

Low bit-rate coding of high-quality audio signals

An introduction to the MASCAM system

G. Theile, G. Stoll and M. Link*

The article explains how the bit-rate of a digital audio signal can be reduced, without impairing the subjective quality, by arranging to transmit only those details of the signal which will effectively be perceived by the listener. There follows a description of preliminary hardware exploiting these principles. Further bit-rate reduction possibilities are then discussed, leading to a suggested codec configuration offering 16-bit linear subjective quality with a bit-rate of less than 100 kbit/s per monophonic channel, including an additional dynamically-varying bit-rate margin averaging about 30 kbit/s which can be used for any purpose such as error protection and/or the transmission of arbitrary data.

1. Introduction

Digital signal processing has progressed over the past few years to the point where high-quality audio signals can be recorded. There is a certain disadvantage, nonetheless, owing to the relatively high data-rate needed for the faithful representation of a digitally-coded audio signal. A typical net bit-rate for PCM recordings on magnetic tape or on Compact Disc is of the order of 700 kbit/s for each monophonic channel.

Whilst this problem is not unduly severe in audio signal storage systems, difficulties can arise with audio signal transmission because of the limited capacity of the available channels. For the West German DSR system for digital sound broadcasting in 12-GHz DBS channels, the net bit-rate is reduced to 448 kbit/s by using a sampling frequency of 32 kHz and a 16/14-bit scale-factor coding technique [1]. However, a much lower bit-rate will be needed for digital sound broadcasting to portable and mobile receivers and the question arises: how low can we go? We know that the desired accuracy of the received signal determines the minimum bit-rate. However, two questions have to be answered:

If the desired accuracy of signal delivery is known, how can the theoretical minimum bit-rate of the signal be attained? The answer, in principle, is to delete all information which is not necessary for the reconstruction of the signal with the desired accuracy. The **redundancy** has to be zero. The redundant elements of the signal, which are already known at the receiver because of what has happened in the past, need not be transmitted. Audio signals generally exhibit stochastic characteristics, although they can be anticipated, statistically, to a certain degree. This is especially the case with speech signals if the source is well-known. There are other audio signals, however, which are entirely stochastic. A further point to be considered is that the sensitivity to bit-errors is a maximum for signals having no redundancy. For these reasons, the reduction of redundancy is not very effective as a means of reducing transmission bit-rate.

The second question is: what accuracy is desired for the signal delivery? The answer, in principle, will be based on consideration of the limited spectral and time-related resolution capability of the receiver — i.e., the human ear. We say that the **irrelevance** has to be zero. The irrelevant parts of the signal — those fine details of the audio signal which cannot be perceived by the human ear — need not be transmitted. This second question is

* The authors are with the Institut für Rundfunktechnik (IRT).

therefore of a psychoacoustical nature, and it is necessary to know exactly how the ear performs in order to minimise the transmission bit-rate without any perceptible loss of quality.

It will be shown here that the answer to the second question is most valuable as the basis for the design of a low bit-rate coding scheme for high-quality audio signals. The required accuracy of signal delivery is surprisingly low, although it is also extremely frequency and time-dependent.

This implies that the quantizing error (quantizing noise) of a digitally-coded signal ought to be frequency-dependent. Ordinary (broadband) PCM systems (such as 16-bit uniform coding or the 16/14-bit scale-factor technique) produce a broadband, almost white, quantizing noise, the spectral distribution of which is very different to the strongly frequency-dependent sensitivity of the human ear. Only a small reduction in bit-rate can be achieved with broadband digital companding techniques without excessive loss of quality, because the broadband companding noise easily exceeds the frequency-dependent **masking-threshold** of the ear. A "broadband PCM channel" is inherently mismatched to the "frequency-dependent coded channel" of the ear.

In order to match the quantizing noise to the ear characteristics, two different source coding techniques are under investigation. They both use spectrum analysis techniques and frequency-dependent bit-allocation and they both achieve a net bit-rate of about 120 kbit/s per monophonic signal without causing perceptible impairments.

Transform coding

This coding method [2, 3, 4] involves the conversion of a block of consecutive samples into the frequency domain (e.g. Fourier transform or cosine transform). This strategy enables a reduction in the redundancy of the audio signal and also a matching of the quantization to the thresholds of perception of quantizing errors. Only those values of amplitude and phase that are relevant with respect to the masking effects of the human ear are quantized.

Sub-band coding

This coding method [5, 6] divides the broadband signal into a number of sub-band signals with a suitable filter bank [7, 8, 9] and into digital frames between 2 and 8 ms long. Once in each frame, the maximum level attained by each sub-band signal (the "scale-factor") is quantized and transmitted. Each sub-band is quantized with a bit allocation based on the masking thresholds [10]. This method avoids time-frequency-time conversion and therefore reduces the degree of signal processing com-

plexity in the receiver. Also, it avoids problems with time domain windowing and it optimises the distribution not only of the quantizing noise but also of the bit-error noise across the spectrum, again with regard to the perceptibility.

The present article is concerned with this second approach. It will explain the basic principles of the so-called "masking-pattern adapted sub-band coding and multiplexing" (MASCAM) system, describe prototype hardware and examine further refinements to the system which are currently under investigation.

2. Sub-band coding: an effective irrelevance-reduction method

2.1. Principles of masking

The optimum adaptation of a sound transmission channel to the human ear would involve simulation of the internal coding process within the ear, but this cannot easily be realised. Instead, we must use coding techniques that are technically feasible and which take advantage of the noise behaviour of the ear. Such coding can be in the frequency or time domain, may be fixed or adaptive, and may or may not use memory. It can be regarded as truly optimum only if the resulting noise spectrum conforms to the masking threshold [10] of the ear. Thus, in terms of signal theory, the frequency-dependence of the masking threshold can be thought of as corresponding to the spectrum of the coding noise of the auditory communication channel.

Some masking thresholds are presented in *Fig. 1*. The lower part of the figure shows the masking of individual tones by masking narrow-band noise of various sound pressure levels, whilst the upper part shows this for narrow-band noise of three different centre frequencies. It can be seen that loud low-frequency sounds mask quiet high-frequency sounds much more strongly than vice-versa.

A sound signal below the masking threshold of the masking sound will not be perceived and is therefore irrelevant as far as the ear is concerned. This effect of "mutual masking" [10] is particularly evident in broadband sound signals with well-defined formant structures. If the mutual masking is taken into consideration for the coding of an audio signal, then it follows that the portion of the signal lying beneath the masking threshold does not need to be coded and need consume no transmission capacity.

It is helpful to regard the ear as a communications channel and to consider a model in which the ear splits the frequency spectrum into sub-bands, the so-called "critical bands" [10]. The critical bands are defined as characteristic portions of the broadband audio spectrum within which

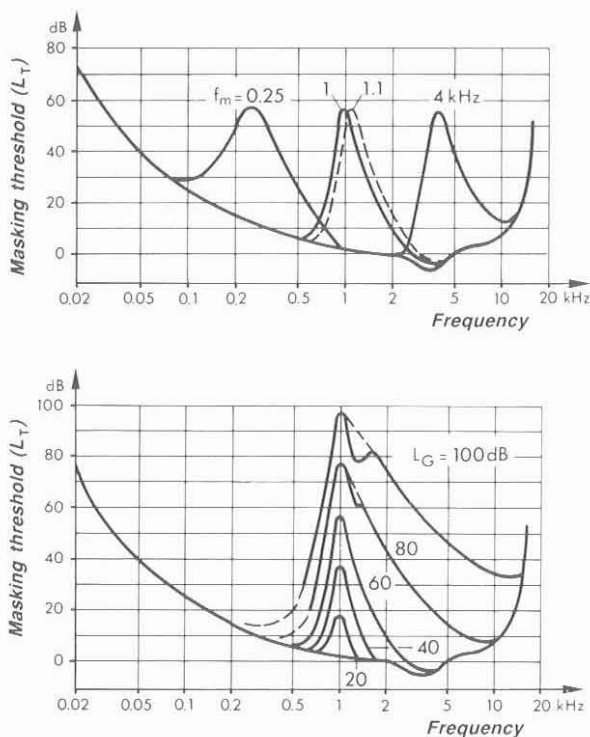


Figure 1
Masking thresholds L_T obtained with narrow-band noise having:

- top : a sound pressure level (SPL) of 60 dB and centre frequencies f_m of 0.25 kHz, 1 kHz and 4 kHz
bottom : centre frequency of 1 kHz and various values of SPL.

specific psychoacoustic regularities are valid, such as frequency selectivity, the masking threshold, etc. If we assume that an optimum coding is performed in each "critical-bandwidth" sub-band, then the complete spectrum of the signal-dependent "coding noise" will correspond, in principle, with juxtaposed partial spectra of the individual critical-bandwidth sub-bands.

From this, an optimum coding concept can be derived. The coding noise can be masked optimally by the sound signal if the signal is not broadband, but instead narrow-band coded since, within each sub-band, the coding can be adapted exactly to the noise performance of the ear.

2.2. Determination of sub-band bit-rates

An example for the required quantizing of a sampled tone is presented in Fig. 2. A 1-kHz tone is present in a sub-band one-third of an octave wide. The critical situation as regards the perception of the quantizing noise occurs when the tone is at the upper cut-off frequency of the sub-band filter, because the slope of the low-frequency flank of the masking threshold is very steep (about 27 dB/critical band) and at the same time is independent of level (see Fig. 1, bottom). This critical case is depicted in Fig. 2. The threshold between detectable and undetectable sub-band limited quantizing noise is determined by the point of intersection of the

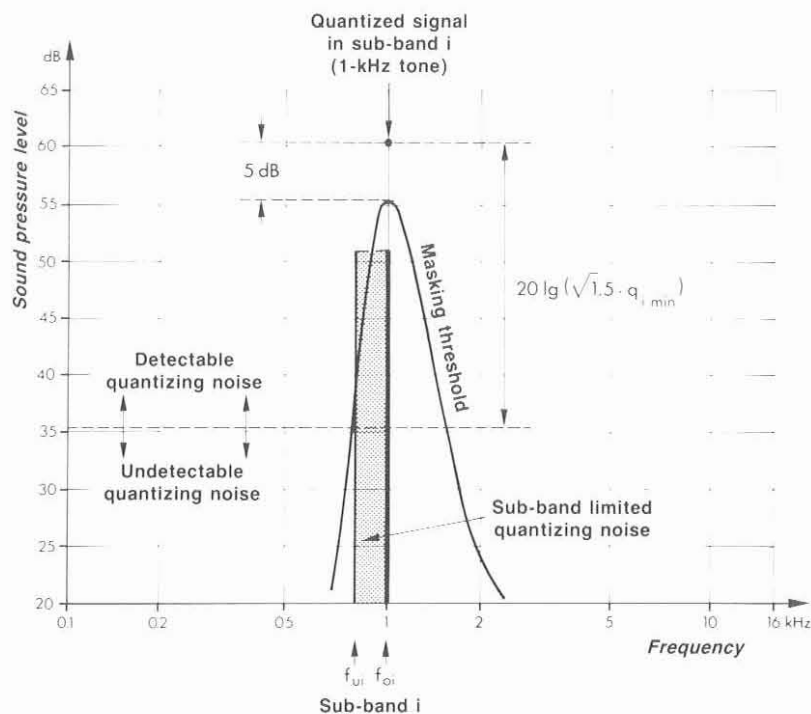


Figure 2
Derivation of the minimum quantizing resolution $q_{l(min)}$ within a sub-band. The most critical case is shown, with a tone at the upper cut-off frequency of sub-band i , one-third of an octave wide.

masking threshold (low-frequency flank) and the lower cut-off frequency of sub-band i . The difference between this threshold level and the signal level will be called the "critical signal-to-noise ratio a_{qi} " of sub-band i . It can be calculated for the most critical case as follows:

$$a_{qi} = 200 \lg (f_{oi}/f_{ui}) + 5 \text{ dB} \quad (1)$$

where f_{oi} and f_{ui} denote the upper (German *oben*) and lower (*unten*) cut-off frequencies of sub-band i .

For q_i quantizing steps, according to [11], the signal-to-noise ratio is given by:

$$a_{qi} = 20 \lg (\sqrt{1.5} \cdot q_i) \text{ dB} \quad (2)$$

The minimum quantizing resolution which is necessary for a sample if the quantizing noise is to be masked can be determined directly from equations (1) and (2):

$$\lg q_{i(\min)} = 32 \lg (f_{oi}/f_{ui}) + 1 \text{ bit/sample} \quad (3)$$

The minimum resolution of a sub-band signal (bits per sample) depends only on the relative bandwidth f_{oi}/f_{ui} of the filter used, and if this decreases towards the higher-frequency sub-bands the required resolution of the corresponding sub-band decreases. Equation (3) demonstrates that this bit-rate could be reduced, in theory, by increasing the number of sub-bands; nevertheless, it will be shown in section 4 that there is another more advantageous way of obtaining additional bit-rate economies.

3. First MASCAM system

Working on the basis of equation (3), a first MASCAM system has been designed and built. The system described in this section is the one used for the EBU demonstrations during the WARC-ORB 88 Conference in Geneva of advanced digital techniques for UHF satellite sound broadcasting, in conjunction with the OFDM transmission system developed by the CCETT [22].

This prototype system uses 24-band partitioning, as shown in Fig. 3. To simplify the construction of the digital finite impulse response (FIR) filter (transversal filter), only two different sub-band bandwidths were used for this first realisation. 16 filters each of 0.5 kHz width cover the range from 0 to 8 kHz, and 8 filters each 1 kHz wide cover the range 8 to 16 kHz. In Fig. 3 the critical S/N ratio of each sub-band is shown (left ordinate), together with the corresponding minimum bits per sample, calculated using equation (3) (right ordinate). At the changeover point between the 16th and 17th sub-bands the absolute bandwidth of the filters is doubled and there is a step in the function defined by equation (3).

The hardware built for the EBU demonstrations functions at a bit-rate of 112 kbit/s, excluding the bit-rate needed for transmission of the scale-factors [1] (24 kbit/s) and for bit-error protection. Schematic diagrams illustrating the send and receive

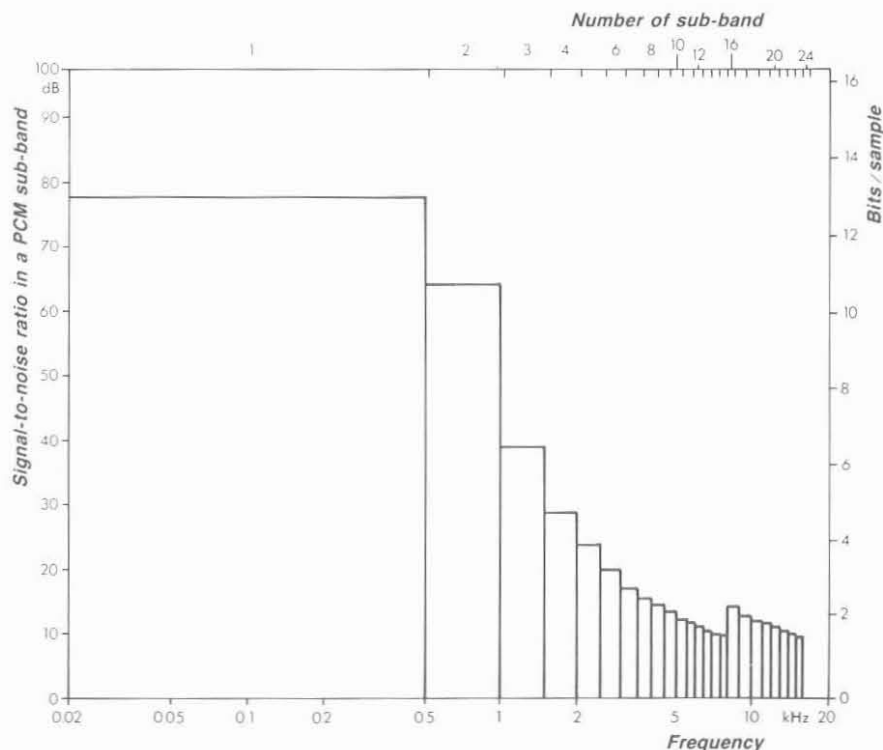


Figure 3
Critical signal-to-noise ratio and required quantization of samples in the sub-bands in a scheme with 24 sub-bands, as derived from equation (3). The preliminary scheme has two constant sub-band widths of 0.5 and 1 kHz.

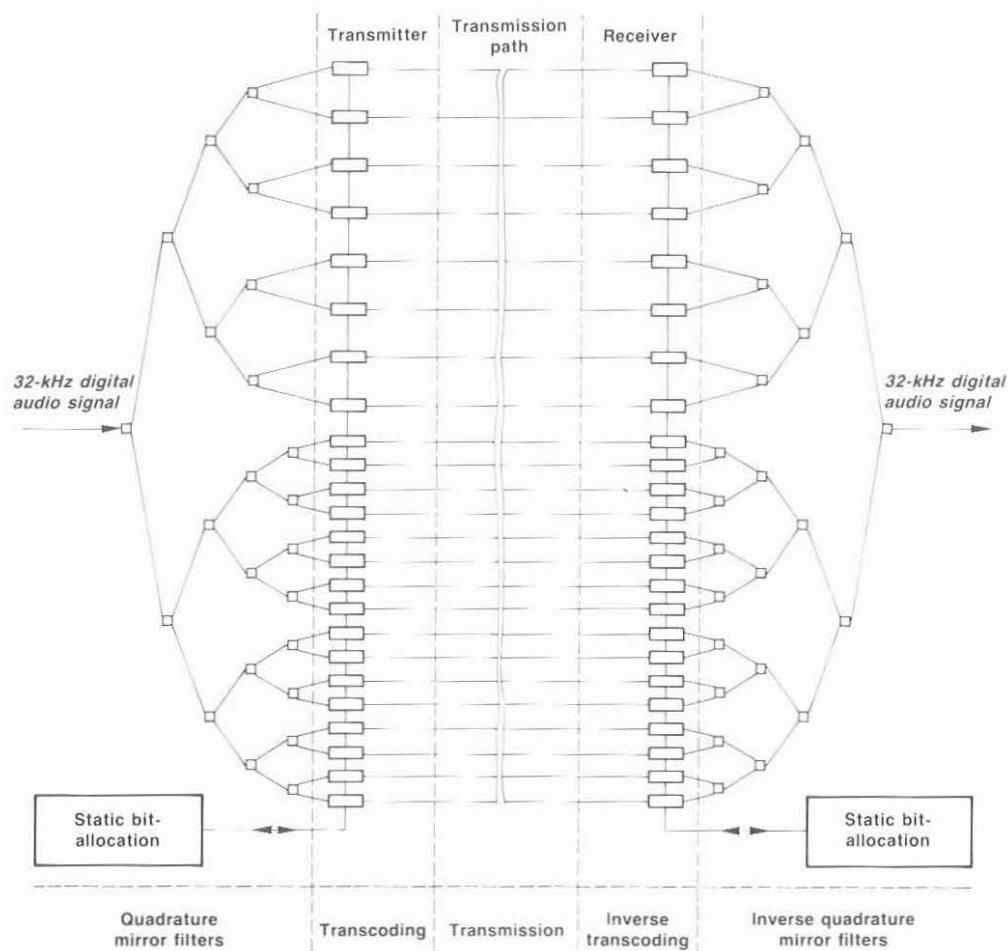


Figure 4
General structure of the preliminary MASCAM scheme, with static bit-allocation to 24 sub-bands.

functions are presented in *Figs. 4, 5 and 6*. The first stage of the encoding process is filtering (*Fig. 4*). The second stage is transcoding, in which the initial data reduction takes place. The band-split audio signal, now largely free of irrelevance, is then transmitted (stage 3). Stages 4 and 5 take place in the receiver, and consist of inverse transcoding and inverse filtering, after which the 16-bit uniform coded audio signal is again available.

3.1. Source and channel coding

3.1.1. Band-splitting

The splitting of the incoming audio signal into the 24 sub-bands is achieved using a quadrature mirror filter bank (QMF) [7, 8, 12, 13]. These filters make complete utilization of the sampling frequency possible, with a minimum of aliasing and exceptionally small ripple. The sampling frequency required for each sub-band filter is exactly twice as large as the bandwidth of the filter. The aliasing distortions

resulting from the selectivity of the filters are completely eliminated by the inverse filtering at the decoder, provided that there is uniform quantizing in all the sub-bands [12]. In principle, this criterion is not satisfied where different quantizing is used for adjacent sub-band signals. However, measurements relating to this problem have shown that the differences in resolution in adjacent sub-bands, as shown in *Fig. 3*, will only produce aliasing distortions which are masked due to the masking threshold. This is valid for any programme signal. With a view to reducing the complexity of the MASCAM decoder, investigations are under way to examine the suitability of the so-called wave digital filter technique [9, 14].

The filter banks use a sampling frequency of 32 kHz and cover an audio base-bandwidth of 16 kHz. This sampling frequency has been chosen only for convenience for the first MASCAM prototype using static bit-allocation. In the final version using dynamic bit-allocation, a sampling frequency of 48 kHz will be used to ensure compatibility with the studio standard [15] (see section 7.2).

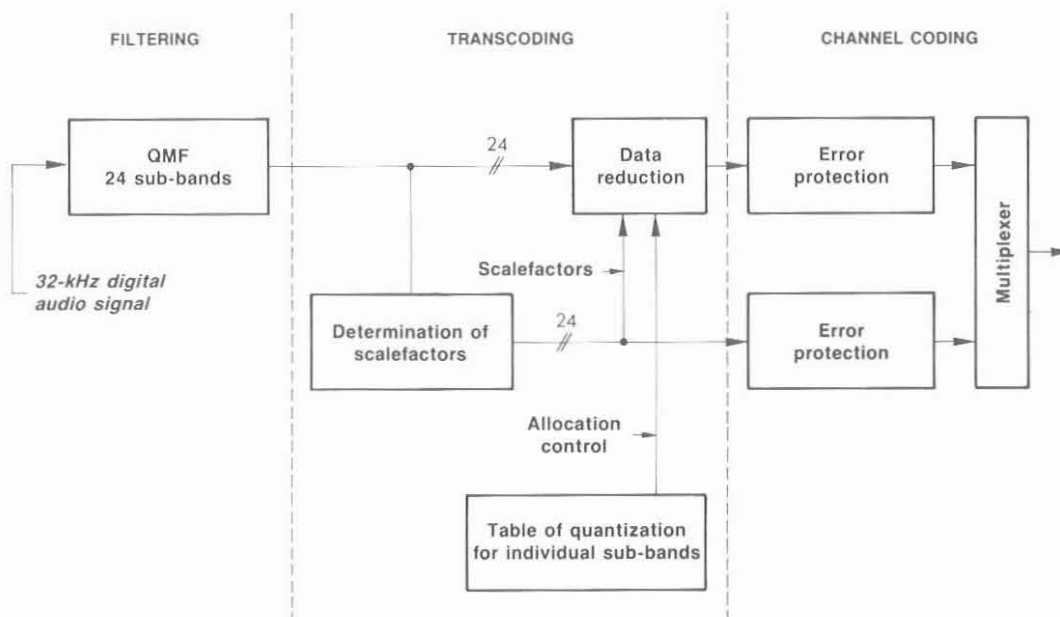


Figure 5

Block diagram of the MASCAM coder with static bit-allocation used for the demonstrations at the WARC-ORB 88.
 Bit-rate of multiplex signal: 168 kbit/s Bit-rate of samples + scale-factors: 136 kbit/s
 Sampling frequency: 32 kHz

3.1.2. Transcoding

The basic operations needed for the transcoding process are shown in Fig. 5. After band-splitting, a scale-factor is determined for each of the sub-band signals and for a block of samples. The block-length is 8 ms for sub-bands 1 to 16 (0 to 8 kHz) and 4 ms for sub-bands 17 to 24 (8 to

16 kHz). The scale-factor indicates the maximum level attained by the sub-band signal within the block. The transcoding of the 16-bit sub-band signals depends on the scale-factor information and on the number of bits allocated to each sub-band. In the case of static bit-allocation, the allocation control information for each sub-band established according to equation (3) is held in a look-up table

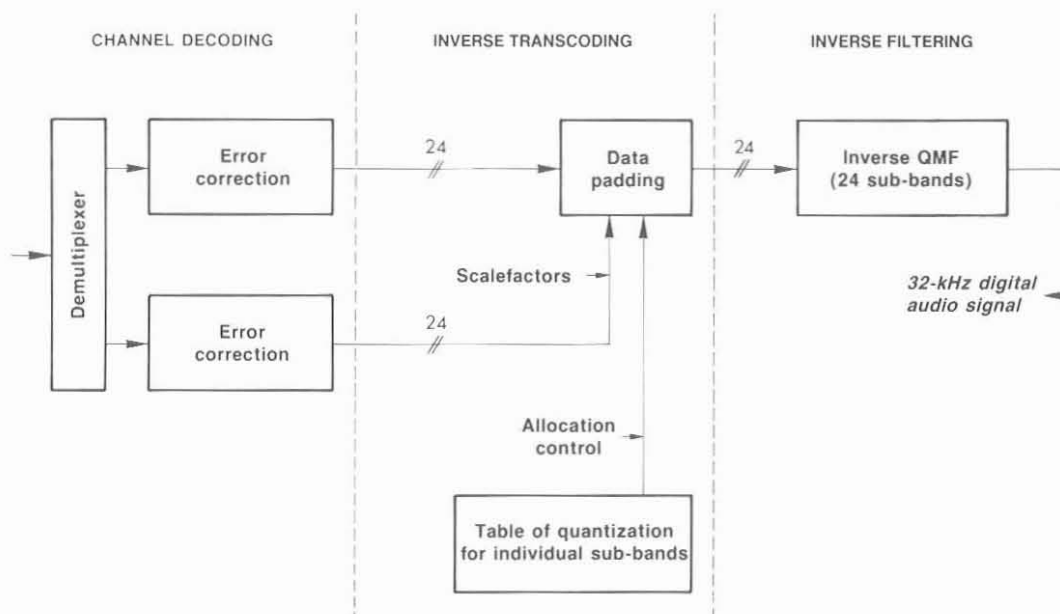


Figure 6

Block diagram of the MASCAM decoder functioning with the coder of Fig. 5.

provided in both the transmitter and receiver (Figs. 5 and 6).

3.1.3. Error protection

In order to obtain an undistorted transmission over a non error-free transmission channel, it is necessary to apply greater protection to the scale-factors than to other parts of the transmitted signal. The prototype channel coder and decoder use a channel code which provides equal error-protection for all bits, so an additional error-protection is implemented for the scale-factors in the source coder and decoder. Furthermore, it has been found that it is useful to apply greater protection to the most-significant bits (MSB) of the lower, rather than the higher, sub-band signals. Therefore the MSBs of the five lowest-frequency bands are protected (in sub-bands 1 and 2, the two MSBs are protected and in sub-bands 3, 4 and 5 the MSB alone is protected). As the maximum bit-rate of the channel coder input signal is 168 kbit/s, the code-rate (ratio of useful information-rate to total data-rate) of the complete MASCAM multiplex signal is 0.81.

For the protection included in the source coding/decoding, a (23,12) Golay code [18] with a Hamming distance of 7 is used. Thus correction is possible for up to 3 bit-errors in a 23-bit word. To achieve detection of up to 4 bit-errors, a parity bit is added to the 23-bit word. As shown in Fig. 7, the scale-factors and the MSBs are put together in 12-bit words which are protected by the additional 11 bits of the Golay code and the parity bit. The whole of this processing is performed inside the

signal processors by using several look-up tables, so no additional hardware is needed for the error-protection.

3.2. Hardware complexity

The filter bank consists of 23 quadrature mirror filters. In each filter the spectrum of the input signal is divided into two data streams, each having one-half of the bandwidth and one-half of the sampling rate of the input signal. The 64-tap filter elements process two input samples at a time to provide one sample at each output; i.e. the average processing effort is 32 multiplications and additions for one input sample. Fig. 4 shows that all audio data pass through four filter stages, and the lower half of the spectrum passes through a fifth one. The first stage is a single filter working at the full sampling rate, the second stage comprises two filters each working at one-half of the original sampling rate, etc.

The total calculating effort for the filtering processes is 4.5 times 32 operations for each input sample. With a sampling rate of 32 kHz, about 4.6 megaflops (mega floating-point operations) are required in the filter banks of the coder. The inverse filter elements in the decoder require the same processing capacity.

After filtering, the following steps are required in the coder :

- find the greatest absolute value among 8 samples, for each sub-band;

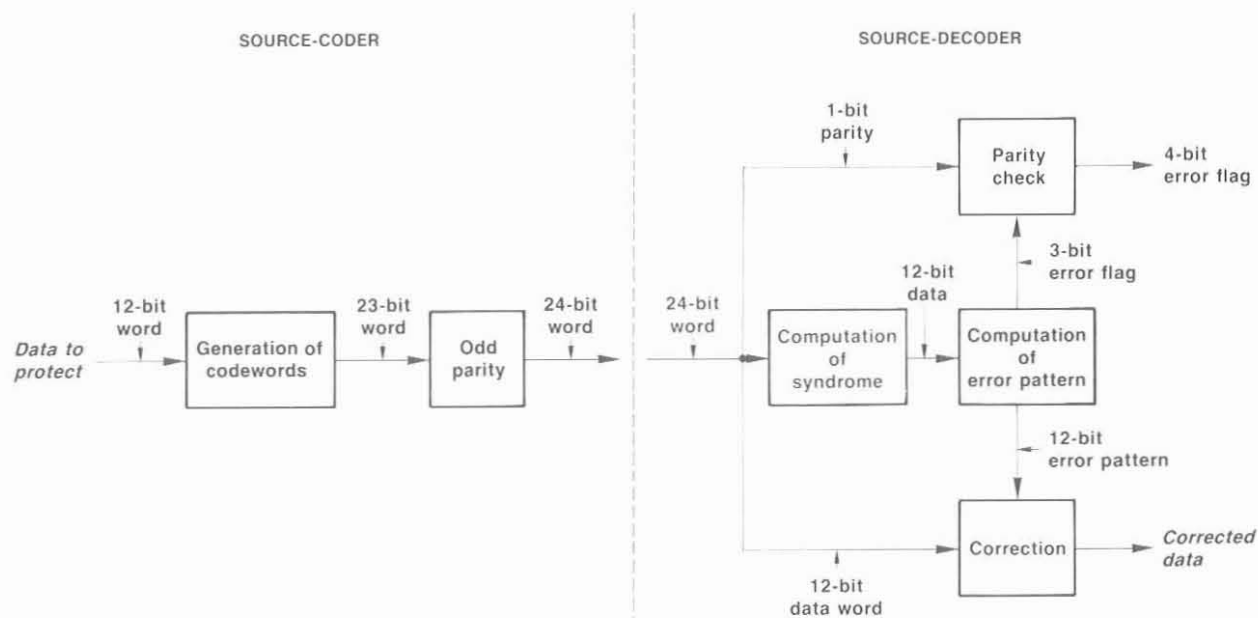


Figure 7
Error-protection of scale-factors and MSBs.

- classify these maxima to 1.5 dB accuracy (6-bit scale-factor);
- classify 8 samples in a block with regard to the scale-factor and to the sub-band-related accuracy of linear quantization;
- protect the scale-factors and MSBs of the 5 low-frequency sub-bands with the Golay code;
- assemble a continuous data stream of scale-factors and numbers of steps from one and two-byte locations in the signal processor's RAM.

In the decoder the data first have to be stored in addressable RAM locations. The next step is error correction and then the scale-factors and numbers of steps must be transcoded to floating-point values before they are passed to the inverse filter bank.

The MASCAM coder and decoder are based on the DSP 32 signal processor manufactured by AT&T, within a development system designed by Fraunhofer Gesellschaft Erlangen. One DSP provides 6.25 megaflops. In addition to the theoretical demand of the filter banks for 4.6 megaflops, a great many processor instructions are necessary for setting up the filter bank structure and for data transfer via the serial ports of the signal processors. About 95 % of the capacity of a signal processor is taken up by the complete filtering function (transmitter and receiver), so a second processor is used in the transmitter for the transcoding, Golay error-protection and data stream assembly, and in the receiver for the dis-assembly, error correction and inverse transcoding. About 30 % of the second processor in the transmitter, and about 20 % of that in the receiver, are occupied by these functions.

The complete hardware therefore comprises four signal processors in the stereo transmitter, and four in the stereo receiver. The average occupation of the processors in the transmitter is 63 %, and in the receiver it is 58 %. This calculation effort could be reduced in the future MASCAM systems by reducing the number of coefficients in certain filters, reducing the number of sub-bands and using wave digital filters (see section 7.4).

Additional hardware components in the prototype transmitter and receiver are a timing circuit giving bit and word clock signals, and a frame synchronizer to lock the MASCAM signal to the modulation system provided by the CCETT for the demonstration in Geneva. The audio data at the source coder input and decoder output are delivered via standard EBU/AES digital audio interfaces.

4. Optimizing the bit-rate reduction

It has already been shown that the elimination of irrelevant detail in a signal is a powerful means of achieving bit-rate reduction. The required accuracy of signal representation is in fact surprisingly low,

even without employing all the techniques that are possible. In this section we will consider how the reduction can be optimized to suit specific system applications.

4.1. Codec used for simulations

The investigations on further bit-rate reduction possibilities are based on the codec shown in Fig. 8. The following characteristics have **not** been optimized as they would be in a final real-time version, such as that described in section 7:

- In order to achieve a frequency distribution of the quantizing noise which can be matched optimally to the masking threshold, a 24-band partition has been realised using quadrature mirror filters. The smallest sub-bands (125 Hz wide) are used over the frequency range from 0 to 1 kHz, and 2-kHz sub-bands are used from 8 to 16 kHz.
- The side information includes the scale-factors but **not** the bit-allocation control data or the channel coding status data. The bit-allocation control data are determined at the coder and at the decoder, on the basis only of the scale-factors.

4.2. Scale-factors

4.2.1. Temporal masking effect

A third step in the bit-rate reduction process may be defined in terms of the temporal masking effects of the ear [10]. In addition to "simultaneous" masking (masking of one sound by another occurring at the same time), two other phenomena may be discerned: pre-masking and post-masking. Exploitation of these phenomena allows the quantizing noise to be spread over a certain time interval without audible degradation. The maximum block length for the validity of a certain scale-factor is defined mainly by the pre-masking, which is the period just before a masking signal appears (see also [5]). The duration of the pre-masking period is about one-tenth of that of the post-masking period, and is typically between 50 and 200 ms.

4.2.2. Bit-rate reduction

Quantization of the scale-factors with 6 bits gives a resolution over the whole dynamic range of a sub-band signal which varies in steps of 1.5 dB. If the scale-factors of every sub-band are transmitted every 4 ms, the data rate is 36 kbit/s. These scale-factors are especially important in the decoding process, so they must be well protected against bit-errors by the introduction of redundancy in the channel coder. The application of bit-rate reduction to the scale-factors is therefore essential, in view of

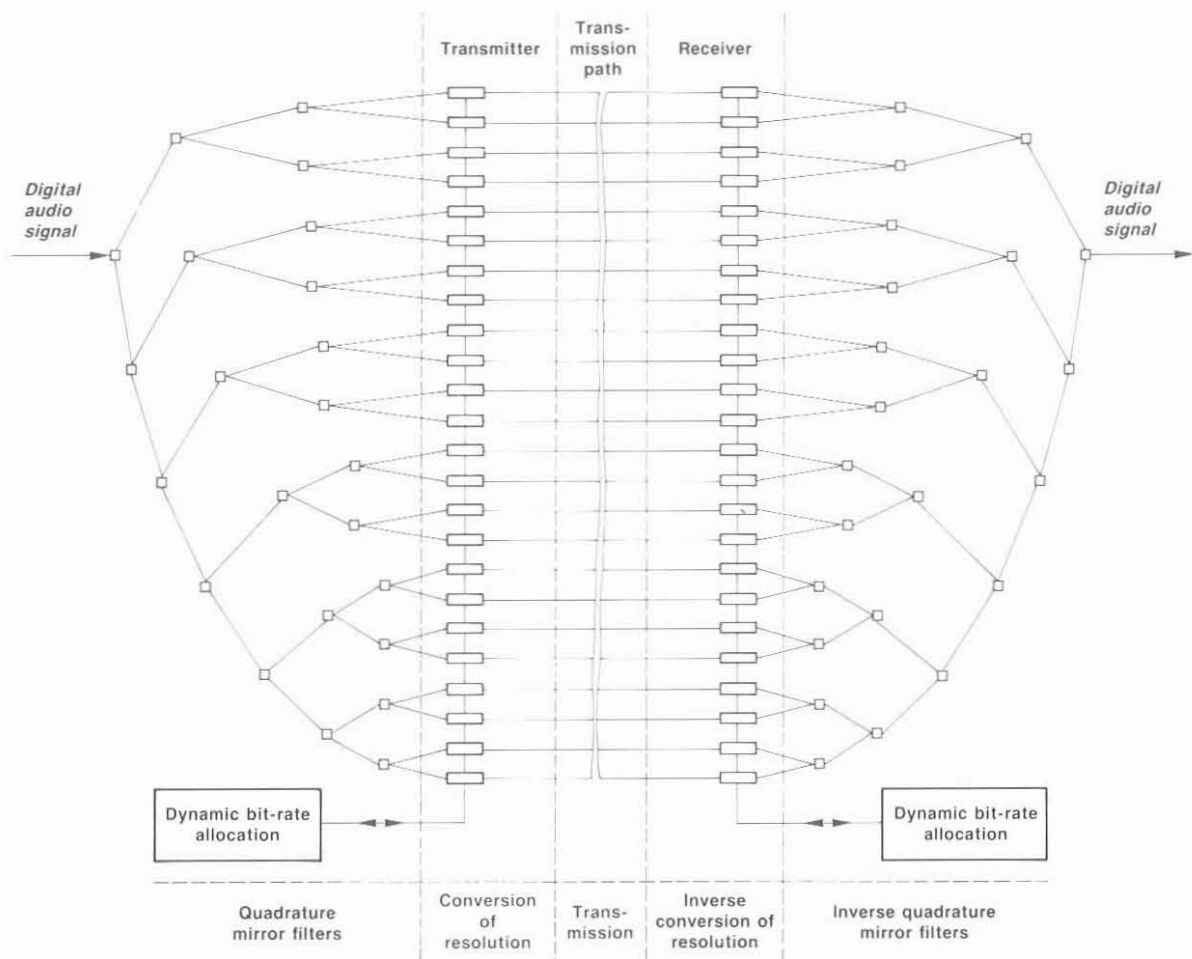


Figure 8
General structure of the MASCAM codec using dynamic bit-allocation to 24 sub-bands, used in the current simulations.

the large proportion of the total channel bit-rate which they will otherwise consume. The temporal masking characteristics of the ear are exploited for this purpose.

In a time window having a length up to 64 ms (this value depends on the specific application of the MASCAM scheme), each scale-factor is examined to determine its "subjective importance". Scale-factors located at fast increases in signal level are the most important ones to transmit because the ear is especially sensitive to fast impulses [20]. The post-masking effect is used by the scale-factor bit-rate reduction algorithm. Generally, fast decay curves of signal levels do not have to be reconstructed as accurately in the decoder.

It is therefore not necessary to transmit all 24 scale-factors in every 4-ms block. A complete new set is nevertheless required every 4 ms in the decoder, which is instructed to interpolate interim values for those that are not transmitted, on the basis of those that are.

Fig. 9 shows typical temporal changes in scale-factors for three sub-bands. Case A shows the variations without data reduction, whereas in case B the bit-rate was reduced from 36 kbit/s to 8 kbit/s. The strong impulses in the scale-factors in the higher-frequency sub-bands represented in Fig. 9 illustrate that a fast increase in signal level can be reproduced accurately. The decreases in level are reproduced only approximately and there is some "smoothing" of the decay which nevertheless corresponds to the post-masking effect [21].

Several algorithms for data-rate reduction of the scale-factors have been devised. Listening tests are necessary to determine the permissible degree of data reduction. It can be expected from the first results with various types of audio signal that a bit-rate of about 15 kbit/s will be sufficient for the transmission of 24 scale-factors, and in the remainder of this article this value will be assumed when calculating the total bit-rate of MASCAM processed signals.

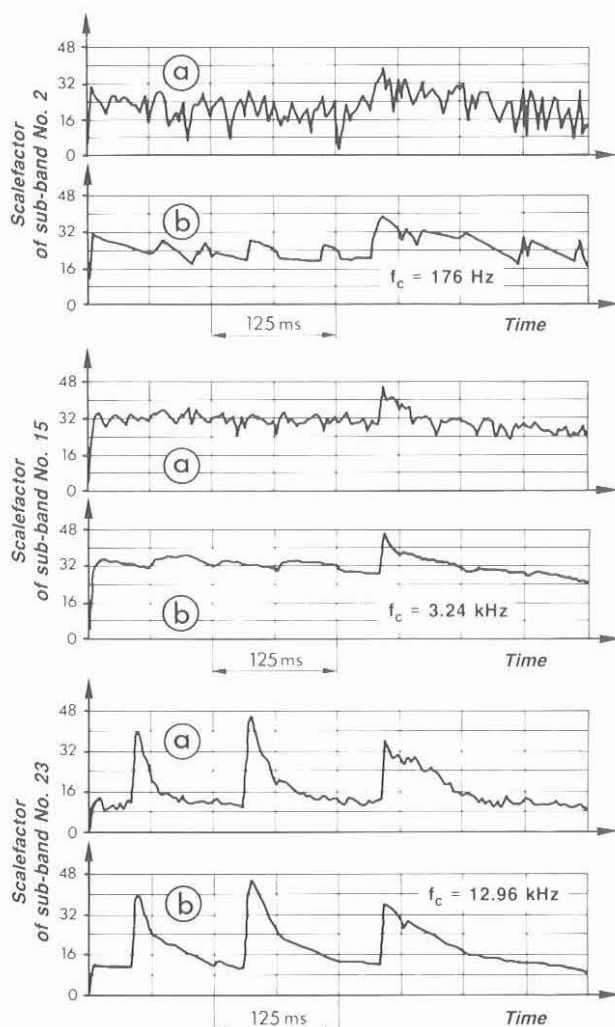


Figure 9

Temporal variations in the scale-factors (64-step resolution, calculated every 4 ms) of three sub-band signals for an excerpt of "pop" music.

A - without data reduction (36 kbit/s)
B - after reduction to 8 kbit/s.

4.3. Dynamic bit-allocation

4.3.1. Principles

A further reduction of the irrelevance within the audio signal can be expected if, in addition to the masking characteristics **within** each sub-band, the masking effect of **neighbouring** sub-band signals is also taken into account. The available information processing capacity can then be allocated optimally to the individual sub-bands.

To illustrate this, *Fig. 10* shows the amplitude spectrum (circles) and the "masking pattern" L_T of a steady vowel | ə | consisting of 20 harmonics having a fundamental frequency of 200 Hz. The masking pattern, which is the result of the mutual masking of all the harmonics, is calculated by an appropriate algorithm [19] based on psychoacoustical data representing the masking effects of the ear.

Two important psychoacoustical phenomena are apparent in *Fig. 10* and these can be exploited in the source coding procedure. First, all the harmonics of the vowel which come beneath the masking threshold do not need to be transmitted. This means that any sub-band which contains masked — i.e., inaudible — components of the wanted signal can be set to zero. Such sub-bands then make no claim on the information processing capacity. Second, the quantizing noise can be distributed across the signal spectrum so that it always comes beneath the resulting masking threshold as shown in the figure. This gives a quantizing noise which is optimally adapted to the masking pattern of the audio signal.

The criteria for the perception of aliasing distortions may be considered with reference to *Fig. 11*. As mentioned in section 3.1.1, aliasing distortions will occur when different quantizing is

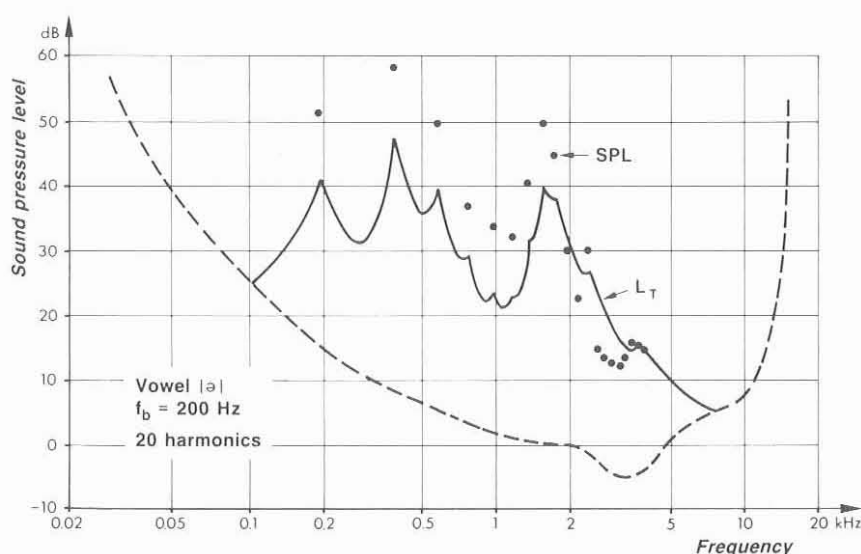


Figure 10

Amplitude spectrum SPL (circles) and masking pattern L_T (full line) of the vowel | ə | containing 20 harmonics.

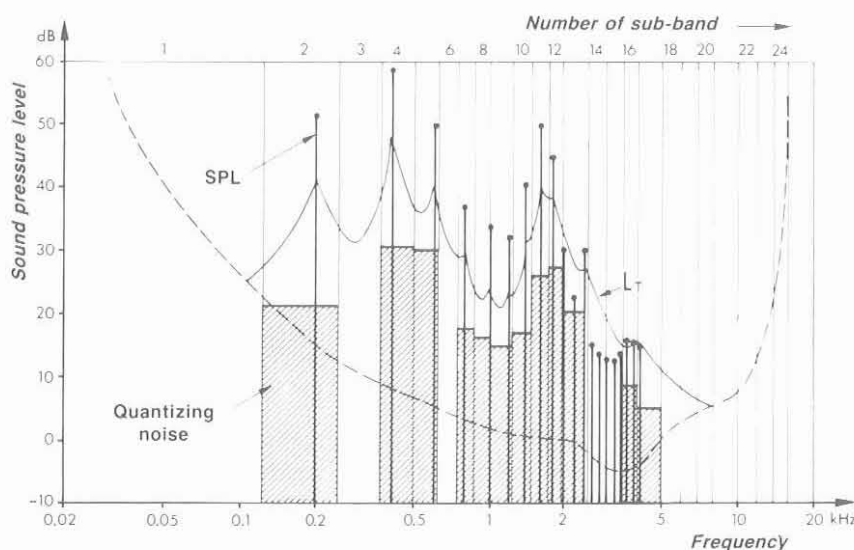


Figure 11
Amplitude spectrum SPL
of the vowel [æ] shown in
Fig. 8, the masking pattern L_T
and the quantizing noise
(hatched bars), with a bit-rate
of 16 kbit/s for the audio samples
and 20 kbit/s
for the scale-factors.

applied in neighbouring sub-bands. This will be especially valid in respect of any sub-bands that are set to zero (sub-bands 3, 6, 14 and 15 in the case in hand). Nevertheless, simulations have indicated that the levels of the resulting aliasing distortions in the sub-bands do not exceed the masking pattern and that they are comparable with the level of the quantizing noise. To illustrate the mechanism leading to this result, let it be assumed that the first harmonic (in sub-band No. 2) ceased to exist; the dynamic bit-rate allocation would ensure an adequate bit-rate in the sub-bands that are carrying relevant information. There would be no penalty as regards the overall bit-rate reduction, because the bit-rate available for the relevant information would be increased by an amount corresponding to the bits that are "saved" in the sub-bands that are set to zero.

4.3.2. Bit-rate reduction achieved through dynamic bit-allocation

Dynamic bit-allocation provides an additional bit-rate reduction over and above that derived from equation (3). The transmission bit-rate r_i required in each sub-band i varies over a limited range. The minimum value is zero and the maximum depends on the sub-band bandwidth, as formulated in equation (3)

$$r_i = 2 (f_{oi} - f_{ui}) + \text{ld } q_{i(\min)} \text{ bit/s} \quad (4)$$

However, the greatest amount of net transmission bit-rate r that is required is less than the sum of the individual sub-band rates, $\sum_1^{24} r_i$, because the relevant masking threshold for sub-band i does not result from sub-band signal i but, instead, from sub-band signal $i-1$. The relevant masking threshold is therefore that on the high-frequency flank, which is less steep (see Fig. 1). Assuming the ratio

of the slopes of the upper and lower flanks to be about 2, equation (3) has to be modified as follows:

$$\text{ld } q_{i(\min)} = 16 \lg (f_{oi}/f_{ui}) + 1 \text{ bit/sample} \quad (5)$$

On this basis, the required net bit-rate can be estimated for the signal containing the greatest amount of relevant information. The estimated value is $r \approx 95$ kbit/s. The value obtained from equation (3) is 112 kbit/s. The dynamic bit-allocation therefore offers a bit-rate saving of about 20 kbit/s.

The 24-band partitioning described in section 3 will serve as an example illustrating the advantages of dynamic bit-allocation to individual sub-bands. Fig. 11 shows the levels of quantizing noise in the sub-bands (hatched bars), in relation to the masking pattern, in the case if an available transmission bit-rate of 16 kbit/s for the samples plus 20 kbit/s for the scale-factors. A higher rate, for example 30 + 20 kbit/s, would reduce the quantizing noise enormously (Fig. 12). The fact that some sub-bands can be set to zero — 12 sub-bands in the present case — demonstrates that the relevant information content of the audio signal is not very high. It also shows that audio signals containing the greatest amounts of relevant information will have the smallest "masked-to-noise" ratios. In most cases, the masked-to-noise ratio in the sub-bands will be much greater than zero [6].

Using this technique, the quality of the coded audio signal, compared to that of a reference audio signal, will not be impaired subjectively. This would be true even with possible (future) digital audio signals having better than 16-bit linear resolution, such as an 18-bit linear system. The required bit-rate of this source coding scheme is determined exclusively by whichever audio signal contains the greatest amount of relevant information.

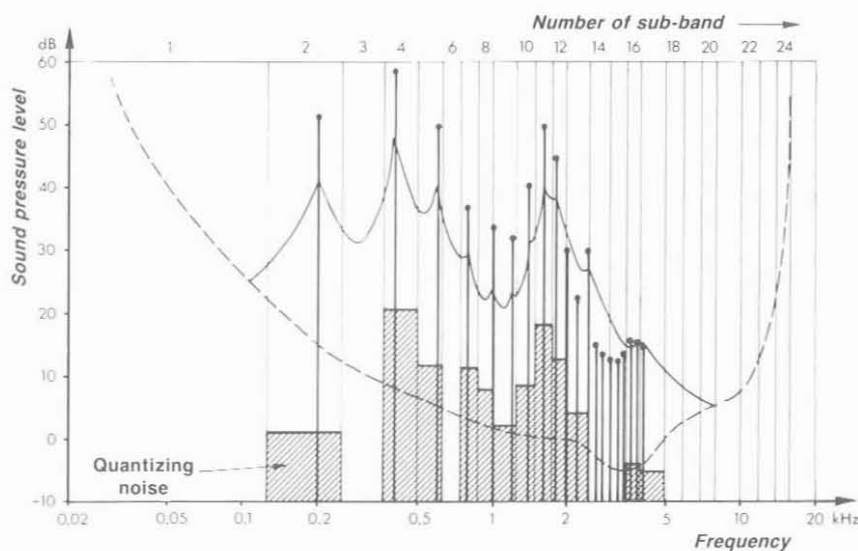


Figure 12

As Fig. 11, but with a bit-rate of 30 kbit/s for the audio samples and 20 kbit/s for the scale-factors.

4.3.3. Dynamic bit-rate

The dynamic bit-allocation controller determines the resolution necessary for the samples of each individual sub-band, on the basis of the calculated masking threshold. The quantizing is arranged to ensure that the ratio of the masking threshold to the quantizing noise level is approximately the same in every sub-band. The so-called "mask-to-noise" ratio will then be approximately frequency-independent. If the noise levels are below the masking threshold, there will be a "mask-to-noise reserve". To achieve a constant mask-to-noise ratio in the case of signals that vary with time, the bit-rate of every sub-band changes with each new 4-ms block. The dynamic bit-rate therefore depends mainly on

the masking thresholds derived from the spectral and temporal structure of the audio signal.

Narrow-band audio signals require only small bit-rates. With such signals many of the sub-band signals will be below the masking threshold and need not be transmitted. Broadband audio signals, and especially those with high levels in the high-frequency range require a higher bit-rate because a larger number of sub-bands will contain relevant information that does need to be transmitted.

A critical case in which minimal masking occurs is represented by a signal with spectral components as shown in Fig. 13. This is an in-harmonic complex tone consisting of 12 partials, and has been identified as the "worst-case" signal. The frequencies of

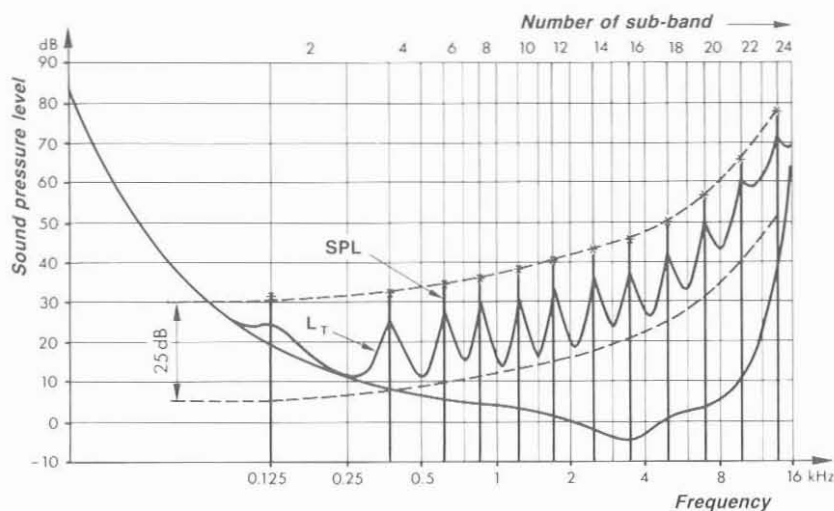


Figure 13

Amplitude spectrum (stars) and masking pattern L_T (line) of an inharmonic complex tone of 12 partials occupying all 24 sub-bands (worst-case signal).

this tone are situated just at the cut-off frequencies of the sub-bands. The levels of the low-frequency partials have been set to low values, so they create minimum masking. The curve in Fig. 13 shows the masking pattern of the complete complex tone. Generally, the upper flank of a masking pattern has a more pronounced masking effect than the lower flank [10]. In the case of the complex tone of Fig. 13, the strong masking effect of the upper flank can be exploited only to a limited extent because the level increases towards the higher frequencies.

It can also be seen from Fig. 13 that a minimum signal-to-noise ratio of about 25 dB is required. In sub-bands at the high frequencies, however, the values of signal-to-noise in the individual sub-bands are lower. On the basis of these individual values, the dynamic bit-allocation controller has to provide individual bit-rates to each sub-band, with the total amounting to 95 kbit/s. The controller provides an additional bit-rate of 15 kbit/s for the side information. Therefore, with a total bit-rate of 110 kbit/s, this critical audio signal can be transmitted without any perceptible quality impairment.

In the case of natural microphone signals, the necessary bit-rate varies with time. Figs. 14 and 15

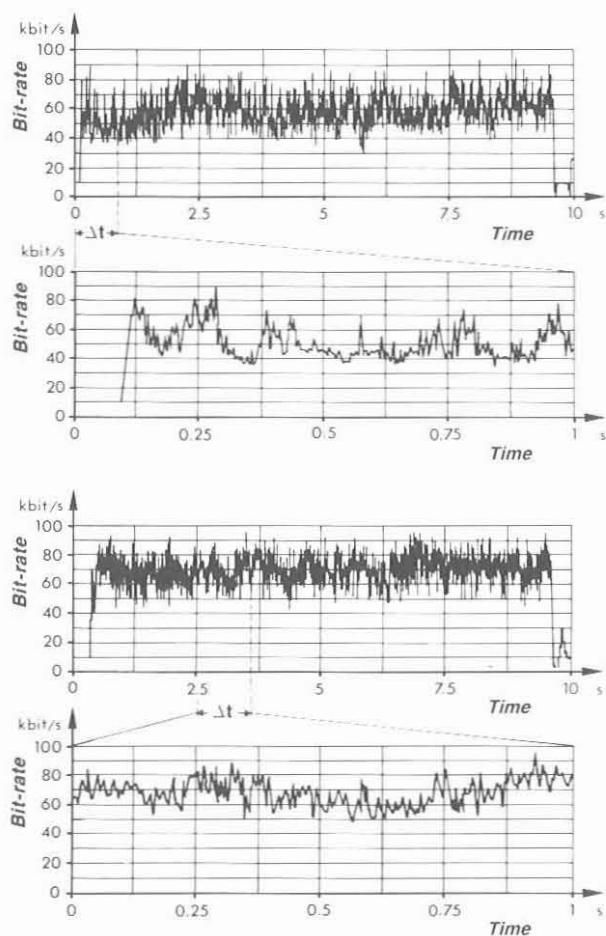


Figure 14

Reduced bit-rate of audio samples (excluding side information) as a function of time.

top : chamber music
bottom: "pop" music.

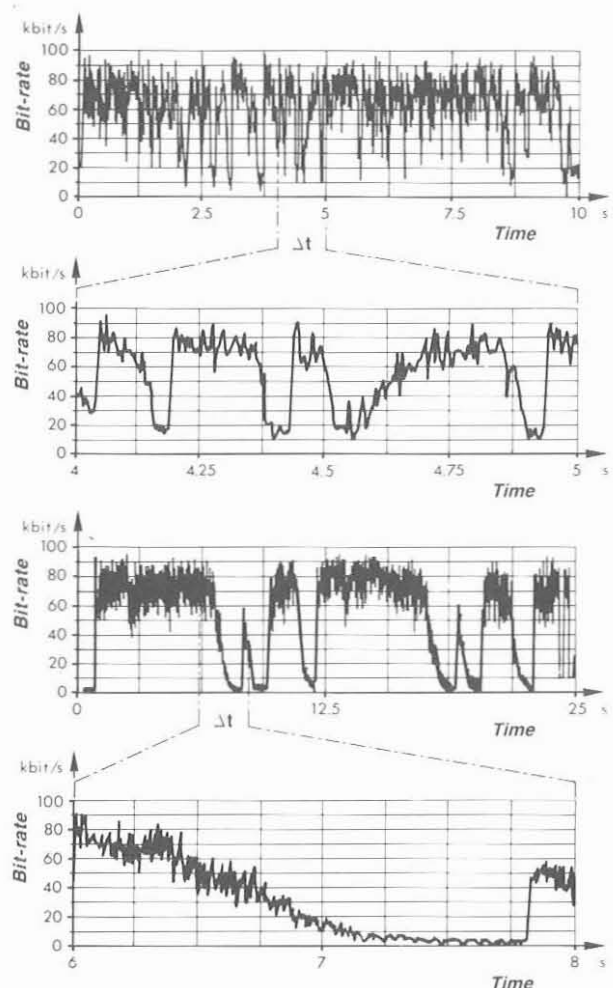


Figure 15

Examples of bit-rate variations with extreme variations.
top : speech (female newscaster)
bottom: solo piano.

show the variation of the required sample bit-rate (excluding the side information) for various forms of sound signal (chamber music, solo piano, pop music and newscaster). Large differences are apparent, and it is also seen that a bit-rate in excess of 90 kbit/s is seldom required.

In the example "pop music", the bit-rate varies only very slightly (Fig. 14, bottom). In contrast, the examples "newscaster" and "piano" produce a wide range of required bit-rate values (Fig. 15). For example, the decay of piano chords is characterized by a spectrum containing no high frequencies. Such high-frequency components require a high amount of bit-rate, so the resulting bit-rate is exceptionally low in this case.

4.4. Buffering

It has been shown that the instantaneous data-rate of the source-coded MASCAM signal will vary quite widely. A data-rate of about 100 kbit/s will occur only occasionally and for very short periods. When a data buffer is used, the spread of the

variable bit-rate can be averaged over a period of time and the data channel capacity may be smaller than that demanded by the peak demand of the unbuffered signal from the transcoder. Typical variations of bit-rate with time for various buffer sizes (i.e. various delay times) are shown in Fig. 16. The bit-rate variation in the one-second excerpt shown, when no buffer delay is used, is 46 kbit/s. The variation decreases to 37 kbit/s with a buffer length of 20 ms. In this particular example an averaging time of 500 ms provides an almost constant bit-rate and a maximum bit-rate of only 60 kbit/s is required (86 kbit/s without any buffer).

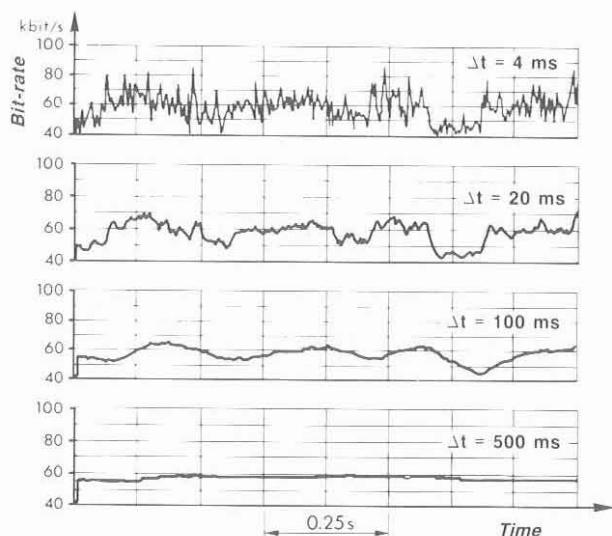


Figure 16

Bit-rate variations with different buffer delays (Δt ms).

In Fig. 17 the maxima of the bit-rate are shown for different buffer lengths. If there is no buffer, the peak bit-rates are independent of the audio signals at about 100 kbit/s. As the buffer length is increased slightly, the peaks fall and certain differences become apparent in the bit-rates for different audio signals. As expected, these differences become more marked as the buffer length is made longer. As shown in Fig. 17, the gain in bit-rate reduction is already quite high (approximately 10 kbit/s) with a buffer giving a 40 ms delay, compared to that which can be obtained with much larger buffers.

Audio signals with high signal levels in the high-frequency range, such as pop music or big-band jazz, require higher bit-rates without buffer delay than other examples without such large high-frequency components (for comparison, see the steady "worst-case" signal in Fig. 13). Other audio signals which have short pauses or a relatively narrow-band amplitude spectrum, such as piano chords or speech, can save up to 30% of the bit-rate if a

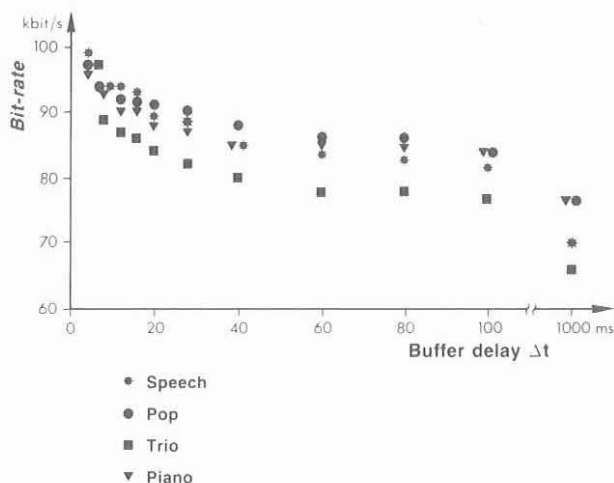


Figure 17

Maximum bit-rate of four audio signals as a function of the buffer delay.

longer delay of up to 1 second is used. An analysis of a radio drama has shown that the maximum bit-rate of 95 kbit/s without a buffer can be reduced to 67 kbit/s when applying a 1-second delay. If a long delay of 20 seconds is used, a bit-rate of only 41 kbit/s results. In the broadcast transmission situation, the delay should perhaps be minimised, whereas in a storage situation even longer delays would be acceptable. In view of the different evaluations of certain parameters according to the application, it seems appropriate to design a VLSI MASCAM decoder chip which can be adapted to a variety of MASCAM formats.

5. Exploitation of the dynamic bit-rate margin

As illustrated above, the necessary bit-rate of MASCAM-processed audio signals depends on the spectral and temporal distribution of the audio signal energy, on the buffer delay, and on the desired quality level. If the buffer delay is about 40 ms, the resulting maximum bit-rate is 90 kbit/s (see Fig. 17). This maximum bit-rate will occur with microphone signals containing the greatest amount of relevant information. With the steady worst-case signal of Fig. 13, the bit-rate is 95 kbit/s (excluding 15 kbit/s for the side information). Most audio signals contain smaller amounts of relevant information and the total bit-rate required is less than 110 kbit/s. Although the bit-rate of the MASCAM source-coded signal is signal-dependent, the bit-rate q_m of the MASCAM multiplex signal which is to be transmitted should nonetheless be constant, say at 128 kbit/s. A so-called "dynamic bit-rate margin" results from this situation.

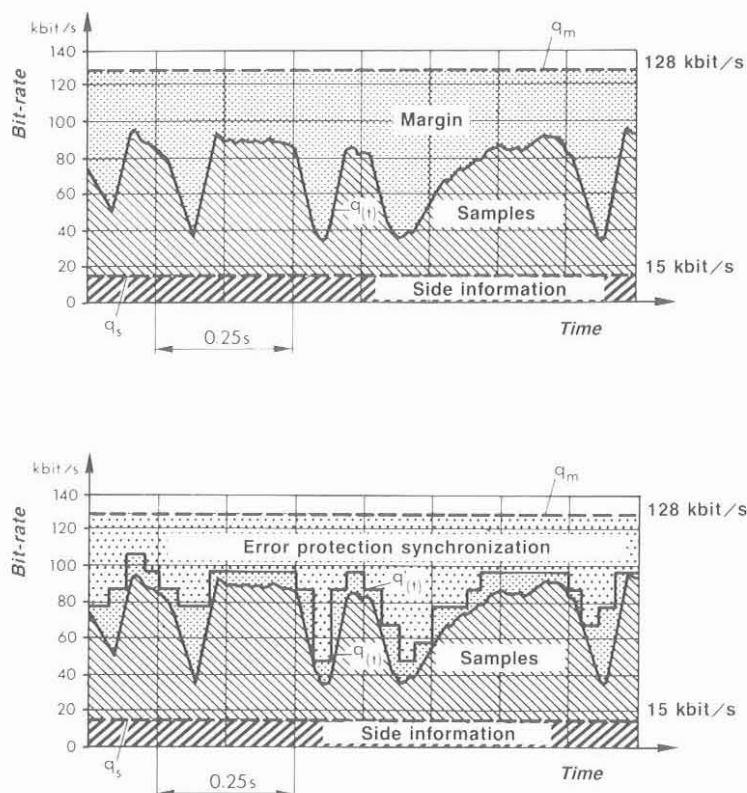


Figure 18
Composition of the MASCAM multiplex signal q_m .

q_s = side information
 $(q_t - q_s)$ = samples
 $(q_m - q_t)$ = dynamic bit-rate margin.

The example for q_t is derived from Fig. 15 (speech) with a buffer delay of 40 ms.

top : Dynamic bit-rate margin (dotted area), averaging 60 kbit/s assuming a transmission bit-rate of 128 kbit/s.

bottom: Insertion of synchronization information and error-protection in the dynamic bit-rate margin. The grade of error protection depends on q_t and varies step-wise.

As an example, Fig. 18 (top) shows the required bit-rate q_t for the samples (test signal: female newscaster, see Fig. 15), and the required bit-rate q_s for the side information. The difference between q_m and q_t , shown as the dotted area, is the dynamic bit-rate margin. It varies according to the audio signal over a range from about 30 to 100 kbit/s. This dynamic bit-rate margin, averaging about 60 kbit/s, may be used for various purposes:

- to obtain a larger ratio of masking threshold to quantizing noise level (increased "mask-to-noise ratio")
- for bit-error protection and the formation of the multiplex signal ("dynamic bit-error protection", "synchronization")
- for the transmission of additional information in the multiplex signal, which are non time-critical and are independent of the sound signal, such as programme information or radio-text data ("additional data channels").

The dynamic bit-rate margin may of course be attributed to any combination of these three possibilities. In the following, increased S/N ratio and dynamic bit-error protection will be described in more detail.

5.1. Increased mask-to-noise ratio

Source coding schemes which eliminate **all** the irrelevance of a digital audio signal (i.e. those which fully exploit the pre-masking, simultaneous masking and post-masking effects) can cause problems in certain applications [5]. For example:

- Where such source coding systems are cascaded, the quantizing noise can accumulate and exceed the masking threshold. If, for example, both the recording and the transmission of a radio programme signal were to be done with such a coding

scheme, the quality impairment could become perceptible on critical sound signals.

- If the levels of certain parts of the signal spectrum are raised or lowered at the receiver, the spectral masking effects could be weakened to the extent that impairments become perceptible in the quality. This may happen if the level of a sub-band signal which normally serves to mask neighbouring sub-band signals, is lowered at the receiver or, inversely, if the level of a sub-band signal which is completely or partially masked by a neighbouring sub-band signal is raised at the receiver.

To prevent quality impairments in such circumstances, the dynamic bit-allocation scheme does not function specifically to achieve the maximum possible reduction in bit-rate but rather to give a certain amount of "mask-to-noise reserve". This will vary in a signal-dependent manner and is approximately proportional to the bit-rate margin. If the audio signal is so narrow-band, for example, that the bit-rate in many sub-bands is set to zero, then an appropriate increase occurs in the bit-rate of those sub-band signals which evoke masking. Such increases occur as far as the bit-rate margin permits. Thus the resolution of the masking sub-band signals is considerably higher, in certain circumstances, than that required on the basis of the masking threshold.

A further advantage of this form of dynamic bit-allocation to the sub-bands resides in the fact that sinusoidal signals, for example, are transmitted with very high resolution (e.g. 16 to 18-bit linear). The transmission of an isolated spectral line with 16-bit resolution in a sub-band 500 Hz wide theoretically requires a bit-rate of 16 kbit/s, although of course a bit-rate about twice as large as this must in fact be considered, owing to the aliasing distortion (see [5]).

5.2. Dynamic bit-error protection

An additional bit-rate is needed in the channel coding to provide bit-error protection of the transcoded samples and the side information, and for synchronization of the data transmission. The total bit-rate of the MASCAM signal therefore comprises contributions from both the source coding and the channel coding.

The dynamic bit-rate margin may be exploited for the channel coding in such a way that the degree of error-protection of the multiplex signal is made to be dependent on the instantaneously-available bit-rate margin. This is called "dynamic bit-error protection". It is advantageous to arrange for the degree of error-protection to be determined in a series of steps. For example, Fig. 18 (bottom) shows a signal-dependent varying bit-rate $q_m - q'_{(t)}$ (dotted zone) which is used for the dynamic bit-

error protection and for synchronization. The bit-rate traced by the step-like curve $q'_{(t)}$ represents the bit-rate available for the source coding, and it is seen that it remains larger than the required bit-rate $q_{(t)}$ by a reasonable margin.

Dynamic bit-error protection leads to an increase in the average degree of bit-error protection which is governed by the average available bit-rate margin. The probability of impairments due to bit-errors is therefore reduced.

Furthermore, the dynamic error-protection functions in such a manner that audio signals which require a low bit-rate will be transmitted with a high degree of bit-error protection; conversely, those requiring a high bit-rate will have a lower degree of bit-error protection. This characteristic is especially advantageous because signals requiring a low bit-rate will give rather limited masking of the bit-error noise, but it is precisely these sensitive audio signals which will be given the greatest protection. This will be relevant, for example, in the case of a speech signal (more especially during the pauses) or during certain passages of music (such as "pp" passages on the piano, see Fig. 15, bottom, or "silence" in a concert hall); the required bit-rate for such signals is particularly low and the degree of bit-error protection is particularly high.

It is a fortuitous feature of the system that the degree of dynamic bit-error protection is determined in such a way that, with a given bit-error rate, the subjective impairment is almost independent of the content of the audio signal.

6. Optimisation of the bit-error resistance

6.1. Source coding

One of the most important applications of MASCAM will be digital broadcasting to portable and car radios. The transmission path in such circumstances is particularly hostile and the channel coding principles must be chosen carefully, on the basis of the path characteristics and the desired failure characteristic, with a view to minimising the channel bit-rate.

The bit-error resistance of the source coding scheme should also be optimised, and examination of this aspect has highlighted three significant factors.

6.1.1. Inherent insensitivity of audio samples to bit-errors

Audio signal distortions do not occur in the broadband spectrum but are instead restricted to the narrow bandwidth of the sub-band containing the defective samples. The narrow-band limitation re-

sults at the receiving end after the inverse filtering operation. Bit-error noise in ordinary wideband transmission systems is perceived as "clicks" whose intensity depends on the significance of the erroneous bit. As the bandwidths of the MASCAM sub-bands approximate to the "critical sub-bands" of the human ear, the loudness of noise resulting from bit-errors is optimally limited. Furthermore, if the scale-factors of all sub-band signals are given a high degree of protection, the maximum amplitude of the bit-error noise is limited by the scale-factor. The noise-level limiting effect occurs individually for each sub-band signal, so the overall bit-error noise is masked under virtually all circumstances. As an example, a bit-error ratio (BER) of 10^{-3} will result in noise which is just perceptible, but not annoying.

To illustrate the bit-error resistance of MASCAM coded signals, we can consider a situation where **all** the samples are totally distorted. In this case, the system will ensure that the spectral curve of the distortion noise coincides very closely with the spectral curve of the undistorted audio signal. The coder would operate like a 24-channel vocoder, in which a broadband noise input is divided between 24 narrow bands which are amplitude modulated. In such a situation, speech remains intelligible, even if all the samples are totally distorted. This is illustrated for the vowel [ə] in Fig. 19. A prerequisite for this characteristic is that there must be no loss of side information or of synchronization bits during the transmission.

6.1.2. Low number of important bits to be protected

In the MASCAM signal the bits in the side information are of paramount importance because they carry all the information about the bit-allocations to the different sub-bands. These bits must be very powerfully protected. It has been found that the importance of the most-significant bits (MSBs) of the sub-band samples depends on the position of the sub-band in the baseband spectrum: in sub-bands below about 1 kHz the two MSBs of each sub-band signal are important, in sub-bands between 1 kHz and 3 kHz the MSB only is important, and above 3 kHz the MSBs are less important.

Taken as a whole, the MASCAM multiplex signal contains a small number of very important bits and a large number which require no protection at all. The scale of importance of the various bits is being established by means of subjective evaluation of signal quality.

6.1.3. Use of the dynamic bit-rate margin for error-protection

As explained in section 5.2., dynamic bit-error protection is a useful means of enhancing the bit-error resistance. One approach to exploiting this, assuming that constant and effective protection is provided for all the important bits (see section

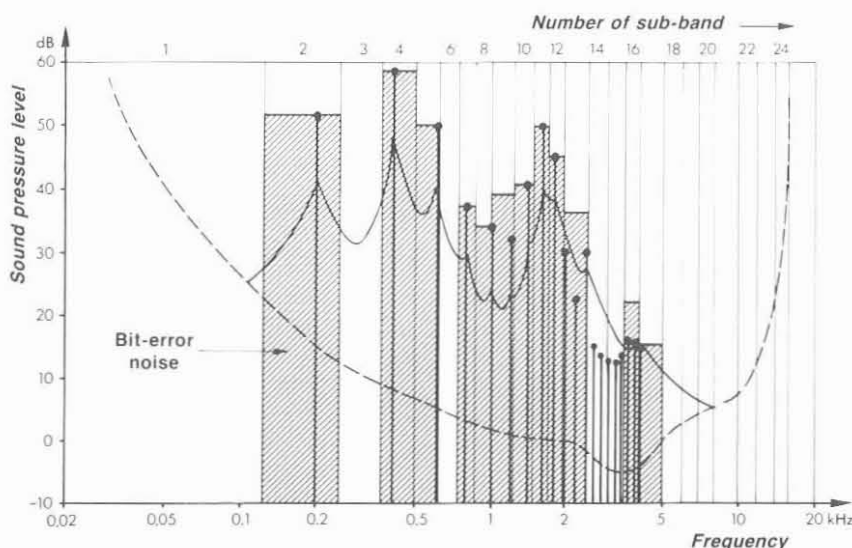


Figure 19
Amplitude spectrum SPL of the vowel [ə], the masking pattern L_T and the noise due to bit-errors (hatched bars) for a case where all samples are totally distorted but the scale-factors are perfectly protected.

6.1.2. above), would be to use dynamic protection to cover the remaining bits of less importance. This would result in high transmission quality even where the bit-error ratio is high.

6.2. Channel coding strategies

The channel coding strategy should be selected not only with respect to the transmission path characteristics and the desired degradation behaviour, but also with due regard to the characteristics of the source-coded signal. This means that the channel codec should be able to derive maximum benefit from the source codec, although other factors such as low decoder complexity, additional data channel capacity or easy adaptability to other applications (e.g. recording) are also important.

As explained in section 6.1., a high degree of bit-error resistance is achieved with MASCAM source coding. Investigations of "source-adapted channel coding" are now under way with a view to adding to these characteristics a channel code offering optimum bandwidth economy and a graceful degradation of performance in the presence of high noise levels on the transmission path. Particular attention is being given to the optimisation of this concept for digital sound broadcasting to vehicle receivers.

Another channel coding concept has been proposed recently [22] which disregards the source characteristics and aims to provide a "virtually error-free channel" allowing the transmission of any sort of data by coding each bit in the same manner, irrespective of its importance. Every bit will consume the same redundancy. The ability of the source codec to minimise the bit-rate to be transmitted would not be fully exploited since the channel bandwidth needed for a certain transmission performance would not be minimised.

Summarising, a transparent channel provides error-free transmission of arbitrary data even when subject to high noise levels but the efficiency is not optimum when transmitting audio signals, whilst a source-adapted channel coding concept will allow a more efficient transmission of audio signals with good subjective quality even when subject to high noise levels on the transmission path, but this performance is unrelated to the BER performance of the channel when transmitting arbitrary data. This suggests that the optimum solution would be a strategy tailored, above all, to give the most efficient transmission of audio signals, despite high noise levels, but allowing also the virtually error-free transmission of arbitrary data. A number of approaches to this question are being investigated. Three of these will be presented briefly in order to assess their likely capabilities and performance.

They are all based on the use of the orthogonal frequency-division multiplex (OFDM) modulation system with coherent demodulation proposed in [22].

6.2.1. Separate error-protection for source and channel

This approach is used in the prototype MASCAM hardware described in section 3. It is based on a separation between the source and channel codecs, and on a progressive performance degradation characteristic of the channel codec when there are high noise levels on the transmission path. The bit-error resistance of the source codec comes into play in this region of progressive degradation of the BER performance.

As an example, we may consider a MASCAM multiplex signal which is channel coded with a convolutional code (4-PSK) using a constraint length $k = 7$ and a code rate $= 0.5$. The corresponding BER performance curves given in [22] show that the BER is degraded from 10^{-5} to 10^{-1} over a range of values for E_b/N_o from 7 to 2 dB in a Rayleigh channel.

The resulting subjective quality impairment depends on the bit-error resistance of the MASCAM multiplex signal, and it can be measured as a function of the BER. With MASCAM signals, first simulations have shown that "excellent" quality can be expected with a BER of 10^{-4} and "good" quality with a BER of 10^{-3} , assuming there is effective error-protection for the side information, that dynamic error-protection corresponding to Fig. 18 (bottom) is used, and that the MASCAM multiplex signal bit-rate is 128 kbit/s.

The performance of this coding scheme can therefore be estimated as follows:

- channel bit-rate (excluding synchronization bits)	256 kbit/s
- average bit-rate margin available	0 kbit/s
- "virtually error-free channel"	$E_b/N_o > 8$ dB
- "excellent" quality	$E_b/N_o > 6$ dB
- "good" quality	$E_b/N_o > 5$ dB
- "failure" (BER = 10^{-1})	$E_b/N_o \approx 2$ dB

The subjective audio signal quality degrades over a range of E_b/N_o from 6 to 2 dB, with the signal remaining "intelligible" at values below 3.5 dB.

6.2.2. Additional transparent data channel

This approach is based on the error-free transmission of any form of data, even with low values of E_b/N_o , as proposed in [22]. Any additional

source-adapted error-protection will be superfluous and to derive maximum benefit from the features of the MASCAM source codec the dynamic bit-rate margin should be used instead to provide an additional transparent data channel. This could carry a variety of broadcast-type services such as programme information, traffic information, radio-text or even still-picture delivery, without any need for additional channel capacity.

Again taking an example of a channel codec from [22], we will consider a concatenation of a convolutional code and a Reed-Solomon code. The resulting code-rate will be chosen as $r = 0.437$. The corresponding BER performance curves for this concatenated code suggest that the following channel performance will be attained:

- channel bit-rate (excluding synchronization bits) 252 kbit/s
- average capacity of the additional data channel (useful bits) 30 kbit/s
- "virtually error-free channel" $E_b/N_o > 4.5$ dB
- "failure" (BER = 10^{-1}) $E_b/N_o \approx 3.5$ dB

These values may be compared with those of the first approach. Assuming that there is in fact a need for an average useful bit-rate of 60 kbit/s accompanying a stereophonic programme, this second approach may be advantageous because the channel bit-rate is not increased and because the bit-error performance is "excellent" at values of E_b/N_o very close to the failure point. If the degradation characteristic is considered, however, the first approach may be preferred because failure occurs at a value of E_b/N_o of about 2 dB instead of about 3.5 dB, and intelligibility of speech is maintained over this range of values, rather than abrupt muting.

6.2.3. Source-adapted dynamic error-protection

This approach is proposed in [23] in order to obtain optimum bit-efficiency and fair subjective quality at very low values of E_b/N_o (i.e., graceful degradation) through the exploitation to the greatest possible extent of the performance of the MASCAM source coder. In this concept all the error-protection is provided by the source-adapted channel coder and the MASCAM multiplex signal is modulated without further channel coding. The source and channel coders are combined in such a way that the channel coder produces different degrees of error-protection to individual bits of the source-coded signal, in a manner dependent on the importance of those bits and on the number of redundancy bits available.

The channel coding is done with a convolutional code and soft-decision Viterbi decoding. The convolutional code is designed to give sufficient bit-error protection of the most-important bits. The

desired degradation characteristic is obtained by puncturing the code bits of the convolutional coder according to the relative importance of the bits in an information block. To ensure correct maximum-likelihood decoding of the convolutional code, the puncturing is done under rate-compatible restrictions [24]. By this means the same coder and decoder can be used for all the bits, but the channel coder remains extremely flexible and is able to provide unequal error-protection to individual bits. It can also provide equal error-protection to all the bits, to permit the transmission of arbitrary data (transparent data channel). The protection level may be selected according to the desired value of E_b/N_o for "virtually error-free" performance.

To illustrate this, it will be assumed that all the bits of one 4-ms block of the source-coded signal are divided into five groups of differing importance, and that each bit in the block can be allocated to one of the five groups. Table 1 shows the code-rates r chosen for each group, ranging from $r = 1/4$ for group 1 (most important) to $r = 2/3$ for group 5 (least important). Assuming that the "worst case" signal (Fig. 13) requiring a source bit-rate of 110 kbit/s has to be coded, the source and channel bit-rates for each group are shown in Table 1. The resulting total channel bit-rate is 255 kbit/s.

Table 1
Allocation of code-rates to five groups of source-code bits of differing importance, assuming a maximum source bit-rate of 110 kbit/s (worst-case signal)

Level of importance	1	2	3	4	5	Σ
Code-rate r	1/4	1/3	2/5	1/2	2/3	
Source bit-rate (kbit/s)	15	15	20	20	40	110
Channel bit-rate (kbit/s)	60	45	50	40	60	255

This example illustrates the increase in efficiency achieved with this graded error-protection scheme. The side information and the MSBs of the MASCAM source-coded signal are given good protection so the graceful degradation characteristic is shifted towards smaller values of E_b/N_o .

Theoretically, failure (corresponding to a BER of 10^{-1}) is at $E_b/N_o = 0$ dB. To obtain the same performance with equal error-protection for all the bits, the channel capacity would have to be 440 kbit/s instead of 255 kbit/s.

Extending this principle, the dynamic bit-rate margin can be exploited even more effectively by varying the code rates according to the instantaneous signal content. For example, if the instantaneous source bit-rate is 80 kbit/s, a suitable set of code rates would be as shown in Table 2. It is clear that the error-protection is enhanced by this process. At very low values of source bit-rate such as

Table 2
Allocation of code-rates to five groups of source-code bits of differing importance, assuming an instantaneous source bit-rate of 80 kbit/s

Level of importance	1	2	3	4	5	Σ
Code-rate r	1/5	1/4	1/3	2/5	1/2	
Source bit-rate (kbit/s)	15	15	12	12	26	80
Channel bit-rate (kbit/s)	75	60	36	30	52	253

60 kbit/s or even 40 kbit/s (see *Figs. 14 and 15*), the grade of error-protection could be extremely high, even for the less-important bits.

However, it may not be appropriate to use all the bit-rate margin for error-protection because the improved code-rate allocation in *Table 2* will in fact have only a small impact on the subjective performance. For practical reasons, failure will occur at higher values of BER than suggested by the theoretical BER performance curves in [22], because the synchronization cannot be maintained if there is a very high noise level on the transmission path. The optimum solution would therefore involve a trade-off between error-protection and the provision of an additional data channel and/or achieving a saving of bandwidth, discussed below.

As the channel coder produces different degrees of error-protection to separate (groups of) bits, some of the bit-rate margin could be used for other purposes, such as the provision of an additional data channel. The corresponding group of data bits could be given good protection, for example by using the code-rate 1/4, as for other parts of the source-coded signal. Then, assuming an average bit-rate margin of 30 kbit/s, the average capacity of this additional transparent channel is 7.5 kbit/s. This is obtained without impairing the channel performance given by the arrangement of *Table 1* and, more particularly, without increasing the bandwidth.

An important way of optimising the channel coder involves saving bandwidth by accepting the use of a more economical code-rate allocation as shown in *Table 1*. This may be found appropriate if it is borne in mind that the worst-case source bit-rate of 110 kbit/s will occur only very occasionally. If the code-rates of *Table 1* are applied to the case where the instantaneous source bit-rate is 80 kbit/s, the channel bit-rate will be 198 kbit/s. A saving of up to 50 % could be worthwhile, even taking account of an increasing degradation of the subjective quality at low values of E_b/N_o . Here again a trade-off needs to be established with other factors such as the desired performance of the broadcast network, tolerable subjective impairment at low values of E_b/N_o , the performance of additional concealment techniques, the synchronization performance, etc.

In view of the above points, the strategy of source-adapted dynamic error-protection appears beneficial in comparison to the approaches considered in sections 6.2.1. and 6.2.2. The channel coding is more efficient than that obtained with equal error-protection to all bits, or that obtained with separate protection for the source and the channel. As far as the MASCAM scheme is concerned, the dynamic bit-rate margin may be used in any manner appropriate to the specific application. Where it is used for dynamic bit-error protection, the resulting BER performance may be unnecessarily high if a channel bit-rate of 255 kbit/s is provided. The required degradation characteristic of the subjective quality, depending on E_b/N_o , must therefore be defined first with regard to bandwidth economy. The flexible channel coder described in [23] would be the ideal tool for these studies, since the failure characteristic can be tailored to suit the requirements and any further developments in the source coding technique can be implemented efficiently without changing the channel coder hardware.

However, the final choice of channel coding strategy will depend also on some other factors not discussed here. To mention just one example, it would appear advantageous to provide a defined interface between the source and channel coding sub-systems in order to simplify adaptation to other applications (e.g., recording) by avoiding the need for cascaded source codecs.

7. Future MASCAM hardware

Any MASCAM system for operational use will need to be optimised with a view to achieving minimum hardware complexity, particularly in the receiver. Relevant parameters when establishing the best trade-off in any particular application are:

- subjective quality
- bit-rate
- delay
- bit-error protection
- post-processing capability
- decoder complexity.

When evaluating these parameters, the specific application of the MASCAM system must evidently be taken into consideration. For example, as described in section 4.4, the maximum bit-rate of the MASCAM source-coded signal depends on the buffer length. By increasing the delay to about 1000 ms an additional bit-rate reduction of about 35 kbit/s is achieved (see *Fig. 17*). For transmission purposes the delay should perhaps be minimised; in recording applications even longer delays may be acceptable. In these contrasting circumstances, it would seem appropriate to develop a universal VLSI MASCAM decoder chip which can be adapted optimally to different applications (digital audio broadcasting, ISDN transmission, recording, etc.) and which is of minimum complexity.

7.1. Coding scheme

Functional block diagrams of the transmitting and receiving ends of a more advanced MASCAM system are shown in Figs. 20 and 21 respectively. The main differences compared to the preliminary hardware system of Figs. 5 and 6 are:

- The bit-allocation control data for each sub-band are not provided in look-up tables based on equation (3), but instead come from a controller within the coder which monitors the masking threshold. The instantaneous masking threshold is calculated in a special sub-section of the coder.
- The scale-factors and the allocation control data are coded and transmitted separately. This enables the decoder to decode any bit-allocation established by the coder. Hence the degree of irrelevance reduction is governed only by the performance and complexity of the coder, and not of the decoder.
- Dynamic error-protection is provided. The dynamic bit-rate margin is used (partially) to extend the bit-error resistance of the MASCAM coded signal and the degree of protection is

governed by the allocation controller in the coder.

- There are fewer than 24 sub-bands. This is possible because the required resolution of each sub-band signal depends not on the sub-band bandwidth according to equation (3) but, instead, on the calculated masking threshold. A certain number of sub-bands are needed in order to match the spectral slope of the quantizing noise to the masking threshold, and it is found that at high frequencies this can be done with relatively few sub-bands.
- The channel coder is also controlled by the dynamic bit-allocation controller, so the channel coding can be adapted to suit the characteristics of both the source coded signal and the channel.

7.2. Sampling frequency

To ensure compatibility with the EBU/AES digital audio standards, the future MASCAM system will use 48-kHz sampling. Two aspects require consideration in this context:

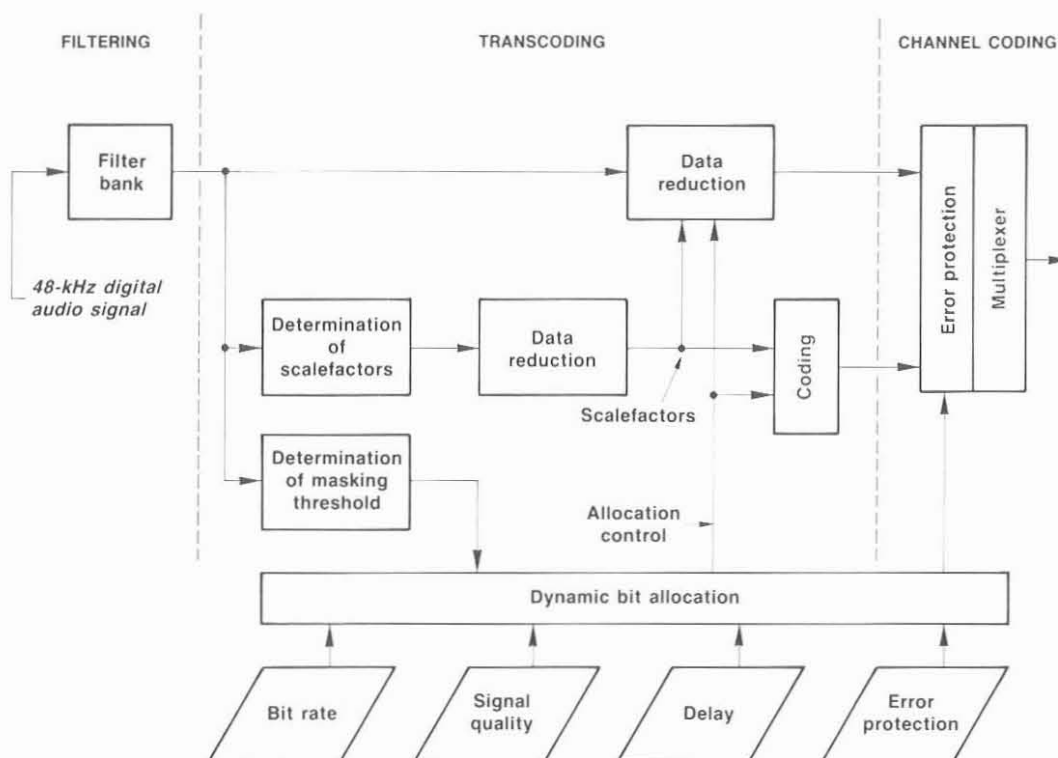


Figure 20

Block diagram of a MASCAM coder with dynamic bit-allocation. The auxiliary information includes scale-factors, bit-allocation control data and channel coding status data. The expected bit-rate of samples + side information is less than 100 kbit/s. Sampling frequency is 48 kHz.

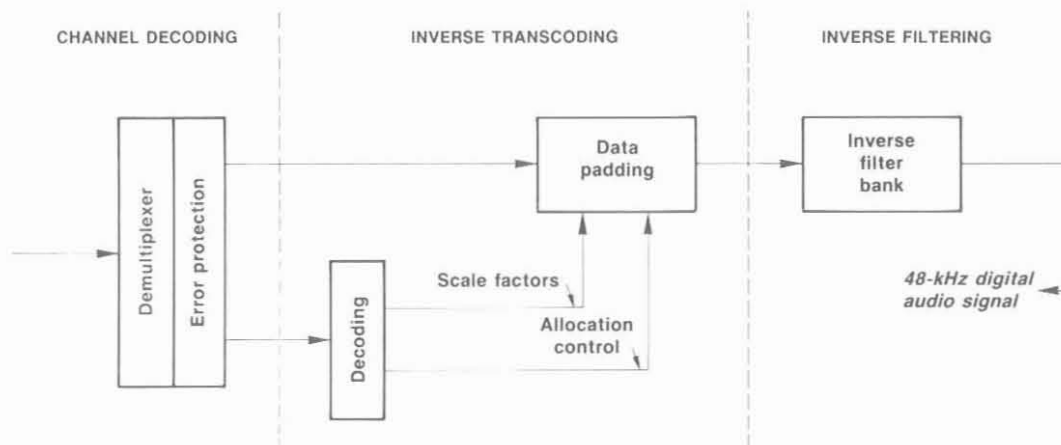


Figure 21
Block diagram of the MASCAM decoder corresponding to the coder in Fig. 20.

- Increasing the base-bandwidth from 16 kHz to 20 kHz will imply either an increase in the bandwidth of the individual sub-bands, or an increase in the number of sub-bands. If static bit-allocation is used this will lead to an increased bit-rate as determined by equation (3) corresponding to the inclusion of additional sub-bands in the range 16 to 20 kHz. The increase in bit-rate would be about 15 kbit/s.
- Where dynamic bit-allocation is used, the inclusion of four additional sub-bands in the range 16 to 20 kHz would barely increase the bit-rate because the spectral components of a **broadband** signal falling in this range will be masked [16, 17]; in other words, there is no relevant information in these bands that needs to be transmitted. If there is a **narrow-band** signal in the range 16 to 20 kHz, all the bits can be used for transmission of this signal.

In principle, if an optimum irrelevance-reduction coding scheme is used, the bit-rate of the coded signal is independent of the sampling frequency. It will be governed only by the amount of relevant information in the source signal. Independence of sampling frequency cannot in fact be achieved when using static bit-allocation, but it will be very nearly achieved in the final version of MASCAM using dynamic bit-allocation.

7.3. Bit-rate

In the codec used for simulations (section 4), the bit-allocation data have been determined as a function of the scale-factors. It has been found that this technique does not result in an optimum bit-rate because a set of 24 scale-factors, renewed every 4 ms, can serve only approximately for the calculation of the instantaneous masking threshold. Therefore, a more exact description of the masking

audio signal is derived in a special sub-section of the coder, as the basis for a more precise calculation (Fig. 20); this takes account of all the relevant pre, simultaneous and post-masking effects. By enhancing the performance of this sub-section of the coder the number of bits allocated to the sub-bands is minimised.

This type of bit-allocation control data cannot be calculated in the decoder, however, so it has to be transmitted from the coder. The bit-rate of the side information increases accordingly. It has been found, however, that this apparent disadvantage is neutralized by two facts. First, the number of sub-bands, and hence the number of scale-factors to be transmitted, is reduced because the sole remaining purpose of the filter bank is to match the spectral slope of the quantizing noise to the instantaneous masking threshold. Second, the scale-factor and allocation control together contain some redundancy; this can be eliminated, for example by using tables. The resulting bit-rate of the side information is about 15 kbit/s.

The strategy of calculating the instantaneous masking threshold as exactly as possible in the coder, and transmitting the resulting control data to the decoder, is advantageous compared to the bit-rate reduction referred to in section 4. It is expected that the bit-rate of the samples plus the side information (scale-factors, bit-allocation control data, channel coding status data) need not exceed 100 kbit/s. If the bit-rate of the MASCAM multiplex signal is maintained at 100 kbit/s, an additional dynamic bit-rate margin of at least 30 kbit/s will be available which can be used freely. In the case of digital sound broadcasting, this margin would be particularly useful for dynamic bit-error protection and/or providing an additional data channel. The resulting total transmission bit-rate of course depends on the choice of channel coding strategy, although values in the range 190-230 kbit/s per

monophonic channel will be achievable in any case (e.g., "transparent channel" or "source-adapted channel coding" philosophy) where the MASCAM technique is adopted.

Furthermore, if the delay of the coded/decoded audio signal is not critical in some applications, an additional bit-rate reduction can be achieved. For example, a 64 kbit/s MASCAM multiplex signal could be obtained by inserting a buffer of suitable length (about 2 s) without impairing the subjective quality of the decoded signal. In recording applications, even lower bit-rates can be expected, with minimised dynamic bit-rate margin (see Fig. 16).

7.4. Decoder

In this future MASCAM concept, the hardware of the source decoding sub-systems (inverse transcoding and inverse filtering in Fig. 21) will be less complex than that discussed in section 3 because the number of sub-bands and the number of coefficients in certain filters is reduced. Also, if wave digital filters [9] can be used instead of quadrature mirror filters, it is expected that a 48-kHz stereophonic MASCAM source decoder would require processing capacity equivalent to less than one DSP 32.

Once the choice of filtering technique has been made, a VSLI MASCAM source decoder chip could be designed without limiting future progress in the calculation and exploitation of the masking threshold, and without restricting the range of possible applications. The performance and characteristics of the MASCAM source coding technique depend only on the investment made in the transcoding part of the coder (Fig. 20).

As regards demultiplexing and error-correction, the hardware complexity will of course depend on the channel coding strategy. For example, if a virtually error-free channel is used, any error-protection is superfluous and a complete MASCAM stereophonic decoder would require barely more processing capacity than the equivalent of one DSP 32. On the other hand, if source-adapted dynamic error-protection is used, as described in section 6.2.3., the complexity of the flexible Viterbi decoder is almost identical to that of an ordinary Viterbi decoder operating at a fixed code rate.

When using source-adapted dynamic error-protection, the performance and characteristics of the channel coding again depend only on the properties of the coder; i.e., on the puncturing of the code bits of the convolutional code. The combined MASCAM source and channel decoder hardware could therefore be designed without restricting the flexibility to optimise the system with regard to specific applications and performance targets.

* * *

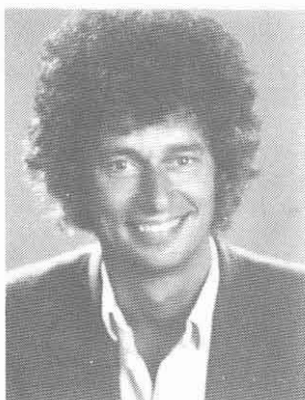
The authors wish to thank Mr. R. Sedlmeyer and his colleagues, in particular Messrs. A. Biefort, E. Ingelsberger, W. Krafft, W. Richter and G. Ulbrich, for their work on the hardware realisation. The authors are indebted to Messrs. G. Schneeberger, C. Weck and D. Wiese for contributing to the investigations and for their comments on the channel and source coding.

* * *

Bibliographical references

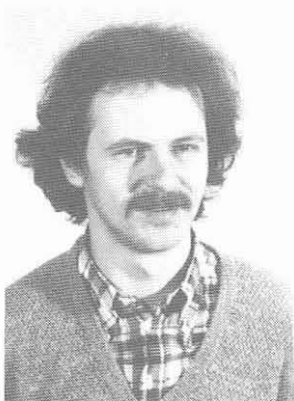
- [1] Jakubowski, H. and Twietmeyer, H.: **The German DBS digital sound-broadcasting system: Audio coding methods.** 77th AES Convention, 1985. Preprint No. 2179.
- [2] Krahé, D.: **Ein Verfahren zur Datenreduktion bei digitalen Audiosignalen unter Ausnutzung psychoakustischer Phänomene** (Bit-rate reduction method for digital audio signals based on psychoacoustic phenomena). Rundfunktechnische Mitteilungen, No. 3, May/June 1986, pp. 117-123.
- [3] Schröder, E. F. and Voessing, W.: **High quality digital audio encoding with 3.0 bits/sample using adaptive transform coding.** 80th AES Convention, 1986. Preprint No. 2321.
- [4] Brandenburg, K.: **High quality sound coding at 2.5 bits/sample.** 84th AES Convention, 1988. Preprint No. 2582.
- [5] Theile, G., Link, M. and Stoll, G.: **Bit-rate coding of high quality audio signals.** 82nd AES Convention, 1987. Preprint No. 2432.
- [6] Stoll, G., Link, M. and Theile, G.: **Masking-pattern adapted subband coding: Use of the dynamic bit-rate margin.** 84th AES Convention, 1988. Preprint No. 2585.
- [7] Jayant, N. S. and Noll, P.: **Digital coding of waveforms.** Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1984.
- [8] Crochiere, R. E. and Rabiner, L. R.: **Multirate digital signal processing.** Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1983.
- [9] Fettweis, A.: **Wave digital filters: Theory and practice.** Proceedings of the IEEE, Vol. 74, 1986, pp. 270-327.
- [10] Zwicker, E. and Feldtkeller, R.: **Das Ohr als Nachrichtenempfänger** (The ear as a receiver of information). S. Hirzel-Verlag, Stuttgart, 1967.
- [11] Jakubowski, H.: **Quantisierungsverzerrungen in digitalen arbeitenden Tonsignalverarbeitungs- und Übertragungssystemen** (Quantising distortion in digital audio signal processing and transmission systems). Rundfunktechnische Mitteilungen, No. 2, April 1980, pp. 91-92.
- [12] Crochiere, R. E.: **Sub-band coding.** The Bell System Technical Journal, Vol. 60, No. 7, September 1981, pp. 1633-1653.
- [13] Esteban, D. and Galand, C.: **Application of quadrature mirror filters to split band voice coding schemes.** Proceedings of the ICASSP, May 1977, pp. 191-195.

- [14] Spille, J.: **Simulation von Datenreduktionsverfahren bei digitalen Audiosignalen** (Simulation of data reduction methods for digital audio signals). Diplomarbeit, Ruhr-Universität Bochum, 1987.
- [15] **Source coding for digital sound signals in broadcasting studios.** CCIR Recommendation 646, Vol. X, Part 1, XVIth Plenary Assembly, Dubrovnik 1986.
- [16] Plenge, G., Jakubowski, H. and Schöne, P.: **Which bandwidth is necessary for optimal sound transmission?** Journal of the Audio Engineering Society, Vol. 28, 1980, pp. 114-119.
- [17] Muraoka T. and Iwahara, M.: **Examination of audio-bandwidth requirements for optimum sound signal transmission.** 64th AES Convention, 1979, Preprint No. 1545.
- [18] Clark, G. C. and Cain, J. B.: **Error-correction. Coding for digital communications.** Plenum Press, New York and London, 1981 ISBN 0-306-40615-2.
- [19] Terhard, E., Stoll, G. and Seewann, M.: **Algorithm for extraction of pitch and pitch salience from complex tonal signals.** Journal of the Acoustical Society, AM 71, 1982, pp. 679-688.
- [20] Penner, M. J.: **Effects of masker duration and masker level on forward and backward masking.** Journal of the Acoustical Society, AM 56, 1974, pp. 179-182.
- [21] Zwicker, E. and Fastl, H.: **Zur Abhängigkeit der Nachverdeckung von der Störimpulsdauer** (Influence of noise-pulse duration on masking). Acustica, Vol. 26, 1972, pp. 78-82.
- [22] Alard, M. and Lassalle, R.: **Principles of modulation and channel coding for digital broadcasting for mobile receivers.** EBU Review-Technical, No. 224, August 1987, pp. 168-190.
- [23] Weck, Ch. and Theile, G.: **Digital audio broadcasting: Optimizing of a combined concept of source and channel coding with respect of subjective criteria.** To be presented at the international Broadcasting Conference, Brighton, 23rd to 27th September 1988.
- [24] Hagenauer, J.: **Rate compatible punctured convolutional codes.** Proceedings of the ICC '87, June 1987, pp. 1032-1036.



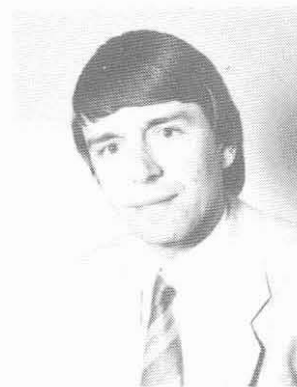
G. Theile

Dr. Gunter Theile studied electrotechnology at Bremen Technical College and worked with the Philips Company in Bremen before going to the Technical University of Berlin in 1970 to study psychoacoustics and communications theory. He spent two years at the Heinrich-Hertz Institute and moved to the Institut für Rundfunktechnik (IRT) in 1977 where, since 1980, he has led the group dealing with audio production and reproduction systems, psychoacoustic research and sound quality assessment.



M. Link

Gerhard Stoll studied communications theory and cybernetics at Stuttgart University and then at Munich Technical University where he was later employed in the Institute of Electroacoustics in the field of psychoacoustic research. He subsequently moved to the IRT where he is involved in the development of source coding systems and sound transmission quality assessments; he now leads the group dealing with psychoacoustics.



G. Stoll

Martin Link studied information theory at Aachen University, where he took a particular interest in computer simulations of acoustic fields. In 1986 he joined the IRT to work on audio production and reproduction systems, psychoacoustics and signal processing.