A Mobile Hi-Fi Digital Audio Broadcasting Scheme

L. Hanzo, X. Lin, R. Steele, W.T. Webb
Dept. of Electr. and Comp. Sc., Univ. of Southampton, SO9 5NH, UK.

Abstract

The audio quality, robustness and complexity issues of a novel mobile digital audio broadcast (DAB) scheme are addressed. The audio codec is based on a combination of subband coding (SBC) and multipulse excited linear predictive coding (MPLPC), where the bit allocation is dynamically adapted according to both the signal power in different subbands and a perceptual hearing model. Typically a segmental signal to noise ratio (SEGSNR) in excess of 30 dB associated with high fidelity (HI-FI) subjective quality was achieved for 2.67 bits/sample transmissions at a mono bit rate of 86 kbits/s. Four different source-matched forward error correction (FEC) schemes were investigated in order to explore the complexity, bit rate and robustness trade-offs. When using 4 bit/symbol 16-level star-constellation quadrature amplitude modulation (16-StQAM) the overall signalling rate became approximately 30 kBaud, accommodating two stereo DAB channels in a conventional 200 kHz analogue FM channel's bandwidth. Our diversity assisted DAB scheme required a channel signal to noise ratio (SNR) of about 25 dB for unimpaired audio quality via the worst-case Rayleigh fading mobile channel, when the mobile speed was 30 mph and the propagation frequency was 1.5 GHz. In case of the stationary Gaussian scenario an SNR of about 20 dB was required.

1 Digital Audio Broadcasting

Analogue frequency modulated (FM) radio broadcasting is antiquated and there is a growing demand for higher quality digital audio broadcasting (DAB) for mobile receivers [1], [2]. Adavanced features, such as five-channel surround sound with ambient-dependent dynamic control, catering for example for reduced dynamic range in a noisy vehicle or traffic- and control-data decoding are also desirable features.

In this contribution we propose a DAB scheme for mobile channels, which is based on a subband split modified multipulse excited linear predictive (SB-MMPLPC) codec studied in Section 2. The audio bits are protected by a variety of block codes and transmitted using 4 bits/symbol 16-level quadrature amplitude modulation (16-QAM) as discussed in Section 3, while performance figures are reported in Section 4.

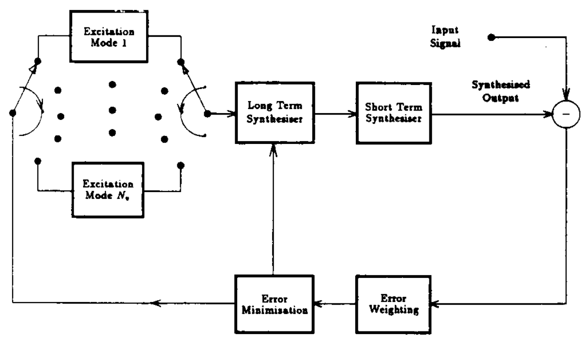

Figure 1: MMPLPC Codec Schematic

2 The Audio Codec

The **MMPLPC** codec's schematic diagram is shown in Figure 1, which is similar to that of a conventional MPLPC arrangement [3], except that it incorporates N_v number of different excitation modes, where we have opted for $N_v = 2$, corresponding to **Mode 1** and **Mode 2**. The audio input signal $s(n)$ is divided into frames of N samples for LPC analysis and the LPC filter parameters are determined by minimising the mean squared prediction error E_w over this interval. Each frame is further divided into contiguous subframes of N_s samples, for which the long term predictor (LTP) delay D and gain G_l parameters minimising the mean squared error E_w for the current subsegment are determined [4].

The MMPLPC codec's efficiency can be further improved, if the human ear's frequency and energy sensitive properties are exploited by dividing the audio bandwidth into subbands corresponding to the critical bands found in hearing. However, after band splitting, the correlation between adjacent time domain samples is reduced, and the more the band is split, the more this correlation is decreased. The MMPLPC codec utilizes linear prediction requiring high correlation between adjacent samples. In order to compromise, we chose four-band splitting.

In the **SB-MMPLPC** codec seen in Figure 2 the input audio signal $s_a(n)$ is split into four subbands: 0-4 kHz, 4-8 kHz, 8-12 kHz, 12-16 kHz, by a *Quadrature Mirror Filter* (QMF) bank, using two cascaded 64th order QMF filters [3]. The four subband signals $\{s_k(n), k = 1, 2, 3, 4\}$ are each encoded by an MMPLPC codec.

Since hearing sensitivity is different for the different sub-

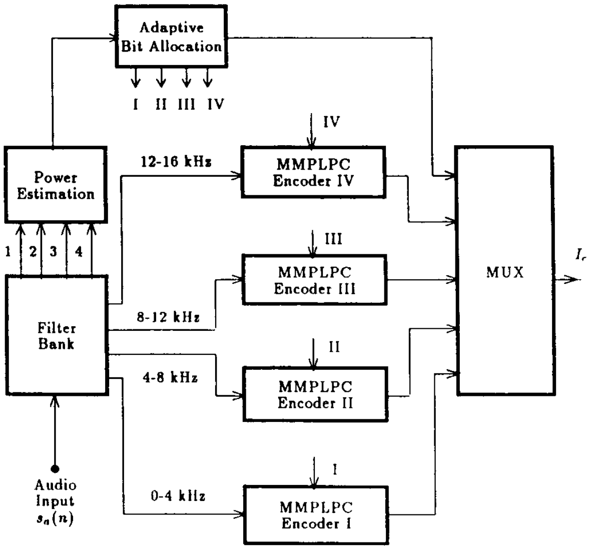

Figure 2: SB-MMPLPC Codec Schematic

bands, the short time energy σ_k^2 in each subband was estimated and subband k was assigned to one of sixteen empirically designed different bit allocation classes $C_j, j = 1 \ldots 16$, as demonstrated by Table 1 designed for subbands SB1 and SB2. Similar tables were constructed for the less significant subbands SB3 and SB4 summarising for both excitation modes the number of excitation pulses N_g, their quantisation accuracies in terms of the number of bits/pulse as well as the number of bits needed for the encoding of their positions, when using the enumerative method [5]. For the same bit allocation class $C_j(k)$ the lower frequency subbands SB1 and SB2 $k = 1, 2$ were typically assigned a higher number of excitation pulses and higher number of pulse amplitude quantisation bits than the high-frequency subbands SB3 and SB4.

The LPC analysis frame size of 20 ms was found to be suitable for every subband. As expected, the LPC prediction gain increased in each subband, when the LPC order was increased. To achieve high fidelity audio, much higher excitation densities were needed than for encoding speech.

The number of LPC filter coefficients was 6,4,4,4 for subbands 1,2,3,4 respectively. The LPC filter parameters were quantized by linear quantization of log-area ratios [3] $LAR_i(k)$. After band splitting to a bandwidth of 4 kHz, the sampling frequency was reduced to 8 kHz, yielding a subsegment length of 5 ms, equivalent to 40 samples. Accordingly, a subsegment excitation frame size of 5 ms or 40 samples was used for our SB-MMPLPC codec. Again, the excitation pulse positions were encoded using the enumerative method [5]. A long term predictor (LTP) was also invoked, as it provided a noticeable increment in subjective and objective quality when the excitation pulse density was low, and even for high excitation densities, it retained the same performance as without the LTP both in terms of bit rate and SEGSNR. For each

	mode 1			mode 2			
	N_g	bit/pulse	position	N_g	bit/pulse	position	total bit
1	2	2	10	2	2	10	14
2	6	3	22	7	2	25	41
3	9	3	29	12	2	29	58
4	13	4	34	17	3	37	89
5	18	4	37	24	3	37	110
6	17	5	37	22	4	37	126
7	20	5	38	26	4	35	140
8	22	5	37	28	4	33	148
9	27	5	37	34	4	34	170
10	23	6	37	28	5	33	176
11	26	6	35	33	5	25	192
12	27	7	34	34	6	22	227
13	33	7	25	33	7	25	256
14	37	7	14	37	7	14	273
15	39	7	6	39	7	6	279
16	40	7	0	40	7	0	284

Table 1: Bit Allocation Scheme for bands B1 and B2 of the SB-MMPLPC Codec

subband, 4 bits were needed to linearly quantize the LTP filter gain, while 7 bits were used to encode the LTP delay.

When quantising the excitation pulse amplitudes g_i, we found that if the number of excitation pulses N_g was less than six, 3, 4, 5, or 6 bit quantization achieved almost the same quality with the segmental SNR differing by only 0.2 dB, while the number of excitation pulse quantisation bits doubled. If N_g was from six to ten, 4, 5, or 6 bit quantization had a similar effect, whereas if N_g was from eleven to sixteen, 5 or 6 bit quantization also got to within 0.2 dB and so on. So when we constructed Table 1 and the equivalent tables for SB3 and SB4, we used less precision to quantize the lower excitation density pulses, while higher precision was employed to quantize the higher excitation density pulses. The excitation pulse amplitudes were normalised by their maximum value within the subsegment and this maximum value was logarithmically quantised using eight bits for each subsegment and each subband before quantisation. The MMPLPC codec structure was identical for all four subbands. The total number of bits per 20 ms became 1707, which yielded a mono bit rate of about 86 kbits/s at a coding rate of 1707 bits/640 samples $\approx$ 2.67 bits/sample.

Having considered the audio codec used in our DAB transceiver we now focus our attention on transmission issues.

3 The DAB Transceiver

The schematic diagram of our audio broadcast transceiver is depicted in Figure 3, where the audio encoder's source bits are source-matched forward error correction (FEC) coded and 16-level quadrature amplitude modulated (16-QAM) on to a 1.5 GHz satellite carrier. Low-complexity non-coherent 4 bits/symbol 16-QAM is used [6] combined with a range of embedded block FEC codes [3] in order ensure high bandwidth efficiency.

The proposed 2.67 bits/sample SB-MMPLPC audio codec was subjected to bit sensitivity investigations by systematically corrupting all of its bits in a 1707 bit frame and evaluating the SEGSNR penalty inflicted. When for example the sensitivity of bit 1 was investigated, this bit was consistently

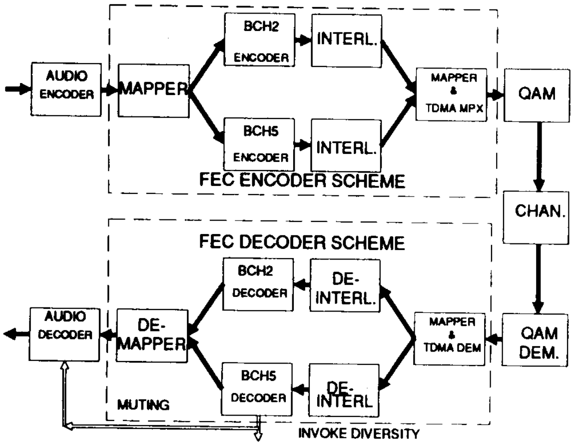

Figure 3: Broadcasting System Schematic

Parameter	$\sum$ SB1-SB4	SB1	SB2	SB3	SB4
Lar_i Bit-position	103	35 1-35	24 36-59	24 60-83	20 84-103
$C_j(k)$ Bit-position	13	4 104-107	4 108-111	4 112-115	1 116
$Mode_i$ Bit-position	4	1 117	1 118	1 119	1 120
LTP delay Bit-position	28	7 121-127	7 128-134	7 135-141	7 142-148
LTP gain Bit-position	16	4 149-152	4 153-156	4 157-160	4 161-164
$\max(g_i)$ Position of g_i g_i Redundant	A B C D				

Table 2: SB-MMPLPC bit allocation for the first subframe

corrupted in every frame, while keeping all other bits intact. The 1707-bit frame is constituted by 103 bits for the LPC parameters and 401 bits for every 5 ms subframe. The detailed bit allocation within a frame is shown in Table 2, where A, B, C, and D represent quantities having a variable number of quantisation bits in the subbands SB1-SB4 that add up to a fixed value of $A+B+C+D = 340$ bits. The overall shape of the SEGSNR degradation versus bit index sensitivity curves, which are omitted here due to lack of space, suggested that basically there are two sensitivity classes, the sensitive C1 and the more robust C2 categories, associated with more than 15 dB and less than 10 dB SEGSNR degradations, respectively. These findings were also confirmed by injecting random errors assuming a given fixed BER, as would be experienced over a Gaussian channel. From these experiments we confirmed that the BER of the more sensitive C1 bits must be below about 10^{-4}, while that of the more robust C2 bits below about 10^{-3}, in order to ensure acceptable audio quality, although even lower BERs are preferable.

Based on the above findings as regards to the bit sensitivities we experimented with a source-sensitivity matched twin-class and three different full-class FEC schemes. Long Reed-Solomon (RS) FEC block codes inherently possess good error randomising properties over bursty channels, but are complex to implement. In our deliberations we will represent this class of codes using the powerful half-rate RS(380,190,95) code over the finite Galois field GF(512), which encodes 190 nine-bit symbols into 380 symbols and can correct 95 symbol errors. This code represents the highest practically acceptable implementational complexity for our system and when encoding the 86 kbits/s audio information generated by the SB-MMPLPC encoder we have a bit rate of about 171 kbits/s. Using our 4 bits/symbol non-coherent 16-QAM modem this yields a signalling rate of 42.75 kBd that requires a bandwidth of about 64 kHz, when employing an excess bandwidth of 50 %.

A very good compromise in terms of implementational complexity and error correcting power is constituted by the family of binary BCH codes of 63 bits length. Often used members of this family are the BCH(63,30,6), BCH(63,36,5), BCH(63,39,4), BCH(63,45,3), BCH(63,51,2) and BCH(63,-57,1) codes, correcting 6, 5, 4, 3, 2 and 1 bits per frame, respectively. The BCH(63,30,6) code has a coding rate of $R = 30/63 \approx 0.5$ and will be used in our experiments as a low-complexity alternative to the complex but similar rate RS(380,190,95) code. The associated bit rate for our DAB transceiver becomes 179.55 kbits/s, yielding a 16-QAM signalling rate of 44.9 kBd and necessitating a bandwidth of 67 kHz.

A third FEC scheme that we will investigate is constituted by a twin-class embedded source-matched un-equal protection arrangement. Our proposed audio codec has a framelength of 20 ms, which is encoded using 1707 bits. Accordingly, these bits are mapped into two sensitivity classes, C1, C2 and matched BCH codes are assigned to them. The number of bits in the sensitivity classes C1 and C2 are 684 and 1020, respectively, yielding a total of 1704 bits, while three bits are unprotected. The FEC codes assigned are the five-error-correcting BCH5 = BCH(63,36,5) and double-error-correcting BCH2 = BCH(63,51,2) codes, respectively. In C1 there are nineteen, while in C2 twenty BCH codewords, respectively, and the total number of FEC-coded bits is 2457 + 3 = 2460.

In the audio transmission frames bits 1-103 represent the LAR parameters, while each of the four subsegments is encoded by 401 bits, yielding a total of $103 + (4 \cdot 401) = 1707$ bits. The LAR bits 1-103 are assigned to C1, along with the most important 145 bits of each subsegment, yielding a total of $103 + (4 \cdot 145) = 684$ C1 bits. The remaining $4 \cdot 255 = 1020$ C2 bits are assigned to the weaker BCH2 codec. The overall coding rate becomes $R = 1707/2460 \approx 0.69$, while the transmission rate is 123 kbits/s, giving a Baud rate of about 30.7 kBd. The required bandwidth is $1.5 \cdot 30.7 = 46$ kHz.

The performance of this scheme will be gauged against that of a similar rate, but less complex single-class code, namely the BCH(63,45,3) code, for which $R = 45/63 = 0.71$, the total number of bits per frame is 2394, the bit rate becomes 119.7 kbits/s and the signalling rate is ≈ 30 kBd. The bandwidth requirement of this scheme is about 45 kHz.

The 'FEC encoder/decoder scheme' drawn in dashed lines in Figure 3 represents the four previously mentioned FEC schemes. If the RS(380,190,95) scheme is used, no mapping and interleaving is needed, since the entire audio frame is encoded by a single $9 \cdot 190 = 1710$-bit codeword, requiring only 3 padding bits. In case of the BCH(63,30,6) code $1710/30 = 57$ codewords constitute an audio frame, allowing for an interleaving depth of 57 words, but again, no source mapping is required. When using the BCH(63,45,3) code, $1710/45 = 38$ codewords encode an audio frame, hence the interleaving depth is 38.

Lastly, when the twin-class source-matched FEC scheme is used, the audio encoder's bits are sorted by the 'Mapper' into two sensitivity classes according to their vulnerability, as described earlier in this Section. Twin-class unequal error protection is deployed using the previously proposed 63-bit binary two- and five-error correcting BCH codes denoted by BCH2 and BCH5, respectively. The FEC coded bits are rectangularly interleaved over the current 20 ms audio frame, namely over 20 and 19 BCH5 and BCH2 codewords, respectively. The interleaved bits are mapped back in one stream by another bit 'Mapper', multiplexed by the Time Division Multiple Access (TDMA) multiplexer (MPX) with three additional mono or another stereo audio program, as it will be explained in the next paragraph, 16-StQAM modulated onto a 1.5 GHz carrier and transmitted via the Rayleigh fading broadcast channel.

The half-rate FEC-coded schemes have a stereo bandwidth requirement of about $2 \cdot 67 = 134$ khz. Single channel per carrier (SCPC) transmissions with a stereo bandwidth of about 134 kHz are convenient in terms of low transmission bitrate, which is well below the coherence bandwidth of the typical mobile broadcast channel, hence ensures that the channel is essentially a narrowband flat fading transmission medium. Hence no 'power-hungry' channel equaliser is required, which is crucial in order to maintain low receiver complexity, low battery drain and hence ultimately light-weight construction. However, at low signalling rates the time between two adjacent signalling symbols is long and hence the channel's fading envelope changes dramatically between two adjacent transmitted samples and due to this the 16-StQAM modem's BER performance suffers from its lack of ability to efficiently trace the received signal's amplitude and phase trajectory. This is particularly true, if the mobile receiver's speed is high.

Although the proposed second order switched diversity receiver efficiently mitigates this problem, it is advantageous in terms of improving the 16-StQAM BER performance to transmit several channels per carrier in a TDMA structure. This increases the transmission signalling rate and hence improves the receiver's BER performance for slowly walking pedestrians. However, the transmission bit rate must not exceed the typical propagation channel's coherence bandwidth in order to avoid using equalisers. A TDMA structure based on two stereo channels constitues a good compromise in this respect.

A further advantage of the TDMA structure is that the receiver can monitor the reception quality of other transmitters during the unused time slot and allow seamless switching between two transmitters broadcasting the same program. The twin-channel stereo TDMA burst rate of the approximately $R = 2/3$-rate schemes becomes about $4 \cdot 30$ kBd=120 kBd, which requires a bandwidth of $1.5 \cdot 120$=180 kHz. This rate allows us to fit two stereo digital audio channels into one conventional analogue frequency modulated (FM) channel while ensuring that the transmission bandwidth is narrow with respect to the fading channel's coherence bandwidth. The $R \approx 0.5$-rate schemes have an approximate signalling rate of 45 kBd, yielding a four-channel TDMA rate of about 180 kBd, which can accommodate a twin-channel stereo scheme slotted into a 270 kHz channel.

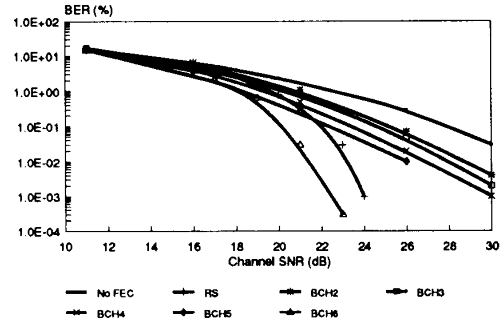

Figure 4: BER versus channel SNR performance of the proposed 16-StQAM/TDMA transceiver at 30 mph with diversity using various FEC codes

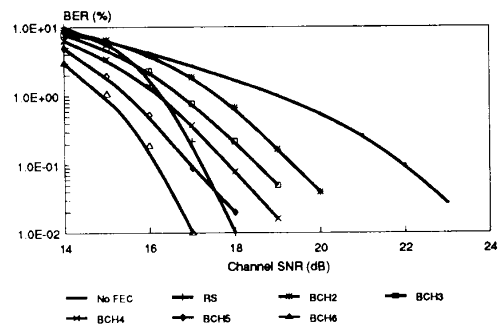

Figure 5: BER versus channel SNR performance of the proposed 16-StQAM/TDMA transceiver via AWGN channels using various FEC codes

4 Results and Discussion

In our experiments linear amplification was assumed, no AGC and carrier recovery were invoked, and the excess-bandwidth was 50 %. Our results are provided for the best and worst case Gaussian and Rayleigh channels, respectively. In Figure 4 we portray the BER performance of the proposed 16-StQAM modem for a signalling rate of 180 kBd assuming four-channel mono TDMA transmissions with diversity (D)

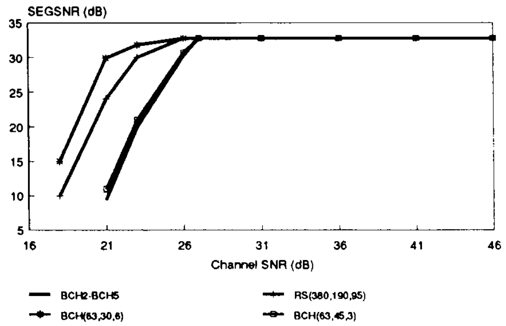

Figure 6: SEGSNR versus channel SNR performance of the proposed 16-StQAM/TDMA transceiver at 30 mph with diversity using various FEC codes

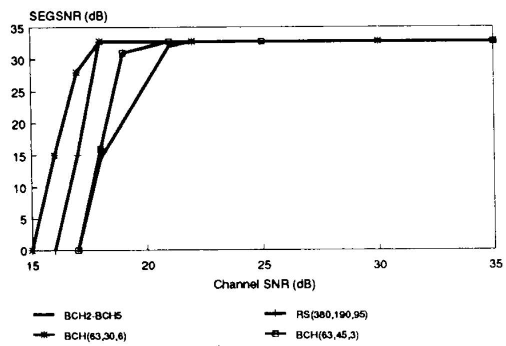

Figure 7: SEGSNR versus channel SNR performance of the proposed 16-StQAM/TDMA transceiver via AWGN channels using various FEC codes

at a mobile speed of 30 mph, and propagation frequency of 1.5 GHz using the set of BCH2-BCH6 codes and the RS(380,190,95) code. Note the gradual BER improvements, when using stronger codes. It is also interesting to observe that the most complex RS codec is outperformed by the lower complexity BCH5 and BCH6 codecs, if the SNR is above about 22-24 dB. Similar tendencies are also true for the AWGN channel, depicted in Figure 5, which is characteristic of the best stationary scenario.

The overall objective SEGSNR versus channel SNR performance of our proposed broadcast scheme is portrayed with diversity at a mobile speed of 30 mph, as well as for the stationary AWGN scenario in Figures 6 and 7, respectively. As anticipated from the BER versus channel SNR curves seen in Figures 4 and 5, best SEGSNR performance is guaranteed by the single-class protected half-rate type systems, namely the BCH6 and the RS coded arrangements. Again, the modest-complexity BCH(63,30,6)-coded scheme outperforms the more complex RS(380,190,95) arrangement. Both of these diversity-assisted higher bit rate schemes require an SNR of about 24 dB at 30 mph for unimpaired audio quality, while this 'corner' SNR value is about 17 dB for the stationary AWGN channel.

In case of the less robust and lower bit rate, lower complexity schemes only about one dB excess channel SNR was required in order to ensure unimpaired audio reception over the 30 mph Rayleigh-fading channel, when compared to the half-rate type systems. For transmissions over the more benign stationary AWGN channel the minimum required SNR was about 20 dB, some 2 dB higher than for the half-rate coded arrangements. The source-matched twin-class BCH2-BCH5 coding scheme failed to outperform the less complex single-class BCH3 codec due to the inherent error sensitivity of the SB-MMPLPC codec, because in case of bad channel conditions the weaker BCH2 code was more frequently overloaded than the BCH3 code. This resulted in a SEGSNR penalty, in spite of the more robust protection of the C1 bits. Based on these experiences in our final proposed DAB scheme we favour the lowest complexity BCH3 codec.

5 Summary and Conclusions

When protecting the proposed 2.67 bit/sample SB-MMPLPC audio codec by the favoured BCH(63,45,3) code and using non-coherent differential 16-QAM, a mono bandwidth of about 45 kHz and a channel SNR in excess of about 25 dB are required for both pedestrian as well as typical urban vehicular speeds in order to maintain Hi-Fi DAB quality. Our current work is targeted at improving the audio quality, complexity, bit rate, bandwidth occupancy and error resilience trade-off using a pilot symbol assisted, block-coded coherent square 16-QAM modem [6].

References

[1] F.Mueller-Roemer: Directions in Audio Broadcasting, J. Audio Eng. Soc., Vol. 41, No. 3, March, 1993, pp 158-173

[2] Proceedings of the 1st Intern. Symp. on DAB, June, 1992, Montreux, Switzerland

[3] R. Steele (Ed.): Mobile Radio Communications, Pentech Publishers, 1992.

[4] X. Lin, L. Hanzo, R. Steele, W.T. Webb: Subband-multipulse digital audio broadcasting for mobile receivers, IEEE Tr. on Broadcasting, Dec. 1993

[5] P. Kroon and E. Deprettere, "A class of analysis-by-synthesis predictive coders for high quality speech coding at rates between 4.8 and 16 kbits/s," *IEEE J. on selected areas in Commu.*, vol. 6, No.2, pp. 353–363, Feb. 1988.

[6] W.T. Webb, L. Hanzo: Quadrature Amplitude Modulation, Research Monograph, to appear Pentech Publishers, London, U.K.