Graceful Degradation of Digital Audio Transmission Systems*

K. A. SCHOUHAMER IMMINK

Philips Research Laboratories, 5600 JA, Eindhoven, The Netherlands

A new technique for encoding and decoding digital audio signals is presented. It offers the advantage of a graceful degradation of the audio performance when the signal is conveyed over channels with a wide range of bandwidths. The flexibility of the new techniques also provides a key to the future evolution of the quality of digital audio or video systems within the same digital format. The class of systems described may well be of particular value in applications involving the storage or broadcast of digital audio signals when the exact bandwidth of the communication system is unknown.

0 INTRODUCTION

The quality and the reliability of digital audio storage or communication systems are so widely appreciated that it is scarcely conceivable that one might dispute their properties. In this report we will not question the perceptual quality of digital audio; we merely wish to impeach some of the fundamental characteristics of the digital communication channel itself.

It is well known that the quality of the decoded signal is, up to certain system thresholds, almost ideal. As an illustrative example consider the Compact Disc signal. The EFM encoded signal requires a bandwidth of at least 1.2 MHz for proper functioning [1]. If, for some reason, the bandwidth is reduced to 1 MHz, for example, the symbol error rate can quite abruptly become so high that the decoding electronics cannot cope with it and will start muting. Basically this should be considered a catastrophic breakdown, while the system should still be capable of supporting 80% of its normal capacity. Even when very smart error correction schemes are included, we will find that when the bandwidth is below the channel's threshold value, the receiver will not be able to retrieve any useful information at all. A similar phenomenon can be observed when additive noise perturbs the communication channel. Up to a certain threshold level of the noise power the system will function properly, but once the noise power goes beyond this threshold, the communication system will suddenly cease to exist. This effect is not normally found in analog systems, which usually show a more gradual or graceful degradation.

Another basic, undesired property inherent in digital communications is the lack of flexibility of digital formats. Once a digital communication or storage system has been designed, it is almost impossible to upgrade or downgrade the quality of the digital audio signal in a manner compatible with the predefined format. Changing to a higher sampling rate or a finer quantization is impossible within the defined digital format. The magnificent improvement of the audio quality of the Compact Cassette, all within the same format, would have been impossible when instead of the analog signal format a digital audio format would have been chosen. It can be argued that, for the Compact Disc system, in the future there will be no need to improve the quality, but the poor man's digital audio systems, which can hardly be called hi-fi, are now on the drawing boards. Actually this is not a real problem but for the fact that many of these new formats have to last for decades. As time goes on and technology improves, more room will become available. When that happens, we will be sorry, since extensions compatible with the current formats are out of the question. For a digital audio broadcast standard it is important that the vast quantities of existing consumer equipment should not suddenly be made obsolete—existing equipment must be capable of processing any improved signals to give the quality now attainable.

By way of example, let us consider the MAC video signal [2]. One of the notable properties of MAC video transmission is the absence of any absolute limitation in bandwidth reduction. Unlike the PAL and NTSC

* Presented at the 82nd Convention of the Audio Engineering Society, London, 1987 March 10–13.

video signals, one can freely convey the MAC video signal over transmission systems covering a wide range of bandwidths with the corresponding quality of the received video signal. This property is lost as soon as the MAC video signal is time multiplexed with a digital component whose rate and coding lead to a minimal bandwidth greater than, or equal to, that of the MAC video signal. The D2-MAC video format has been chosen to accommodate the audio plus video signal in the existing 5–6-MHz cable bandwidth. Once more bandwidth is available, the video quality will improve, but unless the digital audio format is changed, the audio quality will remain the same, thus again fueling the notion that audio is a stepchild of television broadcasting.

In this report we study a new way of encoding and decoding digital audio signals which possesses the virtue of improved bandwidth adaptability compared with conventional digital transmission systems. In Sec. 1 we explain in qualitative terms the basic idea of the frequency domain detection technique with a simple example. In Sec. 2 we apply the new technique to the transmission of digital audio.

1 BASIC CONCEPT

Prior to the development of a quantitative model for the coding and detection technique we consider some basics of digital transmission. There are a number of ways in which digital symbols can be represented by physical parameters. All these involve assigning a range of waveforms of a continuously variable physical function to represent some digital symbol. Most digital systems now in use are binary and synchronous, which means that in each symbol time interval or time slot a condition such as current or no current, pit or no pit, or positive or negative magnetization is transmitted (or stored). (Note that we use the parlance of the communication engineer throughout.) The receiver, under the control of its clock, properly phased with respect to the incoming data, samples the received signal at the middle of each time slot. It is well known that a single pulse transmitted over a bandwidth-limited system is smeared out in time due to convolution with the channel's impulse response. A sample at the center of a symbol interval is a weighted sum of amplitudes of pulses in several adjacent intervals. This phenomenon is called intersymbol interference (ISI). If the product of symbol time interval and system bandwidth is reduced, the intersymbol interference becomes more severe. This effect can become so severe that the receiver, even in the absence of noise, can no longer distinguish between the symbol value and the intersymbol interference and will start to make errors. The basic idea behind the technique to be discussed in the subsequent sections is that the information is not contained in one time interval but is spread out over a plurality of intervals.

We now leave this typical time-domain approach and turn to a frequency-domain discussion of the coding and detection technique. To focus our thoughts, let us concentrate, for example, on a magnetic or optical recorder. It is well known that some parts of the available frequency band of these recording systems are less reliable than others. The low-frequency range of a magnetic recorder has insufficient signal-to-noise ratio since the inductive head does not respond to signals of low frequency. An optical recorder also exhibits an unreliable low-frequency response caused by noise which is induced by reflection variations or fingerprints on the optical disk. The frequency characteristics in the high-frequency range of both recorder types are relatively unreliable due to defocusing, tape-head loss, and so on. We proceed by partitioning the available bandwidth B into m smaller (virtual) subchannels with a bandwidth B/m. We order the subchannels according to their a priori reliability. The subchannel with the greatest reliability is assigned to the most significant bit of the digital audio signal. By modulation of the incoming data on orthogonal time functions, so-called carriers (and demodulation in the receiver), it is possible to fill the total available frequency range [3]. The signals generated in this way are multiple valued or even continuous. Unfortunately many communication or recording channels are "hard-limiting" and accept only two-level full-T pulses. Examples are optical recording, where only pits or lands can be recorded, and magnetic recording, which is normally employed in such a way that the resulting magnetic domains are positively or negatively saturated. To overcome this difficulty we proceed as follows.

We choose candidate codewords of length n with elements from a binary alphabet. The codewords are also selected in such a way that they divide, according to a predefined frequency domain criterion, the available frequency range of the channel into smaller independent subchannels. The frequency domain criterion will be based on the (Walsh)–Hadamard transform, which has the virtues of simplicity and analytical tractability. A selection of codewords based on other frequency domain criteria is certainly feasible. The major advantage of Hadamard domain processing compared with Fourier domain processing is simplified computation, since the Hadamard transform requires only addition operations, whereas the Fourier transform requires complex arithmetic.

We now select the codewords, denoted by $\boldsymbol{x}$, $x_i \in \{-1, 1\}$, $i = 1, \ldots, n$, to have an odd parity, that is, the number of $+1$'s is odd. The codeword set is denoted by S. The reason behind this particular choice of codeword set will become apparent in the following example. It can readily be verified that 2^{n-1} codewords satisfy this condition. We also define the frequency domain representation $\boldsymbol{y}$ of $\boldsymbol{x}$, which results from the $n \times n$ Hadamard transform,

$$\boldsymbol{y} = H_n \boldsymbol{x}$$

where H_n, $H_n(i, j) \in \{-1, 1\}$, is the Sylvester-type $n \times n$ Hadamard matrix. This selection of the frequency domain transform restricts the codeword length to powers of 2 ($n \geq 4$). Due to the particular choice of the codeword set S we find (this can easily be verified) for any $x \in S$ that the elements y_i of the vector $y = H_n x$ are nonzero.

The basic concept of the new technique is now illustrated by a simple example.

1.1 Example 1

Suppose, for the sake of simplicity, that the codeword is of length 4. The Hadamard matrix is

$$H_4 = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \\ +1 & -1 & +1 & -1 \end{bmatrix} .$$

The coefficients of the Hadamard matrix are ordered in correspondence with increasingly rapid variation of the zero crossings (the usual notion of frequency in the case of the Fourier transform). It is inappropriate here to consider in detail the formal establishment of the Hadamard matrix or transform, a comprehensive discussion of which may be found in [3] and its many references, for example.

The $2^{n-1} = 8$ codewords with odd parity are $(+1, -1, -1, -1), (-1, +1, -1, -1), (-1, -1, +1, -1), (-1, -1, -1, +1)$ and their inverses. Table 1 shows the eight source words, their corresponding channel representations x, according to a coding rule to be explained later, with their associated frequency domain representations $y = H_4 x$ and the vector z with elements $z_i = 0$ if $y_i < 0$; else $z_i = 1$.

In this example we find that $m = 3$ source symbols can be mapped onto $n = 4$ channel bits, so we conclude that the rate of this code is 3/4. How do we proceed? We elucidate the new technique by assuming that the physical channel has an undesirable response at the low-frequency end. For example, the received signal is superimposed on an unknown (quasi) dc current. We also assume that transmitter and receiver are perfectly synchronized, that is, the beginnings and endings of the different codewords are known. Suppose we want to transmit the source word 100. The corresponding channel representation found from Table 1, namely $(-1, +1, -1, -1)$, is transmitted. It is tacitly assumed that the symbols of the codewords are transmitted serially. The received codeword, denoted by r, generally distorted and corrupted with noise, is now Hadamard transformed in the receiver into $n = 4$ frequency components, denoted by $\hat{y}_i$, $1 \leq i \leq n$, or

$$\hat{y} = H_4 r . \qquad (1)$$

The estimated values of z_i, $1 \leq i \leq n$, denoted by $\hat{z}_i$, are subsequently determined from the polarity of $\hat{y}_i$, that is, $\hat{z}_i = 0$ if $\hat{y}_i < 0$; else $\hat{z}_i = 1$. Element $\hat{z}_1$ is skipped; the other three decoded symbols $\hat{z}_2, \ldots, \hat{z}_4$ are the received (estimated) source symbols. As a further clarification, the Hadamard transform is only implemented in the receiver. It is now easy to understand why the code set consists of codewords with an odd parity; detection can be kept very simple by just taking the signs of the frequency transform components $\hat{y}_i$. The reader may have noticed in Table 1 that the source word assignment to a specific codeword has been prepared in such a way that the symbols $z_2, \ldots, z_4$ (the three symbols furthest to the right in the column furthest to the right) are equal to the three source symbols (the column furthest to the left). The variable $\hat{z}_1$, which actually contains the unknown dc term of the received codeword, is not used. Columns 2, 3, and 4 of the Hadamard matrix have an equal number of ± 1's so that any unknown superimposed dc term will be canceled here and will have no effect on the decoding.

It is also intriguing to observe that the power density function of the code discussed here does not exhibit a vanishing power at the low-frequency end, while this is a property usually found to be mandatory in channel codes to be applied in dc-constrained channels [4].

It has been seen in this example that an unreliable subchannel situated at the low-frequency range can (by an a priori assignment) be discarded from the available set. In a similar way we can, by an a priori assignment of source words/codewords, eliminate one of the other subchannels. This is made possible by the remarkable fact that all vectors z have an odd parity and that they are all distinct (as can be verified in Table 1). In other words, the eight codewords presented in Table 1 are universally applicable. Irrespective of the particular subset of subchannels used, we employ the same set of codewords. A set with this property is called a key set. We found by computer search that for codeword length $n = 8$ a key set can be assembled. A computer verification showed that this is not possible for $n = 16$.

This introductory section has sought to outline in qualitative terms the main idea of the frequency domain detection technique. In the following section we take a closer look at the more quantitative effects of intersymbol interference and additive noise.

Table 1. An example of a rate 3/4 dc insensitive code.

Source	x	y	z
000	−1 +1 +1 +1	+2 −2 −2 −2	1000
001	−1 −1 +1 −1	−2 −2 −2 +2	0001
010	−1 −1 −1 +1	−2 −2 +2 −2	0010
011	+1 −1 +1 +1	+2 −2 +2 +2	1011
100	−1 +1 −1 −1	−2 +2 −2 −2	0100
101	+1 +1 +1 −1	+2 +2 −2 +2	1101
110	+1 +1 −1 +1	+2 +2 +2 −2	1110
111	+1 −1 −1 −1	−2 +2 +2 +2	0111

2 GRACEFUL DEGRADATION

In the preceding section we showed that a codeword set can be assembled which can be detected in such a way that an unknown dc component does not disturb the detection. The same codeword set can be employed to transmit information over a bandwidth-limited chan-channel with the resulting virtue that parts of the transmitted information can be retrieved when the bandwidth is reduced below a certain value. The codeword/source word assignment is chosen is such a way that the upper channel does not contain user information. Table 2 shows the new assignment. Note that the same codeword set is employed as that in Table 1.

We leave the foregoing time-discrete analysis and turn to the time-continuous case. We assume that the channel has a low-pass spectral characteristic with sinusoidal rolloff [5]. The channel waveform, denoted by $g(t)$, can be written as

$$g(t) = \frac{\sin 2\pi Bt}{2\pi Bt} \, \frac{\cos 2\beta\pi Bt}{1 - 16\beta^2 B^2 t^2}$$

where B is the bandwidth of the transmission system and β, $0 \leq \beta \leq 1$, is the rolloff parameter. The term sinusoidal rolloff becomes clear when we write the corresponding frequency characteristic of the channel $G(f/B)$,

$$G(f/B) = \begin{cases} 1, & 0 \leq f/B \leq (1 - \beta) \\ \frac{1}{2}\left\{1 - \sin \frac{\pi}{2\beta}(f/B)\right\}, & (1 - \beta) < f/B \leq (1 + \beta) \\ 0, & f/B > (1 + \beta) \, . \end{cases}$$

Fig. 1(a) shows the eye patterns of the received signal $r(t)$, which is formed by convolution of the codewords and the channel waveform. The parameters are relative bandwidth $BT_c = 0.5$, T_c being the channel bit interval, and rolloff parameter $\beta = 0.1$, no additive noise. The figure is drawn in such a way that a complete received codeword can be seen. The four eyes at sampling moments $t = -1.5$, -0.5, 0.5, and 1.5 are fully open. Fig. 1(b) shows the eye pattern of the signal $\hat{y}_1(t) = \{r(t) + r(t + T_c) + r(t + 2T_c) + r(t + 3T_c)\}/2$ received in the first subchannel; the sampling moments are at $t = 0$. The scaling by a factor of 2 has been done in order to be able to compare Fig. 1(a) and (b) on the same signal-to-noise basis. In a similar way Fig. 1(c) and (d) shows the eye patterns of $\hat{y}_2(t) = \{r(t) + r(t + T_c) - r(t + 2T_c) - r(t + 3T_c)\}/2$ and $\hat{y}_3(t) = \{r(t) - r(t + T_c) - r(t + 2T_c) + r(t + 3T_c)\}/2$ of subchannels 2 and 3, respectively. The eye pattern of subchannel 4 is not shown.

Table 2. Rate 3/4 low-frequency code.

Source	x	y	z
000	−1 −1 +1 −1	−2 −2 −2 +2	0001
001	−1 −1 −1 +1	−2 −2 +2 −2	0010
010	−1 +1 −1 −1	−2 +2 −2 −2	0100
011	+1 −1 −1 −1	−2 +2 +2 +2	0111
100	−1 +1 +1 +1	+2 −2 −2 −2	1000
101	+1 −1 +1 +1	+2 −2 +2 +2	1011
110	+1 +1 +1 −1	+2 +2 −2 +2	1101
111	+1 +1 −1 −1	+2 +2 +2 −2	1110

Figs. 2 and 3 show the effects of the reduction of the relative bandwidth to 0.4 and 0.3, respectively. (The other parameters are fixed.) Figs. 2(a) and 3(a) correspond to the signal depicted in Fig. 1(a). We observe that the higher a particular subchannel is situated in the frequency band of the channel, the more it is affected by the intersymbol interference caused by the relative bandwidth reduction.

Another general point worth mentioning is that of the measure of degradation, the audibility, of the reconstructed signal. How can we compare quantitatively a channel with a given reliability with another channel consisting of a number of subchannels each with a certain reliability? It is notoriously difficult to specify quantitatively the degree of annoyance experienced by a human observer. Dostis [6] arrived at the following simple mathematical relation that takes account of the effects of both quantization and incorrect decoding of the received data,

$$\mathrm{SNR} = \left[2^{-2N} + 4 \sum_{i=1}^{N} \Pr(E_i) 2^{-2i} \right]^{-1}$$

where SNR is the signal-to-noise ratio of the reconstructed audio signal and $\Pr(E_i)$ is the probability that the ith audio bit is erroneously received. It is assumed that the sound is linearly quantized with 2^N levels and that a natural representation is employed. For more details of the premises of the formula used the reader is referred to the original literature [6]. When $\Pr(E_i)$ is small, we find the well-known relation between SNR and the number of quantization levels,

$$\mathrm{SNR} = 2^{2N} \simeq 6N \quad \mathrm{dB} \, .$$

An illuminating experiment concerning the effectiveness of the new technique is to compare its performance with that of the conventional, uncoded, digital data transmission system. For purposes of illustration we have computed the SNR of an $N = 15$-bit linear quantization system based on a system with an $n = 4$ codeword length as a function of the bandwidth of the transmission system. The choice of the value of N is not arbitrary. We choose the codeword assignment in such a way that five codewords accommodate a 15-bit audio sample. We assume a data format where the five most significant bits of the audio sample are placed in the lowest subchannel, and so on. We now find that the probability $\Pr(E_i)$ of erroneously receiving an audio bit is given by $\Pr(E_i) = \Pr(e_1)$, $1 \leq i \leq 5$, $\Pr(E_i) = \Pr(e_2)$, $6 \leq i \leq 10$, and so on. The probability of receiving a symbol in error in the ith channel is denoted by $\Pr(e_i)$, $i = 1, \ldots, 4$. The channel characteristic used in our computations has a sinusoidal rolloff with rolloff parameter $\beta = 0.1$.

Fig. 4 shows the SNR of both the new and the conventional systems as a function of the relative bandwidth BT of the transmission system (no additive noise assumed). The relative bandwidth of the two systems has been scaled in such a way that they convey, per unit of time, equal amounts of user information $1/T$ (bits per second) at maximum bandwidth. Note that $T_c/T = 3/4$. The diagram shows better than words can explain the sudden loss of quality—the threshold effect—of the conventional system when the transmission bandwidth is reduced below a certain critical value. The diagram reveals quite clearly that the new system is partitioned into three subsystems. At a relative bandwidth $BT = 0.2$, which is a factor of 2 smaller than the critical bandwidth of the conventional system, the new system is still capable of supporting a 5-bit digital audio signal with an SNR of 40 dB. As can be seen, the cost of the new system is fairly low; it requires a 10% larger bandwidth than the conventional system to support maximum user capacity. In the next example we give some numerical data to exemplify the previous theory.

2.1 Example 2

The data bit rate of a digital audio signal is readily found. Assume a stereo channel, 15-bit quantization, and a 48-kHz sample frequency. The data bit rate (excluding error correction, subcode, synchronization, etc.) is $2 \times 48 \times 15 = 1.440$ Mbit/s, or $T = 694$ ns. From Fig. 4 we find that when conventional techniques are used, the minimum requirement of the bandwidth of the transmission channel is $0.42 \times 1.44 = 600$ kHz.

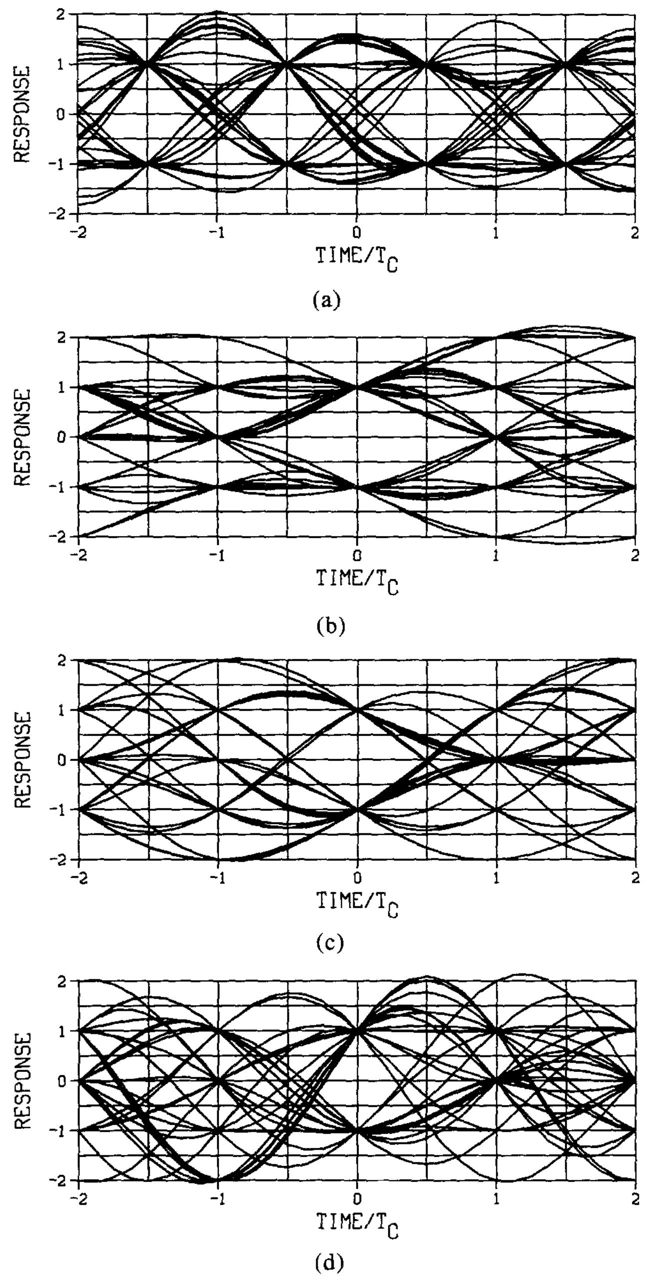


Fig. 1. Eye patterns; relative bandwidth $BT_c = 0.5$.

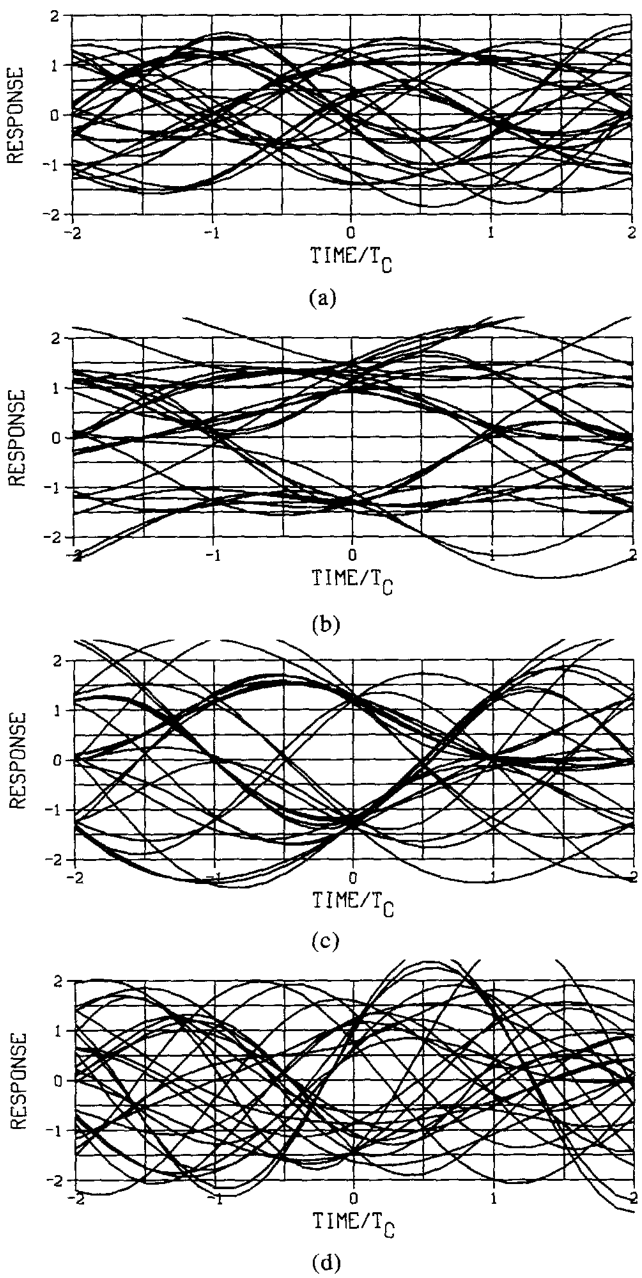


Fig. 2. Eye patterns; relative bandwidth $BT_c = 0.4$.

The diagram clearly reveals that when the channel bandwidth is smaller than this threshold value, no reception at all is possible. When the new technique is used, we find that the channel bit rate, that is, the bit rate after the channel encoder, is 4/3 × 1.440 = 1.92 Mbit/s. The minimum bandwidth for a full 15-bit reception is 0.47 × 1.44 = 677 kHz. When the channel bandwidth is 0.36 × 1.44 = 518 kHz, we are still able to decode a 10-bit audio signal. When the bandwidth is further reduced to 0.18 × 1.44 = 260 kHz, we can retrieve a 5-bit audio signal.

It is possible and indeed preferable in this coding scheme for each of the created subchannels to have a separate error correction and detection control. The error control mechanism can decide whether one or more of the subchannels are beyond their operating conditions and may decide not to use them, whereby

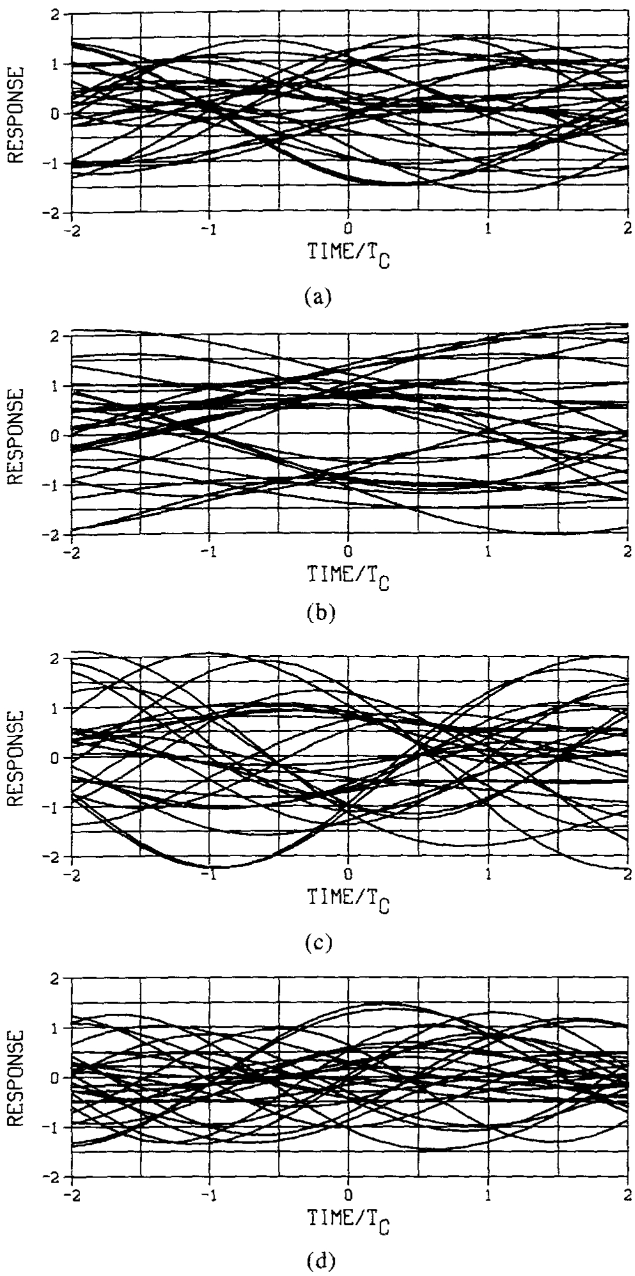


Fig. 3. Eye patterns; relative bandwidth $BT_c = 0.3$.

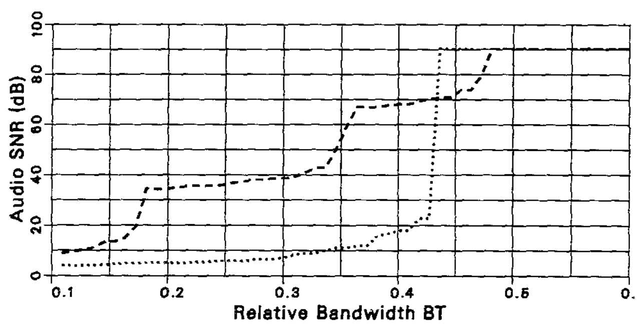


Fig. 4. Signal-to-noise ratio as a function of transmission bandwidth. Dashed curve—new system; dotted curve—conventional system.

it should be noted that this decision can be made entirely in the receiver. Obviously, a subset of the available subchannels can also be used as a tool for design flexibility. For example, we may decide to implement a decoding of the full set or of part of the set of the available subchannels, which leads to a corresponding reduction of the required decoding hardware and of the quality of the decoded audio signal. Though we coined the term "graceful" degradation for the new technique, it can be understood that the new technique provides a key to future evolution, that is, upgrading, of the quality of digital audio systems within the same digital format. If current technology allows a relative bandwidth of $BT = 0.2$, we can receive a 5-bit digital audio signal. As time goes on, technology will provide more bandwidth, which allows the reception of higher quality audio in a manner compatible with the low-bandwidth system.

3 CONCLUSIONS

A new coding and detecion technique was presented which offers the virtues of graceful degradation and the flexibility to up- or downgrade the quality of the transmission system. It has been shown that residual intersymbol interference has a different effect on the error probability of the least and most significant symbols when frequency domain detection of a set of codewords is employed. The class of systems described may well be of particular value in applications involving the storage or transmission of ordered data, such as digital video or audio, when the receiver or the transmitter has no exact knowledge of the channel bandwidth. The flexibility of the new coding technique also provides a key to the future evolution of the quality of digital audio or video signals within the same digital format.

4 REFERENCES

[1] J. P. J. Heemskerk and K. A. Schouhamer Immink, "Compact Disc: System Aspects and Modulation," *Philips Tech. Rev.*, vol. 40, pp. 157–164 (1982).

[2] H. Mertens and D. Wood, "Standards Proposed by the EBU for Satellite Broadcasting and Cable Dis-

tribution," *J. IERE*, vol. 56, pp. 53–61 (1986).

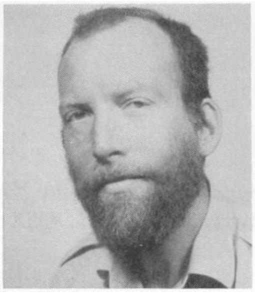
[3] H. F. Harmuth, *Transmission of Information by Orthogonal Functions*, 2nd ed. (Springer, Berlin and New York, 1972).

[4] T. T. Doi, "Channel Codings for Digital Audio Recordings," *J. Audio Eng. Soc.*, vol. 31, pp. 224–238 (1983 Apr.).

[5] J. G. Proakis, *Digital Communications* (McGraw-Hill, New York, 1983).

[6] I. Dostis, "The Effect of Digital Errors on PCM Transmission of Compandored Speech," *Bell Sys. Tech. J.*, vol. 44, pp. 2227–2243 (1965).

THE AUTHOR



Kees A. Schouhamer Immink was born in Rotterdam, The Netherlands, in 1946. He received a B.S. degree from Rotterdam Polytechnic in 1967 and M.S. and Ph.D. degrees from the Eindhoven University of Technology in 1974 and 1985, respectively, all in electrical engineering. He joined the Philips Research Laboratories, Eindhoven, in 1968, where his work involved the signal processing side of optical recording systems. Since 1986, he has been a member of the magnetic recording group at Philips Research.

Dr. Immink is coauthor of a book entitled *Principles of Optical Disc Systems*. He was awarded a fellowship by the AES in 1985 for "his work in the area of optical laser disk and for detailed study on channel codes for the Compact Disc." He serves as a member of the Organizing Committee of the International Conference on Video, Audio, and Data Recording and is on the committee of the AES Netherlands Section and the Benelux Community on Information and Communication Theory. He is a fellow of the IERE.