code. The invalid paths provide a significant problem in the Viterbi decoder minimum distance path search since they may be closer to the ML path than the minimum distance of the code. Unmerged invalid paths are not a problem, since the path distance will exceed the minimum distance of the code for sufficiently long blocks and non-catastrophic codes. However, merged invalid paths may be substantially less than the minimum distance of the code. For example, take the case where the invalid path differs from the ML path only in the final transition and the path distance is given only by the distance in the final branch of the trellis. The decoder must efficiently eliminate those types of paths from consideration to be effective.

The ML decoder [76] is the trivial approach of exhaustively searching all possible valid paths by running the VA for each possible initial state, and comparing the distances from the best path found in each iteration. The exhaustive search eliminates all the invalid paths from the analysis, since there is no ambiguity in determining the initial and final state in each iteration of the VA. However, the complexity of the decoder is greatly increased since the VA must be run once for each state, so the ML decoder is not useful in practical applications.

The Bar-David decoder [76] assumes an initial state, then finds the best path, valid or invalid, using the VA starting at that state. If the best path found is valid, i.e., ends at the same state, then the search is done, otherwise the ending state is used as the initial state and another iteration of the VA is run. This goes on until all possible states are finally tried. This approach has the obvious flaw that even if the initial state is guessed correctly, the VA may not find the ML path since the distance for an invalid path is much less than the code distance. If the initial state is not guessed correctly, the distance between the correct final state and other states is also significantly less than the code distance. Additionally, even if a valid path is found by this procedure, it is not necessarily the ML path. Thus this decoder is likely to not converge or to converge to an incorrect solution, and at low SNR, the complexity will approach that of the ML decoder.

The two step-algorithm developed by Ma [76] uses an ordered list of starting states to prioritize the iterations of the VA, but suffers from the same basic flaw as the Bar-David approach in its path search. A different strategy used by Wang [114] is to iteratively perform the Viterbi Algorithm that allows path metrics for valid paths to be found, then eliminating those paths from future iterations so that all valid paths are eventually searched, or can be shown to have greater distance than the paths already found. This approach does find the ML path, but finding the valid paths among the clutter of invalid paths may require many iterations of the VA under low SNR.

Because the complexity of the iterative algorithms tends to approach the exponential complexity of the trivial ML decoder for poor channels, the use of continuous Viterbi decoding in a Circular Viterbi Algorithm (CVA) was suggested by Cox [19]. Continuous Viterbi decoding corresponds to the case where the encoded blocks are transmitted repeatedly and the decoder operates on the continuous stream. Invalid paths are eventually eliminated because surviving paths in the Viterbi decoder must be merged paths. Continuous Viterbi decoding eliminates the inefficiency associated with numerous iterations of the VA, providing a significant improvement over the iterative algorithms, particularly for poor channels.

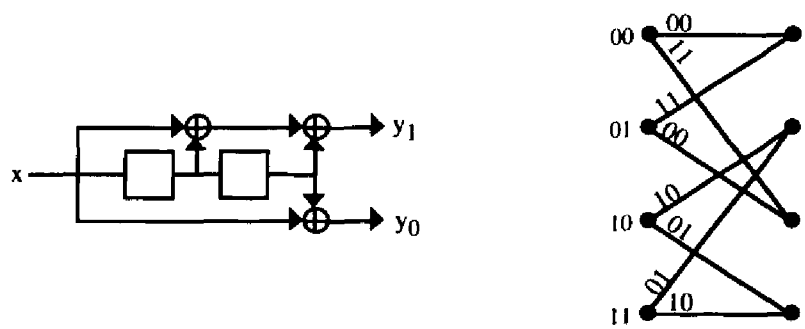
The keys to performing continuous Viterbi decoding are the stopping rules for determining decoder convergence, and traceback rules for determining the best path given the convergent metrics. The CVA uses metric convergence as an adaptive stopping rule, and valid path checking to ensure correct traceback. These are ad hoc methods that allow the ML path to be found, but do not guarantee a limit on the amount of computation required.

The decoding process may be improved by refining the decoding methodology, using the distance properties of tailbiting convolutional codes to constrain the limitations of Viterbi decoding. This work introduces two novel high performance decoders tailbiting convolutional codes. The Tailbiting Viterbi Decoder (TVD), described in section 4.1.2, is a continuous Viterbi decoder similar to the CVA, but with fixed computational workload. The TVD uses the distance properties of the tailbiting convolutional code to achieve asymptotically optimal performance with minimal computation. The Soft Output Bi-directional Viterbi decoder (SOBVD), described in section 4.1.3 is a bi-directional continuous Viterbi decoder combining forward and reverse state metrics to provide reliability measures for the decoder output.

4.1.2 Tailbiting Viterbi Decoder

Continuous Viterbi decoding, in the form of Cox's Circular Viterbi Algorithm (CVA) [19], was shown to be the most attractive approach for decoding tailbiting convolutional codes in terms of complexity and performance. The main advantage of this approach is that it does not "waste" information from trials, but rather uses it to refine future computations so that the algorithm eventually converges to a solution. The Tailbiting Viterbi Decoder (TVD) introduces fixed stopping rules and traceback procedures for the continuous Viterbi decoder by considering the distance properties of the tailbiting codes. Continuous Viterbi decoding is used to first find the most likely initial state, then the best path beginning and ending at that state is found by processing additional symbols through the decoder.

To understand how the TVD operates, consider the rate 1/2, 4 state convolutional code shown in Figure 4.2. The minimum distance is 5, with a multiplicity of 1 when operated in the traditional, continuous manner.



(a) (b)

Figure 4.2 - Encoder (a) and Trellis (b) for rate 1/2, 4 state convolutional code

For the tailbiting convolutional code, the initial and final states are the same, but may be any of the 4 states. To find the minimum distance path, all paths beginning and ending on the same state must be searched. Since the minimum distance and minimum distance multiplicity of the code do not change, it can be decoded with the same asymptotical error probability as in the continuous coding case, with no degradation in decoded error rate at sufficiently high SNR.

With continuous Viterbi decoding, the initial state may be decoded if all paths have merged to a single state. To find the most likely initial state, sufficient transitions must be processed by the decoder such that the state metrics are initialized and the traceback has sufficient length. This requires processing symbols "before" and "after" the initial state in order to achieve decoding based on the full distance of the code.

The initial state decoding error can be analyzed by evaluating the trellis paths as shown in Figure 4.3. The first 7 transitions initialize the state metrics to the full distance starting with no information. Increasing the number of transitions does not increase the distance in state metrics, but fewer than 7 transitions result in decreased distance or increased multiplicity. Once the state metric has been initialized, 7 transitions are required to reach full distance for correct traceback to the initial state. There are thus 2 paths with distance 5 that lead to incorrect traceback of the initial state, which correspond to the 2 wrap-around minimum distance merged paths. The initial state decoding can therefore be performed with the same asymptotic error probability as the decoding of a single state transition through the trellis. If an incorrect initial state decision is made, then the decoded path is likely to be a merged path, which does not degrade the decoding error probability.

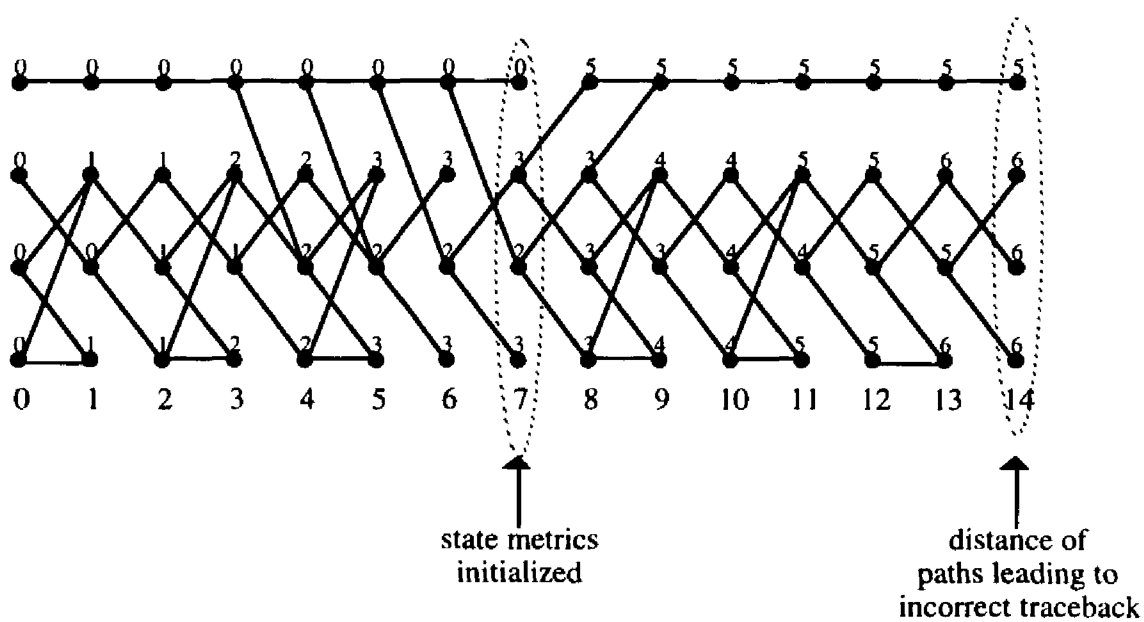


Figure 4.3 - Initial State Decoding for Tailbiting Convolutional Codes

The minimum path search for initial state decoding is analogous to error recovery in the Viterbi decoder. In error recovery, the current state is not known, so a number of transitions must occur before the state metrics can be initialized to their full distance, then additional transitions must occur before the traceback from minimum state decisions converge to the correct decision.

The number of transitions required for state initialization and correct traceback is the truncation length of the code. In order to make a correct decision after an error in decoding, the required number of transitions to guarantee future decisions up to the code distance is twice the truncation length. Similarly, asymptotically optimal continuous Viterbi decoding of tailbiting convolutional codes requires a minimum block length of twice the truncation length of the code.

The decoding process is shown in Figure 4.4. Assuming the block length N is set to twice the truncation length of the code, N/2 symbols must be processed before the state metrics are initialized to the full distance of the code. This corresponds to the effect of transitions "before" the initial state. Once the state metrics are initialized, N/2 symbols must be processed before traceback will converge to an initial state according to the full distance of the code. This corresponds to the effect of transitions "after" the initial state. Thus N symbols must be processed before the "initial" state can be properly decoded. Note that the nature of the tailbiting code is such that any position can be considered the "initial" location, since the code wraps around the end of the block. Once the initial state has been found, the VA processing proceeds as normally. This requires re-processing the received symbols in a second pass. When the "final" state is reached, no state decision is needed since the "final" state is known (similar to termination of the convolutional code with tail bits). Thus all transitions can be traced back from the final state decision and no further processing is required.

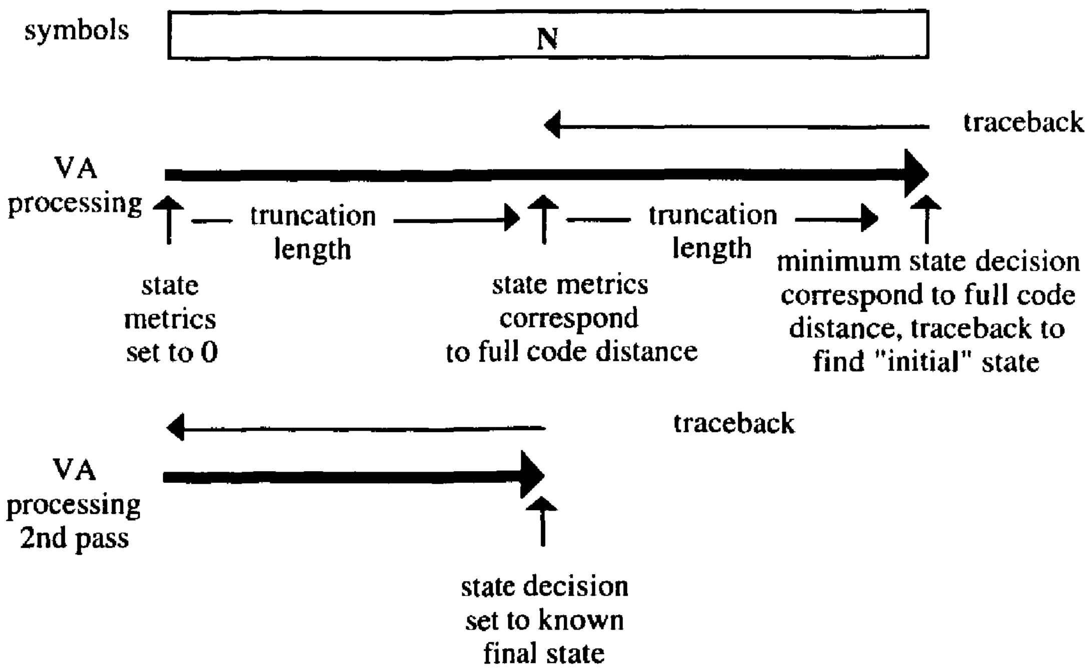


Figure 4.4 - Decoding Tailbiting Convolutional Codes with the Tailbiting Viterbi Decoder

The error performance of the decoder was verified through simulations. The performance over hard decision binary symmetric channels (BSC) and soft decision additive white Gaussian Noise (AWGN) channels for tailbiting blocks of various lengths were compared to tail-terminated blocks. The results shown in Figure 4.5 demonstrate that the performance for tailbiting codes comes very close to the performance of tail-terminated codes as expected from the codeword distance trellis analysis and the asymptotically optimal performance of the decoder.

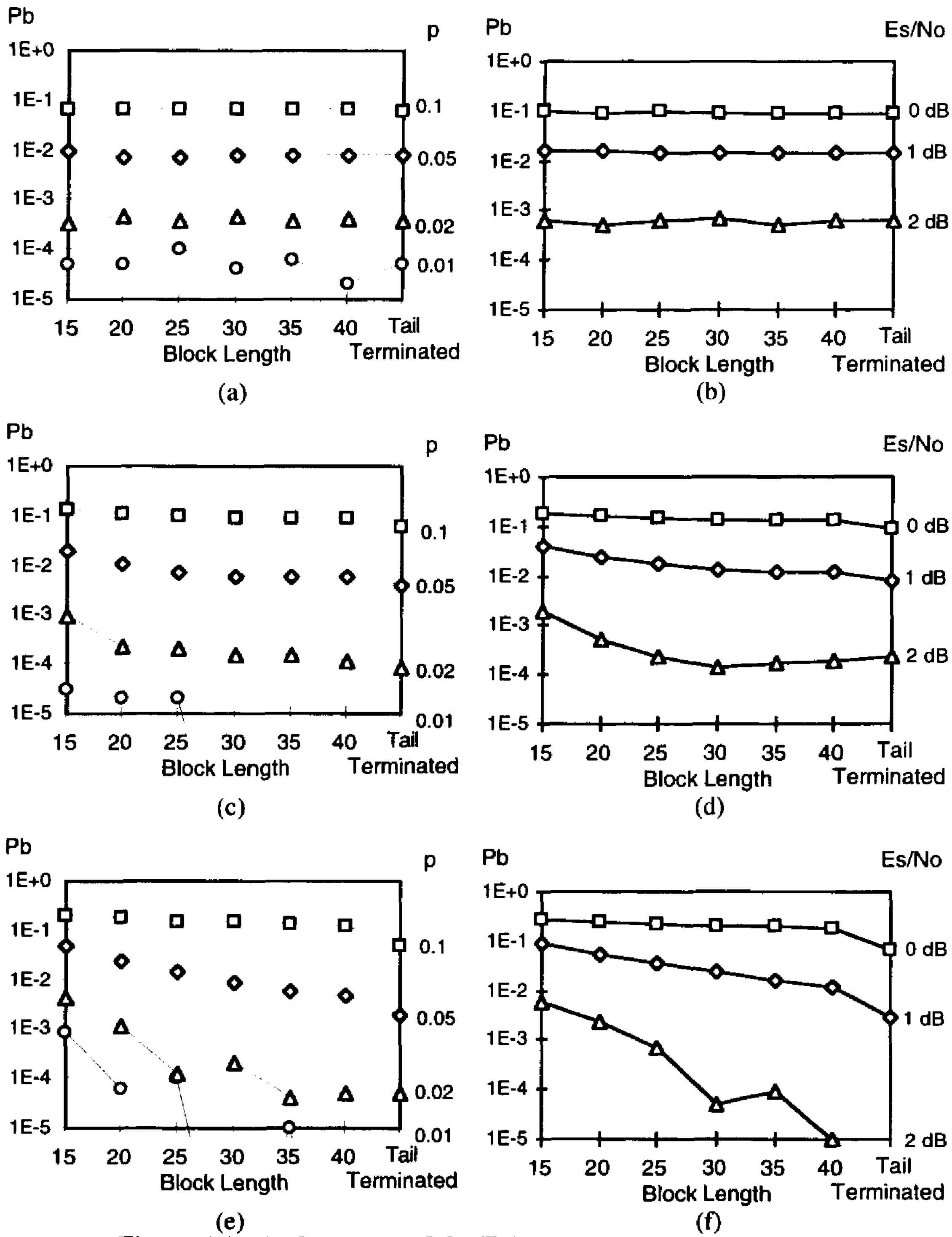


Figure 4.5 - Performance of the Tailbiting Viterbi Decoder for 4 state (a)(b), 16 state (c)(d) and 64 state (e)(f) Tailbiting Convolutional Codes over hard decision BSC (a)(c)(e) and soft decision AWGN channels (b)(d)(f)

The main difference between the TVD and Cox's Circular Viterbi Algorithm (CVA) [19] is the use of fixed stopping and traceback rules. The initial state decoding can be made asymptotically optimal by simply processing sufficient symbols to initialize the state metrics, then processing sufficient symbols for asymptotically optimal traceback. Once the initial state has been decoded, the final state can be decoded without traceback, minimizing the number of symbols that must be processed.

The use of fixed stopping and traceback rules in the TVD reduce the overall complexity to only about 50% higher than traditional Viterbi decoding in the worst case. In contrast, the CVA requires continuous checking of convergence, which increases the complexity of each symbol processed. In addition, state transitions are all obtained from traceback, which requires processing of additional symbols.

The modifications introduced by the TVD result in a consistent, low complexity, high performance decoder for tailbiting convolutional codes. With the TVD, tailbiting RCPC codes become an attractive methodology for embedded channel coding applications in that performance comparable to non-embedded continuous convolutional coding is achieved, with relatively little added complexity.

4.1.3 Soft Output Bi-directional Viterbi Decoder

The previous section discussed how the distance properties lead to an efficient, low complexity, asymptotically optimal decoding structure for tailbiting convolutional codes with the Tailbiting Viterbi Decoder (TVD). The Soft Output Bi-directional Viterbi Decoder (SOBVD) results from an alternate view of the code's distance properties. Consider the problem of determining the most likely or minimum distance state traversed by a path starting at an arbitrary state and ending at another arbitrary state, as shown in Figure 4.6.

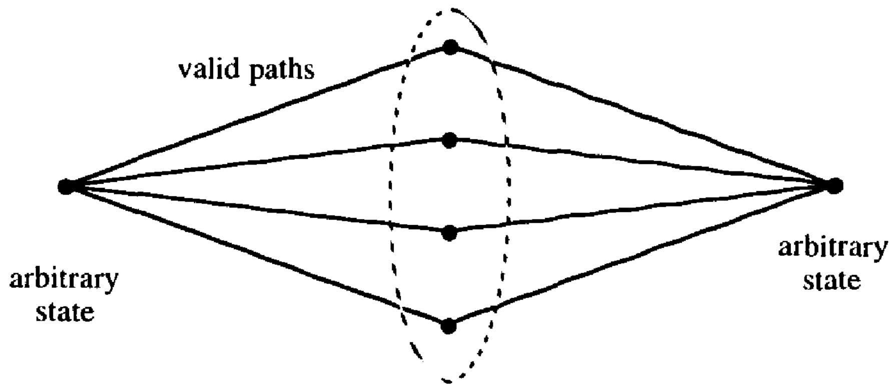


Figure 4.6 - Alternate View of Tailbiting Convolutional Codes Distance Analysis

The most likely state is the one with the minimum distance path among the paths traversing all possible states. The minimum distance paths traversing through the other states determine the distance property of the code. Those paths may be found according to the following procedure.

Starting at an arbitrary state, the valid paths correspond to the forward trellis paths as shown in Figure 4.7 (a). The number of transitions required to reach the full distance of the code, in this case 7, is the truncation length of the code. The minimum distance surviving paths from all states converge after 7 previous transitions. To decode the code to its maximum distance, transitions after the location must be known.

Contributions from transitions after the location can be considered by looking "backwards" from an arbitrary future location, as shown in Figure 4.7 (b). To reach the full distance of the code, a minimum of 7 transitions in the future must be considered, again the equivalent truncation length of the code.

The distance properties of the code can therefore be considered by looking at 7 transitions before and 7 transitions after an arbitrary location. The combined sum of the forward and reverse paths give the minimum distance paths traversing each of the possible states. The combined distances give the overall distance properties of the code, as shown in Figure 4.7 (c). At time location N, by considering observed transitions from N-7 to N+7, there is one path at distance 5 going through state 01 and one path at distance 5 going through state 10. Therefore, the most likely state may be found by processing 14 transitions through the trellis, which are decodable with a minimum distance of 5.

Decoding of the information bits is done directly from the sequence of state transitions. States 00 and 01 correspond to input x(n-1)=0, and states 10 and 11 correspond to input x(n-1)=1. Therefore, if x(n-1)=0, only one of the paths at distance 5 results in incorrectly decoding x(n-1).

This is identical to the result obtained for the initial state search using the TVD algorithm and the distance analysis in Figure 4.3. Bi-directional analysis of the trellis thus allows the analysis of distance properties of convolutional codes to be simplified. It is well suited to tailbiting convolutional codes because it easily deals with the ambiguity of the initial and final state in computing distance properties.

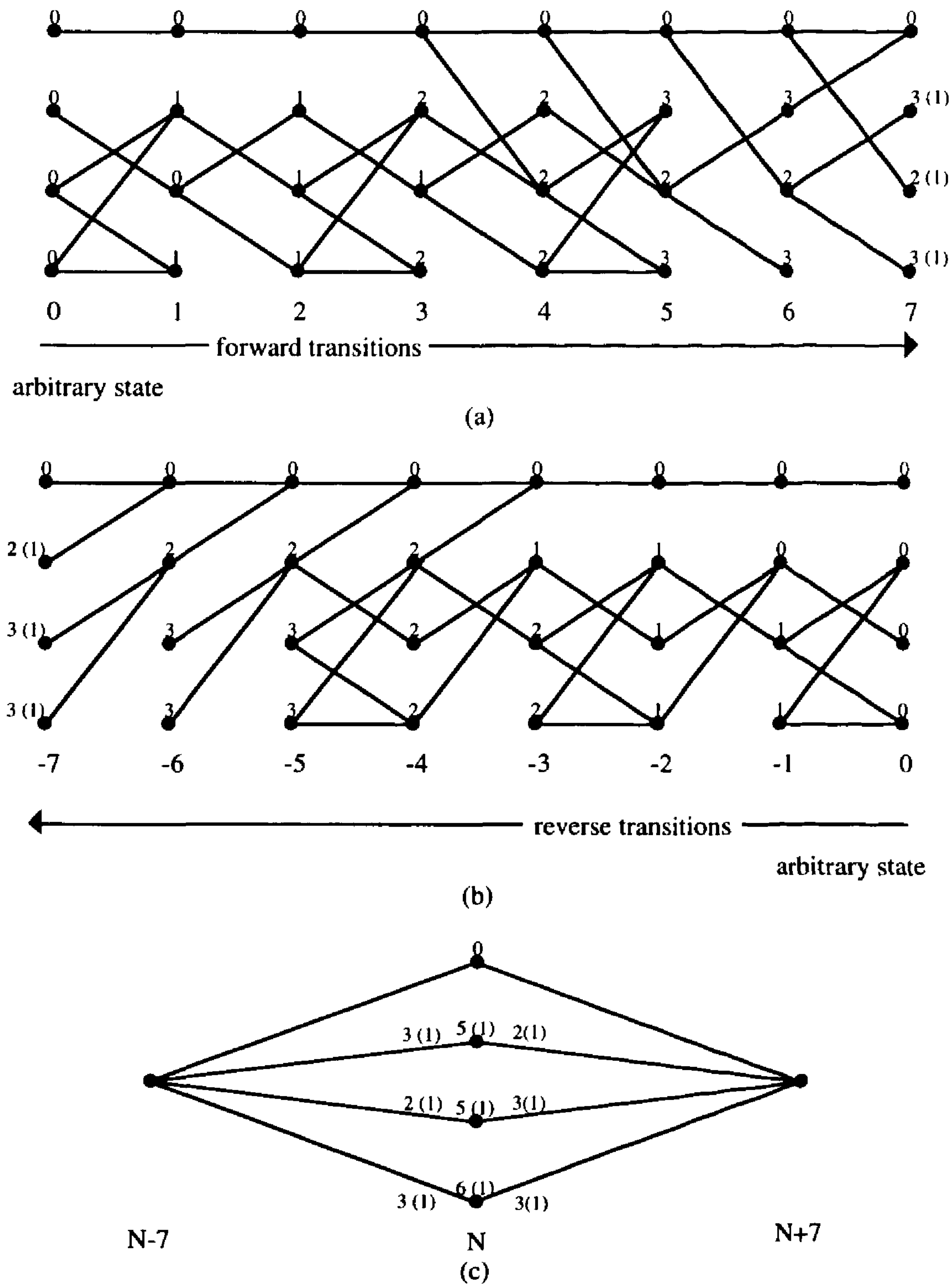


Figure 4.7 - Forward (a), Reverse (b), and Combined (c) Trellis Analysis of Distance Properties

Bi-directional analysis also suggests that if forward and reverse path distances are computed and summed, a total path distance related to the probability of each state being traversed may be obtained. The forward distance corresponds to the effect of input symbols on state transitions and outputs prior to the state being analyzed. The reverse distance corresponds to the effect of the current state on future transitions. Using the Viterbi Algorithm (VA) to analyze both forward and reverse paths allows the maximum likelihood (ML) paths in the forward and reverse direction to be determined. The structure of the VA is such that the ML paths and their distances to all states are computed simultaneously. Combining the distance from both the forward and reverse analysis allows the distance of the entire ML path to be computed. This is in contrast to traditional Viterbi decoding where only forward metrics are computed, as is the case with the TVD. In those decoders, the effect on future transitions, measured by the reverse metrics, are lost in the path merging that occurs in the trellis from later transitions.

The SOBVD decoding algorithm can be described as shown in Figure 4.8. Each block is processed twice: first in the forward direction, and then in the reverse direction. The state metrics for all states are then added, which gives the distance for the ML paths going through each state. The SOBVD is extended to tailbiting convolutional codes, by processing sufficient symbols to initialize the state metrics, similar to the initialization procedure with the TVD algorithm discussed in section 4.1.2. The initialization must be done for both the forward and the reverse analysis.

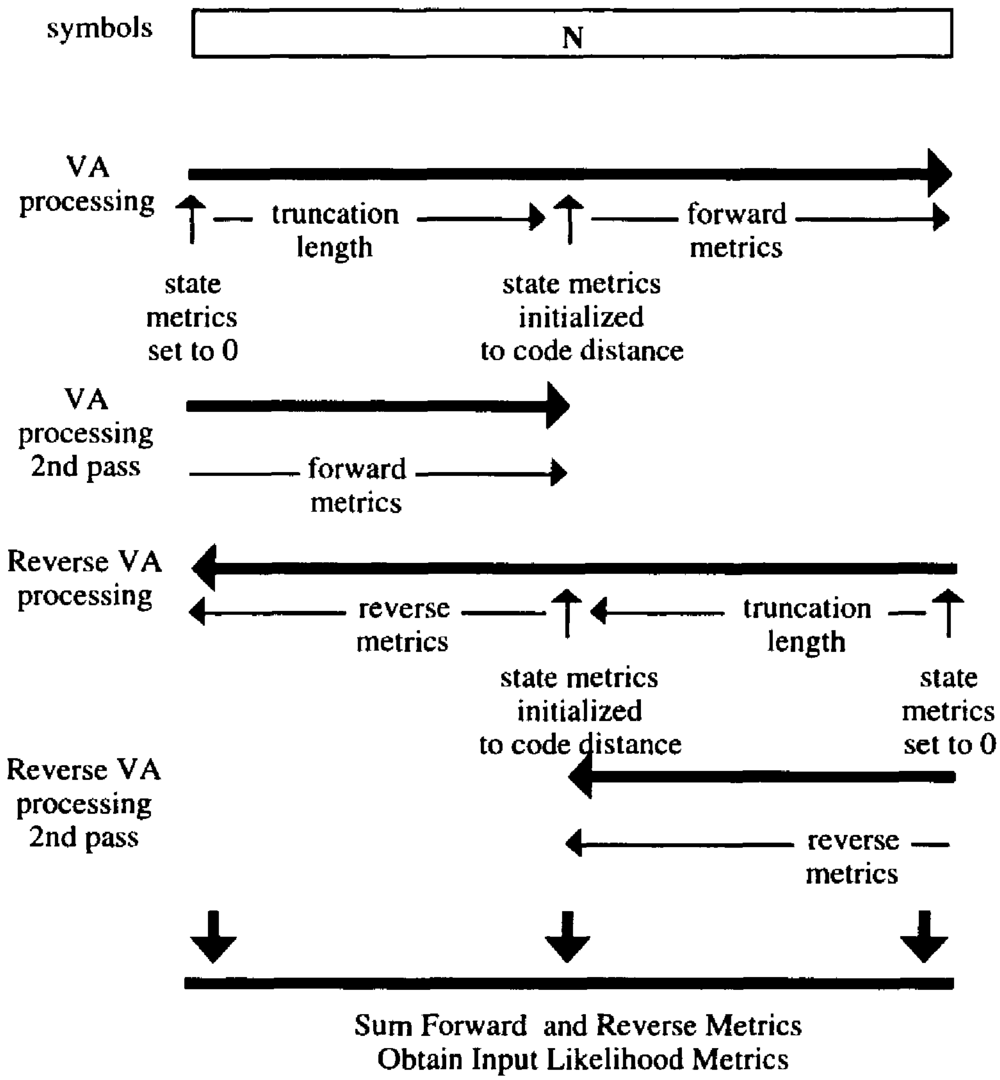
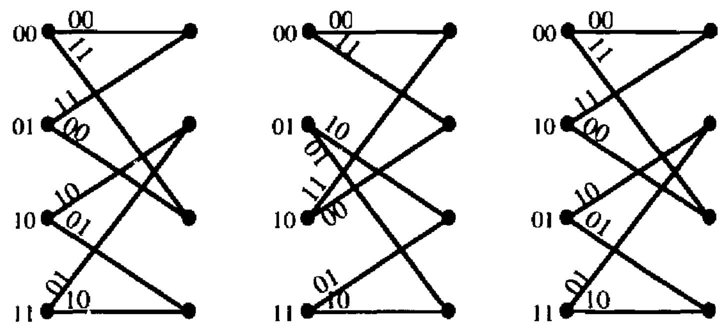


Figure 4.8 - Soft Output Decoding of Tailbiting Convolutional Codes with the Bi-directional Viterbi Decoder

Bi-directional Viterbi decoding is simplified by the symmetric nature of the forward and reverse trellis for shift register based convolutional codes. As shown in Figure 4.9, if the reverse trellis state labeling is bit reversed, the branch transitions are identical to the forward trellis. The branch labels may differ depending on the choice of generating polynomials, but this requires only a simple change in metric calculation in the implementation of the Viterbi decoder. The same basic decoding structure can still be used for both the forward and reverse path searches.



Forward Trellis

Reverse Trellis

Bit-reversed Reverse Trellis

Figure 4.9 - Forward and Reverse Trellis Symmetry

The SOBVD algorithm is asymptotically optimal, in the sense that the soft output obtained asymptotically approaches the likelihood for high SNR. This can be shown by the following discussion. First we formulate the results of running the Viterbi Algorithm (VA) in the forward direction, where the final state metrics correspond to the maximum likelihood (ML) paths ending at each possible state. Then we show that the distance of the ML paths can be obtained by summing the path distance from the best path in the forward direction, plus the best path in the reverse direction. The conditional probabilities for traversing each state, obtained by summing the probability from all possible paths traversing each state, may be approximated by the probabilities of just the ML paths under high SNR, leading to asymptotically optimal behavior. Finally, we show that the best reverse path can be found using modified recursions of the VA in the reverse direction, thus leading to a bi-directional VA implementation.

The Viterbi Algorithm can be described as follows (after Forney [35]):

A convolutional code may be modeled as a Markov process, with L random inputs generating a state sequence x with L transitions from a finite state machine with M states, where

$$x = \{x_0, \ldots, x_L\} \tag{4.1}$$

$$x_i \in \{1, \ldots, M\} \tag{4.2}$$

$$P(x_{k+1} \mid x_0, \ldots, x_k) = P(x_{k+1} \mid x_k) \tag{4.3}$$

The state transitions ξ may be defined with a one-to-one correspondence to the state sequence as follows:

$$\xi_k \equiv (x_{k+1}, x_k) \tag{4.4}$$

$$\xi = (\xi_0, \ldots, \xi_{L-1}) \tag{4.5}$$

$$x \xleftrightarrow{1\text{-}1} \xi \tag{4.6}$$

Given a set of observations $z = \{z_0, \ldots, z_{L-1}\}$, the Maximum Likelihood (ML) state sequence is given by finding $\max_x \; P(x \mid z) \;\; or \;\; \max_x \; P(x, z)$.

$P(x,z)$ may be factored as follows:

$$P(x, z) = P(x) \;\; P(z \mid x) = \prod_{l=0}^{L-1} P(x_{l+1} \mid x_l) \;\; \prod_{l=0}^{L-1} P(z_l \mid x_{l+1}, x_l) \tag{4.7}$$

where the factorization of $P(x)$ is due to x being a Markov process, and the factorization of $P(z|x)$ is due to the channel being memoryless.

The possible state sequences x may be represented as paths through a trellis, where each transition may be assigned the length

$$\lambda(\xi_k) \equiv \; -\log \;\; P(x_{k+1} \mid x_k) \; - \; \log \;\; P(z_k \mid \xi_k) \tag{4.8}$$

The probability then corresponds to the total length of the path through the trellis

$$-log \;\; P(x,z) \;=\; \sum_{l=0}^{L-1} \lambda(\xi_l) \tag{4.9}$$

and the ML path search becomes

$$\min_x \; -log \;\; P(x,z) \;=\; \min_x \; \sum_{l=0}^{L-1} \lambda(\xi_l) \tag{4.10}$$

The VA finds the ML path through the trellis using the following recursive algorithm:

$$\Gamma(x_0) = -log \;\; P(x_0) \tag{4.11}$$

$$\Gamma(x_{k+1}, x_k) = \Gamma(x_k) + \lambda(\xi_k) \tag{4.12}$$

$$\Gamma(x_{k+1}) = \min_{x_k} \;\; \Gamma(x_{k+1}, x_k) \tag{4.13}$$

The VA is initialized in *(4.11)*, generalized for the case where the initial state x_0 is not pre-determined. The branch metric and the path length are calculated in *(4.12)*. The state metric is selected as the shortest path length for each state in *(4.13)*. By tracking the shortest path length to each of the states, the minimum distance path is found by tracing back from the final state selection.

Following this recursion for all transitions, the VA gives the minimum path distance as:

$$\min_x \sum_{l=0}^{L-1} \lambda(\xi_l) \;=\; \min_{x_L} \;\; \Gamma(x_L) \tag{4.14}$$

Moreover, the VA also finds the minimum distances for paths ending at all $x_L=\{1,...,M\}$ states:

$$\min_{\{x|x_L=m\}} \; \sum_{l=0}^{L-1} \lambda(\xi_l) \;=\; \Gamma(x_L = m) \tag{4.15}$$

The SOBVD uses the computations resulting from using the VA in a bi-directional manner to obtain a soft output in the manner described as follows:

A soft output decoder obtains not only the most likely state sequence, but also a reliability measure for each possible state or transition. The reliability measure can be obtained from the likelihood ratios

$$\Lambda_z(i,j,k) \equiv \frac{P(x_k = i \mid z)}{P(x_k = j \mid z)} \tag{4.16}$$

The maximum a-posteriori (MAP) algorithm from Bahl et al. [3] finds the a-posteriori probabilities (APP) $P(x_k = i \mid z)$ for all i and k using forward and reverse recursions on all state and transition probabilities, which gives an exact solution to *(4.16)*, but results in a very complex decoding algorithm.

A simplification of the soft output decoder can be obtained from the following approximation:

$$\Lambda_z(i,j,k) = \frac{\sum_{\{x \mid x_k = i\}} P(x,z)}{\sum_{\{x \mid x_k = j\}} P(x,z)} \cong \frac{\max_{\{x \mid x_k = i\}} P(x,z)}{\max_{\{x \mid x_k = j\}} P(x,z)} \tag{4.17}$$

The log-likelihood may then be approximated as:

$$\log \ \Lambda_z(i,j,k) \ \cong \ \min_{\{x \mid x_k = j\}} -\log \ P(x,z) \ - \ \min_{\{x \mid x_k = i\}} -\log \ P(x,z) \tag{4.18}$$

which can be solved by finding the shortest paths through the trellis according to:

$$\min_{\{x \mid x_k = i\}} -\log \ P(x,z) = \min_{\{x \mid x_k = i\}} \sum_{l=0}^{L-1} \xi_l \tag{4.19}$$

$$= \min_{\{x \mid x_k = i\}} \sum_{l=0}^{k-1} \xi_l + \min_{\{x \mid x_k = i\}} \sum_{l=k}^{L-1} \xi_l \tag{4.20}$$

The first term in *(4.20)* can be computed using the VA for the first k transitions, generating

$$\min_{\{x|x_k=i\}} \sum_{l=0}^{k-1} \lambda(\xi_l) \;=\; \Gamma(x_k = i) \tag{4.21}$$

The second term can be computed using the following recursion:

$$\Gamma'(x_L) = -\log P(x_L) \tag{4.22}$$

$$\Gamma'(x_{k+1}, x_k) = \Gamma'(x_{k+1}) + \lambda(\xi_k) \tag{4.23}$$

$$\Gamma'(x_k) = \min_{x_{k+1}} \; \Gamma'(x_{k+1}, x_k) \tag{4.24}$$

Following this recursion, the second term in *(4.20)* is given by

$$\min_{\{x|x_k=i\}} \sum_{l=k}^{L-1} \lambda(\xi_l) \;=\; \Gamma'(x_k = i) \tag{4.25}$$

The recursion given in *(4.22-24)* is essentially running the Viterbi Algorithm in a time reversed manner, starting from the last transition back to the state transition of interest. To compute $\Gamma(x_k = i)$ and $\Gamma'(x_k = i)$ for all i and k only requires running the VA from beginning to end once in each direction. The state likelihood ratios can then be obtained by summing the state metrics as in *(4.18)*, and it is a simple matter to obtain likelihood ratios for specific transitions and corresponding input bits.

The likelihood ratios obtained by this algorithm are accurate to the extent that the approximation in *(4.17)* is valid. This approximation is necessary in soft output Viterbi decoders because the Viterbi algorithm eliminates less likely paths, keeping metrics only for the most likely paths. For sufficiently high SNR, the likelihood is dominated by the probability of the most likely paths. Thus, the approximation approaches equality so that the results from the SOBVD asymptotically approach the results from the APP. The SOBVD is therefore an asymptotically optimal soft output decoder.

The overall complexity of the SOBVD is approximately twice that of the TVD. Blocks must be processed twice, but no traceback is needed. On the other hand, state metrics must be stored, summed, and then compared. The memory access is fairly straightforward, and the amount of storage can be kept fairly small by using the shortest block length possible for the given code.

The SOBVD offers an attractive option for decoding embedded channel codes based on tailbiting RCPC codes. Soft output decoders are of particular interest in designing robust systems, where the soft output may be used for source aided decoding, concatenated decoding, and channel reliability tracking and estimation [7][52][91][92]. The system can use the reliability information to make rate adaptation decisions without additional overhead, or make use of concatenated coding to protect network overhead information. The choice of decoder is left up to the system, so either the TVD, the SOBVD, or even other decoder implementations, may be used.

4.2 Discussion

The main requirements for the channel coding approach in wireless voice communications are flexibility to variations in the mobile radio environment, good performance in a fading environment, and high overall throughput. In order to meet these requirements, the channel coding scheme must be able to support high throughput under good channel conditions, provide high protection under poor channel conditions, and be effective and efficient for conditions in between.

The channel coding approach should also take advantage of the characteristics of encoded speech. Speech is non-stationary, relatively tolerant to bit errors, intelligible over a fairly large range of reconstruction quality, but sensitive to transmission delay and long periods of transmission blank-outs.

Embedded channel coding is a very attractive technique for wireless applications because of the flexibility in error protection and channel code rates it provides. Embedded channel coding allows rate adaptation, unequal error protection, and multiple decoding options with little added complexity, resulting in substantial system improvements in quality, capacity, and robustness. Embedded channel coding is clearly advantageous over non-embedded approaches in that rate adaptation and decoder architectures are greatly simplified. The embedded channel coding methodology and novel techniques introduced by this research extend the applicability and simplify the implementation of embedded channel coding with tailbiting RCPC codes.

The approach used in this work is fairly typical in its incorporation of block based channel coding with interleaving. Various block based embedded channel coding with unequal error protection schemes have been proposed for wireless voice communication applications using both punctured convolutional and punctured block codes [18][47][79][98]. Block based coding is much more flexible than continuous coding, and the decoding delay disadvantage is mitigated by the need to incorporate interleaving regardless of the approach. Unequal error protection improves both the throughput and robustness of the transmission, by more closely matching the requirements to the effect of errors in perceived quality.

RCPC codes have been previously used for channel rate adaptation and unequal error protection [51][53]. RCPC codes were used by Cox to demonstrate robust speech transmission with a subband speech coder [18][48]. Cox also made use of tailbiting convolutional codes to improve the efficiency of RCPC codes [19]. The approach used in this work is essentially an extension of Cox's approach.

The key innovations in this research are the use of short block length tailbiting RCPC codes and the simplified decoder architectures for both traditional and soft output decoding, which allow embedded channel coding to be implemented with little performance penalty and complexity increase relative to non-embedded schemes.

Short block length tailbiting RCPC codes are very well suited to embedded channel coding, offering good error protection and allowing low complexity, high performance decoders to be used. Short block lengths improve the flexibility in implementation of rate adaptation and unequal error protection, since adaptation can be done on a block-by-block basis. The methodology used in this work demonstrates that decoding complexity and code performance are not detrimentally affected by the use of short blocks.

This work also introduced novel decoders, the TVD and the SOBVD, that allow short block length tailbiting RCPC codes to be efficiently and effectively decoded. The asymptotically optimal performance and fixed computational workload allow both good performance under good conditions, and reasonable attempts at decoding under poor conditions. The low complexity of both decoders is attractive for wireless communications, with its premium on low power and complexity. The compatibility of the two decoders provides a tradeoff in complexity and performance for decoder selection, allowing either soft outputs or hard outputs to be obtained, with little change needed from the basic Viterbi decoder structure.

The TVD takes full advantage of the advantages of continuous Viterbi decoding, making full use of all the information generated from recursions of the VA. The advantages of continuous Viterbi decoding over iterative algorithms were shown by Cox [19]. The TVD improves continuous Viterbi decoding by implementing fixed stopping and traceback rules that reduce the complexity of the decoder, particularly under low SNR. The initial state decoding can be made asymptotically optimal by simply processing sufficient symbols to initialize the state metrics, then processing sufficient symbols for asymptotically optimal traceback. Once the initial state has been decoded, the final state can be decoded without traceback, minimizing the number of symbols that must be processed.

The use of fixed stopping and traceback rules in the TVD reduce the overall complexity to only about 50% higher than traditional Viterbi decoding. In contrast, Cox's approach requires obtaining all state transitions from traceback, and continuous checking of convergence, which not only increases the implementation complexity, but makes it variable as well.

The SOBVD is ideally suited for soft output decoding of short block length tailbiting convolutional codes. Although it requires forward and reverse recursions, the short block length reduces both the decoding delay due to buffering, and the memory required for storing state metrics. Bi-directional decoding offers significant capabilities not possible with just forward decoding.

Bi-directional decoding algorithms for convolutional codes have been previously reported, notably in sequential and recursive decoding methods. Belzile and Haccoun's [5] breadth-first sequential decoding uses bi-directional analysis to alleviate error propagation due to loss of correct path in the decoding trellis. A noisy segment is isolated by allowing reverse decoding instead of forcing the forward decoder through the noisy span. This approach has significant drawbacks in that it does not exploit the full distance of the code and is by design limited to a single noisy burst per block. It does, however, contribute the insight that reverse decoding of convolutional codes is possible and desirable.

The maximum a posteriori (MAP) algorithm from Bahl et. al [3] is an optimal algorithm for computing state transition probabilities It is the optimal method of obtaining reliability information (i.e., soft outputs), but it is complex due to the recursive nature of computation, the need for multiplications, exponentiations, and estimates of the noise variance to compute probabilities, and the need for forward and reverse recursions to compute all the transition probabilities.

Li, Vucetic, and Sato introduced modifications and simplifications to the MAP algorithm with the optimum soft-output algorithm (OSA) and suboptimum soft output algorithm (SSA) [71]. The OSA is an APP algorithm requiring only forward recursions. The SSA is a simplification of the OSA that reduces complexity by making computations in the log-likelihood metrics instead of probabilities. Both algorithms are still significantly more complex than Viterbi decoding based algorithm because they evaluate probabilities for all paths.

The SOVA algorithm from Hagenauer and Hoeher [52] is a modification of the Viterbi algorithm that uses the difference in the path metric comparison and traceback of the two paths to identify the likelihood of each transition. It does not require reverse recursions, so it is particularly useful when the information is not in block form. However, the traceback comparisons greatly increase the complexity. Berrou et al. [7] introduced a simplification to the SOVA by using a conventional Viterbi decoder to first identify the ML path, then doing traceback comparisons only from the states in the ML path. This essentially reduces the complexity of the SOVA to about twice that of the traditional Viterbi algorithm. However, since the SOVA limits the comparisons to only maximum likelihood merged paths, it does not take all possible paths into account, resulting in slightly degraded performance.

The SOBVD differs from the previous work in several ways. It is suboptimal relative to the MAP algorithm in that its output only compares the distances between the best path going through each state, but for high SNR, this ratio approaches the likelihood since the probability is dominated by the best paths, so the SOBVD asymptotically approaches the optimal soft output decoder, at significantly decreased complexity.

The SOBVD is similar to the SOVA algorithm proposed by Hagenauer and Hoeher [52], in that it attempts to compare likelihoods between two paths, but the SOBVD does not place constraints on the compared paths. The SOVA algorithm relies on traceback comparisons of merged paths to determine likelihoods. This limits the path comparisons to the best paths merging with the ML path at each transition (i.e., the merging of surviving paths). If the best comparison path merged with a surviving path, but not the ML path, when the surviving path eventually merges with the ML path, the traceback comparison will fail to identify it as a potential comparison candidate. Thus, the SOVA does not guarantee that the maximum likelihood paths through all possible states are found, as the SOBVD does. The SOBVD offers slightly better performance with similar complexity to the SOVA, which in the simplified implementation of Berrou et al. [7] essentially needs to process the data twice through a Viterbi decoder, but with more complex traceback operations.

The SOVA algorithm does have the advantage in that it may be operated in a continuous manner. However, for block based coding, such as with tailbiting convolutional codes, continuous operation is not required. In addition, for decoding tailbiting convolutional codes, as suggested by Cox [19], the SOVA requires additional checks for validity of the comparison paths, which further increases complexity.

The SOBVD offers enhanced capability and system flexibility over traditional decoding methods. Soft outputs may be used to indicate reliability of the data, which is useful for dropping frames at the speech decoder when the information is not reliable, or request retransmission in a progressive transmission scheme. Soft outputs provide improved channel state estimation, which allows rate adaptation to be more efficiently implemented. Soft outputs improve the performance of concatenated code decoding, allowing both simple schemes with outer parity check codes or with more powerful convolutional, block, or turbo codes [7][52], which may be useful for important network control information. Finally, soft outputs may be used by the source coder in combined source-aided decoding [91].

The embedded channel coding methodology and decoder structures introduced in this research provide high performance, flexible architecture, and low complexity systems that can perform well under adverse conditions with efficient utilization of resources. There is little performance penalty compared to non-embedded designs, validating the embedded channel coding approach as both viable and attractive for wireless communication applications.

5 Adaptive Transmission

In section 3, a methodology for embedded source coding was described that allows efficient digital representation of speech signals. In section 4, a methodology for embedded channel coding was described that allows reliable wireless voice communications over a noisy channel in a flexible and efficient manner. In this section, a methodology for adaptive transmission is described that takes advantage of the flexibility of embedded source and channel coding to improve the quality and robustness of wireless voice communications through rate adaptation and progressive transmission.

Degradation in channel conditions due to fading and interference is a significant problem in wireless communications. The reliability of the received data may be improved under poor channel conditions by using higher output power or greater channel code error protection. Network congestion is also a concern, requiring some data loss or degradation of service when network resources are insufficient for the amount of traffic. Rate adaptation schemes allow improvements in performance and robustness by matching the source and channel rates more appropriately to the existing transmission conditions. Under good conditions, the channel coding can operate at a high rate, since little error protection should be needed. Under poor conditions, lower channel rates must be used to ensure reliable transmission. The source rate can be changed according to the bandwidth constraints and the reliability of transmission requirements.

The performance improvement obtained by rate adaptive over non-adaptive schemes is a major motivation for embedded source and channel coding. The major difficulty with implementation of rate adaptation is the added system complexity to support changing source and channel rates. Embedded source and channel coding addresses this shortcoming by allowing the changes in rate to occur with minimal complexity. Embedded source coding ensures that the data is prioritized, and that the source rate may be changed simply by truncating the data. Embedded channel coding allows the channel rate to be changed simply by defining the puncturing patterns for each of the rates. Thus the source and channel coders always operate at their base rate and the rates can be changed by the system without interaction with the coders.

Two schemes are presented to illustrate the potential of rate adaptation. Section 5.1 describes a scheme for rate adaptive transmission that allows the source and channel rate to be selected based on the transmission channel conditions so that the perceived quality of the received signal is optimized. Section 5.2 describes a scheme for progressive transmission that uses an ARQ scheme to transmit parity bits until enough bits have been transmitted to reliably reconstruct the data.

The benefits of adaptive transmission and the novel techniques employed in this design are discussed in section 5.3. The main accomplishment of this work is the demonstration that adaptive transmission with embedded source and channel coding offers substantial improvements in quality and robustness, without significantly added complexity, compared to non-adaptive implementation of wireless voice communications.

5.1 Rate Adaptive Transmission

In rate adaptive transmission, optimal source and channel rates are selected based on the channel conditions. Improved performance is obtained by varying the channel rate according to the required error rate and channel Signal to Noise Rate (SNR). The source rate is then selected according to the available transmission bandwidth. At high SNR, high source rate and high channel rate is desirable, producing the highest quality reconstruction. At lower SNR, lower channel rate must be used to ensure adequate protection of the information bits, so the source rate must also be reduced. Embedded source and channel coding makes rate adaptation easier to implement, since the rate adaptation does not require interaction with the source and channel coders, only truncation and puncturing of the symbols generated by the coders operating at their base rate.

The channel rate is selected according to the tolerable bit error rate for the information bits. Since the tolerable error rates for each group differ, unequal error protection (UEP) may be used for more effective error protection. Grouping the information bits according to their error tolerance facilitates the implementation of channel coding, using the methodology described in section 4.1. The tolerable error depends on the perceptual significance of the bits in each group. For example, the bits may be grouped such that the acceptable error rates are 10^{-4} for the most significant bits (scale and bit allocation), 10^{-2} for the medium significance bits (subband sample quantization indices MSBs), and 10^{-1} for the remaining bits.

Tables 1 and 2 show the required SNR to achieve Bit Error Rate (BER) of 10^{-2} and 10^{-4} with a punctured, 16 state RCPC with rates of 1/2, 2/3, and 4/5 over an AWGN channel with a soft decision decoder and over an interleaved Rayleigh channel with hard decision decoder without channel state information. These are essentially the two extreme cases of performance over wireless channels. The appropriate SNR measure is Es/No instead of Eb/No, since the rate adaptation is based on truncating symbols, so the symbol energy is fixed.

Table 5.1 - Required SNR for 16 state RCPC Code over AWGN Channel

rate	BER = 10^{-4}	BER = 10^{-2}
1/3	-1.1 dB	-2.5 dB
1/2	1.1 dB	-0.1 dB
2/3	3.1 dB	1.5 dB
4/5	4.9 dB	3.7 dB

Table 5.2 - Required SNR for 16 state RCPC Code over Rayleigh Channel

rate	BER = 10^{-4}	BER = 10^{-2}
1/3	8.6 dB	6.0 dB
1/2	12.1 dB	8.9 dB
2/3	15.8 dB	11.6 dB
4/5	22.2 dB	16.1 dB

Given knowledge of the transmission channel, output power constraint, and appropriate selection of channel coding and decoding algorithms, the channel rate is selected according to the error requirement for each group of bits. The source rate is selected according to the total channel bandwidth constraint, which can be either fixed or variable.

Rate adaptation optimizes the performance and allows channel degradation and congestion to be handled in a graceful manner. With embedded source and channel coding, there is little complexity increase over traditional non-adaptive source and channel coding methods.

The capability of changing source and channel rates after the coding has been performed provides several advantages and unique capabilities. The rate decision does not need to be known by the source or the channel coder, so the frame can be encoded without delay after the rate is selected. The rate selection can occur anytime after the encoding. For example, the rate selection can occur in a packet multiplexer handling multiple traffic sources, or at a different node in a multi-hop network.

5.2 Progressive Transmission

Forward Error Correction (FEC) schemes employ error correcting channel codes to provide robust transmission. When a reverse channel is available, transmission robustness can be further improved by having the receiver request retransmissions if correct decoding was not possible with the original transmission. Many different retransmission schemes can be implemented using Automatic Repeat Request (ARQ) protocols [1].

Progressive transmission combines ARQ with FEC, such that the receiver uses information from all transmissions for decoding. Embedded channel coding allows information from multiple transmissions to be combined and jointly decoded in a simple and effective manner. Progressive transmission systems have been shown to achieve both high throughput and robustness to a poor transmission channel [20][26][51][65].

Source rate adaptation can also be included in progressive transmission schemes to maximize the overall quality of the received transmission. Taking advantage of the prioritization of the encoded information in a congested network environment allows voice traffic to be supported with more graceful degradation [13]. Lowering the overall quality of voice communications allows the network load to be reduced so it can be handled adequately by the resources available. Embedded source coding fits nicely in this scheme since the prioritization of the bitstream allows for both classification by significance and bitstream truncation.

Progressive transmission with ARQ provides significant performance improvement over FEC alone. This is illustrated in a simple example comparing an unequal error protection (UEP) scheme with a progressive transmission ARQ scheme as shown in Figure 13. The data consists of 4 equal size blocks arranged by relative importance, with Class 1 being more significant and Class 3 less significant. The UEP scheme uses rate 1/3 code for Class 1, rate 1/2 for Class 2, and rate 2/3 for Class 3, resulting in an average rate of 1/2. The ARQ progressive transmission scheme first transmits Class 1 and Class 2 at rate 1/2, then depending on whether the first transmission is successful (i.e., no errors), either transmits the Class 3 data at rate 1/2 (Case T2) or transmits the Class 3 date at rate 2/3 and uses the remaining bandwidth to transmit additional parity bits for the Class 1 bits. Thus the overall rate of 1/2 is the same for both schemes.

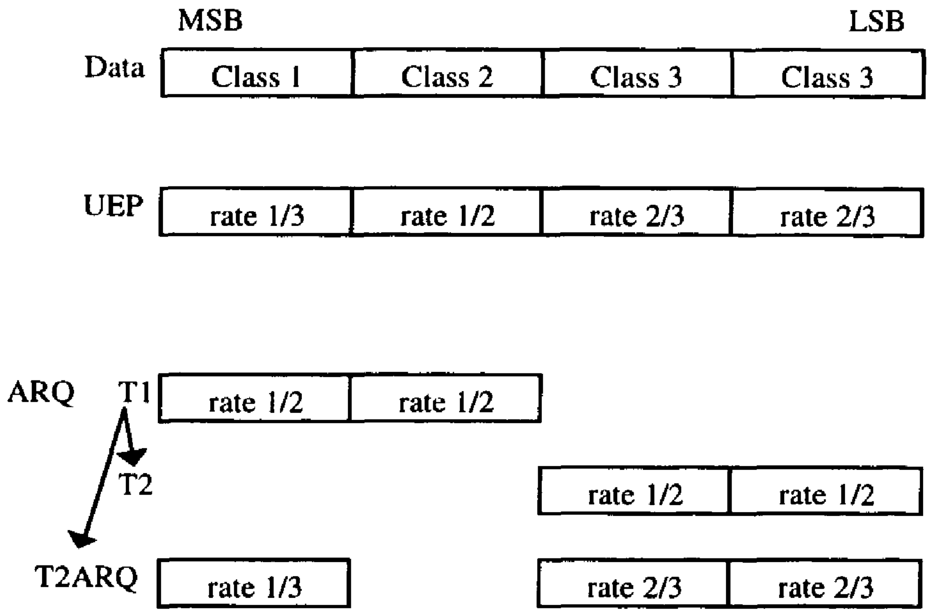


Figure 5.1 - UEP and ARQ Schemes for Simulation Example

The UEP and ARQ schemes were simulated over a hard decision channel using a 16 state tailbiting RCPC with block size of 40 data bits, with the results shown in Figure 5.2. The major benefit of the ARQ progressive transmission scheme is that it allows the Class 1 data to be transmitted at the higher rate most of the time, thus allowing a lower rate to be used for the Class 3 data. There is no change in performance for Class 1 and Class 2 bits since the lowest rate is no lower than the nominal UEP rate. On the other hand, there is also no penalty for transmitting at higher rate first. Since errors at the higher rate are still relatively infrequent, this allows the Class 3 data to be transmitted most of the time at the lower rate. Thus the error performance of the Class 3 data is improved significantly compared to the UEP scheme.

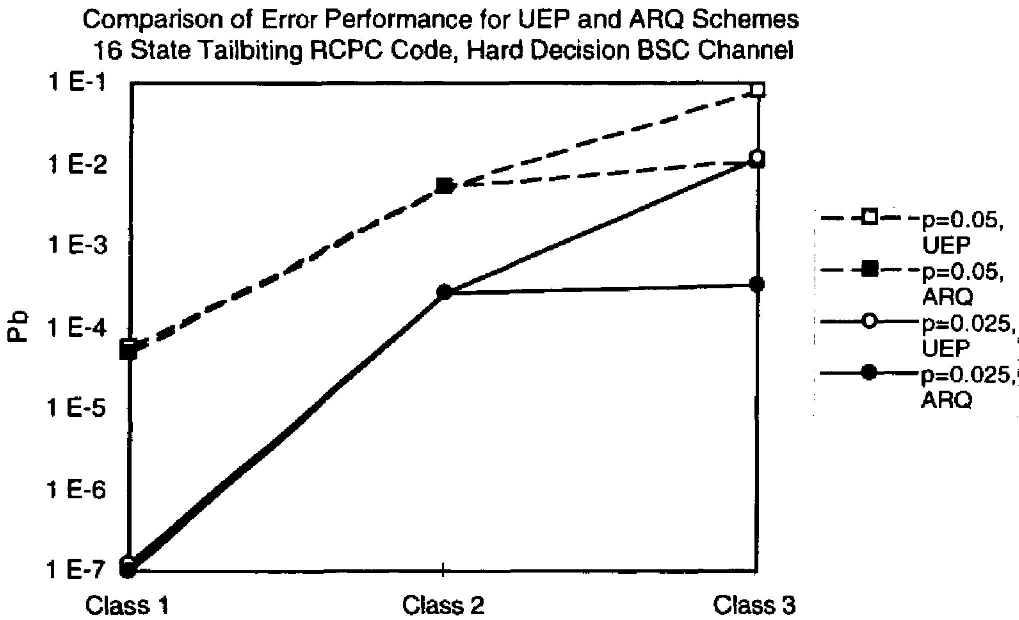


Figure 5.2 - Simulation Results for UEP and ARQ Comparison

This example provides an illustration of the flexibility and performance improvement of the progressive transmission ARQ scheme. Several strategies for transmission are possible. Error rates can be significantly improved by allowing lower peak rates, while the average rates are kept higher. High throughput under good channel conditions can be achieved by transmitting few or no parity bits from a baseline low rate code. In the event of unreliable transmission, additional parity bits are transmitted, which are combined with the original transmission and error correction to improve the decoding performance. Further transmissions may be used to increase the reliability of the MSBs, or if they have already been successfully decoded, to provide additional information about the LSBs. Thus information rate can be traded off with reliability of the transmission depending on the significance of the data, resulting in more graceful degradation in the event of network congestion or poor transmission environment.

The one significant drawback of ARQ based schemes is the additional delay required for the transmitter to receive acknowledgment. ARQ based schemes are typically better suited for applications such as data, where the error rate must be very low and propagation delay is not a big concern.

The progressive transmission ARQ scheme introduced in this work differs from the previous work in that it does not rely on many retransmissions nor requires errorless decoding to be successful, so it is suitable for voice communications. A single retransmission resulted in substantial performance improvements, with a delay penalty comparable to the delay introduced by 2-way interleaving, which is fairly common in wireless voice applications. Rather than attempting to retransmit until perfect decoding is possible, ARQ is used for instant adaptation of the source and channel rate to the existing channel conditions. Decoding errors may still occur, but they are matched to the tolerable error rates for each bit. The overall error rate is lower than what is achievable with FEC alone, resulting in more reliable and robust communications.

5.3 Discussion

Effective implementation of wireless voice communications requires appropriate selection of source and channel coding to match the availability of system resources and transmission environment. The adaptive transmission schemes developed in this work allow matching of the source and channel rates to be done in a simple manner and demonstrate improved performance without significant complexity increase over non-adaptive schemes.

Wireless voice communication systems should be designed to account for the characteristics of its traffic and its transmission environment. Adaptive transmission implementations provide both good system performance and efficiency under good conditions, and deal effectively with adverse conditions. Rate adaptive transmission allows the source and channel rate to be optimally matched to the channel. If a reverse channel is available, progressive transmission allows the channel bandwidth to be utilized even more effectively by taking advantage of data prioritization and tolerable errors.

The benefits of rate adaptive transmission for wireless communications has been illustrated in many different works [17][18][47][79][51][53]. The rate adaptation scheme in this work is similar to the one proposed by Cox et al. [18]. The main difference is that the embedded channel coding methodology in this work provides increased flexibility with the shorter block sizes, allowing improved matching of the channel code rate to the transmission channel conditions and source error tolerance.

The progressive transmission methodology proposed in this work is a significant improvement over traditional hybrid-ARQ schemes. Traditional hybrid-ARQ schemes are designed to provide reliable transmission of data, but are not well suited for voice communications due to the short delay requirements and tolerance to small error rates. The progressive transmission scheme is designed not to try to get the data transmitted at all costs, but to use the channel feedback to decide how bandwidth is best utilized. Data prioritization and unequal error protection is used to maximize throughput at acceptable error rates. Delay is limited by constraining the number of retransmissions allowed. Progressive transmission allows high throughput, high reliability, and good robustness to be achieved under highly variable channel conditions.

Embedded source and channel coding makes it possible for these adaptive transmission schemes to be implemented in a highly flexible system with little added complexity. The embedded coding concept greatly facilitates rate adaptation and frees the system from additional interaction with the source and channel coders. The performance penalty and complexity increase of embedded source and channel coding, compared to non-embedded coding methods, may be minimized as shown by our design methodology. Adaptive transmission may then be implemented with only a slight increase in complexity compared to a non-adaptive system, but with performance over a large range of conditions comparable to using fixed non-adaptive implementation optimized for each condition. The main contribution of this work is the demonstration that adaptive transmission with embedded source and channel coding offers substantial improvements in the quality and robustness, without significantly added complexity, compared to non-adaptive implementation of wireless voice communications.

6. Conclusion

Shannon's introduction of a mathematical model for communications in 1948 has provided engineers a framework for developing and evaluating digital communication systems. The theoretical performance limit, evaluated in the form of rate-distortion and channel capacity, has justified the separation of the digital communications problem into two distinct areas: source and channel coding. In theory, source coding and channel coding may be separately designed, optimized and evaluated. Source coding and channel coding have also evolved into mostly separate disciplines, with distinct methodologies, goals, and research communities. In practice, however, source and channel coding must both be considered within the framework of the overall communication system in order to obtain an efficient implementation.

Digital wireless voice communications present an interesting challenge. The widespread use and importance of voice communications, and the current explosive growth of wireless communications, make it a timely and worthwhile research subject. Efficient and effective technical solutions are needed to accommodate the dynamic nature of the radio channel, limited bandwidth availability, high consumer demand and expectations, and constrained size, weight, and power.

The methodologies developed in this research for adaptive transmission with embedded source and channel coding are significant in that they offer the potential for improvements in the implementation of high quality, efficient, flexible, robust, low complexity wireless voice communications.

The main accomplishment of this work is the added motivation for embedded source and channel coding. Embedded source and channel coding fits nicely in the framework of the digital communication system model. Incremental information and redundancy does not significantly impact the performance of the source and channel coding algorithms. Optimization of the source and channel coding requires only selection of the appropriate code rates. Adaptation of the source and channel coders to match the channel conditions can be easily implemented without significantly adding to the system complexity. Adaptive transmission offers significant advantages for wireless voice communications, offering greater flexibility and higher performance than fixed transmission.

The design methodologies developed in this work demonstrate a number of innovations, among which are:

- Perceptually based subband dynamic bit allocation and prioritization
- Subband spectral analysis
- Embedded quantization
- Embedded channel coding with short block tailbiting rate compatible punctured convolutional codes.
- Asymptotically optimal fixed computational workload decoding of tailbiting convolutional codes.
- Soft output decoding of tailbiting convolutional codes with bi-directional Viterbi decoding.
- Adaptive source and channel rate selection with unequal error protection
- Progressive transmission with ARQ

This research has demonstrated the attractiveness of adaptive transmission with embedded source and channel coding, and provided new methodologies for improved implementation of embedded coding techniques in wireless voice communications. These innovations provide opportunities for meaningful future research in several areas.

The current implementation could be optimized to specific system developments in a manner beyond the scope of this work. This work did not incorporate more complex, "real-world" transmission channels, nor end-to-end subjective evaluation. Such optimization could only take place in the context of actual system implementation, with extensive evaluation under a large range of conditions. It is expected, however, given the flexibility and robustness of the adaptive transmission methodology, that such a system would perform very well under a large range of conditions.

The methodologies developed in this work successfully concentrated on achieving performance and complexity comparable to non-embedded coding methods of similar coding approaches. These methodologies were of general nature, leaving open the possibility of applying them to other coding approaches. They may be particularly suitable for low rate speech coding methodologies, such as CELP and IMBE. These lower bit rate speech coders would also benefit from the perceptually based bit prioritization and embedded quantization to improve performance scalability, and robustness.

The methodologies developed in this work may also be suitable to other applications, notably wireless image and video transmission. Image and video transmission are similar to speech and audio in that they are efficiently encoded using perceptually based transform and subband methods. This results in a bitstream that may be effectively embedded. Adaptive transmission and embedded channel coding would then be advantageous in situations where rate adaptation is necessary to deal effectively with a varying channel.

The advantages of adaptive transmission with embedded source and channel coding, as evidenced by this work, should motivate further research in improving implementations and design methodologies. It provides an attractive framework for optimal realization of wireless communication systems, powerful in its flexibility, efficient in its modularity, elegant in its simplicity.

Acronyms and Glossary

AC-2 (Audio Coder 2) - High fidelity audio coding standard developed by Dolby Laboratories.

Adaptive Transmission - Transmission schemes in which the source or channel coding are modified to improve performance as transmission conditions change.

ADM (Adaptive Delta Modulation) - Adaptive step size Delta Modulation.

ADPCM (Adaptive Differential Pulse Code Modulation) - Adaptive step size and adaptive predictor DPCM. ADPCM is the most common form of speech compression, able to achieve coding rates of 16 kbps - 32 kbps. Various CCITT standards exist based on ADPCM, G.721 (32 kbps toll quality ADPCM), G.723 and G.726 (multiple rate 16-40 kbps ADPCM), G.727 (embedded ADPCM).

A-Law - A logarithmic companding scale used in scalar quantization of speech using PCM.

Aliasing - Distortion introduced by sampling at less than twice the highest frequency component of the source signal. Aliasing causes ambiguity in distinguishing between alias frequencies. The ambiguity is eliminated if the signal is bandlimited such that no alias frequency appears in the signal band.

AM (Amplitude Modulation) - Analog modulation scheme where the analog signal modulates the amplitude of a sinusoidal carrier.

AMPS (Advanced Mobile Phone Service) - The North American analog cellular telephone standard.

Analysis-by-Synthesis - A source coding methodology using closed loop selection of the optimal representation by exhaustively comparing all the possible representations with the input.

Analysis Filterbank - The filterbank which splits the input signal into the subbands.

APCM (Adaptive Pulse Code Modulation) - Adaptive step size PCM.

ARQ (Automatic Repeat Request) - A transmission protocol based on requesting retransmission of bad data blocks through transmitting acknowledgments or retransmission requests from the receiver via a reverse channel.

Array Code - A concatenated block code where the interleaving is arranged such that only one symbol from each block in the inner code is used in each block of the outer code.

ATM (Asynchronous Transfer Mode) - A data network protocol standard for packet based communications.

AWGN (Additive White Gaussian Noise) - A model for a transmission channel where the only degradation is white Gaussian noise added to the transmitted signal.

Bark Scale - The non-linear frequency scale corresponding to the critical bands characterizing the human auditory system frequency selectivity.

Base Station - The hardware in a cellular network that communicates with the handheld units.

BCH (Bose-Chaudhuri-Hocquenghem) Code - A family of cyclic block codes with good distance properties and efficient decoding structures based on using the generating polynomials' zeros in the extension field to evaluate the syndromes and determine the error locations.

Bennett's Formula - A companding approximation to the optimal quantizer.

BER (Bit Error Rate) - A measure of transmission reliability based on the ratio of decoded errors to total transmitted bits.

Binary Symmetric Channel (BSC) - A discrete model for a binary channel with a given probability of a channel error.

CCITT (International Telephone and Telegraph Consultative Committee) - The International Telephone Standards Committee.

CD (Cepstral Distance) - Integrated difference in spectral envelope obtained by LPC analysis, easily calculated from cepstral coefficients. It provides a more "coarse" fit to source spectral shape than spectral distortion, thus modeling auditory system's limited frequency resolution more accurately.

CDMA (Code Division Multiple Access) - A multiple access technique using spread spectrum where each user is assigned a communications channel consisting of a pattern orthogonal to all other users.

Cell - The physical area over which a base station will communicate with the handheld units.

Cellular Communications - Communications network using spatial separation to increase network capacity by allowing frequency reuse.

CELP (Code Excited Linear Prediction) - Hybrid speech coding methodology based on LPC and representing the excitation as a vector source. The excitation is chosen from a VQ codebook that can be stochastic, adaptive, or ordered.

CF (Coherence Function) - Normalized cross power spectrum between the source and coded signal is separated into coherent and noncoherent components and combined via frequency selective thresholds and weighting functions.

Channel Coding - Introduction of redundancy such that the reliability of a transmission through a noisy channel may be made better than a minimum error rate.

Channel Rate - A measure of the amount of redundancy introduced by channel coding correponding to the ratio of information to total encoded data. A low rate channel code provides high redundancy and high error protection, whereas a high rate provides low redundancy and low error protection.

Column Distance Function - For a convolutional code, the minimum weight codeword produced by an information word with non-zero initial segment as a function of the segment length. It represents the minimum distance unmerged path of the given segment length. See also decision depth function.

Companding - A non-linear mapping of the input range to a uniform quantizer used to achieve non-uniform quantization. The most common companding function is a logarithmic mapping, such as A-law or μ-Law.

Concatenated Code - Use of two or more error correction codes in a sequential manner, where the output of one coder becomes the input to the next coder.

CPC (Complementary Punctured Convolutional Codes) - A set of punctured convolutional codes derived from the same base code with puncturing patterns such that each code is individually decodable and when combined yield the base code.

CRC (Cyclic Redundancy Check) - A family of high rate cyclic block code widely used for error detection.

CVA (Circular Viterbi Algorithm) - A decoding algorithm for punctured convolutional codes based on continuous Viterbi decoding and adaptive stopping rules.

DCT (Discrete Cosine Transform) - Orthogonal transform with cosines as basis functions

Decision Depth Function - For a convolutional code, the minimum path length required to achieve the minimum distance. See also column distance function.

Distortion - The error introduced due to processing a signal.

DFT (Discrete Fourier Transform) -Transform with complex exponential basis functions

DM (Delta Modulation) - One bit oversampled differential predictive quantization. The previous quantization value is used as the prediction, and the prediction error or residue is quantized with a one bit scalar quantizer.

DPCM (Differential Pulse Code Modulation) - differential predictive quantization. A linear prediction based on past quantization values is used, with predetermined coefficients according to the source autocorrelation. The residue is quantized with a multi-bit scalar quantizer.

DWT (Discrete Wavelet Transform) - Orthogonal transform with multi-resolution (time-frequency) basis function.

Embedded Channel Coding - A channel coding methodology that allows error-free decoding given only a portion of the bitstream.

Embedded Source Coding - A source coding methodology that allows partial reconstruction given only a portion of the bitstream.

Fading - The characteristics of radio channels where the channel conditions are highly variable due to either movement of the transmitter, receiver, or obstructions.

FDMA (Frequency Division Multiple Access) - A multiple access technique where each user is assigned a communications channel consisting of a portion of the frequency band.

FEC (Forward Error Correction) - Use of an error correction code to allow reliable transmission of digital data over a noisy channel by correcting transmission errors in the receiver.

FFT (Fast Fourier Transform) - A computationally efficient implementation of the DFT based on decimation symmetry of the basis functions.

FIR (Finite Impulse Response) - A class of filters where the response to an input has finite duration.

FM (Frequency Modulation) - Analog modulation scheme where the analog signal modulates the frequency of a sinusoidal carrier.

GSM (Groupe Special Mobile or Global System for Mobile Communications) - The European 2nd generation digital cellular telephone standard.

Hamming Code - A family of simple single error correcting block codes.

Hard Decisions - In a channel decoder, a characterization of the received signal based on a single observation.

Homomorphic Coders - Speech coding methodology based on the source/system model. A deconvolution of the source through calculation of the speech cepstrum, the IDFT of log spectrum, is used to obtain to obtain source/system characterization.

Hybrid ARQ - A transmission protocol combining ARQ and FEC to provide protection from transmission errors as well as allowing retransmission requests of bad data blocks.

Hybrid coder - Speech coders that use both an excitation-filter model for speech production and waveform matching for selecting the optimal representation.

IIR (Infinite Impulse Response) - A class of filters where the response to an input has infinite duration.

IMBE (Improved Multiband Excitation) - see MBE

IMT (Interpolated Masking Threshold) - A cost measure for optimizing quantization noise shaping obtained by log linear interpolation of the Signal-to-Mask Ratio (SMR) required to achieve the Just Noticeable Distortion (JND) masking threshold.

Interference - The interaction between a transmitted signal and other signals such that the ability of the receiver to detect the transmitted signal is degraded.

Interleaving - Changing the order of the bitstream in an established pattern to randomize bursty events.

IS-54 (Interim Standard 54) - The North American 2nd generation digital TDMA cellular telephone standard.

IS-95 (Interim Standard 95) - The North American 2nd generation digital CDMA cellular telephone standard.

ISO-MPEG (International Standards Organization - Motion Pictures Expert Group) - An international standard for digital compression of video and audio.

JND (Just Noticeable Distortion) - Noise masking threshold corresponding to the minimum signal level required to be perceptible in the presence of a masking signal. Thus if the distortion introduced by the coder is below the JND from the speech signal, the listener is not able to detect it.

kHz (kilohertz) - A unit of measure for frequency equal to 1000 cycles per second.

KL (Karhunen-Loeve Transform) - Orthogonal transform with eigenvector basis function.

LBG (Linde-Buzo-Gray) Algorithm - An algorithm for determining optimum codebooks for vector quantization of a given dimension and codebook size by assigning quantization regions and reconstruction vectors. The algorithm is an iterative search to find a codebook such that all input vectors are assigned to their optimal quantization region and the reconstruction vectors are optimized for their quantization regions.

LD-CELP (Low Delay - Code Excited Linear Prediction) - A CELP coder where the linear prediction coefficients are determined entirely from previous samples to reduce the implementation delay.

Likelihood - The ratio of the probabilities of two events.

Lloyd-Max Quantization - Optimal non-uniform scalar quantization with non-uniform levels chosen to minimize the mean squared error given the source probability distribution. The Lloyd-Max criteria are based on each reconstruction value being the centroid of the level, and thresholds between levels being the midpoint between adjacent reconstruction value such that a nearest neighbor assignment is used to perform the quantization.

LPC (Linear Predictive Coding) - Speech coding methodology based on an all-pole (linear prediction) system model and a voiced/unvoiced characterization of the excitation. The coefficients to the all-pole filter are obtained through autocorrelation or covariance methods such as the Levinson-Durbin recursive algorithm. LPC coefficients can also be represented as Log Area Ratio (LAR) and Line Spectral Pair (LSP) coefficients. LAR are lattice filter equivalent coefficients, thus can more simply ensure filter stability. LSP are pole location coefficients, thus can be more simply perceptually coded. LPC analysis can be shown to be equivalent to cepstral analysis, thus it effectively performs a deconvolution of the speech source.

LSB (Least Significant Bit) - In a binary representation, bits which represent the smallest weight among the encoded bits.

Markov Process - A random process modeled by a sequence of state transitions such that the statistical properties depend only on the present state.

Masking - Property of the human auditory system where the presence of one signal may render an otherwise audible signal inaudible.

MBE or IMBE (Multiband Excitation / Improved Multiband Excitation) - Speech coding methodology where the speech is represented by a sum of narrowband components. The spectrum is characterized in multiple bands with the source being voiced or unvoiced in each band, and the excitation and envelope parametrized. The INMARSAT 4.1 kbps speech coder is based on IMBE.

ML (Maximum Likelihood) - A criteria in estimation problems where the desirable result is identification of the most probable event from a set of non-deterministic events.

Modulation - The process of converting information into the actual waveforms to be transmitted over the communications channel

MOS (Mean Opinion Score) - Subjective measure of speech quality obtained through listening tests where subjects rate the perceived quality of the reconstructed speech along a 1-5 scale, where the scores represent 1. Bad, 2. Poor, 3. Fair, 4. Good, and 5. Excellent.

MPLP (Multi-Pulse Excited Linear Predictive) - Hybrid speech coding methodology based on LPC and representing the excitation as a sum of pulses. The Skyphone 9.6 kbps speech coder is based on MPLP.

MSB (Most Significant Bit) - In a binary representation, bits which represent the largest weight among the encoded bits.

MSE (Minimum Square Error) - an error measure based on the squared difference between two signals.

Mu-Law or μ-Law- A logarithmic companding scale used in scalar quantization of speech using PCM.

Multi-Hop Network - Network where multiple wireless communication links may have to be established to allow users access to the network.

NMR (Noise to Mask Ratio) - Integrated noise power above spectral masking threshold. Takes into account masking properties of the auditory system.

NMT (Nordic Mobile Telephone) - One of several European analog cellular telephone standards.

NTT (Nippon Telephone and Telegraph) - Japan's national telecommunications provider, who also is one of the Japanese cellular telephone operators.

Non-uniform Scalar Quantization - Scalar quantization with the quantization levels are non-uniformly distributed over the quantization range.

Packet Network - A communications network where data is routed as individual packets and reconstructed at the receiver.

PCM (Pulse Code Modulation) - Speech coding through companded scalar quantization. The CCITT standard for PCM is given in standard G.711, defining the sampling rate of 8 KHz, the coding rate of 64 kbps, and the u-Law and A-law companding rules. Typically referred to as uncompressed speech.

PCS (Personal Communications Services) - The next generation North American digital cellular telephone standard, currently under development.

PDC (Personal Digital Cellular) - The Japanese 2nd generation digital cellular telephone standard.

Peer-To-Peer Network - Base-stationless cellular network where data is routed through communication links established directly between users.

PM (Phase Modulation) - Analog modulation scheme where the analog signal modulates the phase of a sinusoidal carrier.

Perceptual Coders - Source coders that incorporate a human perception model to obtain optimization criteria.

Probability - For a non-deterministic process, the relative frequency with which a particular event should occur.

Progressive Transmission - Adaptive transmission schemes where information is transmitted in an incremental manner until it is decoded successfully.

QMF (Quadrature Mirror Filterbank) - A Filter Structure allowing subband decomposition such that when the source signal is reconstructed, aliasing due to decimation of the subband samples is completely canceled so that no aliasing error occurs.

Quantization - Process by which a representation from a finite set or codebook is used to represent a 1-dimensional or multi-dimensional variable.

Radio Communications - Electronic communications using electro-magnetic waves over free space.

Rate Adaptive Transmission - Adaptive transmission schemes where the source rate or channel rate are modified to improve performance as transmission conditions change.

Rayleigh - A probability distribution typical of the received signal strength in radio communications where the channel is dynamic and there is no single strong transmission path from the transmitter to the receiver.

Rayleigh Channel - A channel model with a fixed background noise level where the received signal energy is modeled as a Rayleigh variable.

Rician - A probability distribution typical of the received signal strength in radio communications where the channel is dynamic and there is one strong transmission path and many weaker paths from the transmitter to the receiver.

RCPC (Rate Compatible Punctured Convolutional) Code - A family of convolutional codes where the higher rate codes are defined through puncturing of the channel symbols generated from base code.

RPE-LTP (Regular Pulse Excited, Long Term Prediction) - Hybrid speech coding methodology based on LPC and representing the excitation as a sum of periodic pulses. The GSM full rate 13 kbps coder is based on RPE-LTP.

RS (Reed-Solomon) Code - A family of cyclic block codes that are a non-binary subclass of BCH codes. RS codes are maximum distance codes: for the same block length and rate, no other codes have larger minimum distances.

Sampling - The process by which discrete values or samples are used to represent a continuous waveform.

Sampling Theorem - Property discovered by Nyquist that continuous waveforms may be perfectly reconstructed if sampled at higher than twice the highest frequency component in the waveform.

Scalar Quantization - Basic binary encoding methodology where each input value or sample, is represented by a binary word.

seg SNR (segmental SNR) - SNR measurements over short 10-20 ms frames averaged out over time. It provides a better fit to the perceived quality of coded speech than SNR since it localizes effect of errors, thus accounting somewhat for the non-stationarity of speech.

Shadowing - The characteristics of radio channels where there are obstructions between the transmitter and receiver such that the received signal strength falls off as the fourth power of the distance rather than the square.

Soft Decisions - In a channel decoder, a probabilistic measure of a single observation of the received signal.

Spectrum - The range of frequencies usable for radio communications.

Speech Coder - A source coder designed for speech sources.

Spread Spectrum - Communications methodology where the signal is spread over a large bandwidth to facilitate transmission, and de-spread at the receiver.

SMR (Signal to Mask Ratio) - Ratio of signal energy to the masking threshold given by the JND.

SNR (Signal to Noise Ratio) - Ratio of signal energy to noise (error) energy. Classical mathematical method where many methods exist to optimize system to minimize squared error. It does not fit closely to speech quality since it is a global, not local measure, and does not take into account selectivity or masking properties of the auditory system.

SOBVD (Soft-Output Bi-directional Viterbi Decoder) - A low complexity, fixed workload, soft output decoder for tailbiting convolutional codes developed in this work.

SPL (Sound Pressure Level) - Measure of the acoustical energy.

Source Coding - The conversion of a signal to a digital representation such that the minimum number of symbols is required to represent the signal to a given fidelity.

Source Rate - A measure of the amount of information and coding efficiency required to represent a source. A low rate corresponds to less information or better coding efficiency, and a higher rate corresponds to more information or worse coding efficiency.

Spectral Distortion - Integrated difference in spectral magnitude between source and coded speech averaged out over time. It is an improvement over segmental SNR by neglecting less significant phase error and allowing frequency selectivity to be taken into account.

STC (Sinusoidal Transform Coder) - Speech coding methodology where the speech is parametrized as a sum of sinusoids. Parametrization is done through representing the spectrum by a cepstral envelope and pitch/voicing parametrization.

Subband Coding - Source coding methodology where input is decomposed into several spectral subbands to facilitate compression

Subband Spectral Analysis - Transform-based spectral analysis of subband filtered signals obtained by performing the analysis on the subband samples instead of on the input directly.

Synthesis Filterbank - The filterbank which combines the subbands into the output signal.

Vocoders - Speech coders based on modeling the speech signal as either a tonal or white noise excitation and a filter.

TACS (Total Access Communication System) - One of several European analog cellular telephone standards.

Tailbiting - Using convolutional codes to encode blocks of data, such that the initial state of the encoder is set to match the final state, thus eliminating tail bits.

TDMA (Time Division Multiple Access) - A multiple access technique where each user is assigned a communications channel consisting of a time slot in a frequency band.

Trellis - A graphical representation of possible state transitions generated by a Markov process such as a finite state machine or convolutional code.

TVD (Tailbiting Viterbi Decoder) - A low complexity, fixed workload decoder for tailbiting convolutional codes developed in this work.

UEP (Unequal Error Protection) - Channel coding schemes resulting in unequal error protection for different locations in the bitstream.

Uniform Scalar Quantization - Scalar quantization with the quantization levels are uniformly distributed over the quantization range.

VA (Viterbi Algorithm) - The Viterbi Algorithm is a maximum likelihood decoder based on evaluating the likelihood of each possible state transition via an additive metric. The valid state transitions are described graphically as a trellis, such that the computation of the most likely path is done by considering all merging paths at each state. The algorithm keeps only the most likely merging path at each state such that eventually, only the most likely path is retained. The Viterbi Algorithm finds application in systems where the operation may be described in terms of state transitions, such as convolutional codes and trellis quantization.

VQ (Vector quantization) - Source coding method where a digital word representation is used for a block of input values. The digital word represents an index to a codebook representing possible reconstruction blocks or vectors. VQs are also classified according to codebook structure and search methodology. Typical VQs include full search (LBG) VQ, multi-stage VQ, product code VQ, shape gain VQ, and lattice VQ

VSELP (Vector Sum Excited Linear Prediction) - A subset of CELP where the codebook consists of a sum of orthogonal codebooks that can be searched independently, thus reducing the complexity of a full codebook search.

White Noise - Random signal with flat power spectrum.

Wireless - Radio communications over free space, typically refers to on-demand access to a network from portable units not physically connected to the network (e.g., cellular telephone).

Wireline - Communications over an established infrastructure, which may consist of wires, fiber-optics, microwave links, and satellite links (e.g., telephone network).

References

[1] J. B. Anderson, S. Mohan, Source and Channel Coding: An Algorithmic Approach, Kluwer, 1991.

[2] E. Ayanoglu, R. M. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders," IEEE Tr. Information Theory, vol. IT-33, no. 6, November 1987, pp. 855-865.

[3] L. R. Bahl, J. Cocke, F. Jelinek, J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," IEEE Tr. Information Theory, vol. IT-20, March 1974, pp. 284-287.

[4] W. G. Bath, V. D. Vandelinde, "Robust Memoryless Quantization for Minimum Signal Distortion," IEEE Tr. Information Theory, vol. IT-28, no. 2, March 1982, pp. 296-306.

[5] J. Belzile, D. Haccoun, "Bidirectional Breadth-first Algorithms for the Decoding of Convolutional Codes," IEEE Tr. Communications, vol. 41, no. 2, Feb. 1993, pp. 451-457.

[6] W. R. Bennett, "Spectra of Quantized Signals," Bell System Technical Journal, vol. 27, July 1948, pp. 446-472.

[7] C. Berrou, P. Adde, E. Angui, S. Faudeil, "A Low Complexity Soft-output Viterbi Decoder Architecture," IEEE International Conference on Communications, 1993, pp. 737-740.

[8] K. Brandenburg, T. Sporer. "NMR and Masking Flag: Evaluation of Quality Using Perceptual Criteria." Audio Engineering Society Test and Measurement Conference, 1992.

[9] K. Brandenburg, G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio," Journal Audio Engineering Society, vol. 42, no. 10, October 1994, pp. 780-792.

[10] J. A. Bucklew, "Companding and Random Quantization in Several Dimensions," IEEE Tr. Information Theory, vol. IT-27, no. 2, March 1981, pp. 207-211.

[11] J. B. Cain, G. C. Clark, J. M. Geist, "Punctured Convolutional Codes of Rate (n-1)/n and Simplified Maximum Likelihood Decoding," IEEE Tr. Information Theory, vol. IT-25, January 1979, pp. 97-100.

[12] J. Campbell, T. E. Tremain, V. Welch, "The Proposed Federal Standard 1016 4800 bps Voice Coder: CELP," Speech Technology, April 1990, pp. 58-64.

[13] K. K. Chang, I. Rubin, "On Improving Voice Quality Under Dynamic Encoding Algorithms in ATM Networks," submitted to ISDN Systems and Computer Networks.

[14] J. Chen, R. V. Cox, Y. Lin, N. S. Jayant, M. Melchner, "A Low Delay CELP Coder for the CCITT 16 kb/s Speech Coding Standard," IEEE Journal On Selected Areas In Communications, vol. 10, no. 6, June 1992, pp. 830-849.

[15] T. M. Cover, J. A. Thomas, Elements of Information Theory, Wiley, 1991.

[16] R. V. Cox, "The Design of Uniformly and Nonuniformly Spaced Pseudoquadrature Mirror Filters," IEEE Tr. on Acoustics, Speech, and Signal Processing, vol. ASSP-34, no. 5, October 1986, pp. 1090-1096.

[17] R. V. Cox, S. L. Gay, Y. Shoham, S. R. Quackenbush, N. Seshardi, N. S. Jayant, "New Directions in Subband Coding," IEEE Journal On Selected Areas In Communications, vol. 6, no. 2, February 1988, pp. 391-409.

[18] R. V. Cox, J. Hagenauer, N. Seshardi, C. W. Sundberg, "Subband Speech Coding and Matched Convolutional Channel Coding for Mobile Radio Channels," IEEE Tr. on Signal Processing, vol. 39, no. 8, August 1991, pp. 1717-1731.

[19] R. V. Cox, C. W. Sundberg, "An Efficient Adaptive Circular Viterbi Algorithm for Decoding Generalized Tailbiting Convolutional Codes," IEEE Tr. Vehicular Technology, vol. 43, no. 1, February 1994, pp. 57-68.

[20] M. Darnell, B. K. Honary, F. Zolghadr, "Embedded Coding Technique: Principles and Theoretical Studies," IEE Proceedings, vol. 135, Pt. F, no. 1, February 1988, pp. 43-50.

[21] G. Davidson, M. Bosi, "AC-2: High Quality Digital Audio Coding for Broadcast and Storage," Dolby Technical Papers, Publication No. S92/9433, NAB 1992 Broadcast Engineering Conference Proceedings Reprint.

[22] G. Davidson, W. Anderson, A. Lovrich, "A Low Cost Adaptive Transform Decoder Implementation for High Quality Audio," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1992, pp. II-193-196.

[23] I. S. Dedes, D. R. Vaman, C. V. Chakravarthy, "Variable Bit Rate Adaptive Predictive Coder," IEEE Tr. on Signal Processing, vol. 40, no. 3, pp. 511-517.

[24] J. A. Deer, P. J. Bloom, D. Preis, "Perception of Phase Distortion in All-Pass Filters," Journal audio Engineering Society, vol. 33, no. 10, October 1985, pp. 782-785.

[25] R. De Iacovo, D. Sereno, "Embedded CELP Coding for Variable Rate Between 6.4 and 9.6 Kbit/s," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1991, pp. 681-684.

[26] R. H. Deng, M. L. Lin, "A Type I Hybrid ARQ System with Adaptive Code Rates," IEEE Tr. Communications, vol. 43, no. 2/3/4, February/March/April 1995, pp. 733-737.

[27] C. L. Despins, J. Belzile, D. Haccoun, "Bidirectional Decoding of Convolutional Coders for Indoor Cellular Radio," IEEE Tr. Vehicular Technology, vol. 43, no. 2, May 1994, pp. 264-269.

[28] S. Dimolitsas, "Objective Speech Distortion Measures and their Relevance to Speech Quality Assessments," IEE Proceedings, Vol. 136, Pt. I, No. 5, October 1989, pp. 317-324.

[29] S. Dimolitsas, "Subjective Quality Quantification of Digital Voice Communication Systems," IEE Proceedings I, Vol. 138, No. 6, December 1991, pp. 585-595.

[30] S. Dimolitsas, C. Ravishankar, G. Schroder, "Current Objectives in 4-kb/s Wireline-Quality Speech Coding Standardization," IEEE Signal Processing Letters, vol. 1, no. 11, November 1994, pp. 157-159.

[31] J. J. Dubnowski, R. E. Crochiere, "Variable Rate Coding of Speech," Bell System Technical Journal, March 1979, pp. 577-600.

[32] D. Esteban, C. Galand, "Application of Quadrature Mirror Filters to split-band Voice Coding Schemes," IEEE International Conference on Acoustics, Speech, and Signal Processing, 1977, pp. 191-195.

[33] M. V. Eyuboglu, G. D. Forney, "Lattice and Trellis Quantization with Lattice and Trellis Bounded Codebooks - High-Rate Theory for Memoryless Sources," IEEE Tr. Information Theory, vol. 39, no. 1, January 1993, pp. 46-59.

[34] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," IEEE Tr. Information Theory, vol. 36, no. 4, July 1990, pp. 799-809.

[35] G. D. Forney, "The Viterbi Algorithm," Proceedings of the IEEE, vol. 61, no. 3, March 1973, pp. 268-278.

[36] H. Fujiya, K. Saito, "Variable-Rate Voice Communication using ADPCM in ATM Networks," IEEE Intl. Conference on Communications, 1991, pp. 1713-1717.

[37] P. S. Gaskell, "Developing Technologies for Personal Communications Networks," IEE Electronics & Communication Engineering Journal, April 1992, pp. 53-64.

[38] A. Gersho, "Principles of Quantization," IEEE Tr. Circuits and Systems, vol. CAS-25, no. 7, July 1978, pp. 427-436.

[39] A. Gersho, "Asymptotically Optimal Block Quantization," IEEE Tr. Information Theory, vol. IT-25, no. 4, July 1979, pp. 373-380.

[40] A. Gersho, "On the Structure of Vector Quantizers," IEEE Tr. Information Theory, vol. IT-28, no. 2, March 1982.

[41] A. Gersho, "Advances in Speech and Audio Compression," Proceedings of the IEEE, vol. 82, no. 6, June 1994, pp. 900-918.

[42] A. Gersho, R. M. Gray, Vector Quantization and Signal Compression, Kluwer, Boston, MA, 1990.

[43] I. Gerson, M. Jasiuk, "Vector Sum Excited Linear Prediction Speech Coding at 8 kbps," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1990, pp. 461-464.

[44] I. Gerson, M. Jasiuk, M. McLaughlin, E. Winter, "Combined Speech and Channel Coding at 11.2 kbps," Signal Processing V, Amsterdan 1990, pp. 1339-1341.

[45] D. J. Goodman, B. McDermott, L. Nakatani, "Subjective Evaluation of PCM Coded Speech," Bell System Technical Journal, Vol. 55, No. 8, October 1976, pp. 1087-1109.

[46] D. J. Goodman, "Embedded DPCM for Variable Bit Rate Transmission," IEEE Tr. on Communications, vol. COM-28, no. 7, July 1980, pp. 1040-1046.

[47] D. J. Goodman, C. W. Sundberg, "Combined Source and Channel Coding for Variable Bit Rate Transmission," Bell System Technical Journal, September 1983, pp. 2017-2036.

[48] K. Gould, R. V. Cox, N. S. Jayant, M. Melchner, "Robust Speech Coding for the Indoor Wireless Channel," AT&T Technical Journal, July/August 1993, pp. 64-73.

[49] R. M. Gray, D. S. Ornstein, "Sliding-Block Joint Source/Noisy-Channel Coding Theorems," IEEE Tr. on Information Theory, vol. IT-22, no. 6, November 1976.

[50] R. M. Gray, "Vector Quantization," IEEE ASSP Magazine, April 1984, pp. 4-29.

[51] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes and their Applications," IEEE Tr. Communications, vol. 36, no. 4, April 1988, pp. 389-400.

[52] J. Hagenauer, P. Hoeher, "A Viterbi Algorithm with Soft-Decision Outputs and its Applications," IEEE Globecom, 1989, pp. 47.1.1-7.

[53] J. Hagenauer, N. Seshadri, C. W. Sundberg, "The Performance of Rate Compatible Punctured Convolutional Codes for Digital Mobile Radio," IEEE Tr. Communications, vol. 38, no. 7, July 1990, pp. 966-980.

[54] A. Haoui, D. Messerschmitt, "Embedded Coding of Speech: A Vector Quantization Approach," Intl. Conference on Acoustics, Speech, and Signal Processing, 1985, pp. 1703-1706.

[55] J. Hardwick, J. Lim, " The Application of the IMBE Speech Coder to Mobile Communications," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1991, pp. 249-252.

[56] A. Jain, Digital Image Processing, Prentice Hall, Englewood Cliffs, NJ, 1989, pp. 99-119.

[57] N. S. Jayant, V. B. Lawrence, D. P. Prezas, "Coding of Speech and Wideband Audio," AT&T Technical Journal, September/October 1990, pp. 25-41.

[58] N. S. Jayant, "Signal Compression: Technology Targets and Research Directions," IEEE Journal on Selected Areas in Communications, vol. 10, no. 5, June 1992, pp. 796-818.

[59] N. S. Jayant, J. Johnston, R. Safranek, "Signal Compression Based on Models of Human Perception," Proceedings of the IEEE, vol. 81, no. 10, October 1993, pp. 1385-1422.

[60] N. S. Jayant, P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[61] D. Jeong, J. Gibson, "Uniform and Piecewise Uniform Lattice Vector Quantization for Memoryless Gaussian and Laplacian Sources," IEEE Tr. Information Theory, vol. 39, no. 3, May 1993, pp. 786-804.

[62] Z. Jiang, A. Alwan, A. N. Willson Jr. "High-Performance IIR QMF Banks for Speech Subband Coding.", Proc. IEEE International Symposium on Circuits and Systems, June 1994, pp. 493-496.

[63] J. D. Johnston, "A Filter Family Designed for Use in Quadrature Mirror Filter Banks," IEEE International Conference on Acoustics, Speech, and Signal Processing, 1980, pp. 291-294.

[64] J. D. Johnston, "Transform Coding of Audio Signals Using perceptual Noise Criteria," IEEE Journal On Selected Areas In Communications, vol. 6, no. 2, February 1988, pp. 314-323.

[65] S. Kallel, "Complementary Punctured Convolutional (CPC) Codes and their Applications," IEEE Tr. Communications, vol. 43, no. 6, June 1995, pp. 2005-2009.

[66] V. Karanam, K. Sriram, D. O. Bowker, "Performance Evaluation of Variable Bit Rate Voice in Packet Switched Networks," AT&T Technical Journal, September/October 1988, pp. 57-69.

[67] N. Kitawaki, M. Honda, K. Itoh, "Speech Quality Assessment Methods for Speech Coding Systems," IEEE Communications Magazine, October 1984., pp. 26-33

[68] N. Kitawaki, H. Nagabuchi, "Quality Assessment of Speech Coding and Speech Synthesis Systems," IEEE Communications Magazine, October 1988, pp. 36-44.

[69] P. Kroon, E. F. Deprettere, R. J. Sluyter, "Regular Pulse Excitation: A Novel Approach to Effective and Efficient Multipulse Coding of Speech," IEEE Tr. Acoustics, Speech, Signal Processing, vol. ASSP-34, October 1986, pp. 1054-1063.

[70] V. O. K. Li, X. Qiu, "Personal Communication Systems (PCS)," Proceedings of the IEEE, vol. 83, no. 9, September 1995, pp. 1210-1243.

[71] Y Li, B. Vucetic, Y. Sato, "Optimum Soft-Output Detection for Channels with Intersymbol Interference," IEEE Tr. Information Theory, vol. 41, no. 3, May 1995, pp. 704-713.

[72] Y. Linde, A. Buzo, R. M. Gray, "An Algorithm for Vector Quantizer Design," IEEE Tr. Communications, vol. COM-28, Jan. 1980, pp. 84-95.

[73] S. P. Lloyd, "Least Squares Quantization in PCM," IEEE Tr. Information Theory, unpublished memorandum, Bell Laboratories, 1957.

[74] S. P. Lloyd, "Least Squares Quantization in PCM," IEEE Tr. Information Theory, vol. IT-28, March 1982, pp. 129-137.

[75] T. J. Lynch, Data Compression Techniques and Applications, Van Nostrand Reinhold, 1985.

[76] H. H. Ma, J. K. Wolf, "On Tail Biting Convolutional Codes," IEEE Tr. Communications, vol. COM-34, no. 2, February 1986.

[77] B. Masnick, J. Wolf, "On Linear Unequal Error Protection Codes," IEEE Tr. Information Theory, vol. IT-3, no. 4, October 1967, pp. 600-607.

[78] J. Max, "Quantizing for Minimum Distortion," IRE Tr. Information Theory, vol. IT-6, Mar. 1960, pp. 7-12.

[79] J. W. Modestino, D. G. Daut, "Combined Source-Channel Coding of Images," IEEE Tr. Communications, vol. COM-27, no. 11, November 1979, pp. 1644-1659.

[80] D. L. Neuhoff, D. H. Lee, "On the Performance of Tree-Structured Vector Quantization," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1991, pp. 2277-2280.

[81] P. Noll, R. Zelinski, "Bounds on Quantizer Performance in the Low Bit-Rate Region," IEEE Tr. Communications, vol. COM-26, February 1978, pp. 300-304.

[82] P. Noll, "Wideband Speech and Audio Coding," IEEE Communications Magazine, November 1993, pp. 34-44.

[83] H. Nyquist, "Certain Topics in Telegraph Transmission Theory," Trans. AIEE, vol. 47, April 1928, pp. 617-644.

[84] J. E. Padgett, C. G. Gunther, T. Hattori, "Overview of Wireless Personal Communications," IEEE Comm. Magazine, January 1995, pp. 28-41.

[85] S. Quackenbush, "A 7 KHz Bandwidth, 32 kbps Speech Coder for ISDN," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1991, pp. 1-4.

[86] L. R. Rabiner, R. W. Schafer, Digital Processing of Speech Signals, Prentice Hall, Englewood Cliffs, NJ, 1978.

[87] N. Rydbeck, C. W. Sundberg, "Analysis of Digital Errors in Non-Linear PCM Systems," IEEE Tr. Communications, vol. COM-24, no. 1, January 1976, pp. 59-65.

[88] D. L. Schilling, "Wireless Communications Going Into the 21st Century," IEEE Tr. on Vehicular Technology, vol. 43, no. 3, August 1994, pp. 645-652.

[89] M. Schroeder, "Linear Predictive Coding of Speech: Review and Current Directions," IEEE Communications Magazine, August 1985, pp. 54- 61.

[90] N. Seitz, S. Wolf, S. Voran, R. Bloomfield, "User Oriented Measures of Telecommunication Quality," IEEE Communications Magazine, January 1994, pp. 56-66.

[91] N. Seshadri, C. Sundberg, V. Weerackody, "Advanced Techniques for Modulation, Error Correction, Channel Equalization, and Diversity," AT&T Technical Journal, July/August 1993, pp. 48-63.

[92] N. Seshadri, C. W. Sundberg, "List Viterbi Decoding Algorithms with Applications," IEEE Tr. Communications, vol. 42, nos. 2/3/4, February/March/April 1994, pp. 313-323.

[93] C. E. Shannon, "A Mathematical Theory of Communications," Bell System Technical Journal, vol. 27, 1948, pp. 379-423, 623-656.

[94] C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," IRE Nat. Conv. Rec., Pt. 4, 1959, pp. 142-163.

[95] M. H. Sherif, D. O. Bowker, G. Bertocci, B. A. Orford, G. A. Mariano, "Overview of CCITT Embedded ADPCM Algorithms," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1990, pp. 1014-1018.

[96] A. Shen,. "Perceptually-based subband coding of speech signals," Unpublished Master's thesis, Dept. of Electrical Engineering, UCLA, June 1994.

[97] A. Shen, B. Tang, A. Alwan, G. Pottie, "A Robust Variable Rate Speech Coder," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95, Detroit MI, May 1995, pp. 249-252.

[98] H. Shi, K. M. Ho, "Combined Speech and Channel Coding for Mobile Radio Communications," IEEE Tr. Vehicular Technology, vol. 43, no. 4, November 1994, pp. 1078-1087.

[99] S. Singhal, D. LeGall, C. Chen, "Source Coding of Speech and Video Signals," Proceedings of the IEEE, vol. 78, no. 7, July 1990.

[100] A. Spanias, "Speech Coding: A Tutorial Review," Proceedings of the IEEE, vol. 82, no. 10, October 1994, pp. 1541-1582.

[101] R. Steele, "Speech Codecs for Personal Communications," IEEE Communications Magazine, November 1993, pp. 76-83.

[102] B. Tang, G. Pottie, "Embedded Non-Uniform Scalar Quantizer for Variable Rate Applications," International Conference on Signal Processing Applications and Technology, October 1994, pp. 456-460.

[103] B. Tang, A. Shen, G. Pottie, A. Alwan, "Spectral Analysis of Subband Filtered Signals," IEEE International Conference on Acoustics, Speech, and Signal Processing, May 1995, pp. 1324-1327.

[104] B. Tang, A. Shen, A. Alwan, G. Pottie, "A Perceptually-Based Embedded Subband Speech Coder," unpublished.

[105] B. Tang, G. Pottie, A. Alwan, "Adaptive Wireless Voice Communication with Embedded Source and Channel Coding," unpublished.

[106] B. Tang, G. Pottie, A. Alwan, "Progressive Transmission with Embedded Source and Channel Coding," unpublished.

[107] P. P. Vaidyanathan, "Quadrature Mirror Filter Banks, M-Band Extensions and Perfect Reconstruction Techniques," IEEE ASSP Magazine, July 1987, pp. 4-20.

[108] P. P. Vaidyanathan, "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial," Proceedings of the IEEE, vol. 78, no. 1, January 1990, pp. 56-93.

[109] P. P. Vaidyanathan, Multirate Systems and Filter Banks, Prentice Hall, Englewood Cliffs, NJ, 1992.

[110] P. Vary et al., "Speech codec for the European Mobile System," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, April 1988, p. 227.

[111] S. Vembu, S. Verdu, Y. Steinberg, "The Source-Channel Separation Theorem Revisited," IEEE Tr. Information Theory, vol. 41, no. 1, January 1995.

[112] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotic Optimum Decoding Algorithm", IEEE Tr. Information Theory, vol. IT-13, April 1967, pp. 260-269.

[113] A. J. Viterbi, "The Evolution of Digital Wireless Technology from Space Exploration to Personal Communication Services," IEEE Tr. on Vehicular Technology, vol. 43, no. 3, August 1994, pp. 638-643.

[114] Q. Wang, V. K. Bhargava, "An Efficient Maximum Likelihood Decoding Algorithm for Generalized Tail Biting Convolutional Codes Including Quasicyclic Codes," IEEE Tr. Communications, vol. 37, no. 8, August 1989, pp. 875-879.

[115] S. Wang, A. Sekey, A. Gersho, "An Objective Measure for Predicting Subjective Quality of Speech Coders," IEEE Journal On Selected Areas In Communications, vol. 10, no. 5, June 1992, pp. 819-829.

[116] I. Wassell, D. Goodman, R. Steele, "Embedded Delta Modulation," IEEE Tr. on Acoustics, Speech, and Signal Processing, vol. 36, no. 8, August 1988, pp. 1236-1243.

[117] S. W. Wong, "An Evaluation of 6.4 kbps Speech codecs for INMARSAT-M System," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1991, pp. 629-632.

[118] R. C. Wood, "On Optimum Quantization," IEEE Tr. Information Theory, vol. IT-5, March 1979, pp. 248-252.

[119] Y. Yasuda, K. Kashiki, Y. Hirata, "High Rate Punctured Convolutional Codes for Soft Decision Viterbi Decoding," IEEE Tr. Communications, vol. COM-32, March 1984, pp. 315-319.

[120] P. Zador, "Topics in Asymptotic Quantization of Continuous Random Variables," unpublished memo, Bell Laboratories, 1966.

[121] P. Zador, "Asymptotic Quantization Error of Continuous Signals and the Quantization Dimension," IEEE Tr. Information Theory, vol. IT-28, March 1982, pp. 139-149.

[122] K. Zeger, A. Gersho, "Pseudo-Gray Coding, "IEEE Tr. Communications, vol. 38, no. 12, December 1990, pp. 2147-2158.

[123] S. Zhang, "An Embedded Scheme for Regular Pulse Excited (RPE) Linear Predictive Coding," IEEE Intl. Conference on Acoustics, Speech, and Signal Processing, 1995, pp. 37-40.

[124] Proposed EIA/TIA Interim Standard IS-95, Appendix A, April 1992.