ON BIT REDUCTION OF DIGITAL AUDIO SYSTEMS

1549 (H-1)

Toshi T. Doi
Sony Audio Technology Center
Tokyo, Japan

Presented at the 64th Convention November 2 through 5, 1979 New York City

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. For this reason, there may be changes should this paper be published in the Journal of the Audio Engineering Society. Additional preprints may be obtained by sending request and remittance to the office of Special Publications, Audio Engineering Society, 60 East 42nd Street, New York, New York 10017, USA.

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

AN AUDIO ENGINEERING SOCIETY PREPRINT

"ON BIT REDUCTION OF DIGITAL AUDIO SYSTEMS"

by Toshi T. DOI

Sony Audio Technology Center,

Tokyo, Japan

ABSTRACT

Various methods of bit reduction for digital audio systems are discussed in this paper. The conditions for high quality digital audio systems are;

(1) Any degradation in quality should not be permitted.

(2) The reliability against code errors should be extremely strong.

Because of the reason (1), a method of variable length word is recommended.

According to the reason (2), Huffman-code, Shannon-code, or Fano-code are very difficult to adopt, and a new code named "Segment Code" is proposed, which can provide the strong reliability by the sacrifice of efficiency.

Modulation methods are also discussed with regard to increase packing density.

1. INTRODUCTION

In this paper, a possibility of bit reduction for digital audio systems is discussed. First of all, digitized data of orchestra music is analyzed, and Entropy without correlation is calculated. Then various variable length codes are compared from the view point of the efficiency and the stability of the synchronization against code errors.

Let me take our speech communication as an example of redundancy and bit reduction. It is well known that the bit reduction more than ten times can be easily achieved by LPC (Linear Prediction Coding). In other words, 90% of our speech is considered as redundancy, and it is why we can communicate even in the noisy circumstances (like in the car of subway) with terrible signal to noise ratio.

In our speech communication, we utilize this natural redundancy of 90% for error correction in our brains.

On the other hand, in digital audio systems, we use a huge number of bits like 16×50,400 bit/sec for quantization of music, the redundancy of which is considered to be pretty large (a signal without any redundancy is called as 'white noise'). We can not use this natural redundancy for error correction, and another plenty of bits should be added for error detection and error correction of digital audio recorders or disc systems. Supposing the natural redundancy of a music is 90%, and another artifitial redundancy of 50% (counting from the total bits) is necessary for error correction, then we have to record 100 bit of signal in order to keep 5 bit of real information of music.

The ideal system is, ofcause, to achieve maximum bit reduction to the original music information, and then add minimum redundancy for error correction.

The methods for bit reduction are classified as follows.

(1) Parameter transformation
$\{$LPC, PARCOR, Vocoder, etc$\}$

(2) Differential coding
$\left\{\begin{matrix}\Delta\text{M} \\ \text{DPCM}\end{matrix}\right\}$ $\left\{\begin{matrix}\text{adaptive} \\ \text{non-adaptive}\end{matrix}\right\}$ $\left\{\begin{matrix}\text{prediction} \\ \text{non-prediction}\end{matrix}\right\}$

(3) Orthogonal transformation
$\{$Adamar, Fourier Vocoder, etc$\}$

(4) Non-linear (Non-uniform) quantization

(5) Variable length word coding

(6) Modulation method
$\{$Run Length Limited Code$\}$

For high quality recording of musics, (1) can not be applied. (2), (3) are dangerous because the informations which rarely happens are neglected. That means a sharp edge of percussion might be changed to be duller, and it might kill the merit of digital audio. By using the method (4), the instantaneous signal to noise ratio becomes worse, and some of the professional people do not like that quality.

In this paper, therefore, the method (5) is mainly described, which has the possibility of bit reduction without any degradation in the quality.

The method (6) is to achieve higher packing density by the sacrifice of jitter margin, and the rough idears are described in this paper.

2. ANALYSIS OF MUSIC

Fig.1~5 show the examples of level distribution of music (Beethoven No.9th Synphony, 4th Movement)*. Number of word is shown from 640 to 40000 word, and the distribution is shown to be smoother as the number of the words increases.

Three curves show the following coding.

(1) PCM (16-bit linear quantization)

(2) DPCM (difference of PCM)

(3) D^2PCM (difference of DPCM)

Any prediction technique is not applied in DPCM and D^2PCM, and they are the simple difference of adjacent words.

The distribution is narrower in DPCM or in D^2PCM comparing to the linear PCM, which is shown in terms of "Entropy [1]" in Fig.6. Entropy H is calculated by the following formula without considering correlation.

$$H = -\sum_{i=32768}^{32767} P_i \log_2 P_i \qquad (1)$$

Where, P_i is the probability of ith level. The Entropy of linear PCM is approximately 10.5 bits, and those of DPCM or D^2PCM are between 7.5 bits and 8 bits.

It should be noted that the Entropy calculated within 8000 bits is not accurate because the signal is not considered as "Ergodic [1]", which can be roughly guessed from level distribution of Fig.1~5.

*Recorded by Sony PCM-1600 digital audio recorder

3. VARIABLE LENGTH CODE

3-1. HUFFMAN CODE [1]

Optimum variable length code is studied by Shannon, Huffman, and Fano. [1] The mean word length of each code can be very close to Entropy.

Table.1 shows an example of Huffman Code with levels of eight.

Table.1 An Example of Huffman Code

No. of Levels	Probability	Code	Word Length
i	0.4	1	1 bits
ii	0.2	000	3
iii	0.15	001	3
iv	0.1	011	3
v	0.07	0101	4
vi	0.04	01000	5
vii	0.03	010010	6
viii	0.01	010011	6

Mean Word Length = 2.47 bits
Entropy H = 2.4064 bits

The mean word length of this example is 2.47 bits, which is effectively close to the Entropy H=2.406 bits.

In this example, the bit reduction of 0.53 bits is achieved comparing to the linear PCM of three bits. Word synchronization of this code is not necessary, because each word is decoded uniquely by looking up a table. For example, if the following bit stream is supplied to the decoder, each word can be decoded one by one.

$$\underbrace{0\ 1\ 0\ 0\ 1\ 1}_{\text{viii}}\ \underbrace{0\ 1\ 0\ 0\ 1\ 0}_{\text{vii}}\ \underbrace{0\ 1\ 0\ 0\ 0}_{\text{vi}}\ \underbrace{0\ 1\ 0\ 1}_{\text{v}}\ \underbrace{0\ 1\ 1}_{\text{iv}}\ \underbrace{0\ 0\ 1}_{\text{iii}}\ \underbrace{0\ 0\ 0}_{\text{ii}}\ \underbrace{1}_{\text{i}}$$

On the other hand, if one bit of this stream is eroneous, several words will be completely misunderstood.

CODE ERROR OF 1 BIT

$$\underbrace{0\ 1\ 0\ 0\ 1\ 1}_{\text{viii}}\ \underbrace{0\ \overset{0}{\ast}\ 0}_{\text{ii}}\ \underbrace{0\ 1\ 0\ 0\ 1\ 0}_{\text{vii}}\ \underbrace{0\ 0\ 0}_{\text{ii}}\ \underbrace{1}_{\text{i}}\ \underbrace{0\ 1\ 0\ 1}_{\text{v}}\ \underbrace{1}_{\text{i}}\ \underbrace{0\ 0\ 1}_{\text{iii}}\ \underbrace{0\ 0\ 0}_{\text{ii}}\ \underbrace{1}_{\text{i}}$$

MISS DECODED AREA

This phenomena is known as error propagation.

In digital audio systems, any error after correction and concealment can not be permitted, and therefore the code format of fixed length frame is usually adopted.

Huffman code has extremely high efficiency, but fixed frame structure is very difficult to adopt, and error correction schemes are hard to be applied.

3-2. SEGMENT CODE

Segment code is an idea of variable length code which is possible to form a fixed length frame structure. Each segment consists of fixed number of bits, and each word is variable length of segment. Supposing one segment be four bits, one word can be 4 bits, 8 bits, 12 bits, 16 bits, or 20 bits. In order to show the number of segment belonging to one word, the first bit of each segment expresses conjunction; for instance, "0" means the first segment of the word, and "1" expresses the segment connecting the the preceding segment.

Table.2 Segment Code

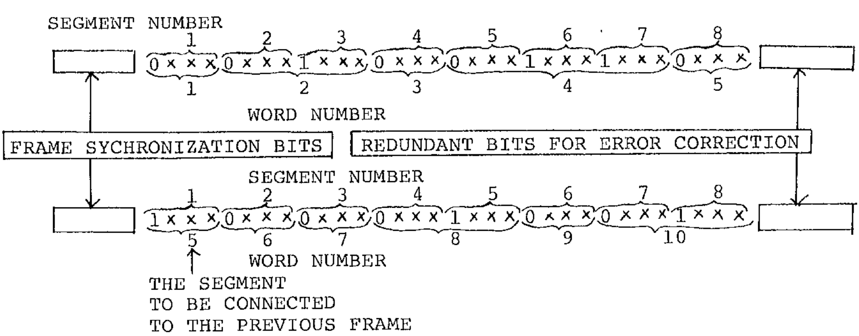
1 segment word : $0\underbrace{x\ x\ x}$
imformation bits (3 bits)

2 " " : $0\underbrace{x\ x\ x}\,1\underbrace{x\ x\ x}$
information bits (6 ")

3 " " : $0\underbrace{x\ x\ x}\,1\underbrace{x\ x\ x}\,1\underbrace{x\ x\ x}$ conjunction bit
information bits (9 ")

In this case a frame of fixed number of segments but of variable number of words is easily formed, as is shown in Table 3.

Table.3 A Frame with Fixed Number of Segments (8) but Variable Number of Words



In this example, each frame consists of eight esgments, but the number of words is variable, and sometimes a word is separated into different frames as is shown in the word #5 in Table.3.

The efficiency of this code is not very high because a bit for conjunction is necessary for each segment, but the great advantage is that the system design like bit synchronization, frame synchronization, and error correcting schemes can be just the same as fixed length-word coding. The digital audio system with this segment code is nothing diggerent from ordinal system with fixed length word except an appropriate buffer memory.

In this code, one bit in each segment can not be used for information, and the effective bit length is expressed as follows.

$$b_1 = \log_2 \sum_{i=1}^{\ell} 2^{i(s-1)} \qquad (2)$$

Where s is segment length (bits), and ℓ is maximum word length (segments).

While, if we just omit "O's" of linear coding, the transformation of variable lingth word becomes very simple. In this case, the effective bit length is;

$$b_2 = \ell(s-1) \qquad (3)$$

The code corresponding to eq.(3) is named as "simplified segment code". Table.4 shows an example of the segment code and the simplified segment code.

Table.4 Segmet Code (s=3, ℓ=3)

No	LINEAR CODE	SEGMENT CODE	NUMBER OF SEGMENT	SIMPLIFIED SEGMENT CODE
0	0000000	0 0 0		0 0 0
1	0000001	0 0 1		0 0 1
2	0000010	0 1 0	1	0 1 0
3	0000011	0 1 1		0 1 1
4	0000100	0 0 0 1 0 0		0 0 1 1 0 0
				0 1
15	0001111	0 1 0 1 1 1		0 1 1 1 1 1
16	0010000	0 1 1 1 0 0	2	0 0 1 1 0 0 1 0 0
				0 1 1
19	0010011	0 1 1 1 1 1		0 0 1 1 0 0 1 1 1
20	0010100	0 0 0 1 0 0 1 0 0		0 0 1 1 0 1 1 0 0
				0 1 1
63	0111111	0 1 0 1 1 0 1 1 1	3	0 1 1 1 1 1 1 1 1
83	1010011	0 1 1 1 1 1 1 1 1		

Fig.7 shows mean word length of simplified segment code and Huffman code comparing to Entropy of the signal, where the level distribution is assumed to be Gaussian.

The mean word length of Huffman code is approximately equal to Entropy, excluding a small error at the lower variance. While those of segment code are worse in 1~3 bits, this is the sacrifice of the efficiency in order to get the stability against code errors which is essential in digital audio systems.

4. MODULATION

The packing density can be increased by adopting run length limited code, which is not bit reduction but the results might be the same (the longer playing time or smaller recording media).

3PM[3] is well known and is adopted to stationary head digital audio recorders[4] and digital audio disc systems.[5] The principle is to reduce the frequency of the inverse of the signal by the sacrifice of jitter margin.[5]

Table.5 shows a comparison between several run length limited code.

Table.5 RUN LENGTH LIMITED CODE

	MFM	3PM	FTM
MINIMUM TRANSITION	T	1.5T	2T
MAXIMUM TRANSITION	2T	6T	$8^2/_3$T
JITTER MARGIN	$^1/_2$T	$^1/_2$T	$^1/_3$T
JITTER MARGIN / MINIMUM TRANSITION	50%	33%	17%
JITTER MARGIN / MAXIMUM TRANSITION	25%	8.3%	3.8%

Fig.8 shows the relationship between jitter margin and packing density, which indicate that by improving tape support mechanisms or disk rotation, there is a possibility to increase packing density in fairly large amount.

5. CONCLUSION

A bit reduction using variable length word method is discussed is this paper. A reduction of 30~40% of recording bits is expected by the combination of DPCM and simplified segment code with keeping the excellent reliability against code errors.

Correlation of word values is not included in this initial study, but it is evident that further more reduction is possible by using prediction method.

Modulation method is another point to achieve higher packing density, which can be increased 50%~100% by using appropriate run length limited code and with sacrificing jitter margin.

The details of newly developed modulation method named FTM will be reported at the next chance.

REFERENCE

[1] for instance; R.Ash: "Information Theory" John Wiley & Sons (1965)

[2] Toshi T. DOI: "A System Theory of Real-Time Unequal-Length Transmission (Recording)", ASJ Conf. 3-P-3 (1978, Oct.)

[3] G.V.Jacoby: "A New Look Ahead Code for Increased Data Density" IEEE Trans., Vol.MAG-13, No.5 (1977 Sep.)

[4] Y.Tsuchiya et.al.: "A 24-Channel Stationary-Head Digital Audio Recorder" AES 61st Conv. No.1412 (F-2) (1978 Nov.)

[5] Toshi T. DOI, et.al.: "A Long Play Digital Audio Disc System", AES 62nd Conv. No.1442 (G-4) (1979 Mar.)

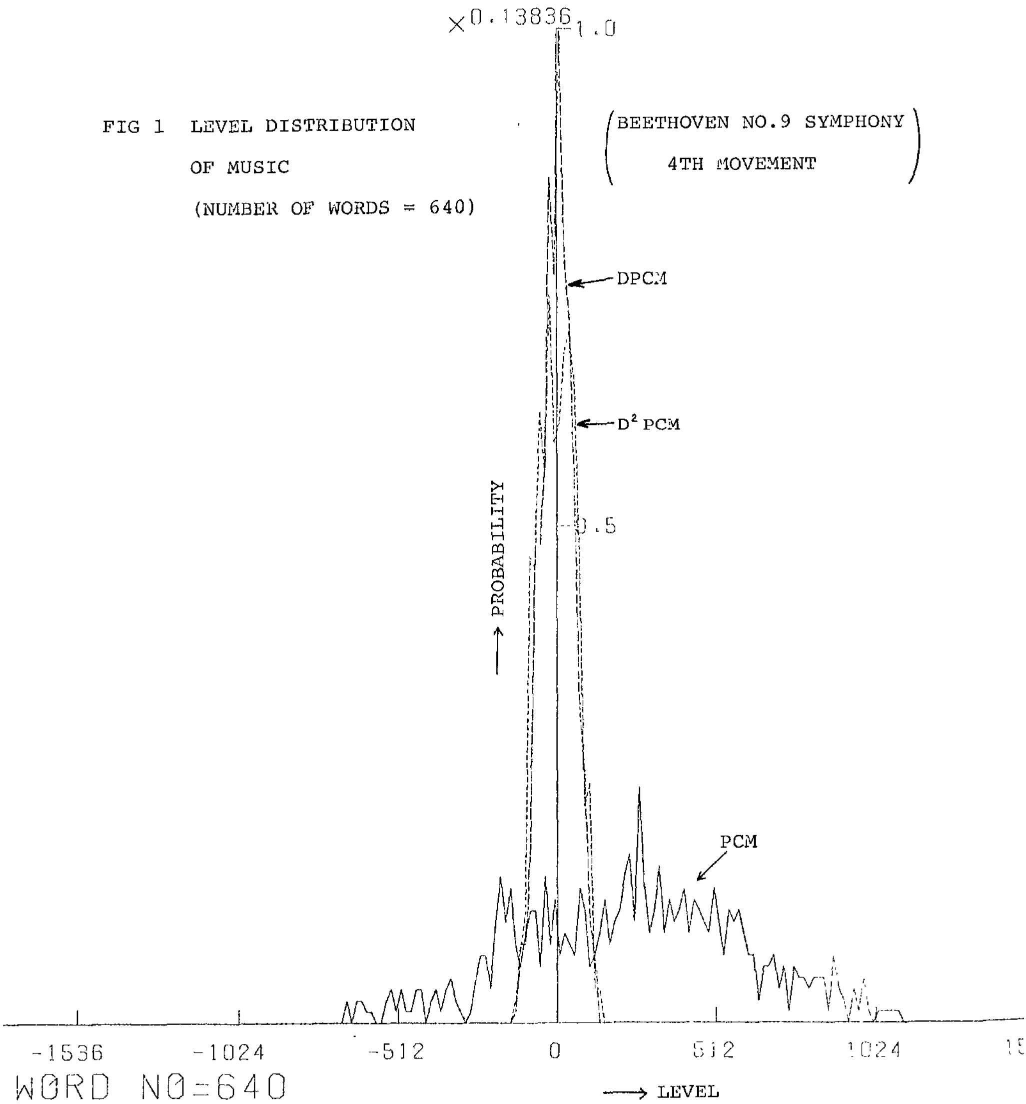


FIG 1 LEVEL DISTRIBUTION
OF MUSIC
(NUMBER OF WORDS = 640)
BEETHOVEN NO.9 SYMPHONY
4TH MOVEMENT
×0.13836
1.0
DPCM
D² PCM
0.5
PROBABILITY
PCM
-1536
-1024
-512
0
512
1024
WORD NO=640
LEVEL

FIG 2 LEVEL DISTRIBUTION OF MUSIC (NUMBER OF WORDS = 1600)

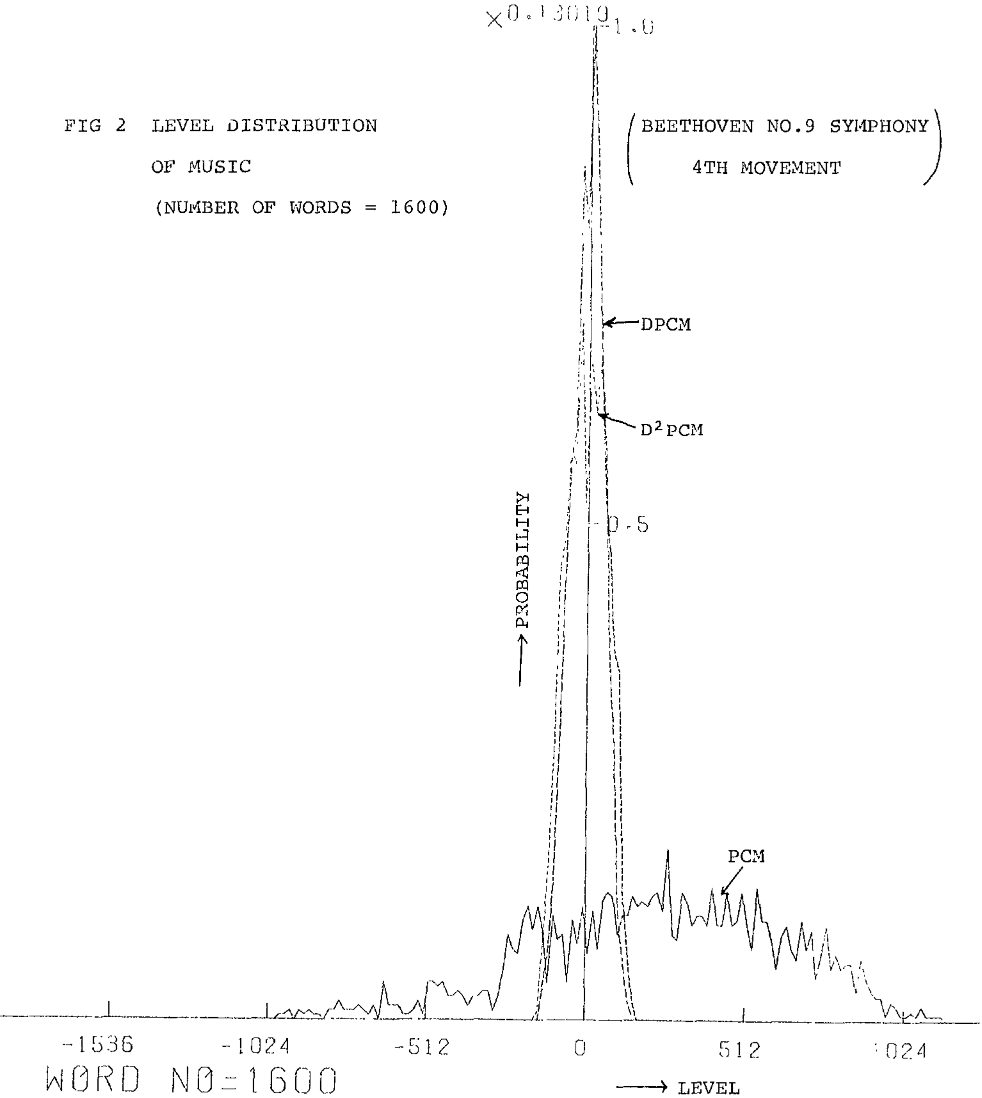


FIG 3 LEVEL DISTRIBUTION OF MUSIC

(NUMBER OF WORDS = 8000)

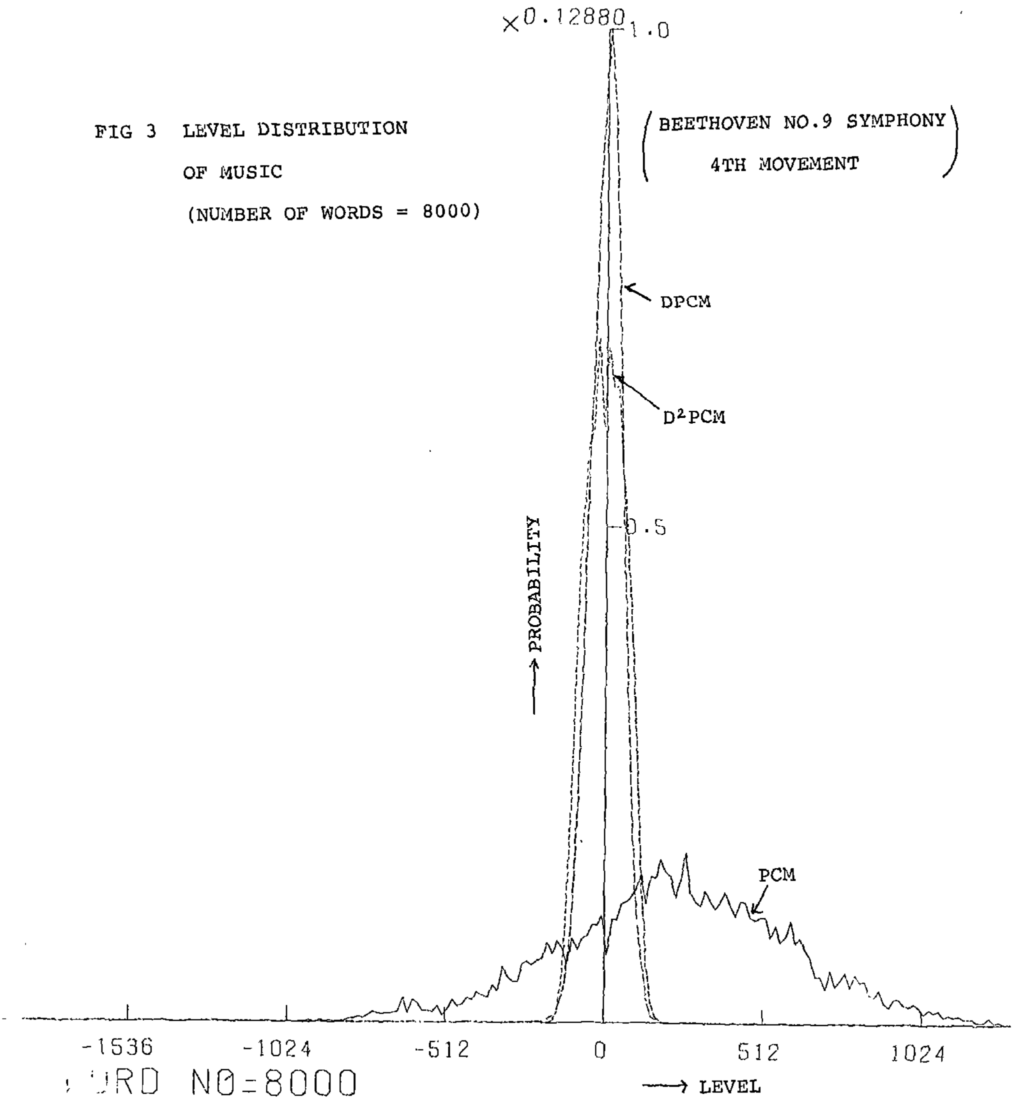
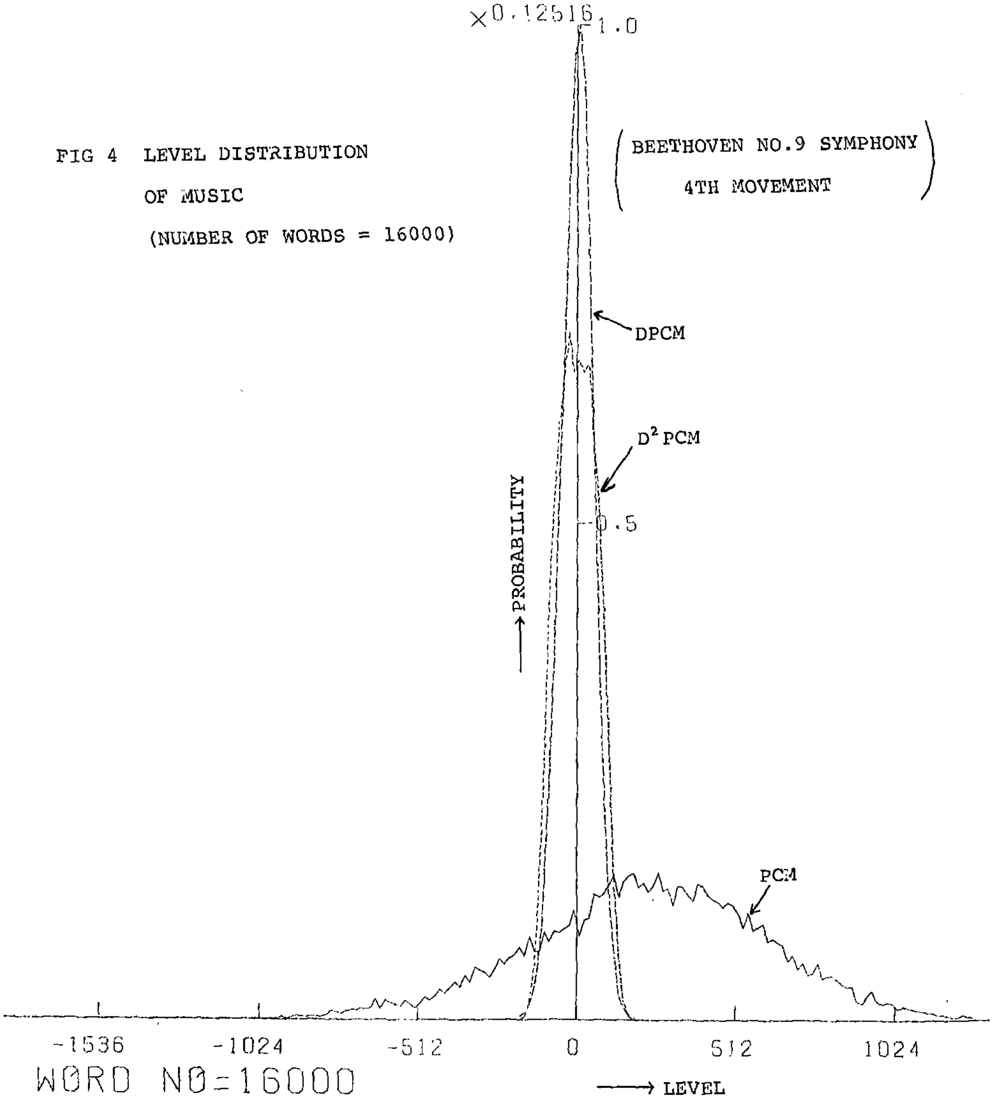


FIG 4 LEVEL DISTRIBUTION
OF MUSIC
(NUMBER OF WORDS = 16000)



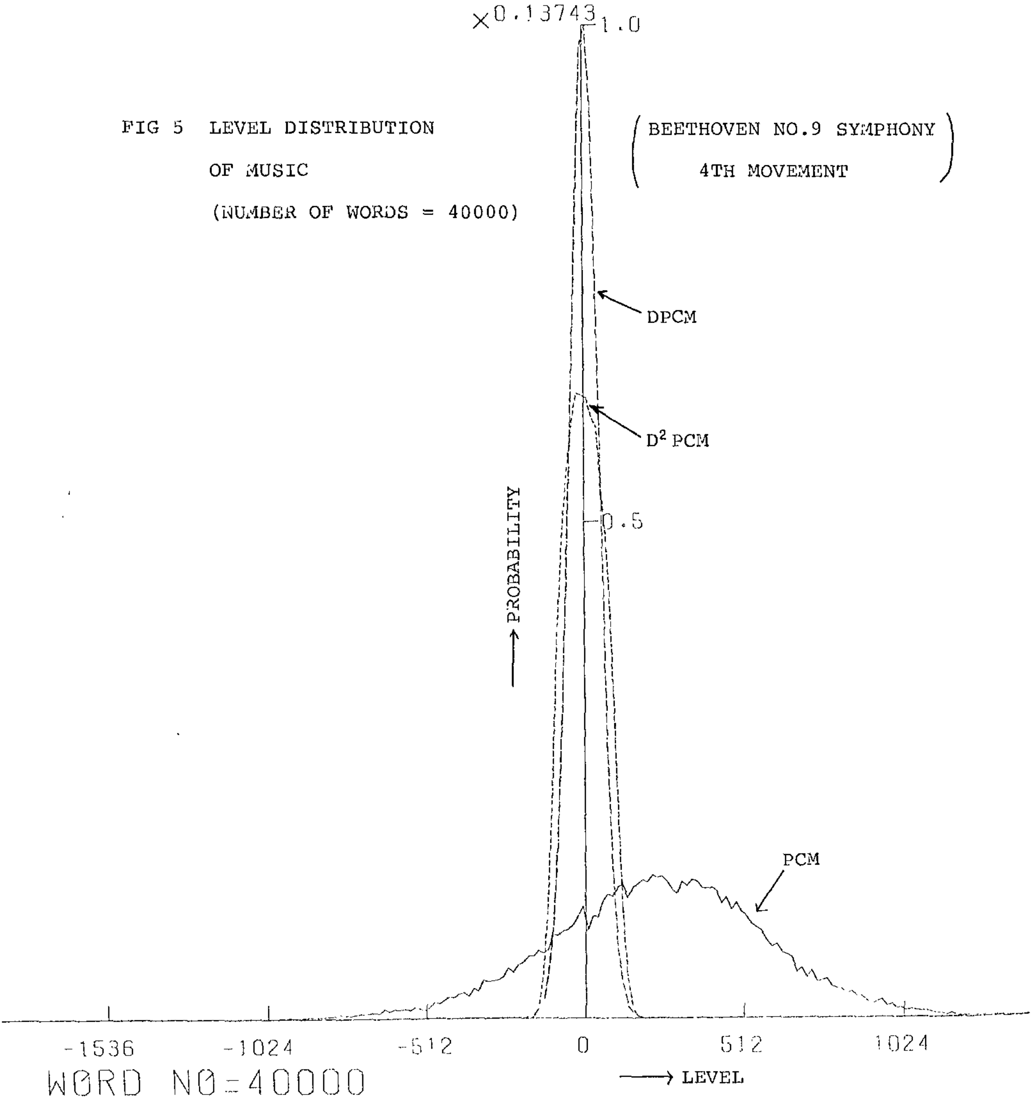


FIG 5 LEVEL DISTRIBUTION OF MUSIC

(NUMBER OF WORDS = 40000)

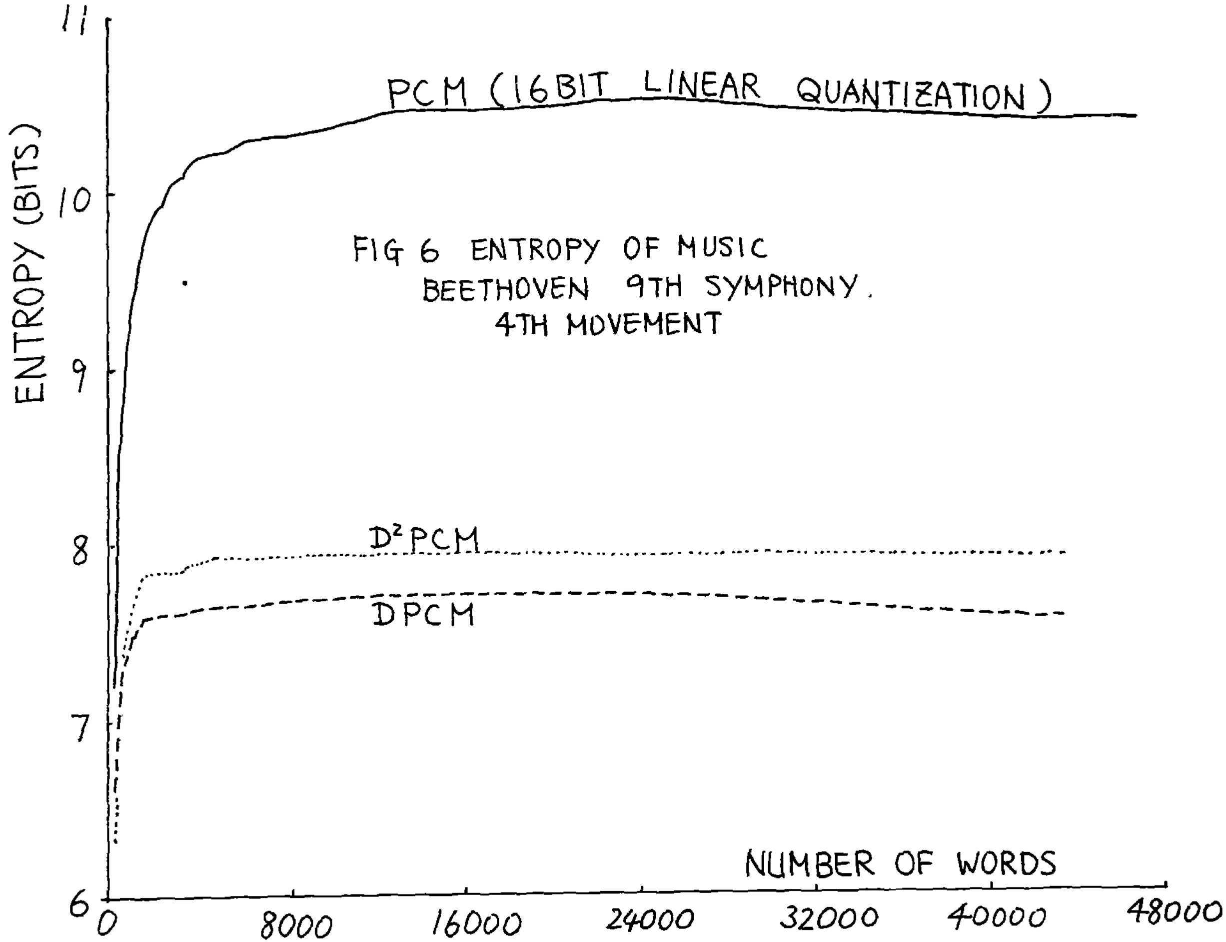

PCM (16BIT LINEAR QUANTIZATION)
FIG 6 ENTROPY OF MUSIC
BEETHOVEN 9TH SYMPHONY.
4TH MOVEMENT
D^2PCM
DPCM
ENTROPY (BITS)
NUMBER OF WORDS
11
10
9
8
7
6
0
8000
16000
24000
32000
40000
48000

FIG 7. MEAN LENGTH OF HUFFMAN CODE AND SEGMENT CODE

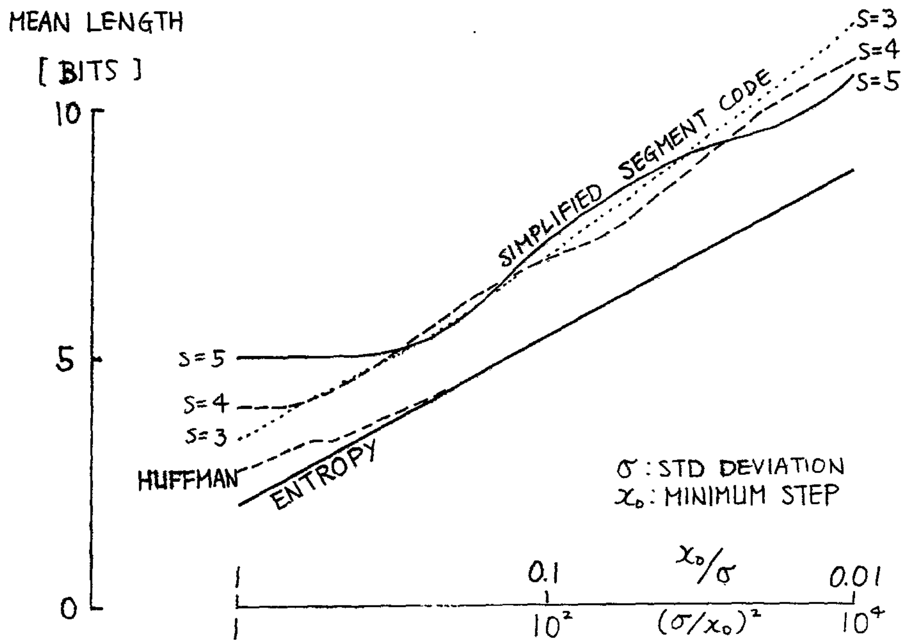


FIG 8 JITTER MARGIN vs PACKING DENSITY