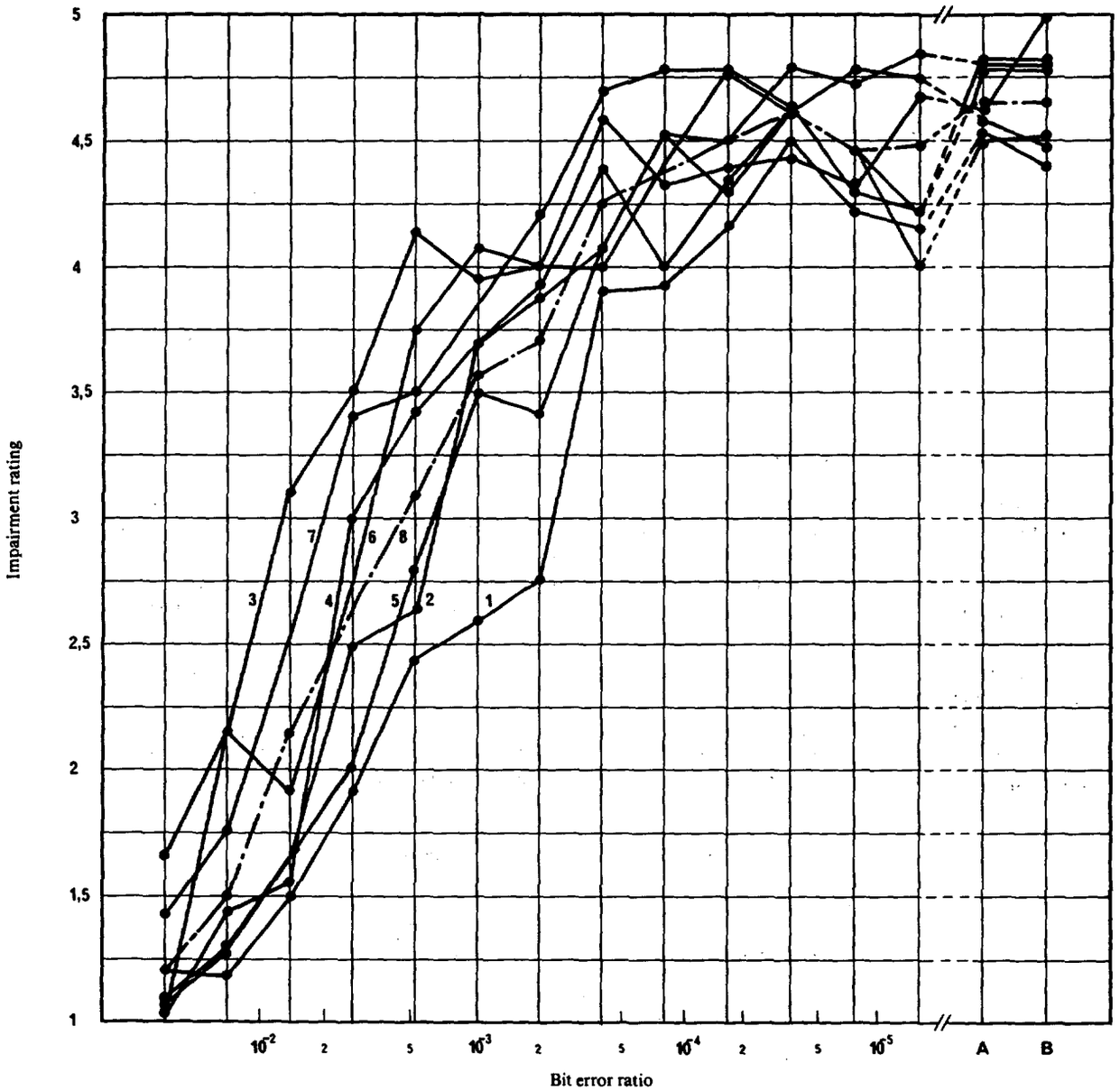


FIGURE 12 – *Impairment rating versus switch position, item by item*

Note 1. – Curves No. 1-7 are from programme selections indicated in Table III. Curve 8 is for overall (average) result.

Note 2. – In scale position A, signal is transmitted through Dolby ADM system without errors.

Note 3. – In scale position B, signal comes direct from analogue source (e.g.: PCM on VCR).

2. Subjective measurements conducted in Canada

In the process of identifying the relative merits of the various proposed coding schemes, a test programme was conducted in Canada.

The following coding schemes were covered by the measurement programme:

- 10-14 semi-instantaneous companding (NICAM 3) with CCITT Recommendation J.17 pre-emphasis;
- 10-14 instantaneous companding (A-law) with CCITT Recommendation J.17 pre-emphasis;
- 14 bit linear coding with CCITT Recommendation J.17 pre-emphasis;
- adaptative delta modulation (sampling rate = 330 kHz); and
- adaptative delta modulation (sampling rate = 204 kHz).

The first three coding schemes were modelled on a computer working in conjunction with a vectorial processor.

The pre-emphasis applied to the signal was changed through digital filtering. The sampling frequency was also changed to 32 kHz through digital processing. Companding and truncation to generate the proper number of bits was then performed to finally obtain the companded digital stream. Proper re-formatting was made to store the test sequence on video tape.

This process was not used for the ADM system which could not be modelled on the computer. Conversion to analogue signal was necessary before processing by the ADM. At the output of the ADM the signal was converted back from analogue to 16 bits/sample for insertion in the recorded test sequence.

The source material was either read from compact laser discs or generated by the computer itself:

- Cardozo, Pájaro Campana (Indian Harpsichord)
- Orff, Carmina Burana (Lyrics)
- synthetic gong generated at 44 kHz simulating the sound of a triangle on the first 4 notes of "Frère Jacques" (4.2 kHz, 4.7 kHz, 5.3 kHz and 4.2 kHz). (Attack: T = 3 ms, decay: T = 300 ms.)
- synthetic gong generated at 44 kHz simulating a bass guitar on the first 4 notes of "Frère Jacques" (65.4 Hz, 73.4 Hz, 82.4 Hz and 65.4 Hz). (Attack: T = 3 ms, decay: T = 300 ms.)

Tests were made according to the suggested CCIR comparison method (Recommendation 562) where sequence A-B is presented twice and then 15 s is left for scoring also according to the CCIR comparison scale. Audiometric measurements were taken with each subject before the subjective tests. On average, three listeners were accommodated in a well-calibrated listening room. A total of 25 listeners performed the tests.

All comparisons were repeated in reverse order to have a measure of the listener's consistency. Accordingly, 4 listeners were discarded from the analysis of the results. Furthermore, when the two scores of a same listener for the same comparison differed by more than two grades, these were discarded.

Due to uncontrollable vibrations and rattles induced in the listening room by the high-level low frequency synthesized gong, the test sequence was also assessed by 8 listeners using headphones. The results of this group are used for the low frequency gong.

In addition to the mean opinion score and the standard deviation obtained from the filtered data, a study of the statistical significance of the results was conducted using the one tail Student-t distribution. The level of confidence that, on average, the normal population will prefer codec "A" to codec "B"* was obtained.

Table IV summarizes the results. The mean opinion score expressed on the −3, 0, +3 scale, the standard deviation as well as the level of confidence that "codec A is better than codec B" are given in each case. In the table, the 10-14 near-instantaneous companding as described in Report 953 is noted as NICAM 3.

As can be seen from the table, no significant difference was found between these two encoding systems when the ADM was sampled at 330 kHz except in the case of the critical low frequency gong where a noise spectrum shaping was perceived. This was mostly noticeable during the decay time of the gong. When the ADM was sampled at 204 kHz, a degradation in the reproduction of the high frequency gong could also be perceived, giving worse performance than the NICAM 3. For these two cases, however, no significant difference could be perceived between the two codec for normal programme material.

* "Codec" = "coder-decoder".

TABLE IV – *Results of the subjective tests conducted in Canada*

Codec "A"	Codec "B"	Harpsichord	Lyrics	HF gong	LF gong
		$\bar{X}$ / S / % Conf	$\bar{X}$ / S / % Conf	$\bar{X}$ / S / % Conf	$\bar{X}$ / S / % Conf
NICAM 3, CCITT Rec. J.17	ADM (330 kHz)	0.1 0.7 88%	0.2 0.9 91%	−0.1 1.3 28%	1.1 0.8 99.95%
NICAM 3, CCITT Rec. J.17	ADM (204 kHz)	0.0 0.8 50%	0.2 0.9 92%	1.8 1.0 99.999%	1.4 0.5 99.995%
ADM (330 kHz)	A-law, CCITT Rec. J.17	0.0 0.6 50%	−0.2 0.8 6%	0.9 1.4 99.99%	−1.2 0.5 0.01%
NICAM 3, CCITT Rec. J.17	A-law, CCITT Rec. J.17	0.0 0.7 50%	−0.1 1.0 27%	0.8 1.1 99.975%	0.1 0.3 88%
NICAM 3, CCITT Rec. J.17	14 bits linear, CCITT Rec. J.17	−0.1 0.7 18%	−0.1 0.8 19%	−22 0.8 0.001%	0.1 1.0 60%

$\bar{X}$: mean opinion score on a scale −3 to +3 as per Recommendation 562.
S: standard deviation.
% Conf: statistical confidence level that codec "A" is preferred to codec "B".

REFERENCE

AuBC [May, 1985] Australian Broadcasting Corporation Engineering Research and Development Report No. 139. Subjective Tests of Adaptive Delta Modulation with B-MAC for HACBSS.

REPORT 953-1 *

DIGITAL CODING FOR THE EMISSION OF HIGH-QUALITY SOUND SIGNALS IN SATELLITE BROADCASTING (15 kHz NOMINAL BANDWIDTH)

(Question 2/10 and 11, Study Programmes 51C/10, 2F/10 and 11)

(1982–1986)

1. Introduction

Technological developments presently allow the emission of digitally encoded audio signals to the public to be envisaged. It would be possible in particular to introduce this technique in the broadcasting-satellite service. This will require the prior definition of characteristics concerning coding, multiplexing and modulation of audio signals. This Report is concerned only with digital sound coding. Multiplexing, bit-error protection strategy and modulation aspects are discussed in Reports 954 and 632 respectively. Information dealing with methods of error correction or concealment can be found in Report 1073.

* This Report should also be brought to the attention of the CMTT.

Study Programme 51C/10 considers that the introduction of digital techniques for emission should allow an improvement in the quality of the signals transmitted. Recommendation 651 deals with digital PCM coding. Results of subjective tests are given in Report 632.

In addition to the improvement of quality, the following should be considered:

- the compromises between quality objectives and bit rate may be different for sound services which may have various quality requirements and planning constraints; they may also vary according to the requirements of individual countries;
- there are clear advantages for broadcasters, receiver manufacturers and the public in using a single standard for each application.

2. Digital PCM coding

2.1 *Linear coding*

The audio signal is encoded into digital form by a high precision analogue-to-digital converter after being passed through an appropriate anti-aliasing filter and possibly through a pre-emphasis network. The uniform coding resulting from this process leads to a linear representation with a minimum of 14 bits per sample. In that case, a 2's complement coding is used for each sample. A possible scale factor can possibly be associated with blocks of successive samples.

2.2 *Floating point or near-instantaneous coding*

A 16 bit linearly encoded sound signal can be transmitted on 14 bits per sample using a floating point coding system where the scale factor is based on 64 consecutive samples. Pre-emphasis may or may not be used.

When a lower bit rate per audio channel is required, the near-instantaneous companding allows for a reduction in the number of bits per sample from 14 bits to 10 bits. This companding is applied on blocks of 32 successive samples with a complementary scale factor on five ranges. A 2's complement representation is still used for each sample. In that case, the use of pre-emphasis is recommended to reduce the programme modulated noise.

2.3 *PCM emphasis*

The quality of the transmission system, in particular the noise and distortion characteristics, depends largely on the signal statistics and the coding law used. Also, the audible noise detection response of the receiving system, including the human ear, can be considered to be non-uniform. Taking these factors into account, the characteristic could be improved by adopting appropriate pre-emphasis prior to the coding process and the corresponding de-emphasis at the receiver following the decoding process. For PCM systems two such emphasis methods have been exhaustively studied and evaluated. Annex I contains more detailed information on the two systems, i.e., one based on CCITT Recommendation J.17, the other based on 50/15 μs.

3. Adaptive delta modulation (ADM)

3.1 *Background*

The adaptive delta modulation (ADM) system is based on delta modulation (the decoder is mathemetically defined in § 3.2). The audio is encoded into digital form by a simple delta modulator after being modified by two processes:

- the audio is passed through a variable pre-emphasis network which alters the audio spectrum; and
- the audio is compressed in level, based on its slope. The compression is "infinite" over a 48 dB input signal level; that is, the audio is compressed to the same level for conversion to digital form. Optimum loading of the digital channel is achieved for most audio signals.

The digital bit stream representing the encoded audio signal is transmitted to the receiver, along with two very low data rate bit streams containing control signal information so that the decoder can precisely reverse the processing performed by the encoder. The decoder recovers audio by integrating the audio bit stream; performing a dynamic range expansion based on the compression (slope) control signal; and performs spectral de-emphasis based on information in the emphasis control signal.

The following points may be noted concerning the use of delta modulation:

(a) the bit rate of delta modulation may be reduced with a minor effect on audio signal quality (S/N degrades by 9 dB for a halving of bit rate);

(b) reproduced errors are tolerable;

(c) a digital-to-analogue converter for delta modulation is very simple, requiring no precision components;

(d) the de-emphasis provides adequate output filtering and sharp cut-off low pass filters are not required.

Items (c) and (d) affect the required decoding circuitry.

3.2 *ADM decoder definition (Fig. 1)*

3.2.1 *Audio decoder*

The audio decoder consists of a leaky integrator fed with pulses derived from the audio data bit-stream so that data 1's and 0's at a rate in the order of 250 kbit/s cause the output to move positively or negatively by equal steps. The size of the pulses is linearly proportional to an applied control signal over a range of about 50 dB, and the leaky integrator has a frequency response defined by:

$$(1 + sT_0)^{-1}$$

where $T_0 = 0.5$ ms.

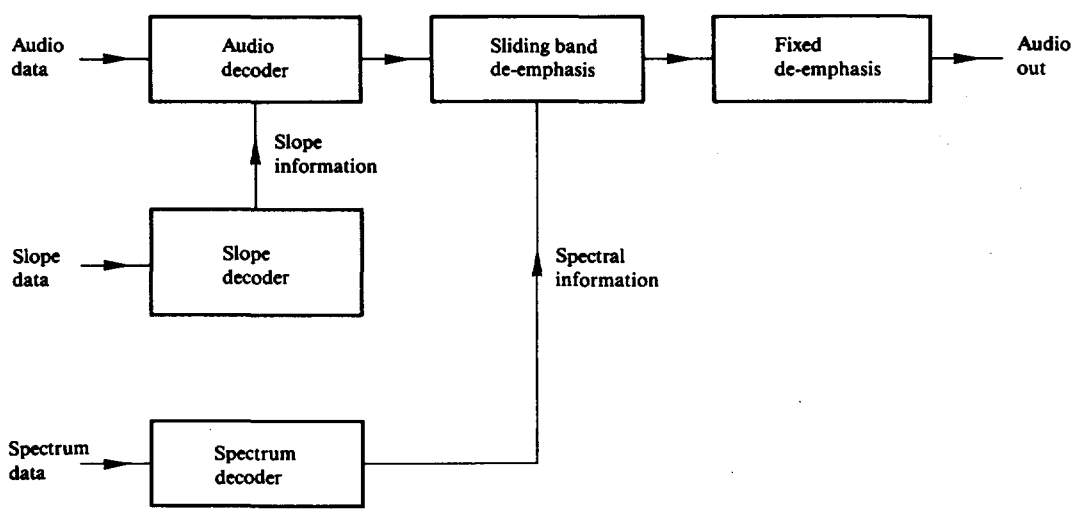


FIGURE 1 – *Adaptive delta modulation decoder*

3.2.2 *Sliding band de-emphasis (Fig. 2)*

The de-emphasis has the variable frequency response:

$$\left[\frac{10sT_1}{1+sT_1} + \frac{1+sT_2}{1+sT_3} \right]^{-1}$$

where $T_2 = 5$ µs, $T_3 = 50$ µs, and T_1 is variable under the control of the spectrum data.

Note that the first term in the bracket is a high-pass characteristic at frequency f_1 where:

$$f_1 = \frac{1}{2\pi T_1}$$

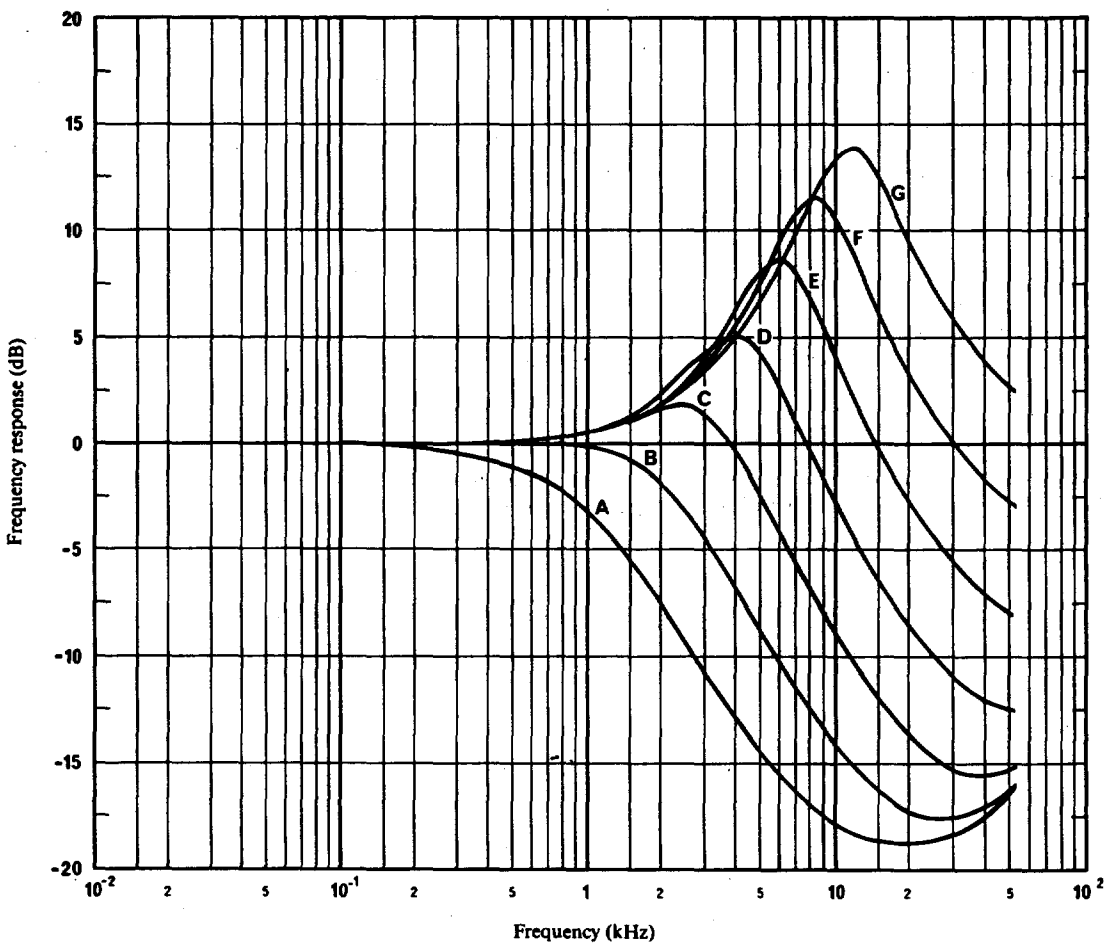


FIGURE 2 – *Adaptive delta modulation (ADM) de-emphasis filter*

3.2.3 *Spectrum decoder*

The de-emphasis control information is contained in a data bit stream at about 8 kbit/s, typically but not necessarily an integral sub-multiple of the audio data bit-rate. The mean level V_m of the bit-stream is derived by feeding the data via a three-pole low-pass filter with the characteristic:

$$(1 + sT_4)^{-3}$$

where $T_4 = 2$ ms.

If the data pulses have height V_p, a controlling parameter x is given by:

$$x = \frac{V_m}{V_p}$$

The output of the filter is fed into an exponentiator to produce a control signal to operate on the variable emphasis. The circuit constants are such as to provide the relationship:

$$f_1 = \frac{1}{2\pi T_1} = 4000\,(2^{10x}) \qquad \text{Hz}$$

With this definition, a change in x of 0.1 moves f_1 by one octave.

3.2.4 *Slope decoder*

The signal slope information is contained in another data bit stream at about 8 kbit/s. This bit stream is converted to a control signal by means of a low-pass filter and exponentiation exactly as in the spectrum decoder. The height of the pulses integrated in the audio decoder is linearly proportional to this control signal.

As in the spectrum decoder, the mean level V_m of the bit stream is derived using a three-pole low-pass filter with the following characteristic:

$$(1 + sT_4)^{-3}$$

where $T_4 = 2$ ms,

and is then fed into an exponentiator to provide the step or pulse size control signal V_{ss} for use in the audio decoder. This exponentiator has the characteristic:

$$V_{ss} = V_0(2^{10y})$$

where y is the normalized mean level of the pulses (as x above) and V_0 is a constant scaling factor to suit the audio decoder. With this definition, the pulse height changes 6 dB for each change in y of 0.1.

3.2.5 *Fixed de-emphasis*

The fixed de-emphasis is a single pole low-pass filter with the characteristic:

$$(1 + sT_5)^{-1}$$

where $T_5 = 25$ µs.

In consumer decoders no further output filtering is necessary.

BIBLIOGRAPHY

FIELDER, L. D. [September, 1985] Pre- and post-emphasis techniques for audio recorders, *J. Audio Eng. Soc.*, Vol. 33, 9.

FIELDER, L. D. [October, 1985] Audible modulation noise in floating point conversion systems, *J. Audio Eng. Soc.*, Vol. 33, **10**.

FIELDER, L. D. [October, 1985] Assessing the low level sound quality of digital audio systems. Presented at the 19th Convention of the Audio Engineering Society, Preprint No. 2268.

TODD, C. and GUNDRY, K. [March, 1984] A digital audio system for broadcast and prerecorded media. Audio Engineering Society, Preprint No. 2071, Paris, France.

TODD, C. and GUNDRY, K. [1984] A digital audio system coding method for DBS, cable and terrestrial broadcasting. Conf. Publ. 240, Ninth International Broadcasting Convention (IBC 84), Brighton, UK.

A COMPATIBLE DIGITAL AUDIO FORMAT FOR BROADCAST AND CABLE TELEVISION

Craig C. Todd
Dolby Laboratories
San Francisco

0. Abstract

The addition of a digitally modulated carrier to the M-NTSC broadcast television signal would make it possible for television broadcasters to deliver digital audio and data. In laboratory tests, a QPSK carrier modulated with 512 kbits/sec of data has been added at a frequency of 4.85 MHz above the vision carrier. At a carrier level of −20 dB with respect to vision carrier, the new signal has been shown to be compatible with the BTSC stereo and SAP signals.

I. Introduction

The BTSC multichannel television sound system (ref 1), which is being adopted by some U.S. broadcasters, is based on analog techniques. While BTSC sound is theoretically capable of excellent performance, the realities of inter-carrier analog sound detection prevent its full potential from being realized. The output of a BTSC stereo decoder invariably contains numerous 'birdie' and 'buzz' products. The resulting quality is inferior to all other common consumer audio formats: compact disc, compact cassette, FM radio, LP disc, VTR, and LaserDisc. Numerous problems affect the carriage of the BTSC signal in cable television (ref 2). While many cable systems are successfully carrying the signal, success is deemed achieved when the received audio quality may be judged 'adequate' (ref 3). Since stereo TV sound is such a novelty, the quality of BTSC audio is not presently an issue. Television sound quality in the US has typically been poor, and with a poor quality source, the quality of the delivery is not very important. The adoption of the BTSC stereo system has, fortunately, stimulated improvements in all areas of TV sound production. Much more care is taken in the production of stereo audio, and the use of noise reduction around the audio vtr is becoming more common. The improvements are audible to both the mono and the stereo listener. Eventually the improved source material and the quality conscious consumer will begin the meet in mass, and the analog BTSC stereo TV sound system will become a limitation.

The limitations of analog multichannel sound methods were recognized in the UK, Sweden, and Finland. Those countries discovered that in order to obtain even fair sound quality, it was necessary to use carriers and levels which began to create interference into the TV picture. In 1983, the BBC showed that it was possible to add a digitally modulated carrier to the I-PAL broadcast signal (ref. 4) *(Figure 1a)*. Later it was shown that a similar digital carrier could be added to the B-PAL broadcast signal *(Figure 1b)*. Although received by inter-carrier methods, the digital signal is sufficiently rugged to withstand this process with little ill effect, and a transparent link between the broadcaster and the receiver is achieved. These countries rejected analog techniques, and selected a stereo system based on digital methods.

The author noticed the similarity between the B-PAL broadcast spectrum, and that of M-NTSC *(Figure 1c)*. Both systems have a 1.5 MHz spacing between the FM sound carrier and the adjacent channel vision carrier. The lower vestigial sideband extends 750 kHz at full amplitude below the vision carrier in both systems. It became apparent that many of the results obtained testing a new digital carrier added to system B-PAL, could be directly applied here in the U.S. The major difference between the systems is the modulation of the FM aural sound carrier which, in the U.S., often carries BTSC stereo and SAP channels. The wider bandwidth of the BTSC signal, and more significantly the wider bandwidth of the BTSC sound receiving circuitry, means greater susceptibility of interference from a new digital sound carrier. In order to determine whether a digital carrier could be introduced into the M-NTSC broadcast signal and not interfere with the BTSC signal, laboratory tests were conducted.

0098 3063/87/0800 0297$01.00 © 1987 IEEE

Manuscript received June 5, 1987.

III. QPSK Modulation

QPSK modulation involves sending a carrier which takes on one of four phase states (0, 90, 180, or 270) degrees during each data symbol period. Since there are four states to choose from, each data symbol carries two data bits. It is perhaps easiest to consider a QPSK modulator as a pair of bi-phase modulators working in quadrature as shown in *Figure 2*. Each of these modulators handles half of the data, or 256 kbits/s.

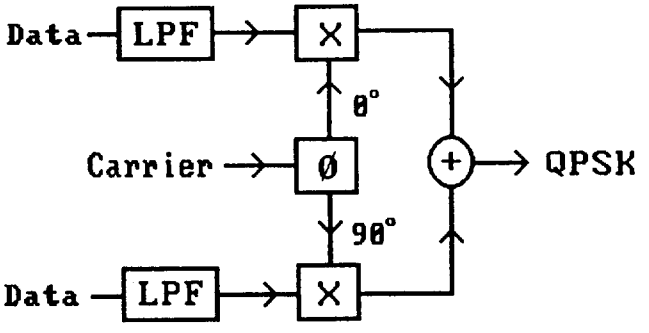
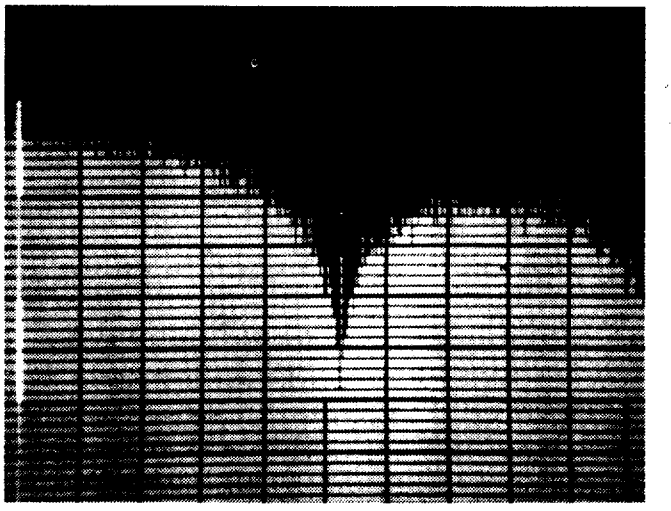


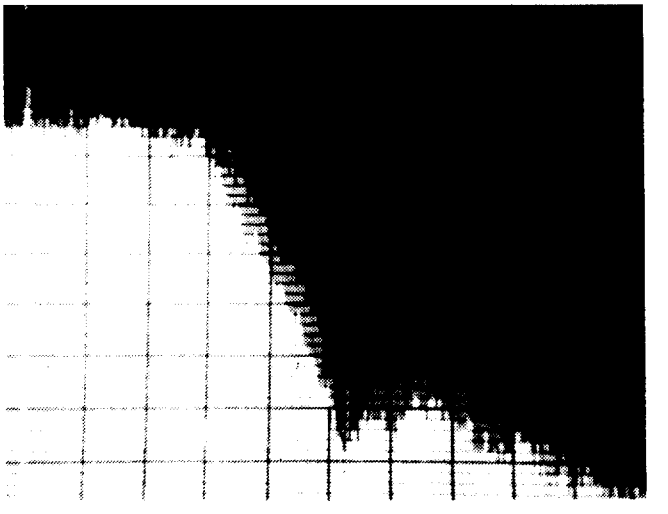
Figure 2 *QPSK Modulation*

A digital data stream is low pass filtered to constrain the data spectrum. The spectrum of a random 256 kbits/sec data stream is shown in *Figure 3a*. The spectrum extends to infinity with nulls every 256 kHz, and repeating lobes of diminishing amplitude. Only the first two lobes are shown. In theory, it is only necessary to transmit half of the main lobe, or up to 128 kHz (achieving 2 bits/Hz bandwidth efficiency). In practice it is necessary to use somewhat more spectrum, and a parameter known as Alpha specifies the fractional excess bandwidth over the theoretical minimum. In *Figure 3b*, the effect of two different data filters are shown; Alpha = 0.3 (the narrower spectrum) and Alpha = 0.7. The sharper the filtering used, the narrower the RF spectrum which will be transmitted and the less interference which will be caused.

Sharply filtering the data has the undesirable effect of making the system less tolerant of some transmission impairments, and requires tighter tolerances on circuit elements. This is due to the time domain effect of filtering on the transmitted digital pulses. *Figure 4a* shows the 'eye pattern' (the eye pattern is the overlap of many pulses) of an unfiltered



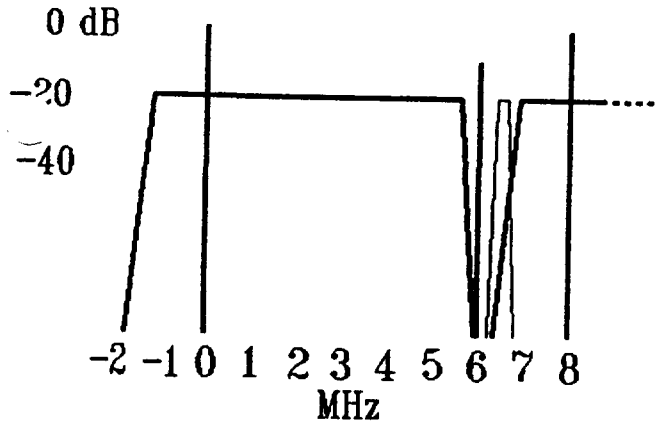
(a) *No filtering*



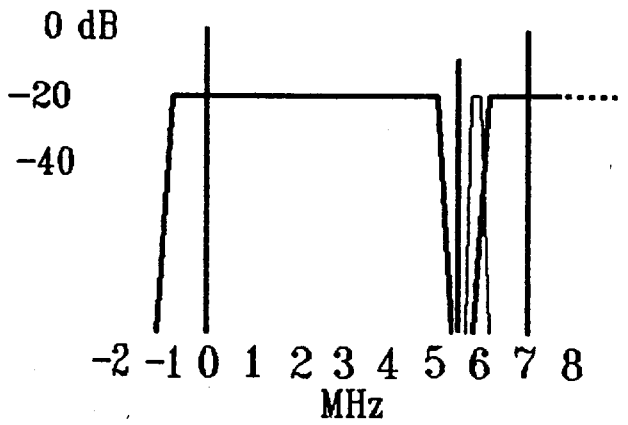
(b) *Alpha = 0.3 (bright) and Alpha = 0.7 filtering*

Figure 3 *Baseband data spectrum of 256 kb/s psuedo-random data. Vertical scale 10 dB/div. Horizontal scale 50 kHz/div.*

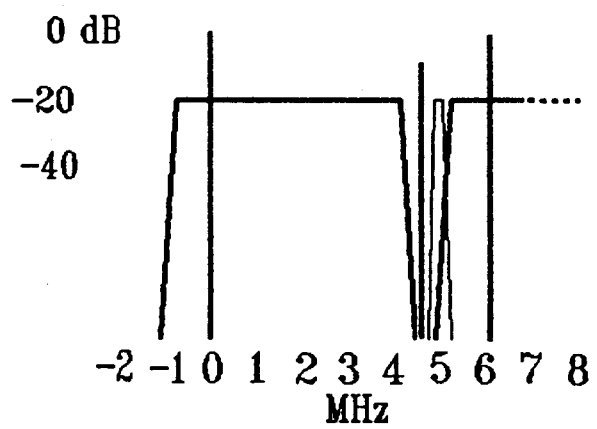
baseband data stream. This is what a decoder looks at in order to demodulate the data. This pattern has a 'wide open' eye and no data transition jitter. *Figure 4b* shows the effect of an Alpha = 0.7 data filter. The eye is now only maximally open at one point, and there is some data transition jitter. The eye pattern of an Alpha 0.3 filter is shown in *Figure 4c*. The pattern is quite a bit more complex, the maximum opening is narrower in time, and there is a lot of data transition jitter. The reason that the eye pattern becomes more complex with tighter



(a) *I-PAL Broadcast Spectrum*



(b) *B-PAL Broadcast Spectrum*



(c) *M-NTSC Broadcast Spectrum*

Figure 1 *RF Spectra of TV Systems showing additional QPSK data signal.*

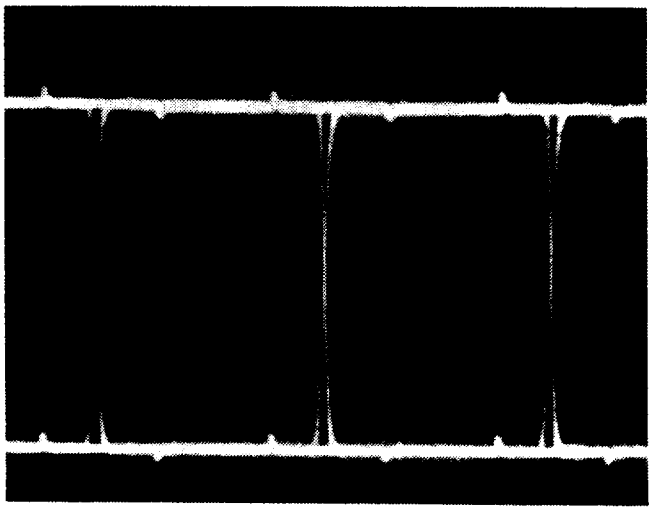
II. Outline of a New Digital Sound System

The most important requirement for the new system is compatibility. The new digital carrier must not interfere with the existing vision or sound signals. The spectrum of the new signal will resemble bandlimited white noise. The transmitted data will appear random due to the addition of a pseudo-random binary sequence at the source. Any interference will manifest itself as additive noise. There is the potential for interference into the video, and into the BTSC sum, difference, and SAP channels (the 'pro' channel is being ignored in this paper). On order to minimize the interference, the bandwidth of the new signal must be kept to a minimum.

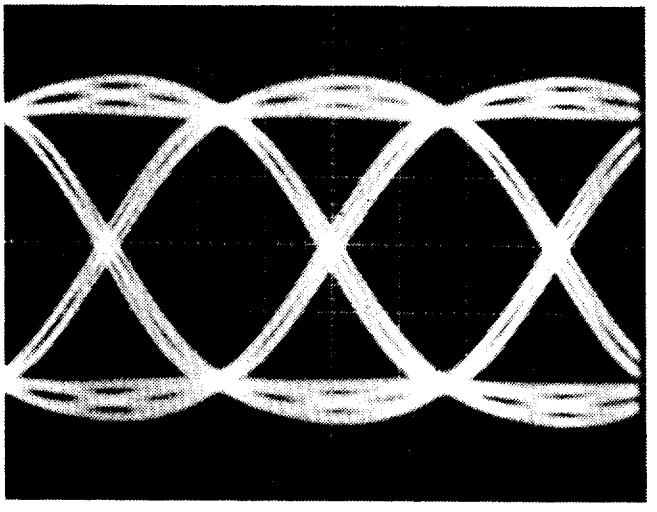
The bandwidth of a digital signal may be kept to a minimum by: minimizing the transmitted bit rate; using sharp band limiting filters; and employing complex multilevel modulation methods. The bit rate chosen must be sufficient for 2 channels of high quality audio. The use of an advanced form of adaptive delta-modulation sound coding (ref. 5,6,7,8,9) allows very high quality sound to be coded into as little as 440 kbits/sec. The audio decoder for this system is very simple, and a stereo IC decoder has already been designed (ref. 10). Allowing for some extra housekeeping and auxiliary data, a data rate of 512 kbits/sec is a good target. Complex bandwidth efficient digital modulation methods trade off bandwidth against ruggedness. They also involve additional circuit complexity which translates into higher cost. A good compromise between these factors is offered by quadrature phase shift keying (QPSK). Bandwidth efficiency can approach 2 bits/Hz, and several IC chips to demodulate QPSK are already available.

Based on work done in Scandinavia, and that reported on here, these are the tentative parameters of the new system:

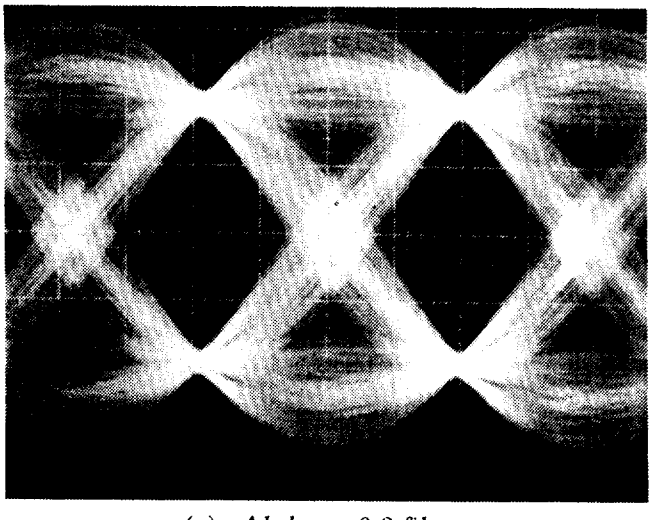
Carrier Frequency	4.85 MHz above vision carrier
Carrier Level	−20 dB relative to vision carrier
Modulation	Differential QPSK
Bit Rate	512 kbits/sec
Audio Coding	Adaptive Delta-Modulation



(a) *No filtering*



(b) *Alpha = 0.7 filter*



(c) *Alpha = 0.3 filter*

Figure 4 *Eye patterns at data demodulator*

filtering is that each data pulse 'rings' for a longer time, and thus has the potential to affect the demodulation of more of the other pulses. With ideal filtering this ringing can be controlled so that it does not create a problem, but any deviation from the ideal degrades a narrow band system sooner than a wide band system. Another problem with narrow filtering is that the longer ringing pulses create more havoc if multipath is present, because more pulses can interfere with the pulse being detected. The complexity of the eye pattern correlates directly to the ruggedness of the data when passed through channels with imperfect amplitude flatness and non-linear phase response. The more complex eye pattern will degrade rapidly when exposed to any additional filtering, or effects such as multipath. For a consumer system, it is desireable to keep things simple and non-critical and thus the Alpha = 0.7 filter is preferred.

In the QPSK modulator of *Figure 2*, one filtered data stream modulates a 0 degree carrier which results in a 0 or 180 degree phase shifted output. The original baseband data spectrum is shifted up to the carrier frequency and becomes doubled sided. Since each sideband contains the same information, half of the bandwidth is wasted. The other data stream modulates a 90 degree quadrature carrier. Its' output is either 90 or 270 degrees. When the two carriers are summed, the result will have one of 4 phase states. The two quadrature carriers occupy the same bandwidth, and the composite signal now has different upper and lower sidebands, thus fully utilizing the spectrum. The absolute phase is not available to the decoder. The decoder can only detect phase changes, so the data must be encoded differentially. The phase change from one data symbol the next (a change of 0, 90, 180, or 270 degrees) carries the two bits of information. This requires some additional digital circuitry at each end.

IV. BTSC Compatibility Tests

As previously mentioned, the effect of the new digital signal on the BTSC audio will be one of additive noise. The interference will be worst into the SAP channel since the SAP spectrum extends farthest from the 4.5 MHz sound carrier. There will be less interference into the L-R channel, and less again into the L+R channel. Testing for BTSC compatibility involves measuring the additional noise caused by

the new signal. The test setup is shown in *Figure 5*. A pair of high quality CATV modulators generate adjacent channels 3 and 4. Channel 3 is outfitted with BTSC stereo and SAP and with the new digital ier. The RF level is controlled before entering various TV sound receivers, which included high-end consumer TV sets and a specialized BTSC receiver. *Figure 6* shows the RF spectrum delivered to the receivers.

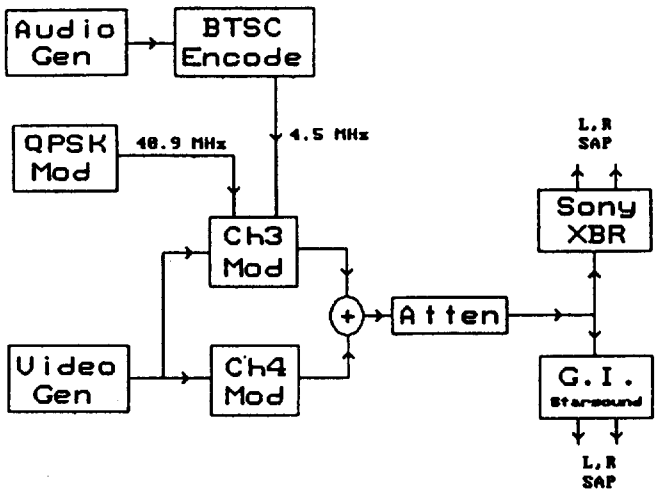


Figure 5 *Test Setup*

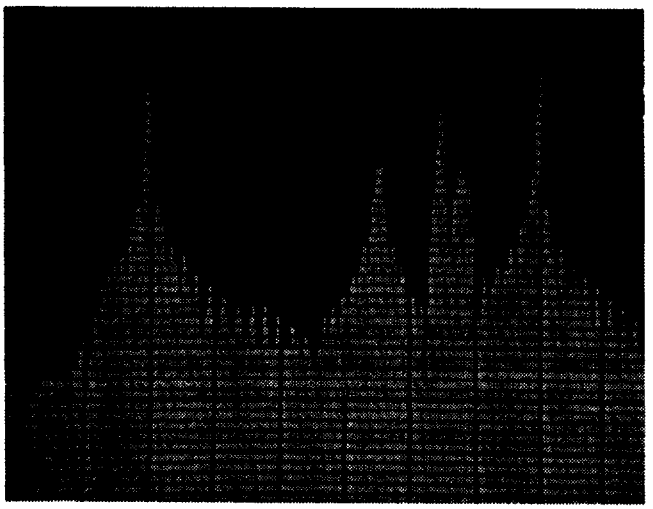


Figure 6 *RF Spectrum of signal at receiver showing Ch3 vision carrier, color subcarrier, FM sound carrier, QPSK data carrier, Ch4 vision carrier.*

Due to the companding applied to the L-R and SAP channels, it is necessary to use a test signal to 'open up' the receive expander so that the noise will be apparent. The test signal chosen was 5 kHz at 100% modulation. In order to be most critical of the interference caused by the QPSK signal we want to obtain the best possible performance from the BTSC system. This is achieved by using the best receiver operating with plenty of RF signal level. The RF level for the tests was 27 mVrms. *Figure 7* shows the SAP output spectrum of the high end consumer TV with video (color bars) modulation on (upper trace) and off (lower trace). To give BTSC the best chance, compatibility measurements were done with no video modulation. *Figure 8* shows the performance of a high end consumer TV (upper trace) compared with the specialized BTSC receiver (lower trace). Since the specialized receiver gives much better performance, it was used for the tests. *Figure 9* shows what happens when the QPSK signal is turned on (upper trace) and off (lower trace). Since the distortion components still dominate it is not really feasible to measure noise with a SINAD (notch out the fundamental and read what's left) measurement. With an HP3561A FFT based signal analyzer it is feasible to set up a band noise measurement. The vertical dotted lines in *Figure 9* show the band 6200 Hz to 9600 Hz. Since there are no significant distortion products in this band, the total energy in the band can be used as a relative noise indicator.

To test for interference we measure the noise in the 6k2Hz to 9k6Hz band as a function of:

1) QPSK carrier freq. [4.80, 4.85, 4.90 MHz above vision carrier]
2) Data filtering applied. [Alpha = 0.7, 0.3]
3) QPSK carrier level. [swept down from −10dB rel vision carrier]

The absolute numbers obtained are not of much significance. What we are looking for is the amount of noise increase caused by the presence of the QPSK signal, which will be a measure of the compatibility. If the QPSK signal increases the noise by only a little bit then it can be considered compatible. If it increases the noise by a great deal, then it would be considered incompatible. The judgement of what constitutes compatibility is a subjective one and different judges might come to different conclusions given the same data.

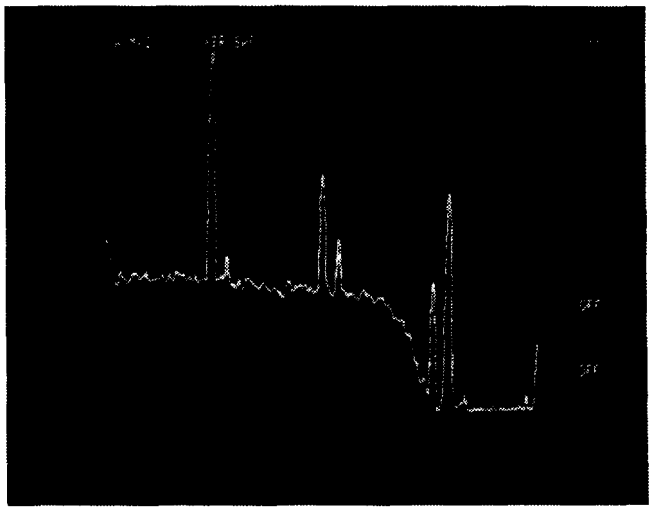


Figure 7 *TV receiver SAP output. SAP channel modulated with 5 kHz, 100%. Video modulated with color bars (upper trace) and unmodulated (lower trace).*

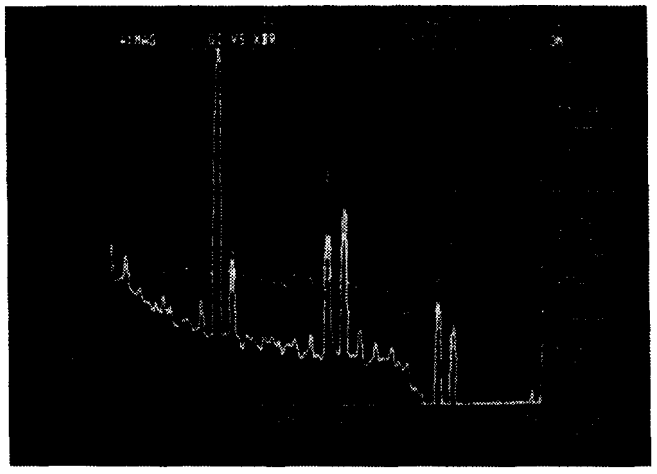


Figure 8 *SAP output of consumer tv (upper trace) and specialized BTSC receiver (lower trace). SAP modulated with 5 kHz, 100%. Video unmodulated.*

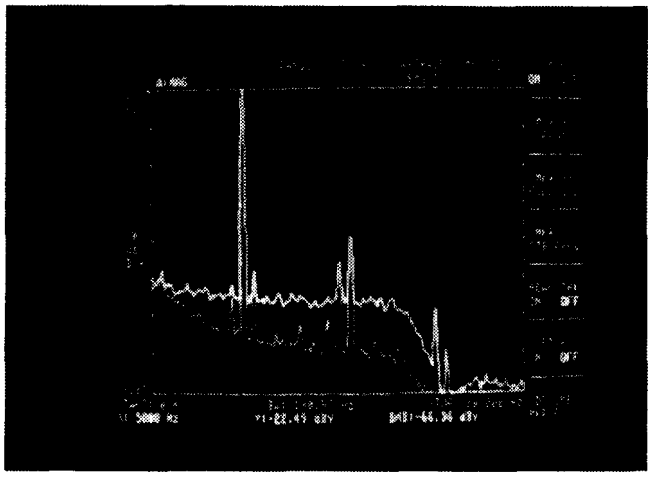


Figure 9 *SAP output of specialized receiver. SAP modulated with 5kHz, 100%. QPSK data on (upper trace) and off (lower trace). Noise measurement band delineated by dotted vertical lines.*

The results of the SAP tests are shown in *Figure 10*. As expected the noise level is worst with the QPSK carrier closest to the FM sound carrier (4.8 MHz) and with the wider filter (Alpha = 0.7). Using a narrower filter (Alpha = 0.3) moves about half way to the 4.85 MHz Alpha = 0.7 curve, which indicates that the extra filtering is narrowing the low side of the transmitted data spectrum by about 25 kHz. The difference between Alpha 0.7 and 0.3 filters would imply a narrowing of 50 kHz but since the filtering is split between the transmitter and receiver, the actual narrowing in the channel is less than might be expected. At the target operation point of 4.85 MHz, Alpha = 0.7, −20 dB, the degradation to SAP is approximately 9 dB. If the carrier frequency is lowered, the interference rises rapidly so 4.85 MHz appears to be the closest the QPSK carrier can be placed to the 4.5 MHz sound carrier. Operation at 4.9 MHz would only create a few dB of degradation, but may create problems with adjacent channel operation.

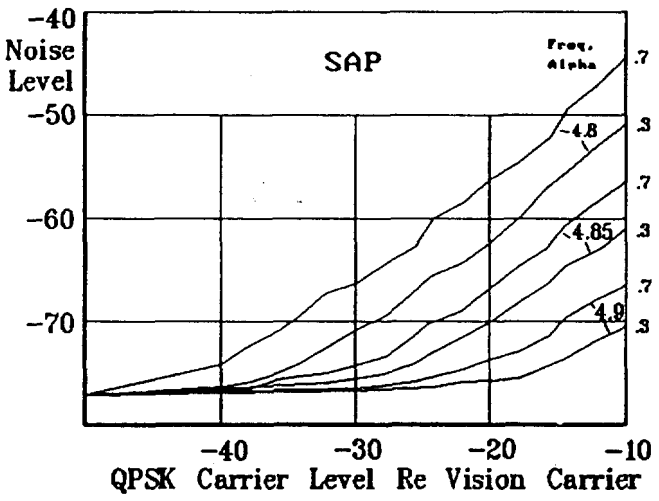
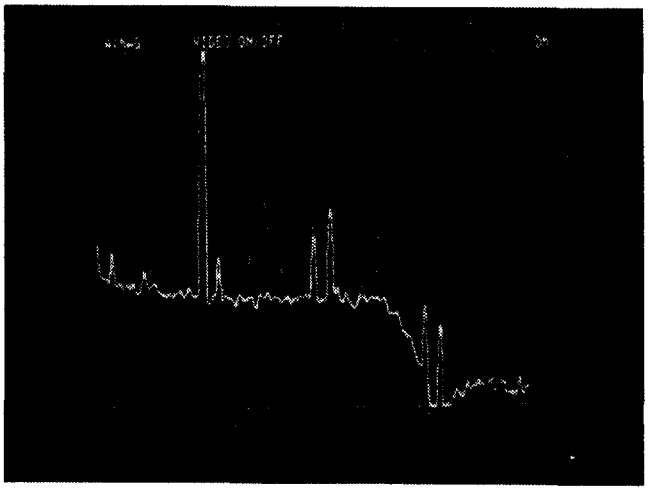


Figure 10 *SAP noise level in band 6200 Hz to 9600 Hz. Specialized BTSC receiver. FM sound carrier level -10 dB relative to peak vision carrier. Vision carrier level 20 mV rms. SAP modulated with 5 kHz, 100%. Video carrier unmodulated. QPSK data rate 512 kb/s. Frequency at 4.80, 4.85, 4.90 MHz above video carrier. Data filtering with Alpha = 0.7, 0.3 filters.*

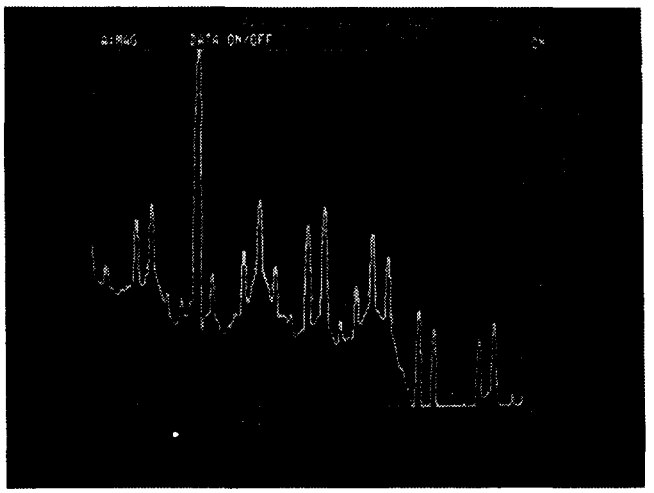
A 9 dB noise penalty might be considered large, but it must be put in perspective. This noise penalty only occurs when:

1) A high RF signal level is used (20 mVRMS).
2) A very good receiver is used.
3) No video modulation is present.

Figure 11a shows the SAP with the QPSK on (4.85 MHz, −20 dB, Alpha = 0.7) and with video modulation on (upper trace) and off (lower trace).



(a) *QPSK data on. Video modulated with color bars (top trace) and video unmodulated (lower trace).*



(b) *Video modulated with color bars. QPSK data on (upper trace) and off (lower trace).*

Figure 11 *SAP output of specialized BTSC receiver. SAP modulated with 5kHz, 100%. QPSK data on. Video modulated with color bars (upper trace) and unmodulated (lower trace).*

Even with the noise made 9 dB worse, the output spectrum is still totally dominated by distortion products. This might lead one to question whether the extra noise is even audible. *Figure 11b* shows the same situation but with the video left on and the QPSK turned on (upper trace) and off (lower trace). The increased noise in the spectral holes between the distortion products is apparent and can be heard.

The penalty for the addition of the QPSK signal is a noise penalty in the SAP channel which is audible ideal laboratory conditions. Under practical conditions the noise change would probably not be noticeable.

Of more concern is the change in noise level in the stereo signal. This is shown in *Figure 12*. While it is difficult to make out the individual curves at the bottom, the significant finding is that if the carrier frequency is 4.85 MHz or higher, there is no effect on the BTSC signal. At the target operating point, the interference caused less than 1 dB degradation in L ch noise.

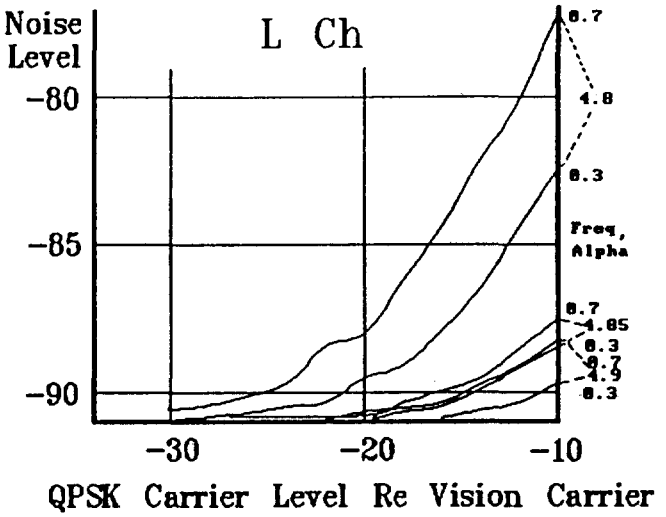


Figure 12 *Left channel noise level in band 6200 to 9600 Hz. Specialized BTSC receiver. FM sound carrier level -10 dB relative to vision carrier. Vision carrier level 20 mV rms. L channel modulated with 5 kHz, 100%. Video unmodulated. QPSK carrier frequency at 4.80, 4.85, 4.90 MHz above video carrier. Data filtered with Alpha = 0.7, 0.3 filters.*

V. Cable TV Compatibility

Cable television systems make use of adjacent television channels. There is the potential for interference from the digital signal into the adjacent

channel vision signal. Tests in Scandinavia, and preliminary tests here in the U.S., indicate that the interference into the adjacent video is not a problem. TV sets designed to work with adjacent channels have adequate rejection. TV sets not designed to work with adjacent channels show interference from the adjacent channel sound and color signals before they show interference from the digital sound signal. This is due to the noiselike spectrum of the new signal which, unlike analog carriers, does not create any patterning on the screen. Further testing will be required to confirm these preliminary findings.

Of more concern is the interference from the adjacent channel vision signal into the digital signal. *Figure 13* shows the QPSK spectrum and the peak level of an adjacent channel spectrum. This display was produced by modulating the upper adjacent channel with a 100 IRE frequency sweep, and setting the spectrum analyzer to peak hold. The shape of the modulator SAW vestigial sideband filter is revealed. The bright area is the overlap of the spectra. With this particular modulator, this did not produce data errors, but the margin against error was reduced. It may be necessary to add some additional lower sideband filtering to assure a good margin against error.

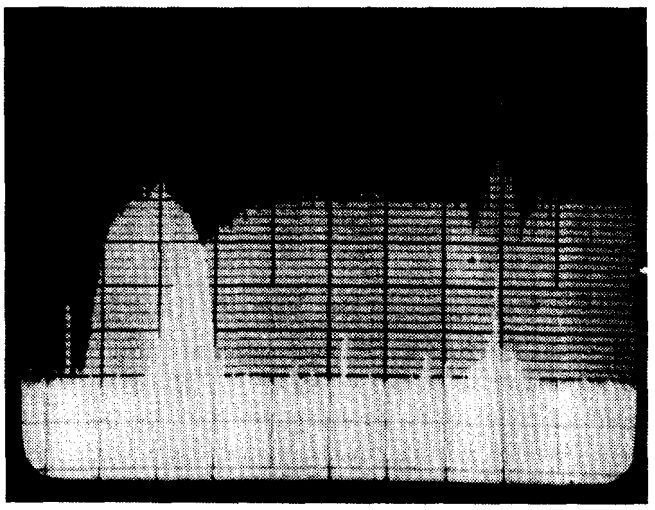


Figure 13 *Overlap of adjacent channel lower sideband and QPSK signal. Vertical scale 10 dB/div. Horizontal scale 200 kHz/div.*

Cable televisions systems often use various types of video scrambling systems. It is possible that the techniques used by some video scrambling systems will create interference into the digital signal. The worst case appears to be sync suppression scrambling systems in which the sync information is carried by a pulse amplitude modulation of the 4.5 MHz FM sound carrier. The pulsing of the FM carrier level at the horizontal rate creates a comb type spectrum which is bandlimited by a band pass filter in the scrambler. *Figure 14* shows the spectrum of such a system, with the QPSK data spectrum in the background. The high side of the comb spectrum overlaps the QPSK data spectrum. Even with this substantial overlap, the data was correctly received without errors, although the margin against error was reduced. In order to handle this situation with sufficient marging against errors, it may be necessary to remove some of the comb energy above 4.65 MHz.

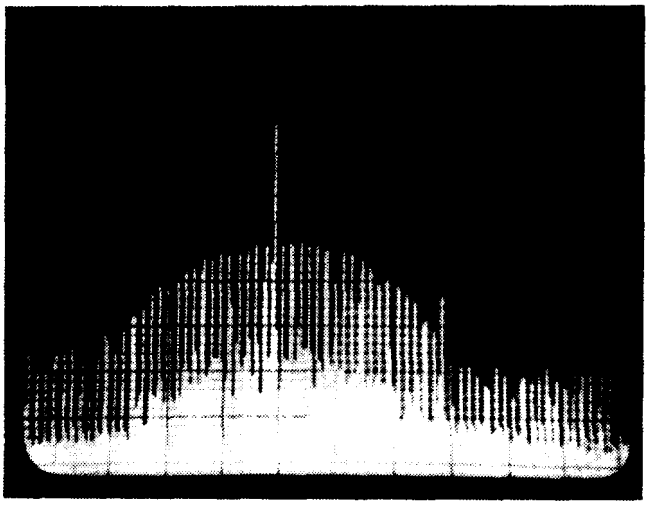


Figure 14 *Spectrum of pulsed FM sound carrier overlapping QPSK data signal. Vertical scale 10 dB/div. Horizontal scale 100 kHz/div.*

VI. Conclusion

A QPSK modulated digital carrier with sufficient capacity to convey digital stereo audio may be added to the M-NTSC broadcast signal. The new signal is very compatible with the existing BTSC stereo and SAP channels. Further tests are required to confirm compatibility with adjacent channel operation. The new signal may contain purely digital data, or a combination of one or two channels of high quality audio and auxillary data. The audio may be program audio related to the video signal, or may be a completely unrelated audio program. The new signal may be thought of as an 'SCA' for television: it allows the broadcast of more information and services in a compatible manner.

References

1. EIA, "Multichannel Television Sound: BTSC System, Recommended Practices", EIA Television Systems Bulletin No. 5, July 1985.

2. Joseph Van Loan, "The Impact of Multichannel Sound on CATV Systems", 1983 NCTA Technical Papers, pp. 245-260.

3. William F. Arnold, "BTSC STEREO: Implementation on Normal and Scrambled Channels", 1987 NCTA Technical Papers, pp. 165-169.

4. S. R. Ely, "Experimental Digital Stereo Sound with Terrestrial Television: Field Tests from Wenvoe", Oct. 1983., BBC Research Dept. Report No. 1983/19.

5. C. Todd, K. Gundry, "A Digital Audio System for Broadcast and Prerecorded Media", 75th AES Convention, preprint no. 2071, March 1984.

6. C. Todd, "Efficient Digital Audio Coding & Transmission Systems", NAB Proceedings, 39th Annual Broadcast Engineering Conference, 1985.

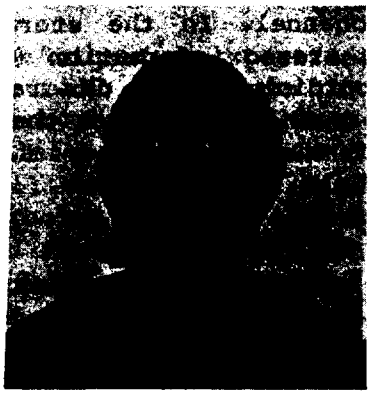
7. C. Todd, K. Gundry "A Digital SAudio System for Broadcast, Cable, and Satellite Delivery Media, 1984 NCTA Technical Papers, pp. 100-108.

8. K. Gundry, "An Audio Broadcast System Using Delta Modulation", SMPTE Journal Vol. 94, No.11, pp.1185-1190, Nov. 1985.

9. S. Forshay, "An Economical Digital Audio System for Consumer Delivery", IEEE Trans on Consumer Electronics, Vol. CE-31, No. 3, pp. 269-277, Aug. 1985.

10. T. Dhuyvetter, C. McClive, M. DeLurio, "A Delta-Modulation Decoder IC for High Quality Audio", ICCE Digest of Technical Papers, June 1985.

Craig Todd received a B.S. degree from the California Institute of Technology in 1970. While with the Signetics Corporation, he designed low noise amplifier and compandor integrated circuits. He joined Dolby Laboratories in 1977, and has designed the 4 channel matrix system used in the Dolby Surround sound system, audio cards incorporating noise reduction for 1" C format VTR's, and digital audio systems. His work with the EIAJ committee to standardize the 8mm VTR format led to the first demonstration to that body, of digital audio (using ADM sound coding) played from an 8mm tape. More recently he has worked on the development of an advanced form of ADM sound coding, for broadcast applications, which has found application on 5 continents. He has contributed to various U.S. groups such as the DBSA, the ATSC, and the NCTA engineering commmittee. He is a member of IEEE, AES, and SCTE, and has served on the US delegation to the CCIR meetings in Geneva.



A DIGITAL AUDIO SYSTEM FOR BROADCAST AND PRERECORDED MEDIA

PREPRINT N° 2071 (C7)

BY
C. TODD
K. GUNDRY
DOLBY LABORATORIES
SAN FRANCISCO CA
U. S. A.

Presented at the 75th Convention 1984 March 27-30 Paris

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd Street, New York, New York 10165 USA.

All rights reserved. Reproduction of this preprint, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

AN AUDIO ENGINEERING SOCIETY PREPRINT

A Digital Audio System for Broadcast and Pre-Recorded Media

C.C. Todd and K.J. Gundry

Dolby Laboratories, San Francisco and London

Abstract

The requirements for a digital audio system to be used for broadcast or pre-recorded media differ from those for recording in that the major economic consideration is the cost of the playback circuitry. This fact has been utilized in the digital audio system to be described. The system offers high quality sound at a relatively low data rate (on the order of 200-350 k bits/sec). No precision components are required in the decoder minimizing cost. The performance, cost and data rate advantages are achieved by placing more sophisticated circuitry in the encoder. Applications being pursued include DBS, cable, and terrestrial broadcast systems.

1. Introduction

The system described in this paper arose from Dolby Laboratories' continuing work on exploiting digital transmission and recording techniques without the high costs inherent in pcm. A previous paper (1) concerned a proposed use of delta modulation for recording television sound on magnetic tape; a consumer VTR incorporating that system would contain both an ADC and a DAC (although much of the circuitry could be in common). In broadcasting only the DAC appears in the consumer's home. This paper describes a digital audio broadcasting system in which the DAC has been further simplified, at the expense of greater complexity in the (professional) ADC. The values assigned in the paper to various circuit constants reflect the probable first application, direct broadcasting of television sound via satellites, (525 line, 60 field/sec) but it will be apparent that the principles lend themselves to many other media.

Although much of the discussion concerns two-level differential quantizing (delta modulation) the noise reduction techniques could equally be applied to multi-level differential systems. However, this would not satisfy one of the design aims, elimination of the multi-level DAC, a component necessitating high precision in manufacturing.

2. Perceived Objectives

We set ourselves the following goals:

A. We demanded a subjectively transparent system for delivery of the highest quality audio from the broadcast center to the consumer. The system should

not introduce audible degradation to audio program material of the highest quality (not only the highest quality which is available today but the highest quality which can be expected to be available in the future).

B. The system should have a high tolerance of errors, such that only a modest degradation of audio quality shall be perceived when "worst case" error conditions occur.

C. The system should be economical. The receiving equipment should be very low in cost. The system should be efficient in usage of channel capacity since then more excess capacity will exist for flexibility to add additional channels, or revenue generating services, or more bandwidth will be available to the video signal.

D. The system should be practical in operation. The transmission equipment should not require any special attention (such as very accurate level adjustment to prevent clipping) which might exceed the capabilities of the broadcast personnel, or require the use of noncomplementary signal processing.

3. Possible Solutions

Considering these goals, we compared various digital encoding schemes.

A. High bit-rate linear pcm with sophisticated error corection.

This brute force approach can satisfy objectives A, B and D easily. However it falls short on objective C. It requires bit-rates of 700 kbit/s or more per audio channel, and is inherently expensive.

B. Efficient PCM with modest error correction.

Digital companding may be used to reduce the bit rate of solution A. Digital compandors resemble analog compandors in that the level of quantizing noise rises as the signal level rises; that is, digital compandors are fundamentally susceptible to noise modulation (defined as a modulation of the perceived background noise level by the audio signal). If the transmitted code has at least 10-bit resolution in the presence of full scale signals, and high frequency pre- and de-emphasis are used, then acceptable performance can be achieved. The resulting loss of high frequency headroom is acceptable with audio program material. The error correction system may be simplified if modest degradations in audio quality at high error rates are accepted and error concealment is employed to lessen the requirements for error correction. Bit rates can then be reduced to perhaps 350 kbit/s.

However the cost saving of this approach is only modest (if any) compared with solution A because the precision of the required DAC has not been reduced. The costs are dominated by the requirements for precise 14 to 16-bit D-A converters, sharp cut-off low pass filters, and the error correction circuitry.

C. Adaptive Delta-Modulation

Delta modulation has some significant virtues. The circuitry does not require any high precision components, and can be manufactured very economically with today's technology. Since all bits have essentially

equal significance, isolated bit errors have a minor audible effect. One can consider operation without an error correction system, yielding a saving in bit rate and cost. However, linear delta modulation at the bit-rates under consideration (a few hundred kbit/s) has an inadequate dynamic range for high quality audio. We are therefore led to consider adaptive delta modulation (ADM).

Adaptive Delta-Modulation is a companded system so noise modulation must be considered. In contrast to companded PCM, noise modulation in an ADM system is caused not by high amplitude signals but by high slope signals. Noise modulation is worst in the presence of high frequency signals but the fact that high frequency signals effectively mask noise makes this characteristic acceptable and even preferable to that of companded PCM where high amplitude low frequency signals will produce noise modulation which may not be masked by the signals.

The usual technique for ADM involves coding into a single bit-stream both the audio information and the step-size or scaling information, so that the adaption of the delta modulator is necessarily output controlled. The significance of a bit then varies in accordance with the characteristics of the adaption algorithm, leading to a "gain blipping" effect on a small percentage of errors (which have hit "critical" gain control bits); this is the most noticeable degradation caused by errors on such a system. The effect can be reduced in magnitude by making the control signal move more sluggishly at the expense of poorer transient response. The adaption characteristic is therefore a compromise between speed of response (necessary for handling transient signals) and tolerance of errors; that is, objectives A and B above cannot be met together.

A judicious choice of fixed pre- and de-emphasis characteristic can provide a good compromise between audible noise modulation with mid frequency signals and high frequency handling capability, but is unsatisfactory when program material contains predominantly high frequency energy. The problem is not so much that the signals might overload the system, but that in this situation the de-emphasis curve in the decoder is more accurately described as a low frequency emphasis characteristic which actually increases low frequency noise (see section 7 below). Output controlled ADM with fixed emphasis may satisfy objective C, but is unlikely to meet objectives A and B.

4. Outline of New System

During our work on digital audio for video tape we became very well acquainted with the promises and pitfalls of delta modulation.

A fundamental aspect of this new approach is that we have significantly increased the complexity and cost of the encoder (of which very few are required) in order to lower the cost of a decoder and to remove the performance limitations of a simple ADM system.

The "gain blipping" effect of a conventional ADM system has been substantially eliminated by extracting the step-size control information and then generating from it a separate low data rate bit stream. A simple algorithm is employed to convert this bit stream into a control signal of limited bandwidth and with this algorithm all bits have equal weight. It is thus impossible for an occasional bit error to cause much of a gain variation.

A low data rate control signal suggests a sluggish response which would yield poor transient performance. We have avoided degrading transient response by employing a delay line in the encoder. This technique allows a sluggish control signal to begin to rise in anticipation of an oncoming audio transient. The conventional tradeoff is not necessary and perfect transient performance is achievable with no additional cost in the decoder.

The performance compromise inherent in fixed pre- and de-emphasis is removed by the use of a variation on our proprietary "sliding band" pre- and de-emphasis. This powerful technique gives a larger improvement in noise modulation than fixed emphasis yet does not incur a penalty of reduced high frequency headroom or low frequency noise emphasis in the presence of predominantly high frequency program material.

The method of controlling the sliding band is similar to that used for the step-size. A circuit in the encoder analyzes the spectral distribution of the program material to determine the optimum placement of the sliding band. After a delay this emphasis is applied to the audio, and the control signal is encoded into a separate bit stream which is handled identically to the step-size bit stream. Again, because of the delay line, perfect dynamics are achieved by the encoding method. The decoder complexity is not affected.

Before explaining in greater detail, it is necessary to cover some theoretical points.

5. Use of Limited Bandwidth Control Information

In adaptive delta modulation, the step-size or unit of quantization is made variable, increasing with increasing slope of the input audio signal. The operation of the adaptive modulator is equivalent to sampling and quantizing an audio waveform which has been multiplied by a further waveform representing the variation of step-size.

When one waveform is multiplied by another, as in amplitude modulation, the output signal has an extended spectrum; in this case, each spectral line of the modulating (step-size) waveform fm adds a pair of modulation side-bands to each spectral line fa of the input audio, spaced from fa by ±fm.

Similarly the action of the adaptive demodulator is to introduce side-bands. In a perfect adaptive system, the modulation and demodulation products will have exactly equal amplitudes but opposite polarities, and will therefore cancel leaving only the desired audio. This discussion is true whether the adaptation occurs "instantaneously" that is, the step-size can change from one value to a distinctly different one between two adjacent sampling periods, or whether it occurs smoothly and is continuously variable.

In a real-world system, the encoder and decoder will not track perfectly because of component tolerances and/or transmission bit errors. In an instantaneous system, discrepancies between the encoding and decoding step-sizes leave modulation products spreading across the whole audio spectrum; if the noise and distortion of the overall system are not to be degraded audibly, the step-sizes must be defined with an accuracy comparable with the minimum resolution of the quantization.

In a continuously variable system, the modulation products resulting from mistracking are the side-bands mentioned above, attenuated by partial cancellation. The lower the bandwidth of the step-size control waveform (the

smaller fm), the more tightly the modulation products are confined to the immediate vicinity of each audio spectral component.

The masking properties of any particular audio frequency are greatest near that frequency. For example, a 1 kHz signal makes low level noise and distortion components within a few hundred Hz of 1 kHz inaudible, but components at the extremes of the audio spectrum remain audible. It is this property of human hearing that makes it possible to design noise reduction systems without audible noise modulation.

Hence the narrower the bandwidth of the step-size control, the less audible will be the modulation products of mistracking, or the greater amount of mistracking can be tolerated. With the control bandwidth employed in our new system, approximately 50 Hz, mistracking of several percent remains inaudible on real program material, and component tolerances can be greatly relaxed. Furthermore error rates up to perhaps 10^{-2} in the control bit-stream can be accepted without correction.

6. Optimized Adaptive Delta Modulation

The noise and distortion emerging from a complete ADM codec depends on (among other things) the nature of the audio input signal and the step-size; both these factors are varying all the time.

Consider a codec receiving and reproducing the simplest audio waveform, a constant amplitude sine-wave. As a function of step-size, the output noise and distortion will vary as shown qualitatively in figure 1.

In the region labelled a, the step-size is unnecessarily large, giving excessive quantizing noise. In region b the step-size is too small and the system is therefore in slope overload, giving high distortion. There is an optimum value of step-size for the particular input conditions, labelled c.

For each short time segment of real audio there is a curve like figure 1, and there is a best step-size. With discrete instantaneous adaptation in which the step-size can only adopt one of a limited number of values, it is clearly impossible to operate at the best values all the time, since they will inevitably lie between the available values.

With continuously variable adaptation it is possible to operate very near the best values. In a system employing limited bandwidth control information, the relevant time segment is a window related to the rise-time of the step-size control signal, in our case around 10 ms. From examination of the audio within a moving window of duration 10 or 15 ms, an optimum step-size can be generated which minimizes the noise and distortion from the codec.

Note that a conventional output controlled ADM system, in which the bit-stream carries two pieces of information, the audio and the step-size, cannot achieve optimum step-size, but remains in region a most of the time, moving into region b on signal transients. Also of course adequate response time for signal transients indicates that the bandwidth of the step-size adaptation must be wide (many kHz), implying a need for much greater precision.

7. Variable Pre- and De-Emphasis

In an audio system whose noise is independent of the audio (for example, an amplifier with thermal noise or an FM broadcast system), the effects of fixed pre-

and de-emphasis are easy to understand and to calculate. The choice of emphasis characteristic is usually a compromise between noise reduction and overload characteristic, and in some applications noise reduction effect is offset by the need to reduce program level to the extent that little improvement in unweighted signal-to-noise ratio is observed. However the change in the spectrum of the noise may be audibly valuable.

In an optimized ADM system, the noise varies with the signal. At any fixed input frequency, the noise is directly proportional to the signal amplitude (that is, the instantaneous signal-to-noise ratio is constant), and at any fixed input amplitude the noise is directly proportional to signal frequency. Furthermore, unlike pcm, there is no clear system-defined maximum signal level which can be transmitted, although practical instrumentation may impose one.

The effects of fixed high frequency pre- and de-emphasis on such a codec are quite different. When the predominant audio input spectral components are at low or middle frequencies, that is in the unboosted area, emphasis does not change the demand on step-size in the ADC or DAC, the basic noise output of the demodulator is unchanged by emphasis, and hence subsequent de-emphasis reduces high frequency noise.

However, when a predominant audio input spectral component is within the area of frequencies boosted by emphasis, an increase step-size is demanded, with the result of increased noise output from the demodulator. The de-emphasis then does several things:

i) It restores the audio component to its correct (unboosted) level.

ii) It lowers the noise in the region of that component, but only back to the level it would have had without pre- and de-emphasis.

iii) It lowers the noise at frequencies above the audio component, but starting from the increased level.

iv) It has no effect on the low frequency noise, which therefore remains at its increased level.

Hence for high frequency signals, the effect of fixed pre- and de-emphasis is not primarily to degrade headroom or give distortion, but to increase low frequency noise, that is, the noise which is least likely to be masked by high frequency signals.

Consideration of the noise levels and spectra delivered by an optimized ADM codec operating at 200 or 300 kbit/s shows that there are three (overlapping) regimes to be considered if satisfactory noise shaping is to be employed.

A. When the predominant audio spectral components lie below roughly 500 Hz, a large high frequency pre- and de-emphasis will reduce noise sufficiently that audible noise modulation will not occur. An example of a practical emphasis characteristic is a 20 dB shelf starting at 800 Hz (see curve 1 on figure 2).

B. As the predominant spectral component is increased up to 2 or 3 kHz, it is necessary to slide the emphasis upwards in frequency so as to retain high frequency noise reduction relative to the spectral component (curves 2 and 3). Low frequency noise is not yet an audible problem.

C. When the predominant spectral component is above about 3 kHz, noise both at low and at very high frequencies must be reduced. An emphasis curve with a dip at the predominant component will reduce the step-size and hence the broad-band noise emerging from the codec, while the subsequent complementary de-emphasis peak will pick out the wanted signal component, while attenuating high frequency noise, and retaining the reduced low frequency noise level which resulted from the smaller step-size. Fox example, if the predominant signal lies at 6 kHz, curve 5 is a desirable emphasis characteristic.

This explanation has assumed that the predominant components of an audio signal at a particular moment are concentrated in a narrow region of the spectrum; such a signal is in fact the most critical case. When the spectral components are more distributed, their masking properties cover more of the noise, and the emphasis shape is less critical.

Thus a variable emphasis circuit giving a family of response curves of the form shown in figure 2 preceding the modulator, with the complementary de-emphasis following the delta demodulator, will provide efficient subjective noise reduction under all signal conditions, provided that circuitry can be designed to analyse the input audio and to give suitable instructions to the emphasis and de-emphasis.

8. Practical Details of the Decoder

Figure 4 illustrates the decoder required in the consumer's home. Each audio channel of decoding receives three data bit-streams.

The audio data is at a moderately high rate, 200 kbit/s or more. The precise rate depends on the application; for television it is convenient to operate with an integral number of bits per horizontal picture line. The two control bit-streams are at much lower rates; for television a practical rate is half the television horizontal frequency.

8.1 Basic Adaptive Delta Demodulator

In order to achieve the required dynamic range from adaptive delta modulation, it is necessary to be able to adapt the step-size or unit of quantization over a range approaching 50 dB. The basic demodulator takes a step-size voltage Vss (or an equivalent current) which can vary over this range and switches it with one polarity or the reverse in accordance with the audio data to a leaky integrator. The leak time-constant corresponds to a few hundred Hz; below this frequency the system is strictly not delta but delta-sigma modulation.

8.2 Step-Size Decoder

The step-size control bit-stream contains the logarithm of the required step-size coded as delta-sigma modulation. It is therefore decoded by passage via a low-pass filter, which determines the bandwidth (and hence rise-time) and ripple of the step-size voltage, and an exponentiation or anti-log. circuit (for example, a bipolar transistor, which inherently has the appropriate characteristic). If the normalized mean level of the bit-stream (or the duty-cycle measured over the rise-time of the low pass filter) is written as x, then

$$Vss = Vo \exp kx$$

where Vo is a constant suitable for the particular instrumentation, and $k = 10 \ln 2$

This definition gives an increase of 6 dB in step-size for every increase of 0.1 in x. Since by definition x is confined to the range 0 to 1, the resultant maximum possible range of Vss is 60 dB. Of this, about 50 dB is useful.

The transmission of step-size information in logarithmic form reduces the dynamic range conveyed in the bit-stream, in this case from about 50 dB to about 9 times, or 19 dB, and spreads the effect of bit errors more uniformly across the dynamic range. Since Vss is confined by the low-pass filter to a bandwidth of about 50 Hz, bit errors lead to slow random amplitude modulation of the output audio. The audible disturbance produced by errors in the control bit-stream is quite negligible compared with the effect of similar error rates in the audio data. The control system is so robust that uncorrected error rates up to 10^{-2} or so produce nearly imperceptible disturbance of music or speech.

8.3 Sliding-Band De-Emphasis

The requirements have been discussed in section 7; figure 3 shows a representative set of de-emphasis curves, complementary to the emphasis of figure 2.

There are obviously many ways of synthesizing responses of this nature. Figure 4 shows one practical technique; here the cut-off frequency of the high-pass filter formed by the capacitor and variable resistor should be directly proportional to the control signal decoded from the sliding-band control bit-stream.

Those familiar with Dolby noise reduction systems (such as A- or B-type) will recognize the dual-path configuration in which a main path with fixed characteristics is paralleled either with feedforward or feedback by further paths having variable characteristics. However in noise reduction the further paths constitute compressors, while in this new system the further path (the variable high-pass filter of figure 4) is ultimately controlled by the spectrum of the incoming audio; there is no systematic compression or expansion of the dynamic range.

8.4 Emphasis Control Decoder

The emphasis control decoder is substantially identical with the step-size decoder. The emphasis control bit-stream contains the logarithm of the cut-off frequency of the variable high-pass filter in the de-emphasis circuit; the virtues and benefits of a log. function here are similar to those in the step-size control. The cut-off frequency follows the relationship

$$f_1 = fo \exp ky$$

where fo is a constant determining the scaling, $k = 10 \ln 2$, and y is the normalized mean level of the bit-stream

This definition results in a movement of one octave for every increase of 0.1 in y. The control of emphasis in the presence of errors is if anything more robust than that of step-size.

8.5 Output Filter (not shown on figure 4)

Since in delta modulation the sampling frequency is vastly greater than the minimum required by information theory, there is little probability of aliasing components in the output. Non-audio spectral components in the output are at frequencies well above the audio band, and hence only an elementary two or three pole low-pass is necessary.

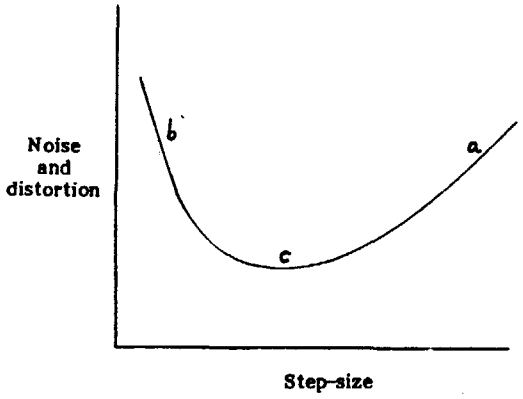


Figure 1

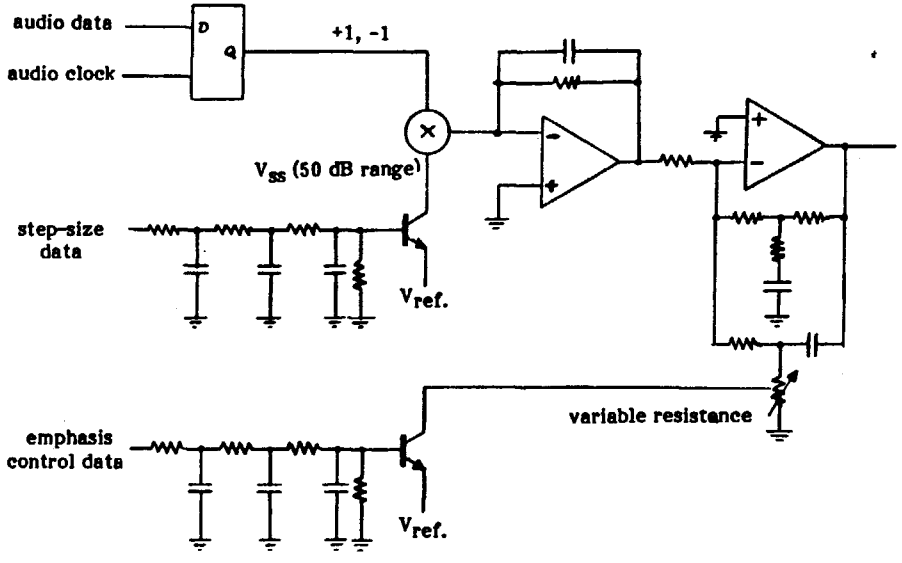


Figure 4 Consumer decoder

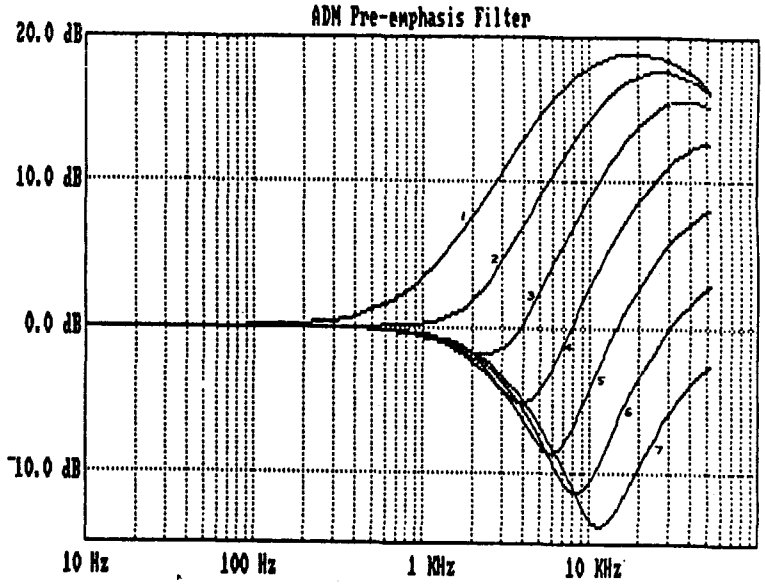


Figure 2

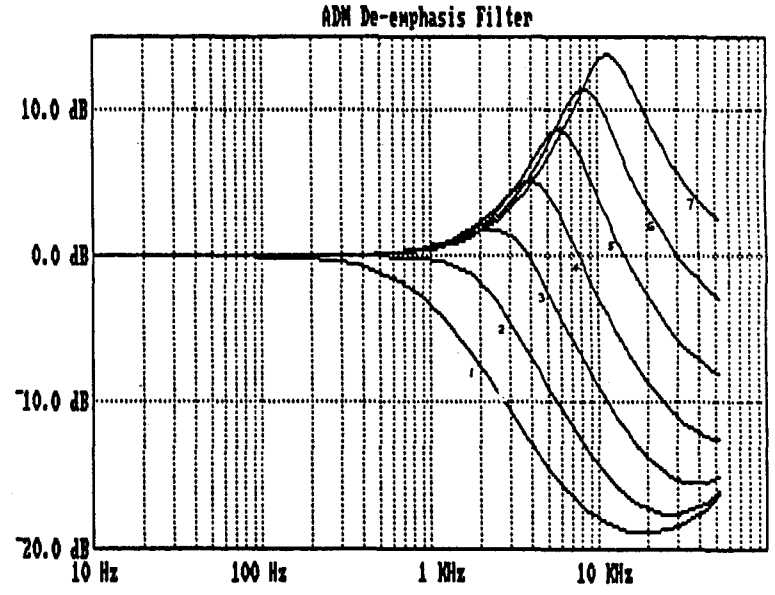


Figure 3

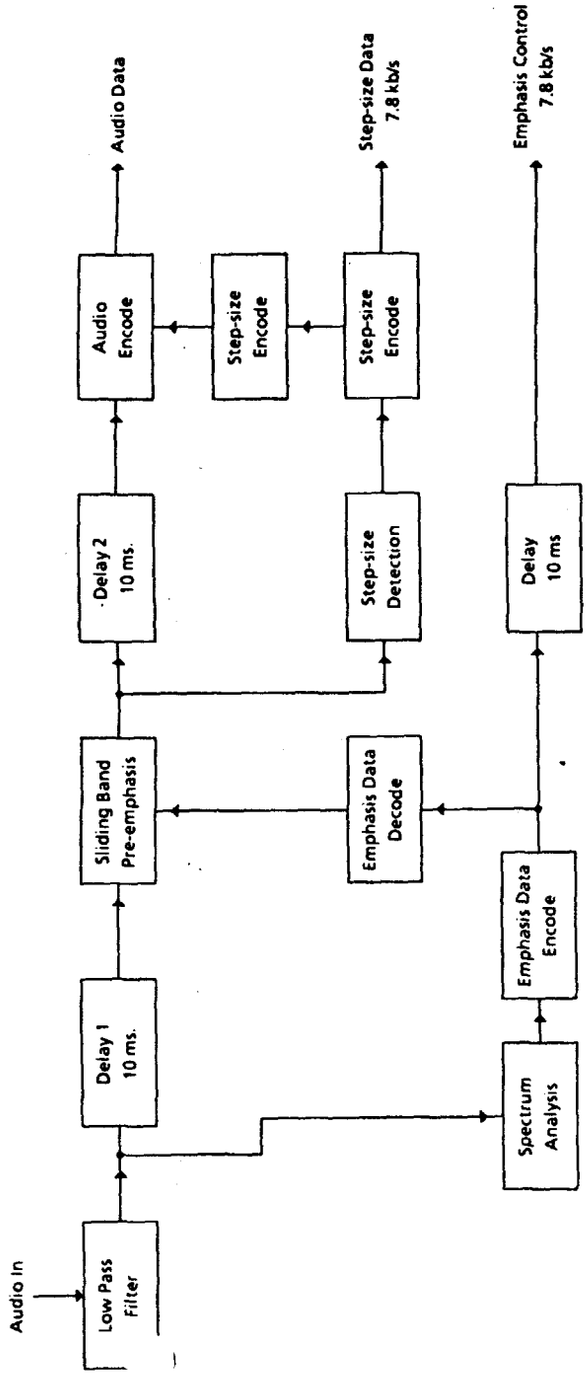


Figure 5. Professional Audio Encoder. Block Diagram

9. Requirements of the Encoder

As will be clear from the block diagram in figure 5, the encoder is very much more elaborate and expensive than the decoder; this is not a serious objection since it does not need to be mass produced for the consumer market.

The basic requirement of course is to deliver bit-streams which can be interpreted correctly by the decoder.

The emphasis control block analyzes the spectrum of the input audio to determine which of the family of available emphasis and de-emphasis curves will minimize the audibility of codec noise in the presence of that signal spectrum. This information is coded into the low-bit rate emphasis control data stream. The conversion of this bit-stream back to a signal suitable for operating on the variable filters of the emphasis and de-emphasis includes limitation of the bandwidth to about 50 Hz, corresponding to a rise-time of about 10 ms. Hence the input audio is delayed by this time before entering the variable emphasis circuit.

After emphasis, the audio is passed to the step-size detection block which measures the slope of the signal to determine the optimum step-size (point c on the curve in figure 1). The logarithm of this value is coded into the step-size control data. Again conversion back to actual step-size voltage Vss restricts the rise-time, and so the emphasized audio is delayed by about 10 ms before application to the main encoding delta modulator.

Since the audio is subject to this further delay after emphasis, the emphasis control bit-stream needs delaying also so that all the data will arrive synchronously at the decoder.

10. Error Concealment

Our previous work (1) illustrated the nature of the difference between a bit error reproduced by a PCM or DM system. In the PCM case the reproduced error takes the form of a narrow impulse, the height of which depends on which bit is hit (being very large for an MSB hit). In the case of delta modulation, the error takes the form of a small step of significant length (decaying only because the integrator has a "leak" with a 0.5 msec time constant); small because there is no possibility of hitting an MSB (there isn't one-all bits are equal!). If the error rate is low (on the order of 10^{-5}) a delta modulation system is quite usable without any error correction or concealment, while a PCM system is not.

PCM systems intended for operation at moderate to high error rates (10^{-4} to 10^{-3}) must (to be usable) carry some overhead for error concealment. No attempt is made to correctly reproduce the audio, but using a small amount of redundant data, bad words are identified and an interpolated value substituted. A very substantial audible improvement can be realized with a modest amount of data overhead and a reasonable amount of additional circuitry (storage of a previous sample, addition of the next sample and shift right for divide by two). The result of concealment is to greatly reduce the amplitude of the impulse which is added to the reproduced audio.

Delta modulation systems can benefit from a somewhat analogous error concealment scheme which yields a similar result with simpler circuitry. The scheme involves detection of the approximate location and polarity of the error, and the creation of an error of opposite polarity in nearly the same location. This has the effect of terminating the duration of the step the error has created yielding an impulse similar to the concealed PCM impulse.

The implementation of the scheme involves blocking the data into N bit blocks and performing a modulo 4 summation of the number of data 1's in the block. This two bit result (0, 1, 2, 3, 0, 1, 2, 3 . . .) is transmitted along with the data block as redundant information. If a single bit error occurs in the data block (for moderate error rates nearly all errors will be single errors) the number of data 1's in the received block will differ from that sent. If the same modulo 4 summation is performed in the decoder and compared to the result received from the transmitter, the presence and polarity of a single bit error is detected. If an error is detected, the concealment involves creating an error of opposite polarity in the block. Since we have only located the error to a particular block and we want our artificial error to be located as close as possible to the real error, we insert the artificial error into the center of the block. The distance between the real and artificial errors will then be between 0 and N/2 bits, with an average distance of N/4 bits. The step introduced into the audio will be (on average) this length. Note that some errors (2/(N + 2)) will hit the concealment data and the system may misconceal. This possibility can be virtually eliminated by carrying along a 3rd bit which is a parity on the other two, but the audible improvement doesn't really justify the added overhead unless the block size is quite small (2/(N + 2) becoming large).

By way of example, suppose we block the data into 32 bit blocks and send 2 redundant bits for each block for an overhead of about 6%. For approximately 32/34 ths of all errors (94%) we correctly receive the concealment data and can conceal the error, terminating it to an average length of N/4 or 8 bit periods. For a system operating at a 250 kHz data rate, the step would have a length of 32 μsec (about the same length as an impulse in a PCM system sampled at 32 kHz). In practical sytems the data is often organized in blocks of some size (in a video application a block of data may be sent every horizontal scan interval), so these same blocks are used to implement the concealment.

11. Conclusion

This paper has described a digital audio system particularly suited to broadcasting. It offers a bandwidth in excess of 15 kHz, a dynamic range potentially around 100 dB (although initial embodiments may be limited to perhaps 85 dB because of the limits of present IC technology and of the available means of producing 10 ms delays), and a level of program-modulated noise better than most of the digitally-companded pcm systems proposed or in use. All this with a total bit-rate potentially less than any existing pcm system, and with a decoder having a fraction of the cost.

Reference

(1) K.J. Gundry, D.P. Robinson and C.C. Todd, "Recent developments in digital audio techniques", presented at 73rd AES Convention, 1983 March (preprint 1956).

SESSION VII - ENHANCED TV AUDIO

WPM 7.3: *DIGITAL MULTI-CHANNEL SOUND FOR TELEVISION*

Anders Nyberg

Swedish Broadcast Company

Stockholm, Sweden

INTRODUCTION

Conventional television broadcasting has over the last decades been the major domestic audio-visual source. However, developments in consumer electronics have led to a situation where new products like the Hi-Fi video recorder and the Compact Disc player will set new standards for sound quality in the homes. One well known and sometimes accepted fact is that television broadcasts, in particular, suffer from poor sound quality due to distortion, noise and buzz-on-sound. Most of these problems are due to practical limitations in receiver and transmitter design which can not be avoided completely with the present analog sound system.

When a new system for stereo and multichannel sound is introduced, one evident step would be to also improve sound quality. However, some countries have already started stereophonic television broadcasts and three different systems are in use, all being based on analog technology giving no improvement but rather a degradation in terms of sound quality.

In the Nordic countries stereophonic and bilingual transmissions with television are about to start and, to this end, different proposals for a suitable system have been evaluated. In 1982 tests were carried out in Sweden with a PAL system, modified to carry digital sound code in the horisontal blanking interval (sometimes referred to as "B-PAL"). This system was found to perform well in a DBS concept, but a terrestrial application was not feasible due to incompatibility with existing receivers.

As opposed to satellite applications, where time-division-multiplexed systems have been fully accepted in the form of MAC signals, further evolution of digital sound for terrestrial television has to be based on the introduction of a new carrier. The major reason for this is of course the demand for compatibility with the existing service.

The major operational requirements for the new service are:

- High quality sound
- Stereo and multilingual transmission
- Compatibility with existing receivers
- Sound failure after vision

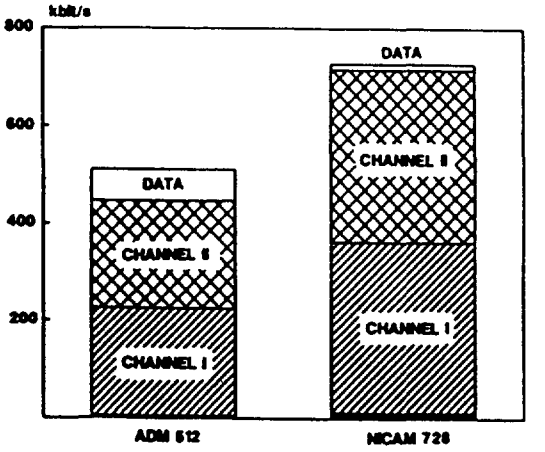
The most important constraint to be taken into account is the 7 MHz channel spacing used in the VHF-band (System B), where the difference between the analog sound carrier and the vision carrier of the upper adjacent TV channel is 1.5 MHz. In System B the vestigial sideband is specified to 0.75 MHz so the digital carrier has to be introduced slightly above the analog sound. One interesting fact is that the same situation also applies for System M (NTSC)

DIGITAL CODING OF SOUND

Two possible system candidates were studied, the first being based on near-instantaneous-companded PCM (NICAM) with a bit-rate of 704 kbit/s. This system had already been developed and tested by the BBC but was later changed to operate at 728 kbit/s with a baseband format similar to that used in the MAC/packet systems.In the latter case an additional data capacity corresponding to 11 kbit/s is available for future use.

The second proposal which was based on adaptive delta modulation (ADM) requires less transmission capacity and is less susceptible to transmission errors than PCM. As a concequence ADM was chosen to be used in our tests and led us to consider a digital carrier system operating at 512 kbit/s which more easily would fit into the restricted bandwidth available.

A very simple baseband data format was used with the two sound channels occupying 222 kbit/s each and an additional spare data capacity of 64 kbit/s for future use such as identification of services, control signalling, encryption related data etc.



ALLOCATION OF CAPACITY FOR SOUND AND DATA

 87CH2454/7/87/0000-0092$01.00 © 1987 IEEE