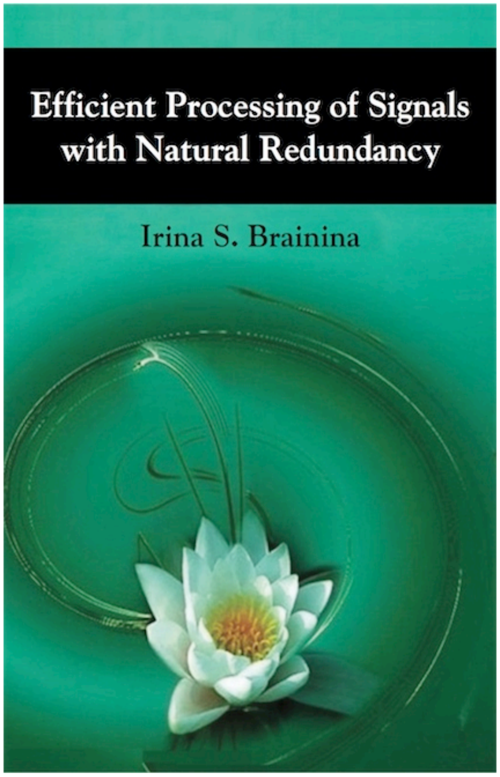


This book offers new effective digital algorithms for processing audio and tonal signals, having a high natural redundancy. The redundancy preserved in headers of compressed audio frames in MPEG-1 Layer 2 format. Computer modeling proves an opportunity to detect and correct all catastrophic errors in three fields of a frame header.

Efficient Processing of Signals with Natural Redundancy

by Irina S. Brainina

Efficient Processing of Signals with Natural Redundancy

Irina S. Brainina

ISBN: 978-1-63492-961-5

Published by BookLocker.com, Inc., St. Petersburg, Florida.

Printed on acid-free paper.

BookLocker.com, Inc.
2018

First Edition

Contents

Preface

This monograph looks at adaptive methods for processing and recognition of signals with natural redundancy. Speech, music and tones are examples of such signals. One of the effects of redundancy is that it provides for higher noise immunity in digital speech transmission systems that transmit pulse-code modulation (PCM) speech over communication channels with a high level of noise.

High natural redundancy of audio signals is present in both time, and frequency domains. In addition, mono-tone and dual-tone signals have distinctive shapes that are different from that of noise.

Redundancy in time domain is especially high in PCM signals, which makes abnormal bit errors in higher order digits of a non-linear code easy to recognize and correct in real time (such errors are perceived as annoying pulse noise).

When a multiple PCM signal is processed in a discrete multichannel system where error bursts in a multiplex group turn into single errors in each voice channel, error correction becomes very easy due to the natural interleaving of multiplex signal samples.

Redundancy is also characteristic of mono-tone and dual-tone tone-dialing signals with adaptive delta modulation (ADM) or adaptive differential pulse-code modulation (ADPCM). Their analog waveform in the time domain is very different from that of other random signals. Due to their quasi-sinusoidal nature, the level of dual-tone signals changes more rapidly near zero crossings and more slowly near extremums.

Knowing the signal shape allows correction of sign bit errors in a stream of identical digits at the delta modulation decoder input. Similarly, it is possible to correct errors in 4-digit ADPCM codes at the decoder input. Such correction results in a higher noise immunity of dial-tone signals in electronic automatic exchanges using ADM or ADPCM.

With the information of the signal waveform available it is easy to recognize the content transmitted (e.g. speech, music, tone signals). The concept of the so-called 'interval spectrum' is introduced for digital signals in the time domain, which is similar to that of a frequency spectrum in the frequency domain.

The distribution law for the number of quantized intervals between successive zero crossings is analyzed. The sign of any particular sample is indicated by the most significant bit in the code. Once it is known, you can roughly estimate the frequency spectrum width and reliably detect the signal type (i.e. speech, music etc.).

Broadband speech and music signals are characterized by a wide range of intervals between successive zero crossings (the range includes both very short and very long intervals). Narrowband mono-tone and dual-tone signals, on the other hand, have a limited ‘spectrum’ of zero-crossing intervals.

Once sign bits are known, it is easy to tell whether the stream of digits represents speech or music. The frequency spectrum of music signals is usually several times as wide as that of speech signals. At the same time, music signals are characterized by a wider ‘spectrum’ (i.e. variety) of zero-crossing intervals compared to speech signals: long intervals are found in both, but only music demonstrates the presence of a large number of short intervals between zero crossings. The distribution of zero-crossing intervals can also be used to determine whether the voice dealt with is male, or female/childish. In a signal representing a high female or childish voice short zero-crossing intervals prevail over long ones.

Strange as it may seem, even with the modern effective audio signal compression format MPEG-1 which is used in digital radio broadcasting and digital TV sound broadcasting, the headers of compressed frames are characterized by notably high redundancy. Audio signal samples themselves are strongly compressed in the frequency domain and practically do not correlate with each other, but the headers of two subsequent compressed frames repeated about 38 times per second, are not much different from each other. As the sound cannot change significantly over a short period of time (slightly over 26 ms), instantaneous spectra in the same frequency band or adjacent bands undergo little changes. It should be noted that, when a header contains the so-called catastrophic digital errors (a certain type of single bit errors), this may result in the loss of the entire sound frame, which, in this case, is simply turned into noise.

A catastrophic error in the header of a particular frame is relatively easy to detect, considering that the previous frame and its header are always recorded in the decoder’s memory: it is sufficient to compare the two headers and look at the difference between them. Once detected, the error can be corrected in the decoder, using the slight redundancy in the frame.

High natural redundancy of speech makes it possible to implement a simple algorithm of digital speech compression in time domain. Using one algorithm described in this book one can compress speech data to one thirtieth of their size as long as one is prepared to sacrifice high quality of sound and does not have to be able to identify the speaker by their voice, in other words, in the case where discernibility of words is the only thing that matters. By making this algorithm slightly more complicated one can improve the quality of sound at the cost of lower compression ratio.

Redundancy of audio signals also allows for statistical multiplexing of a broadband digital communication channel: the multiplexing is achieved by adapting the PCM signal sampling rate to the signal spectrum width. The contents of radio programs broadcast (e.g. speech, pop music, classical symphony, instrumental or vocal music performances) determines changes in the signal waveform: the wider the signal spectrum, the faster the changes and the higher the first derivative of the waveform.

In some cases a change in the content of radio programs allows for a several-fold reduction in the required frequency bandwidth, which means that the sampling rate can be reduced accordingly and that samples of transmitted audio signal can be interleaved with other information.

Audio signals are not the only ones that involve redundancy. One example of highly redundant data is medical data transmitted over radio communication channels. Such data include cardiograms, encephalograms and tomograms. When transmitted over low-quality channels in the presence of high noise, such data are sometimes distorted. However, due to quasiperiodicity and redundancy of signals, the distortions can be detected end corrected.

It is obvious that any additional information on the signal waveform and its derivatives, the distribution of zero-crossing intervals and intervals between extremums, and the spectrum width (the correlation interval) is quite useful for digital processing. As long as the signal in the telecommunication channel is stronger than noise (albeit only slightly), the most outstanding errors can be detected and corrected.

It is worth noticing that the author of this monograph was strongly influenced by some useful ideas suggested by **professor A. M. Zayezdny**, who opened a new field for research by setting forth in mid-1960's ***The Foundations of Signal Division and Measurement Based on Their Structural Properties.*** By structural properties of a signal he meant the relations between the waveform and its linear transformations – derivatives and integrals. Using those relations, (for the high signal-to-noise ratio condition) enabled the implementation of a number of new effective signal processing algorithms.

As we showed in this book, even in the case where the signal-to-noise ratio is low, the preliminary information on the signal's structural properties combined with its natural redundancy allows us to make the signal processing much less sensitive to noise.

This monograph shows that redundancy can be advantageous not only in signals, but also in multipath propagation which affects radio waves of HF and VHF bands, as well as UHF and SHF bands allocated for satellite communication. Multipath propagation of radio waves causes their interference at the receiver side, thus inducing fading. Radio signals carrying the same information reach the receiving antenna by two or more paths with a delay in between. When they occur in different parts of the spectrum at the same time, such delays may result in selective fading.

On the one hand, multipath propagation makes signal processing more difficult. On the other hand, it creates natural redundancy of information, which is transmitted over different paths and thus repeated. This property of multipath propagation is used by designers of adaptive systems that can detect unaffected parts of the frequency spectrum by testing continually the channel condition. During intervals between tests the transmitter power is efficiently distributed among the allocated frequency band parts that are not being affected by fading or interference.

The book looks at an adaptive communication system of the above mentioned kind proposed by the author in late 1960s, which provides for maneuvering within the frequency band allocated for a shortwave channel. The concept is still relevant, as it can be used for building more efficient digital systems of mobile telephone communication.

This monograph contains the results of many years of the author's work in the Research Department of Samara Institute of Telecommunications, presently the Povolzhye State University of Telecommunications and Informatics (PSUTI). The research was, for the most part, carried out at the chairs of multi-channel telecommunication and theoretical foundations of radio engineering and communication (TFRC).

The author is heartily grateful to her unforgettable teacher D. D. Klovsky, D. Sc. (engineering), full professor, Head of Chair (TFRC), for his interest in this work and continuous support.

Big thanks go to Dmitry Arch for translating this book into English. His meticulous and conscientious efforts helped bring the book to a large English-speaking audience.

My thanks go to associate professor Oleg Petrov, a talented programmer, who designed the software for simulating the process of correcting errors in the headers of compressed audio frames in MPEG-1 Layer-2 format. The software designed by Oleg Petrov corrects introduced errors and estimates the gain in the signal-to-noise ratio.

I am also grateful to my former post-graduate student Stanislav Gladyshev for designing programs for simulating the process of introducing errors at a predetermined rate into different fields in the header of a compressed audio frame, and detecting those errors afterwards.

Irina S. Brainina

Introduction

This monograph describes some new adaptive algorithms for receiving and processing random and deterministic signals. These algorithms exploit the natural redundancy of signals.

Some new algorithms help detect and correct single abnormal digital errors in audio signals transmitted using pulse-code modulation (PCM), whereas others solve problems such as distinguishing between various types of audio signals such as speech, music, harmonic mono-tone and dual-tone signals, detecting and correcting catastrophic errors in the headers of compressed sound signal frames in MPEG-1 Layer-2 format, compressing speech in the time domain and a number of other problems.

The first chapter studies in detail the unique algorithms for detection and correction of three types of catastrophic errors in the headers of compressed sound frames in MPEG-1 Layer-2 format (this format is widely used in digital radio and TV broadcasting). It is believed that, in most cases, such errors can only be detected, but not corrected, and for them to be detected, the bit error rate of a satellite channel must not be higher than one per 100,000 bit.

As a rule, errors in the MPEG decoder input turn out to be single due to interleaving of binary digits at the encoder output. After de-interleaving, burst of errors that occur in a satellite communication channel turn into single errors, which can be detected with a code CRC-16.

In the case where a single catastrophic error resulting in the loss of the entire frame is detected, the corrupted sound frame is usually replaced with the preceding frame stored in the decoder memory. In the case of two catastrophic errors, the decoder cuts off the sound at the output until the required quality of communication is restored. On the receiving end this produces the ‘stutter’ effect.

The presence of undetected single catastrophic errors in the header immediately becomes obvious; they grate on the ear. The listener can hear clicks or “stutter”, or the sound is going in and out. Catastrophic errors include the ones in the field that contains the number of bits per sample for each frequency band, the ones in the field that indicates the number of scaling ratios (one to three) in a particular band, and the ones in the first three most significant bits of the logarithmic code that is used to encode the ratios whose dynamic range exceeds 124dB.

The first type of errors may result in the loss of certain spectral bands or appearance of new spectral bands. Errors of the first type also change the frame length, making frames longer or shorter than they are under normal conditions. In addition to that, an error of either of the first two types in any particular frequency band results in misinterpretation of sound data in all subsequent bands, in which case the entire frame is decoded incorrectly and is perceived as noise.

This monograph proposes an indirect method for detecting catastrophic errors of the first two types. The method is based on the difference in the distribution of sample levels in corrupted and uncorrupted frames. The distribution of momentary values for sound signal samples is close to exponential. For the most part, relative values for samples in each frequency band are quite small, whereas high values are rare. A corrupted frame looks like a sequence of noise; for any particular band the likelihood of a sample having a certain value is the same, and the distribution of momentary values is close to a rectangular one.

It is suggested that the signal peak factor, i.e. the ratio of maximum sample level to the average modulus of samples for several high-frequency bands in the highest part of the spectrum, should be monitored. If the average value for the peak factor is significantly underestimated, this will be an indication of the presence of a catastrophic error. The effectiveness of the above method has been proved by numerous computer simulations.

Big errors in scaling ratios are less dangerous, as they do not result in the loss of the entire frame. On the other hand, they are also conspicuous and ear-grating as they cause abrupt changes of the signal level in corrupted bands, perceived as clicks.

The author carried out numerous experiments with computer simulations using quite lengthy sound sequences (files) to represent music or speech. The relative number of digital errors introduced into frame headers was increased to a value exceeding the standard acceptable maximum by a factor of 100 (one error in a 1000-bit sequence). Quite simple algorithms have been worked out which, due to the natural redundancy of sound sequences, allow detecting and correcting (by means of inversion of the false bit) all most dangerous catastrophic errors. The effect of applying those algorithms is equivalent to a nearly 20dB gain in signal-to-noise ratio. Error detection was based on using the slight redundancy of the frame, whose format remains unchanged. A nearly 20dB gain allows for a significant relaxation in satellite communication quality and cost requirements (satellite communication is known to be expensive), and the use of less sophisticated equipment.

The findings presented **in the Appendix 1** and **Appendix 2**, which were obtained with computer simulations, allow evaluating the gain in the resistance of compressed audio frame to noise.

The implementation of the method proposed for detection and correction of catastrophic errors in the headers of compressed audio frames requires using two relatively simple digital boards: one at the encoder output, the other at the MPEG decoder input.

Without changing the frame format, it is necessary to place check bits in an error-correcting Hamming code with parity checks in the positions of 16 bits used for CRC-16 codes. It is also necessary to use the 80 bits that follow each audio frame and allow transmitting supplementary data at a low bit rate. According to recommendations given in [100], these bits can be used, among other things, for making the audio frame less susceptible to interference. Out of the total 96 security bits 24 is used for detecting errors in Bit Allocation field, another 22 bits is used for protecting data on scaling ratios number, and 32 bits will be used for protecting data on the values of scaling ratios. The remaining 18 bits can be used for protecting the most relevant data transmitted in the first 32 bits of the frame header. These 32 bits include twelve “1” bits of the synchronizing sequence and one bit indicating whether the frame has or has not been shortened by one byte.

It should be noted that a few bits at the beginning of the header are redundant as their values remain the same for all frames (in other words, those bits do not bear any new information) or they have a standby status. Those bits can also be used to better protect the frame from errors.

As simulations show, errors in digital sound samples are practically inaudible. Even if bit error probability increases hundredfold, errors occur, on average, in three out of 15 to 20 frequency bands used, resulting in corruption of only one sample out for in each band (in the frequency domain). For noticeable sound distortion to happen bit error probability has to increase thousand fold and there has to be one or two corrupted samples in each band.

This finding is very important, considering that it is impossible to correct errors in digital samples themselves.

Another interesting result is that errors in three least significant bits of the 6-digit logarithmic code used for scale ratios remain unnoticed. This is even more important than the inaudibility of errors in sound samples.

Due to residual noise the actual gain in signal-to-noise ratio is likely to be a little lower than expected (20 dB instead of 25 dB) and we can expect bit error probability to increase 30-fold rather than 100-fold. This is equivalent to the reduction in bit error probability from 10^{-3} before error correction to about $3*10^{-4}$.

It is suggested that a special microprocessor should be placed at the MPEG audio decoder input, which will use given algorithms for rapid processing of the current compressed audio frame pre-recorded to RAM.

After detecting and correcting catastrophic errors in the frame header the compressed audio data is decoded and presented to the user.

Chapter 2, in Section 2.1 presents one method for increasing noise immunity of audio signal reception in communication systems with non-linear PCM. We managed to reduce discernibility to a human ear of pulse interference (crackling noise) caused by abnormal digital errors (i.e. errors in 3 or 4 most significant bits of a non-linear 8-bit code).

The error-detecting method is based on monitoring two opposite jumps of the signal first derivative beyond two adaptive thresholds. The thresholds with opposite signs are proportional to the average modulus of the signal first derivative, calculated for the previous measurement interval. When an error is detected, the current sample of the signal is replaced with the previous one with the change hardly noticeable to the listener.

Computer simulation was extensively used for experimenting with five male and three female voices. The relative number of bit errors inserted was 10,000 times the normal, which is roughly equivalent to 4dB reduction in signal-to-noise ratio. After the correction of abnormal errors the sound quality did not deteriorate significantly; undetected errors in the less significant bits of the code were perceived as slight ‘rustling’.

The simulation results fully confirm the pre-calculated optimal thresholds for detecting jumps in the first derivative. By using these adaptive thresholds with the opposite signs we can minimize residual noise caused by undetected errors or wrong corrections resulting from natural fast changes of the signal level. Simulation also helped to minimize residual noise level by choosing the optimal length of the measurement interval. When the measurement interval is too short, the average modulus of the first derivative cannot be calculated accurately enough. When the interval is too long, non-stationary nature of speech signals begins to show, and the optimal threshold value obtained for the previous interval becomes obsolete. It was found that the optimal range for the measurement interval duration is 8–12 ms, i.e. the interval covers two or three periods of the main speech tone for most male and female voices.

Section 2.2 studies the distribution of momentary values of a vocalized speech signal as well as its first and second derivatives. The knowledge of these parameters allows choosing appropriate adaptive thresholds for speech signal derivatives, which are not to be exceeded during measurement intervals. Jumps beyond the set threshold in the first or second derivative function are caused by abnormal digital errors in the senior-order bits of a non-linear PCM code (when such an error occurs, it is perceived by the listener as 'crackling' noise). In the case of an abnormal digital error, the first derivative makes two consecutive jumps, one upwards and one downwards, whereas the second derivative makes only one jump which is double the size of the first two.

Section 2.3 analyses the relation between the distributions of momentary values in two speech signals: the one at the input and the one at the output of the logarithmic PCM encoder. Various kinds of random encoder input signals with flat power are considered, including the ones with rectangular, exponential and Gaussian distribution of momentary values, and a random-phase sinusoidal signal. To a great extent, the logarithmic nature of the encoder irons out the differences between encoder input signals of the same power level.

It is concluded that the distribution of momentary values in the input signal has little, if any, effect on the probability of abnormal errors or the optimal adaptive thresholds for detecting such errors. The thresholds for detecting jumps of the first derivative do not depend on the unknown distribution of the input signal: they are determined by the average modulus of the input signal and the effective width of its instant spectrum. These data can be obtained from the average modulus of the signal first derivative calculated for the previous measurement interval.

Section 2.4 presented results of modeling with PC of an adaptive method for correction of abnormal digital errors.

Chapter 3, Section 3.1 gives an approximation for the distribution of intervals between successive zeros in the average clipped speech signal. The notion of "the spectrum of intervals" is introduced by analogy with the frequency spectrum. All information on the "spectrum" of intervals is rendered by the sign bit of the PCM code. The number of intervals between successive zeros in a sampled signal is rounded to integer. The set of such numbers ("spectrum of intervals") gives an idea of how rapidly the signal changes in the time domain, and allows rough estimation of the broadband frequency spectrum width. Knowing the "spectrum" of intervals for a clipped signal allows us to distinguish between different types of signals. For example, in telephony, the frequency spectrum of mono- and dual tone signals overlaps with that of a vocalized speech. This makes it difficult to distinguish between signaling data and speech data; the frequency filtering cannot adequately protect the former from the latter.

If, however, we compare spectra of intervals for signaling data and speech, it will be easy to tell them apart. Dual tone signals are characterized by a limited periodically recurrent spectrum of intervals, whereas the speech signal has a rich "spectrum" of intervals in which some intervals are either too long, or too short to occur in the spectrum of intervals for signaling data. The presence of a sufficient number of intervals that are too short or too long is a reliable indicator of speech.

The proposed method of recognizing signals in the time domain is an example of non-linear digital filtration.

Section 3.2 devoted to the evaluation the distribution law for zero-crossing intervals a speech signal.

Chapter 4 sets forth design principles for adaptive group receivers of multitone dialing signals used in telephony (such receivers make part of interface equipment of analog and digital electronic telephone exchanges using pulse-code modulation (PCM), adaptive differential PCM (ADPCM) or adaptive delta modulation (ADM).

The signal-to-noise ratio can be improved by using at the decoder input a device for correcting errors in the random digital data stream. Each error corrector is an adaptive code converter using the apriori data on the shape of tone signals.

Tone signals are characterized by more rapid changes of their level near zero crossings and much slower changes near extremums. It becomes obvious if, for example, you look at the delta pulse stream. In the areas of fast rise or fall of the signal we see series of two or more pulses of the same sign, whereas the pulses near extremums alternate in signs. As the amplitude quantization step increases beyond a certain value, the delta pulse stream changes: series of pulses of the same sign shorten and signs alternate more regularly. As a result, we have a reduction in the noise immunity of correlation reception for a given frequency component. The delta pulse stream is corrected by inverting pulses of the opposite sign; as a result, longer series of pulses of the same sign are formed in areas of rapid rise or fall of the signal. This procedure is equivalent to reducing the amplitude quantization step, restoring the original delta pulse stream and increasing the noise immunity of the dialed digit tone.

Chapter 5 describes a simple, but effective speech compression algorithm that guarantees acceptable quality of sound. The algorithm takes advantage of high redundancy of speech data stream in the time domain.

Redundancy is especially notable in the stream representing vowels and voiced consonants. It shows through quasiperiodic repetition of nearly the same signal segments with the main speech tone frequency. Being the vocal cords oscillation frequency, it is unique for each human voice.

Each vocalized sound is represented by dozens of signal fragments similar in waveform and amplitude, whose period equals that of the main speech tone. For each sound the most typical one-period fragment patterns can be stored in memory to be repeated until the end of the locally stationary segment is reached while replaying the sound.

The sound data stream can be further compressed by reducing the ‘description’ of the chosen fragment patterns. The reduction is achieved by shortening the signal sampling code. The code can be shortened due to adaptation of amplitude quantizing step (it is chosen to be proportional to the average modulus of the signal for the measurement interval).

It is also possible to further compress the sound data stream by shortening vocalized sounds and description of pauses made both within and between words. Modeling of the compression algorithm with a PC simulation demonstrated the possibility of a 30-fold speech compression in the case where words are pronounced separately. In the experiments the words remained highly discernible, and the quality of sound scored 3 on a 5-point scale. By making the algorithm a little more complicated it is possible to further improve the quality of sound at the cost of a lesser compression degree.

The proposed algorithm can be used in devices such as the answering machine, e.g. by help desks or call centers, or in replaying audio commands when the sound quality requirement do not go beyond good discernibility and naturalness of speech.

The compression algorithm is also applicable to detecting continuous speech with pauses between some words. In this case the detector can be set to recognize a particular voice or any voice.

An algorithm for recognition of separate words is described, which is equally effective for all volume levels, voice pitch and speech tempo. The information on volume, voice pitch and speech tempo is redundant and only hampers recognition of words. By eliminating such redundancy one can achieve more reliable word recognition.

Chapter 6 discusses the possibility of statistical multiplexing based on finding the best sampling rate for the given bandwidth of the transmitted audio signal.

In digital PCM-technique-based broadcasting of radio programs, a high quality of sound can only be achieved using an expensive broadband channel. Such channels, however, are only used to their full capacity in the case where vocal or instrumental music (e.g. symphony music) is broadcast. When this is not the case, the channel can be multiplexed and used for transmitting additional information.

One example of such multiplexing is a digital broadcasting channel using multichannel PCM equipment that combines four standard 64kbit/s speech channels. The best sampling rate takes one of three values, depending on the type of sound signal broadcast: 8 kHz for speech, 16 kHz for vocal music and pop music, and 32 kHz for classical symphony music.

In real life, the entire bandwidth of the digital channel is used, the signaling rate being 256kbit/s, and the sampling rate being 32 kHz.

The information on the effective bandwidth of the broadcast signal is essential for finding the optimal sampling rate. The key information on the high-frequency components of a broadband signal can be obtained from the first derivative of the signal. A convenient method of estimating the first derivative of the discrete signal is to take as an estimate the first difference, which is directly proportional to the first derivative, i.e. the difference between two successive samples of the signal divided by the sampling interval. The ratio between the average modulus of the initial random signal and that of the first difference can be expressed through the signal self-correlation ratio taken for the sampling interval. By changing continually the values of first ratio for locally stationary intervals, one can monitor the trends in the self-correlation ratio and the effective bandwidth for different types of the broadcast signal.

It is obvious that the stronger the correlation between successive samples is, the lower average modulus of the first derivative corresponds to the average signal modulus. A decrease in the average modulus of the first derivative indicates the signal spectrum narrowing and the necessity to reduce the sampling rate once a certain threshold has been reached. Using a lower sampling rate allows for longer intervals between successive samples and a larger volume of additional information transmitted.

When the broadcast signal type changes and the signal spectrum widens, the correlation between successive samples weakens and the average modulus of the first derivative increases, causing automatically an increase in the sampling rate up to its maximum value. The increased sampling rate should be applied for the next measurement interval.

On the one hand, the measurement interval duration of 0.1–0.2 sec. is long enough to fit hundreds of non-correlated samples, which ensures sufficiently reliable figures of the average signal modulus and the average first difference. On the other hand, a delay of 0.1–0.2 s in making a decision to change the sampling rate is not significant, as the audio signal type (or the ‘genre’ of the broadcasting program) cannot change instantly.

A more efficient multiplexing scheme for the source traffic in a broadband channel can be obtained by adding an audio silence detector at the transmitter side. Two algorithms can be used for audio silence detection: one is based on comparing amplitudes, whereas the other is based on distinguishing different signal waveforms.

In the first case the signal modulus averaged for the measurement interval is compared with the threshold value corresponding to the highest noise level allowed in the channel. This comparison indicates whether the channel is occupied or not. If it is not occupied, i.e. a long audio silence is detected; the channel is then fully designated for the transmission of additional data. If, however, no silence in broadcasting is detected, then we estimate, at the intervals of 0.1 to 0.2 sec., the signal’s power spectrum and the optimal sampling rate, using the above technique.

Another technique of detecting broadband noise during pauses in broadcasting is based on the indirect monitoring of correlation between successive samples of the signal at the maximum sampling frequency.

Considering that the broadband noise power spectrum is nearly homogeneous within the channel bandwidth, correlation between successive samples of the noise is likely to be weak. At the same time, in the presence of the broadcast signal with the widest frequency band and a non-homogeneous power spectrum successive samples of the signal taken at the maximum sampling rate should be strongly correlated with each other. It was estimated that, in the presence of a broadcast signal, the ratio between the average moduli of the samples and the first difference increases 3 to 4-fold compared to the value calculated for the case where the noise is present.

Finally, measuring the intensity of sign changes (i.e. zero crossings) in the signal at the maximum sampling rate enables detection of noise or silence periods. Extensive use of PC for simulating long sections of speech represented by several male and female voices brought us to the conclusion that the intensity of sign changes for noise samples during silence periods is at least five times as high as the intensity of sign changes for speech samples, especially in the case of broadcasting low-pitched male voices.

The proposed digital telecommunication system with statistical multiplexing appears to be economical and relatively easy to implement, especially with the use of an existing digital broadcasting channel created in the operating PCM equipment by combining four standard telephone channels. The system development and production costs may be quickly recovered from operating an expensive broadband telecommunication channel.

Unlike the previous chapters, **the seventh chapter** of the book is dedicated to exploiting the natural redundancy of multipath channels rather than signals themselves. Multipath channels are characterized by frequency-selective fading. Multipath propagation is common for communication in short-wave (SW) and ultra-short-wave (USW) ranges where various paths of the same signal experience different delays. Due to interference the signal fading becomes selective within the frequency band used for transmission. Certain parts of the spectrum are subject to fading, whereas no fading is currently observed in its other parts.

Considering that radio waves in all propagation paths reflect from various layers of ionosphere (or troposphere) and carry the same information, one may talk natural redundancy as an inherent feature of a multipath channel. This redundancy can be exploited in both time and frequency domains.

The book describes an adaptive communication system proposed by the author, which allows maneuvering in the frequency bandwidth (MFB). A mechanism of using K out of L (L>K) narrow-band data-transmission channels (here L is the highest feasible number) for transmitting over a multiplexed radiotelephone channel is described. The number K is subject to change depending on the channel condition.

Continual probing of the SW channel is used to reveal the part of its spectrum affected by selective fading, and the K channels are designated in the remaining part. Instead of evenly distributing its power across all L channels of the assigned frequency bandwidth, the transmitter distributes it across the K channels that do not experience fading.

As most errors in data transmission occur during signal fading intervals characterized by a sharp decrease in signal-to-noise ratio, the use of the proposed communication system will make reception a lot more resistant to noise. Based on the probing results, the number K can be adjusted to the channel condition, making it possible to keep the bit rate more or less stable throughout the transmission and reducing the bit error probability at the receiver side.

Despite the fact that since the advent of new broadband channels the role of shortwave communication has declined, MFB-type communication systems similar to the one proposed in this book retain their relevance and significance. We are talking about satellite channels characterized by the same phenomena as decameter range channels. Radio waves transmitted in decimeter and centimeter ranges are also affected by channel-selective fading, allowing for the use of unaffected parts in the bandwidth.

Recently there has been a rising tide of interest in the development, on a new basis, of adaptive communication systems that can be used for mobile communication. Such systems will help increase the capacity of expensive satellite channels and the number of subscribers to any particular frequency band without impediment to the quality or reliability of communication.

The **Appendix 1** and **Appendix 2** presents a software package allows simulating introducing, detection and correction of three types of catastrophic errors in the headers of compressed sound frames in MPEG-1 Layer-2 format.

As a bonus, the software package implementing those algorithms along with 12 compressed audio files that represent different types of content can be downloaded for free from **Appendix 2**

https://www.dropbox.com/sh/s6ll5nxyga6vrnu/AACsuq8ehSK67adZZoctDD_-a?dl=0

This book is addressed to researchers, practicing engineers and graduate students, interested in adaptive digital processing of telecommunication signals with natural redundancy.

Chapter 1 Detecting and correcting catastrophic errors in the headers of compressed frames in MPEG-1 Layer-2 format

1.1 Formulation of the problem. Using statistics on relations between headers of two adjacent frames for detection and correction of errors

The audio data compression algorithm described in ISO/MPEG, which is used in broadcasting television programs MPEG-1(Layer 2) and is also implemented in equipment for satellite digital audio broadcasting (DAB), is based on the principle of differentiated error protection [100]. In line with this principle, the header of each segment/frame in a compressed audio signal is protected from digital errors by an error-correcting code known as Cyclic Redundancy Check (CRC) while the rest of the frame remains unprotected. After detecting a fatal error in the header, the decoder repeats the previous frame instead of using the current one that contains errors. If errors are found in the headers of the next frame or subsequent frames, the decoder mutes the sound of its output. As long as the bit error ratio (BER) at the decoder input does not exceed 10^{-5} a human ear 99does not perceive any distortions in the audio signal, because the affected frame is replaced with the previous one no more than once a minute. However, an increase in error ratio in the satellite digital communication channel may result in the need to replace distorted frames much more often: thus, with the increase of BER to 10^{-3}, the rate of replacement of distorted frames increases to nearly two times per second. It is obvious that, in this case, short mutes at the decoder output will often occur, resulting in the unacceptably poor quality of sound.

The redundancy introduced into the frame header by means of a 16-bit CRC is rather small and is mainly aimed at detecting and correcting some rare single errors in high quality digital communication channels. In low-quality channels with high noise level, correcting code redundancy is too small. An increase in redundancy reduces the digital audio signal transmission rate unless a wider communication bandwidth is used. Reduced transmission rate implies a loss in the quality of sound reproduction, whereas using a wider band increases the cost of communication channel lease.

Despite the fact that MPEG is a compression format and therefore, by definition, redundancy in the digital stream at the encoder output should be insignificant, it is still found in recurrent frame headers. In the process of compressing, the audio spectrum is divided into 32 frequency subbands, each

500 to750 Hz wide. In view of the fact that the processing of one frame by the encoder takes a long time (50 ms or longer) and because of the need to transmit sound through the communication channel in real time, information on the previous frames is not used in frame processing. We can, therefore, talk about redundancy resulting from strong correlation between sufficiently short successive frames and closely spaced frequency subbands within a given frame. This natural redundancy can be used to detect and fix the most dangerous (catastrophic) errors in the frame header.

Figure 1.1.1 shows the structure of a frame formed by the encoder. The digital PCM stream at the audio input of an MPEG encoder is divided into segments (frames), each containing 1,152 samples. The number of samples is constant, so the duration of one frame depends on the sampling frequency of the input PCM stream. The encoder uses one of the following three sampling frequencies: 48 kHz, 44.1 kHz or 32 kHz. The frame length varies from 24 to 36 ms, which means that the width of frequency subbands varies from 750 down to 500 Hz. For each frame the encoder processor calculates the input signal spectrum and masking border (the latter serves as a psychoacoustic model of a human ear) [100].

Preamble 32 bits	CRC-code 16 bits	Bit Allocation 94 bits	SCFSI 2 bits per subband	Scale Factors 6 to18 bits per subband	Subband Samples 96 groups of 12 samples (each sample represented by 1.67 to 16 bits)

Fig.1.1.1 Frame structure typical of MPEG-1 Layer 2

The 32 bits of the preamble include 12 bits of the synchronizing sequence, bits representing the sample frequency and bit rate, one bit indicating that the frame has been truncated, and some other data. It is clear that, in a recurrent header, about 50% of all bits are redundant (in fact, all bits are redundant except for the synchronizing sequence and important data on the frame length).

The preamble is followed by 16 protective bits of the CRC-code. The Bit Allocation field contains the numbers of subbands used for transmission and the number of bits per sample allocated for each subband. Scale Factor Selection Information (SCFSI) indicates how many (one, two or three) scale factors (each represented by 6 bits) are used in a given frequency subband. Scale factors (SCF) approximate the envelope of the instantaneous frequency spectrum characterizing the frame.

Errors affecting different parts of the frame unequally affect the quality of the decoded signal. Errors in those header bits that define the position of particular data fragments and the values of scale factors are called catastrophic, as they are especially perceptible to a human ear [100-102].

The most severe catastrophic errors are found in the Bit Allocation field, which contains the key to the interpretation of the entire frame. The change of any bit (senior or junior) in this field results in subband samples misinterpretation: audio signal samples will be "mixed" randomly in the process of decoding. This, in turn, will lead to distortion of the entire frame, a kind of distortion that cannot pass unnoticed.

A similar situation arises in the case of an error in the SCFSI field.

Table 1.1.1 [100] shows how sensitive an audio frame is to errors in individual bits; the immunity to errors is evaluated on a scale from 0 to 5, indicating the degree of sound quality deterioration caused by one single error: 5 – catastrophic, 4 – very annoying, 3 – annoying, 2 – slightly annoying, 1 – perceptible, 0 – imperceptible. The evaluation is subjective; it is based on the assumption that no error detection system is used.

As we can see from **Table 1.1.1**, the parts of the digital sequence that need to be protected in the first place are Bit Allocation and SCFSI fields, as well as three senior digits of Scale Factors, which is a standard requirement.

Although the error protection requirement provided for by the European standard ETSIEN 300 401 v.1.3.3 [100] is very lenient, it is still strong enough to ensure that the probability of errors p_{er} in data transmitted over the communication channel does not exceed the acceptable maximum ($p_{er} \leq 10^{-5}$). Therefore, the length of the correcting CRC-16 code to protect the 94 bits of the Bit Allocation field and 44 bits of the SCFSI field, was limited to 16 bits only (see **Fig. 1.1.1**).

The Scale Factors field represented by a sequence with the average length of 160 bits, is protected by a 32-bit CRC-code, which is added to the audio frame at the encoder output.

There are three types and values of redundancy r introduced into a compressed audio stream [16, 17]:

1. Non-uniform redundancy ($r \approx 1.5\%$) which is used to protect only the most important information in the header (the header contains 43 bytes of information). Non-uniform redundancy is typical of MPEG-1 Layer 2 format

used in high quality communication channels with bit error ratio no greater than 10^{-5}.

2. Low-level uniform redundancy ($r \approx 2.5\%$) which is used to protect the entire frame including samples of audio signal compressed in the frequency domain. This kind of redundancy can be used in medium quality channels with bit error ratio ranging from 10^{-5} to 10^{-4}, where the prevailing type of errors in code words are single digital errors, i.e. errors whose multiplicity equals one ($t = 1$).

3. High-level uniform redundancy ($r \approx 10\%$) is used in low quality communication channels with bit error ratio ranging from 10^{-4} to 10^{-3}. In those channels, along with single errors, double errors (i. e. errors whose multiplicity equals two ($t = 2$)) are quite common.

Table 1.1.1 Immunity of different audio frame parts to errors

Frame section	Digit number	Immunity level
Bit Allocation	All digits	5
SCFSI	All digits	5
Scale Factors	5 (the most significant digit)	4
	4	4
	3	4
	2	3
	1	2
	0 (the least significant digit)	1
Subband Samples	15-8 (the most significant digit)	3
	7 – 5	2
	4 – 3	1
	2 – 0 (the least significant digit)	0

Table 1.1.2 shows the probability p_{dc} of errors in decoded codewords, calculated for the header. The header occupies about 10% of the frame and

constitutes one codeword with the average length N = 43 bytes (344 bits) and the maximum length 418 bytes (3344 bits).

Calculations demonstrate that p_{dc} depends on the probability of error p_{er} in the communication channel, the codeword length N (bytes) and the multiplicity t of errors detected. In its turn, the multiplicity t is determined by the introduced redundancy r (when r=0, t=0).

The probability p_{dc} of decoding an error in a codeword consisting of N bytes (8N bits) is determined by the probability that no errors will be detected (t=0), the probability that only single errors (t=1) will be detected and errors of higher multiplicity will be passed unnoticed, and the probability that both single and double errors (t=2) will be detected and errors of higher multiplicity (t> 2) will be passed unnoticed.

If t = 0, then $p_{dc}(8N) = 1 - (1 - p_{er})^{8N} \approx 8 * N * p_{er}, \quad p_{er} << 1$.

If t =1, then

$$p_{dc}(8N) = C_{8N}^{2}\,(p_{er})^2 * (1 - p_{er})^{8N-2} = \frac{8N * (8N-1)}{2} * (p_{er})^2 * (1 - p_{er})^{8N-2} \approx$$

$$\approx \frac{8N * (8N-1)}{2} * (p_{er})^2, \; p_{er} << 1.$$

If t =2, then

$$p_{dc}(8N) = C_{8N}^{3} * (p_{er})^3 * (1 - p_{er})^{8N-3} = \frac{8N * (8N-1) * (8N-2)}{6} * (p_{er})^3 * (1 - p_{er})^{8N-3} \approx$$

$$\approx \frac{8N * (8N-1) * (8N-2)}{6} * (p_{er})^3, \; p_{er} << 1.$$

It is clear that the probability p_f of decoding a frame in which a particular word occurs L times can be obtained using the formula $p_f = p_{dc}(8N * L)$.

As we can see from **Table 1.1.2**, in the case where $p_{er} = 10^{-6}$ the quality of communication is so high that there is no need to increase redundancy of the signal. The probability $p_f = 3.44 * 10^{-4}$ that there will be errors in a frame which lasts $T \approx 26.122$ ms corresponds to less than one error per minute. The highest acceptable bit error ratio for a frame this long is $p_f = 4.4 * 10^{-4}$.

Table 1.1.2

	p_{er}	10^{-6}	10^{-5}	10^{-4}	$5*10^{-4}$	10^{-3}
$N=43$, $L=1$ $t=0$, $r=0$	p_{dc}	$3.44*10^{-4}$	$3.43*10^{-3}$	$3.38*10^{-2}$	$1.58*10^{-1}$	$2.91*10^{-1}$
	p_f	$3.44*10^{-4}$	$3.43*10^{-3}$	$3.38*10^{-2}$	$1.58*10^{-1}$	$2.91*10^{-1}$
$N=43$, $L=1$ $t=1$, r=1.5%	p_{dc}	$5.9*10^{-8}$	$5.89*10^{-6}$	$5.7*10^{-4}$	$1.24*10^{-2}$	$4.19*10^{-2}$
	p_f	$5.9*10^{-8}$	$5.89*10^{-6}$	$5.7*10^{-4}$	$1.24*10^{-2}$	$4.19*10^{-2}$
$N=43$, $L=10$ t=1, r=2.5%	p_{dc}	$5.9*10^{-8}$	$5.89*10^{-6}$	$5.7*10^{-4}$	$1.24*10^{-2}$	$4.19*10^{-2}$
	p_f	$5.9*10^{-6}$	$5.7*10^{-4}$	$4.18*10^{-2}$	$2.65*10^{-1}$	$1.9*10^{-1}$
$N=43$, $L=10$ t=2, r=10%	p_{dc}	$6.7*10^{-12}$	$6.68*10^{-9}$	$6.48*10^{-6}$	$7.09*10^{-4}$	$4.7*10^{-3}$
	p_f	$6.7*10^{-9}$	$6.68*10^{-6}$	$4.8*10^{-3}$	$1.52*10^{-1}$	$2.15*10^{-1}$

For $p_{er} \geq 10^{-5}$ the required quality of sound can only be obtained by increasing the redundancy r introduced into the frame at the decoder output. Thus, in medium quality channels where bit error ratios range from 10^{-5} to 10^{-4}, the percentage r of redundant bits will have to fall within the range from 1.5 to 2.5% to enable the detection of single digital errors (multiplicity t=1), which mainly occur in the frame header. In low-quality channels where $p_{er} > 10^{-4}$, redundancy r should be further increased to 10% or even 25%, to enable the detection of common double (t = 2) and triple (t = 3) errors in the codeword which contains N bytes. As we can see from Table 1.1.2, in the channels of very low quality, i. e. channels with p_{er} ranging from $5*10^{-4}$ to 10^{-3}, ten-percent redundancy (r=10%) cannot guarantee the required sound quality at the decoder output.

For example, *Radian* company based in the city of St. Petersburg, Russia, estimated the real channel quality at $p_{er} = 5*10^{-4}$.In this case MPEG frame is additionally protected by a powerful corrective CRC-code in compliance with the standard ITU J.52. This leads to a 25% increase in the length of the frame and increases the data rate from 128 to 160Kbit/s, thus making the communication band 1.25 times as wide as it was. Ultimately, the introduction of such a high level of artificial redundancy ($r = 25\%$) makes expensive satellite and radio relay communication channels less effective. In this case, only 4 radio stations instead of 5 can operate in the frequency band allocated for digital broadcasting. At the same time, reserves created by high natural redundancy of audio frame headers remain untapped.

The use of natural redundancy of compressed frame headers for detecting and correcting catastrophic errors at the input of an MPEG digital audio decoder can significantly reduce the channel quality requirements. Experimenting with computer models shows that, by using a low redundancy Hamming correction code with parity check instead of a Reed-Solomon code, it is possible to maintain the preset high-level of error immunity with the relative level of interference increased up to 15 dB. The number of check bits will remain the same, and artificial redundancy will not exceed a given value $r \approx 3\%$ of the frame length.

The results obtained through the extensive use of computer-based models of 12 segments of digital sequences representing various types of audio content fully confirm that it is possible to detect and correct single and rare double catastrophic errors in three fields of any frame header: the Bit Allocation field, the SCFSI field and the Scale Factor field.

Computer modeling allowed establishing the relation between p_{er} at the decoder input and signal-to-noise ratio at the decoder output for different fields of the frame header.

The simulation results are presented in **Table 1.1.3**. They were obtained without detecting and correcting introduced errors. As you can see from the table, a 10-fold increase in p_{er} causes the signal-to-noise ratio to decrease almost to the same extent.

The standardized level of noise in compressed audio frames, measured in dB, was estimated as the ratio of the mean square difference between the corresponding samples of the source signal frame and the affected frame to the power (variance) of the audio signal. The measurement time interval was approximately 27 minutes long (it contained about 62,000 frames). The compressed audio signal under analysis had the form of a non-stationary random process, as it consisted of several shorter segments representing various types of signals (classical music, pop, jazz, vocal music, speech). Thus, the experimental conditions were made as realistic as possible. As we can see from table **1.1.3**, even in the case where the standard bit error ratio requirement is met (p_{er} = 10^{-5}), all attempts to protect the frame from residual noise (i.e. reach the SNR of 25dB) fail due to errors in the Bit Allocation field.

The task of obtaining the required quality of sound without using a correcting code to protect the header is only feasible for $p_{er} < 10^{-6}$. A real digital broadcasting system, therefore, usually introduces redundancy to detect catastrophic errors; the frame in which a catastrophic error is detected is then replaced with the previous frame stored in the decoder memory. When $p_{er} = 10^{-5}$, such replacement only occurs once or twice in a minute and has no effect on the quality of sound [100].

Table 1.1.3 Immunity of the frame to various kinds of errors

p_{er} in decoder input	Bit Allocation dB	SCFSI, dB	Scale Factors, dB	Signal samples, dB	The entire frame, dB
10^{-5}	+14.55	+26.92	+26.15	+ 41.20	+14.05
10^{-4}	+4.53	+16.93	+16.13	+31.21	+4.04
$3*10^{-4}$	-0.24	+12.16	+11.38	+26.42	-0.73
10^{-3}	-5.47	+6.94	+6.16	+21.19	-5.96

In a situation where $p_{er} = 10^{-3}$, however, corrupted frames would have to be replaced no less than 100 times per minute, which is totally unacceptable. It follows from **Table 1.1.2** that, in this case, the introduction of 10-percent redundancy would not help either, as frame replacements would take place every 5 or 6 seconds. When the above-mentioned *Radian* Company conducted its tests of a real satellite channel, it had to increase redundancy to 25% and use a higher bit error ratio.

The length of the twelve segments representing different audio content, which were extensively used in computer simulations, varied from 16 to 500 seconds. Numerous random single digital errors were introduced in the segments, using different bit error ratio settings.

The results obtained from computer modeling fully confirmed that it is possible to detect and correct catastrophic errors in the frame header. Natural redundancy of frame headers made it possible to lower significantly the required level of artificial redundancy in Hamming codes with parity check, enabling not only detection, but also correction of errors.

The algorithm for detecting and correcting the first two types of catastrophic digital errors is described in [16, 17]. It is based on the random mixing of audio signal samples in each frequency subband (the mixing results from misinterpretation of data in the Bit Allocation and SCFSI fields).

It was proved [16] that the distribution of momentary values in audio signal samples is characterized by a high peak factor (*PF*), which is the peak amplitude of the waveform divided by the RMS value of the waveform. As the audio signal amplitude takes small values much more often than big ones, its *PF* ranges from 3 or 4, in the case of speech, to 10, in the case of symphony music.

When a catastrophic error occurs in the frame header, the data in the Subband Samples field are misinterpreted, and audio signal samples in each frequency subband get randomly "mixed". It is obvious that, in this case, the distribution of sample momentary values moduli for each frequency subband will be close to a rectangular one, with $PF = \sqrt{3}$. Knowing the peak factor of an audio signal, therefore, allows for the corrupted frame detection. To simplify the peak factor calculation, the average modulus can be used instead of the RMS value of the signal, thus giving the average peak factor value $PF_{av.}$

Table 1.1.4 shows PF_{av} values calculated using the normalized signal modulus which takes values within the interval (0; 1). The computing was done for four differential distribution laws all of which have the maximum in the area of low values for the signal. The four distribution laws are listed below:

(1) truncated exponential distribution
(2) truncated normal (Gaussian) distribution
(3) triangular distribution
(4) rectangular distribution

In the case of the first two distributions, truncation of tails was done for the "unit" standardized level; the probability that this level will be exceeded never gets greater than 0.003. For the Gaussian distribution the respective peak factor values are PF = 3 and PF_{av}= 3.75. At this maximum level in the case of the exponential law, the average peak factor PF_{av} equals 5.81.

Table 1.1.4

Distribution law	1	2	3	4
Peak factor PF_{av}	5.81	3.75	3	2

Table 1.1.4 shows a decrease in the peak factor in the case of a rectangular distribution of the signal momentary values: the factor by which PF_{av} decreases varies from 1.5 to 2.3.

Computer modeling used a real fragment of a compressed music signal. The fragment lasted over 20 sec. consisted of 800 frames with over 500K samples. The peak factor varied from frame to frame ranging from 2.5 to 3.9, the average value being three ($PF_{av} = 3$).

Another simulation was used to obtain data for 100 distorted frames of the same signal into which catastrophic digital errors had been introduced. In this case the peak factor varies from 1.6 to 2.3, averaging at PF_{av}=1.9.

The bar-charts shown in **Fig.1.1.2** were copied from a PC monitor screen; they represent the modulus distribution of the normalized signal samples for undistorted and distorted frames. The bar-chart on the left clearly shows that, in the case of undistorted frames, the signal modulus distribution law resembles the Gaussian law, or a triangular distribution. In the case of distorted frames we see a higher probability that large values will occur, hence the resemblance of the bar-chart on the right to a rectangular distribution with fluctuations about the average value.

Comparing experimental results obtained from computer modeling with theoretical computations shows their close proximity. That makes it possible to ascertain, at least indirectly, the presence of a catastrophic error by looking at the distribution of sound sample momentary values.

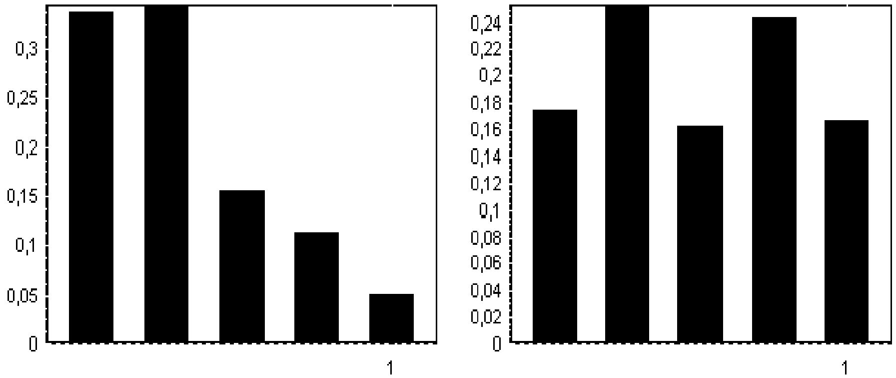


Fig.1.1.2. Distribution of the normalized signal samples moduli for undistorted (on the left) and distorted (on the right) frames

Later in this chapter, we will show that the peak factor will be easiest to maintain in the highest frequency subbands. That part of the frequency spectrum is almost imperceptible to a human ear; therefore, binary codes representing audio samples in those subbands are rather short; each codeword only has two or three digits. Most of the time samples are taken at levels 0, 1 and 2, the average code value being 1, or at levels 0,1,2,3 and 4, the average code value being 2. Average code values correspond to samples with low values, whereas deviations from the average indicate an increase in the signal sample modulus.

Undistorted frames are, for the most part, characterized by low average values of samples and high peak factor values. A catastrophic error in the frame header results in a significant increase in the average sample moduli for all frequency subbands. This increase is particularly noticeable in the area of higher frequencies where a smaller number of code words can be possibly used to encode samples. As the average modulus increases, the number of low value samples goes down, causing a significant reduction in the peak factor. In the case where codes 0, 1 and 2 are used, codes 0 and 2 start to prevail over code 1, whereas in the case where four codes are used, codes 0 and 4 become more common than code 2. If such deviations occur in several subbands at the same time, that is a good indication of the presence of a catastrophic error.

Table 1.1.5 Evaluating the effectiveness of digital error correction algorithm

Errors in Bit Allocation field		
p_{er} at the decoder input	With no error correction algorithm applied, dB	With some error correction algorithm applied, dB
10^{-5}	+14.55	+75.87
10^{-4}	+4.53	+65.98
$3*10^{-4}$	-0.24	+61.24
10^{-3}	-5.47	+56.01
Errors in SCFSI field		
p_{er} at the decoder input	With no error correction algorithm applied, dB	With some error correction algorithm applied, dB
10^{-5}	+26.92	+84.82
10^{-4}	+16.93	+74.91
$3*10^{-4}$	+12.16	+70.26
10^{-3}	+6.94	+65.04

Errors in Scale Factors field		
p_{er} at the decoder input	With no error correction algorithm applied, dB	With some error correction algorithm applied, dB
10^{-5}	+26.15	+46.85
10^{-4}	+16.13	+36.83
$3*10^{-4}$	+11.38	+32.04
10^{-3}	+6.16	+26.82
Errors in the entire frame		
p_{er} at the decoder input	With no error correction algorithm applied, dB	With some error correction algorithm applied, dB
10^{-5}	+14.05	+39.74
10^{-4}	+4.04	+29.93
$3*10^{-4}$	-0.73	+25.23
10^{-3}	-5.96	+19.9

Protecting to some extent the data in the most important fields of the frame header makes it possible to identify the subband in which a catastrophic error occurred. Monitoring the peak factor is a reliable technique for detecting the error and correcting it by inverting the error bit.

Existing error correction methods use error masking based on interpolation of signal samples [100, 102]. Masking makes errors less noticeable but fails to recover them in full. The earlier described method of correcting catastrophic errors has the advantage that it guarantees the recovery of all detected single errors.

The effectiveness of using the proposed algorithm for correcting digital errors in the headers of compressed frames can be evaluated by comparing figures shown in two columns of **Table 1.1.5**. The data presented in the table were obtained in the long process of computer modeling by averaging the results for 12 audio files in which various types of signal (various broadcasting program genres) are presented.

It follows from **Table 1.1.5** that, even if the probability of errors p_{er} in the Bit Allocation field equals 10^{-5}, error correction will still be necessary; otherwise the listener will hear irritating crackling twice a minute.

Errors in the SCFSI field occur much less often. Although the probability of error in the Bit Allocation field is 2 to 2.5 times as high as the probability of error in the short SCFSI field, errors that occur in the latter also have to be corrected.

Table 1.1.5 shows that, in the case where p_{er}=3∗10^{-4}, correcting catastrophic errors in all audio signal samples and all fields of the frame header makes it possible to ensure that SNR=25dB, the noise immunity level recommended in [41]. This figure is comparable to the one obtained for the case where p_{er} = 10^{-3} (see **Table 1.1.3**); in that case correcting errors in the Bit Allocation field alone raised the power ratio of signal to residual noise by 27dB. Once errors had been introduced to all fields in the frame header, the average SNR for 12 audio signal wave segments came to 19.9dB.

Given that 25dB is the minimum level required, the probability of errors needs to be lowered by a factor of 3.33 to p_{er} = 3∗10^{-4}. It means that using the natural redundancy of frame headers allows you to weaken the satellite channel quality requirement by a factor of 30.

It should be noted that residual noise is caused by undetectable errors that occur in audio signal samples or the three least significant digits of scale factor codes. Should the frame have any artificial redundancy, it would have been possible to reduce residual noise, but no such redundancy was introduced in the frame. Reducing the acceptable probability of errors from 10^{-3} to $3*10^{-4}$ is, therefore, the only solution.

Nevertheless, we have to admit that the result obtained by implementing that solution is quite good. Indeed, the required quality of sound was ensured despite the 30-fold increase in the probability of errors compared to its nominal value $p_{er} = 10^{-5}$. We also managed to bring the power ratio between the signal and the residual noise to the desired level of 25dB. All we did was use a low-redundancy correcting code; there was no need to increase the bit error ratio or widen the frequency band of the expensive satellite communication channel.

What makes the proposed error correction method particularly effective is the use of high redundancy of the original uncompressed digital audio signal. Some of that redundancy is retained even after the compression, as it is impossible to get rid of it completely without using a much more complicated encoder and increasing the signal processing time. The frame header is never compressed. The need to broadcast in real time limits acceptable signal delays in the encoder to a very short interval (between 50 and 100 ms).

The time delay caused by signal compression is too short for the encoder to process the information on the previous audio frames; therefore it only compresses the currently transmitted frame. Frames are compressed by the encoder independently, without taking into consideration statistic relations between consecutive frames and between parts of the same frame.

The frames are relatively short; in most cases, one can obtain high quality sound, using frames that last 24 to 26 ms. The average duration of vowel speech sounds and music sounds ranges from 100 to 200 ms, which means that each sound of that kind is represented by 4 to 8 frames.

Statistic relations ascertained for time and frequency domains should be present in several (usually 2 to 4) subsequent frames preceding the frame in question.

In the case where $p_{er} \le 10^{-5}$, high quality sound can be secured by using a error-correcting code to detect the above-mentioned two types of catastrophic errors and by replacing the current frame that contains errors with the preceding one. However, in communication channels that fail to meet the above standard requirements frequent replacements of error-affected frames will be perceived by a human ear, which is absolutely unacceptable.

On the face of it, the problem appears to have only one solution; namely, introducing high-level artificial redundancy in the frame header and using check bits to detect and correct errors. Implementation of this solution will require increasing the bit error ratio and using a wider frequency band, thus making the use of an expensive satellite channel less efficient.

An alternative solution consists in tapping the reserve and is based on using the natural redundancy of audio frames. Computer modelling fully confirmed that we made the right choice when we took this approach. Knowing the properties of sound in time and frequency domains makes it possible to predict the type and magnitude of the detected error. The time taken by any two subsequent frames is too short for the sound to undergo any significant changes, hence strong correlation between the frame headers in the time domain.

The audio frequency spectrum is divided into a number of relatively narrow subbands. The power spectrum envelope changes slowly as you go from one subband to the next. One cannot expect significant changes of the signal level in any particular subband as you move from one frame to the next. At the same time, the signal level in any particular subband correlates with the signal level in the adjacent subbands within the same frame. Hence strong correlation between signal level values in the frequency domain.

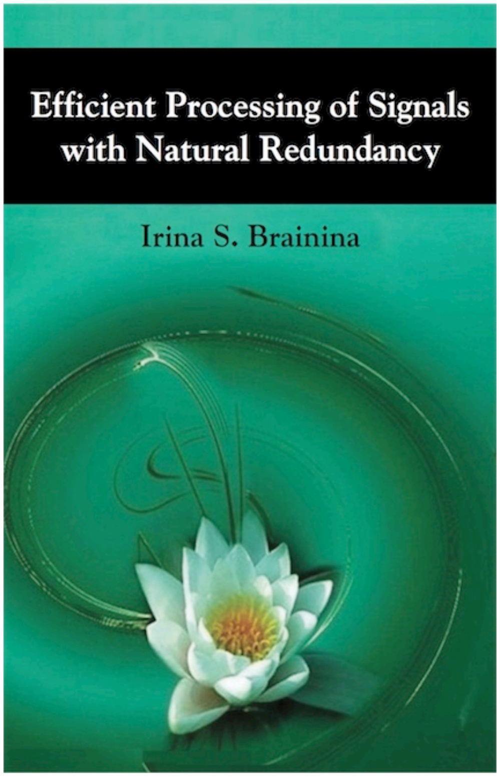
A big difference between the headers of two frames – the one in question and the preceding one – indicates the presence of an error in the frame being processed. Considering that the decoder can store the preceding frame in its memory, no complicated modification of decoding algorithm will be required.

By using an error-correcting Hamming code with parity check one can detect and localize different types of errors. Once the differences between the current frame and the preceding one have been established, errors can be corrected by inverting the erroneous bits.

It is assumed that various types of errors that occur in the frame header are single and independent (in other words, there is no correlation between them). Bursts of errors that occur in a satellite radio channel are turned into single errors by interleaving binary digits at the encoder output. Calculations and experiments show that the probability of two or more errors occurring within the same frequency subband in a compressed frame is rather small, even if $p_{er} = 10^{-3}$ therefore, as a rule, we only deal with single errors.

Digital Audio Broadcasting (DAB) standard [100] suggests using a back-up mechanism to protect information against errors. It provides additional protection of frame headers by using a low bit error ratio channel to transmit data from the encoder to the decoder.

To each compressed audio frame 80 bits are added, which are transmitted to the decoder. 32 of them are used to protect the Scale Factor code, whereas



This book offers new effective digital algorithms for processing audio and tonal signals, having a high natural redundancy. The redundancy preserved in headers of compressed audio frames in MPEG-1 Layer 2 format. Computer modeling proves an opportunity to detect and correct all catastrophic errors in three fields of a frame header.

Efficient Processing of Signals with Natural Redundancy

by Irina S. Brainina