



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
15.03.2000 Bulletin 2000/11

(51) Int Cl.7: **G10L 19/14**

(21) Application number: **99202927.2**

(22) Date of filing: **08.09.1999**

(84) Designated Contracting States:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE**
Designated Extension States:
AL LT LV MK RO SI

(72) Inventors:
• **Duenas, David**
Bournemouth, Dorset BH4 9JN (GB)
• **Bennett, Jeremy**
Eastleigh, Hampshire SO53 4QA (GB)

(30) Priority: **11.09.1998 GB 9819920**

(74) Representative: **Anderson, Angela Mary**
NDS Limited,
Gamma House,
Enterprise Road
Chilworth, Hampshire SO16 7NS (GB)

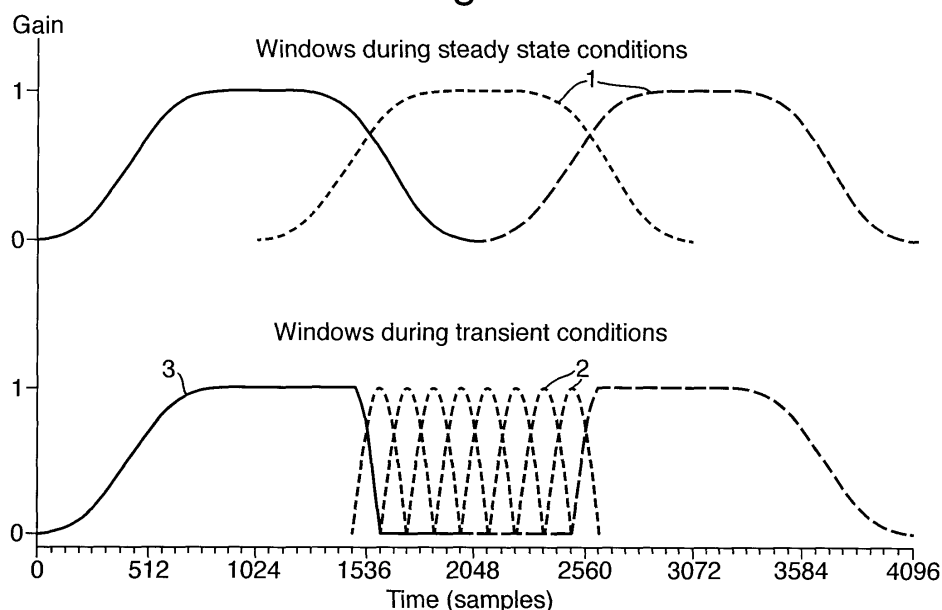
(71) Applicant: **NDS LIMITED**
West Drayton, Middlesex UB7 0DQ (GB)

(54) **Audio encoding system**

(57) A method of switching between long and short block encoding of frames of audio data with improved noise reduction in relation to the number of bits required for encoding. The method comprises the steps of calculating the difference in the long block perceptual entropy of a present frame and the preceding frame and switching to short block encoding of the present frame if the difference is greater than a first threshold. The method

may further comprising the steps of comparing the long block perceptual entropy of the succeeding frame with a second threshold if the difference is greater than the first threshold, and switching to short block encoding of the succeeding frame if the long block perceptual entropy of the succeeding frame is greater than the second threshold. The method may additionally or alternatively involve the study of the perceptual entropy of short blocks.

Fig.1.



Description

[0001] This invention relates to a method of encoding audio data and, more particularly, to a method of switching between long and short block encoding.

[0002] It is well known in the field of audio encoding to switch between long and short blocks of samples of audio data. A psycho-acoustic model can be used to determine the amount of quantisation required for the samples of each short or long block to keep noise to a suitably low level whilst using as few bits as possible for encoding. This information can then be used to switch between long and short block encoding such that as few bits as possible are used for encoding and a high level of data compression may be achieved. Generally, the use of a psycho-acoustic model will result in switching being carried out according to characteristics of the audio data signal. For example, if the audio signal contains sharp increases in energy known as attacks, short block encoding is often preferable as this has better time resolution and reduces the number of bits required for encoding. When the audio signal is stationary, or varies only slowly, long block encoding is preferable as fewer bits are then required than for short block encoding. In other words, switching is generally dependent on determining whether or not there are any attacks within a given block.

[0003] It is known to use a given psycho-acoustic model in combination with perceptual entropy (*Estimation Of Perceptual Entropy Using Noise Masking Criteria*, James D Johnston, IEEE 1998, pages 2524 to 2527) to determine the number of bits required to encode a block. Perceptual entropy is a simply measure of the number of bits required to encode the spectrum of a signal to a resolution sufficient to mask audio noise.

[0004] It has been suggested in the MPEG-2 Advanced Audio Encoding Standard, for example, to measure the perceptual entropy of long blocks of samples of an audio signal in successive frames and to switch to short block encoding for a frame when the perceptual entropy of the long block for that frame is above a certain threshold. There is, however, a memory effect in the calculation of perceptual entropy, introduced in order to ensure smooth encoding between successive frames. This means that succeeding frames may have long block perceptual entropy above the given threshold despite an attack only actually occurring in an earlier frame, and short block encoding may therefore be used unnecessarily when long block encoding would be preferable.

[0005] Objects of the present invention are to overcome such problems, to provide more accurate switching between long and short block encoding, to reduce noise and to reduce the overall bit rate of the encoded signal.

[0006] According to the present invention there is provided a method of switching between long and short block encoding of frames of audio data, the method

comprising the steps of:-

calculating short block perceptual entropy of short blocks of a present frame and the sum of the resulting short block perceptual entropy values;
calculating the long block perceptual entropy of the present frame and applying a predetermined weighting to the long block perceptual entropy value; and
switching to the use of short block encoding of the present frame if the sum is less than the weighted long block perceptual entropy value.

[0007] Thus, short block encoding may be used when there is a sufficient change in the perceptual entropy of the long blocks of succeeding frames. This is advantageous as the human ear is particularly sensitive to pre-echo, which is noise that occurs just before the energy of an acoustic signal and the perceptual entropy of the relevant blocks, undergo a large change, and less sensitive to noise during periods in which the acoustic signal is constant or has a similar pattern and perceptual entropy of succeeding blocks is similar.

[0008] Preferably the method further comprises the steps of:

comparing the long block perceptual entropy of the succeeding frame with a second threshold if the difference is greater than the first threshold; and
switching to short block encoding of the succeeding frame if the long block perceptual entropy of the succeeding frame is greater than the second threshold.

[0009] This enables short block encoding to be used for two succeeding frames both having high perceptual entropy despite the fact that the difference in the perceptual entropy of the frames is small.

[0010] Also according to the present invention there is provided another method of switching between long and short block encoding of frames of audio data, the method comprising the steps of:

calculating short block perceptual entropy of short blocks of a present frame and the sum of the resulting short block perceptual entropy values;
calculating the long block perceptual entropy of the present frame and applying a predetermined weighting to the long block perceptual entropy value; and
switching to the use of short block encoding of the present frame if the sum is less than the weighted long block perceptual entropy value.

[0011] Also according to the present invention there is provided another method of switching between long and short block encoding of frames of audio data, the method comprising the steps of:

calculating the long block perceptual entropy of a present frame in order to determine if short block encoding of the present frame may be preferable; when short block encoding may be preferable, calculating the short block perceptual entropy of short blocks of the present frame; and switching to the use of short block encoding if any of the short block perceptual entropy values exceeds a third threshold.

[0012] Preferably, the method further comprises the step of switching to short block encoding of the succeeding block if none of the short block perceptual entropy values exceeds the third threshold.

[0013] The third threshold may be dependent on the difference between long block perceptual entropy of the present and preceding frames.

[0014] An example of the present invention will now be described with reference to the accompanying drawings, in which:

Fig. 1 is a graphical illustration of short and long block sampling windows according to the MPEG-2 Advanced Audio Coding standard;

Fig. 2 is a flow chart illustrating the switching process of the invention.

[0015] There are several fields in which audio data compression is used. For example, digital recording on media such as Minidisc® uses data compression, as do certain types of digital audio broadcast. An example of the invention will be described in relation to the MPEG-2 Advanced Audio Coding (AAC) Standard for digital television, although the invention may be applied in other fields.

[0016] Referring to Figure 1, an audio signal is sampled in successive frames using particular windows or blocks of samples. Referring to Figure 1, according to the MPEG-2 AAC Standard, each frame may be sampled using long blocks 1 of 2048 samples or short blocks 2 of 256 samples. There are either 8 short blocks or one long block in each frame, and the short blocks of a frame will be referred to as SB0 to SB7. As known from the prior art and explained, for example, in the MPEG-2 AAC standard, intermediate blocks 3 are also used in switching between long and short blocks. It can be seen from Figure 1 that gain varies across each window or block with the result that the central samples have greater statistical weight than those at the beginning and end of the block.

[0017] All blocks, both long and short, overlap into the respective preceding and succeeding blocks. Other than during transition between block types, each block overlaps half way into the preceding and succeeding block such that the audio signal is sampled twice.

[0018] Greater time resolution of the requisite bit rate can be achieved by using short blocks 2 as illustrated in the lower part of Figure 1. In other words, the bit rate

can be set more frequently so that sharp changes in audio energy or attacks can be coded without introducing noise and, particularly, pre-echo.

[0019] An example of a method of switching between long block and short block encoding is illustrated in the flow diagram of Figure 2. Firstly, the perceptual entropy LPE of long blocks, and the perceptual entropy SPED to SPE7 of short blocks SB0 to SB7, of the required frames of the audio signal are calculated. A number of known methods of calculating perceptual entropy may be used. This enables the difference Diff in perceptual entropy of the long block of the present frame LPE_n and that of the previous frame LPE_{n-1} to be calculated and stored, ie.

$$\text{Diff} = LPE_n - LPE_{n-1} \quad (1)$$

where n is a number 1,2,3.. of successive long blocks.

[0020] Also, SPE 0 to SPE 7 are added together and the result SUM stored.

[0021] In this example there are two flags, SPE_flag and LPE_flag, which are also used in the switching process. These will be described in more detail below.

[0022] Generally, Diff is compared with a set threshold LTH1, and if Diff is less than LTH1 long block encoding is used for the present frame. If Diff is greater than LTH1 then short block encoding may be used for the present frame but, in this example, a number of other criteria are also examined.

[0023] Also, if the perceptual entropy of the long blocks of successive frames is similar but large, Diff may not be greater than LTH1, but short block encoding may be preferable. Thus, when Diff is greater than LTH1 the

[0024] LPE_Flag is turned on (ie. set equal to 1) and when the succeeding frame is considered LPE is compared to another threshold LTH2. If LPE is greater than LTH2 then short block encoding may be used, but, in this example, because of the memory effect of perceptual entropy calculation, the other criteria are again examined.

[0025] In the case that Diff is greater than LTH1 or LPE_n is greater than LTH2, SUM is compared with LPE_n . If SUM is lower than a weighted LPE_n ($C.LPE_n$) then short block encoding may be used or, as in this example, SPE0 to SPE7 are examined in more detail. If SUM is greater than $C.LPE_n$ then, excluding other factors, long block encoding is generally used. The weighting C sets the importance of coding efficiency in relation to pre-echo noise reduction. In other words, better audio quality may be provided by setting C such that the system ignores a requirement for a certain number of extra bits for coding particular frames using short blocks rather than long blocks.

[0026] To examine SPE0 to SPE7 in more detail, the value of each of SPE2 to SPE7 is compared with a threshold STH which, in this example, is dependent on

Diff, ie.

$$STH = STH_{\text{constant}} - \frac{\text{Diff-LTH1}}{A} \quad (2)$$

where A and STH_{constant} are constants which can be adjusted and STH varies within a range STH_{MIN} to STH_{MAX} .

[0027] Only SPE2 to SPE7 are examined as any signal discontinuity that would cause high perceptual entropy in SPE0 or SPE1 should be detected during the previous frame. This enables the number of operations to be reduced. It can be seen from Figure 1 that SB0 to SB3 are all in the previous frame. The samples in SB0 and SB1 have a reasonably high gain in the previous frame so, as mentioned above, SPE0 and SPE1 can be ignored. The samples of SB2 and SPE3 have very small gain in the previous frame, so a signal artifact which may result in high perceptual entropy may not be detected in the long block of that frame and these samples must be studied again in the present frame.

[0028] If any of the values of SPE2 to SPE7 are greater than STH, short block encoding is used for the present frame. In order to decide whether or not to use short block encoding for the succeeding frame the short block in which the perceptual entropy first exceeds the threshold (referred to as First in Figure 2) is considered.

[0029] If the first block to have perceptual entropy greater than the threshold occurs before SB7 in the present frame, then short or long block encoding is used for the succeeding frame dependent only on the study of the succeeding frame.

[0030] If SPE7 is the first value to be greater than STH, it is preferable to use short block encoding for present frame and for the succeeding frame because the attack is at the end of present frame and the signal will be coded more smoothly and with less noise by using short block encoding for both of these frames.

[0031] It may be the case that none of the values SPE0 to SPE7 is greater than STH because the last 448 samples used for calculating LPE_n are not used for calculating SPE0 to SPE7 and the signal artifact or attack affecting LPE_n may be in those samples. These samples will therefore need to be studied in the succeeding frame and this is achieved by turning the SPE_flag on (ie. Setting it equal to 1). When coding the succeeding frame the system reads the SPE_flag and, if it is on, SPE0 to SPE3 will be studied.

[0032] Thus, in the succeeding frame, when the SPE_flag is on, if any of the values SPE0 to SPE3 is greater than STH for that frame and SUM is less than the weighted LPE (C.LPE) for that frame then short block encoding is used for that frame. If the Diff or LPE are greater than their thresholds then the system will also study the SPE4 to SPE7 and proceed as described above.

Claims

1. A method of switching between long and short block encoding of frames of audio data, the method comprising the steps of:-

calculating short block perceptual entropy of short blocks of a present frame and the sum of the resulting short block perceptual entropy values;
calculating the long block perceptual entropy of the present frame and applying a predetermined weighting to the long block perceptual entropy value; and
switching to the use of short block encoding of the present frame if the sum is less than the weighted long block perceptual entropy value.

2. The method of claim 1, further comprising the steps of:-

calculating the difference in the long block perceptual entropy of the present frame and a preceding frame; and
switching to short block encoding of the present frame if the difference is greater than a first threshold.

3. The method of claim 2, further comprising the steps of:-

comparing the long block perceptual entropy of the succeeding frame with a second threshold if the difference is greater than the first threshold; and
switching to short block encoding of the succeeding frame if the long block perceptual entropy of the succeeding frame is greater than the second threshold.

4. The method of claim 1, further comprising the steps of:-

calculating the long block perceptual entropy of the present frame in order to determine if short block encoding of the present frame may be preferable;
when short block encoding may be preferable, calculating the short block perceptual entropy of short blocks of the present frame; and
switching to the use of short block encoding if any of the short block perceptual entropy values exceeds a third threshold.

5. The method of claim 4, further comprising the step of switching to short block encoding of the succeeding block if none of the short block perceptual entropy values exceeds the third threshold.

6. The method of claims 4 or 5, wherein the third threshold is dependent on the difference between long block perceptual entropy of the present and preceding frames. 5
7. An apparatus for switching between long and short block encoding of frames of audio data, the apparatus comprising:
- means for calculating short block perceptual entropy of short blocks of a present frame and the sum of the resulting short block perceptual entropy values; 10
 - means for calculating the long block perceptual entropy of the present frame and applying a predetermined weighting to the long block perceptual entropy value; and 15
 - means for switching to the use of short block encoding of the present frame if the sum is less than the weighted long block perceptual entropy value. 20
8. Apparatus according to claim 7, further comprising
- means for calculating the difference in the long block perceptual entropy of a present frame and the preceding frame; and 25
 - means for switching to short block encoding of the present frame if the difference is greater than a first threshold. 30
9. The apparatus of claim 8, further comprising means for comparing the long block perceptual entropy of the succeeding frame with a second threshold if the difference is greater than the first threshold; and
- means for switching to short block encoding of the succeeding frame if the long block perceptual entropy of the succeeding frame is greater than the second threshold. 35
10. Apparatus according to claim 7, further comprising: 40
- means for calculating the long block perceptual entropy of the present frame in order to determine if short block encoding of the present frame may be preferable; 45
 - means for when short block encoding may be preferable, calculating the short block perceptual entropy of short blocks of the present frame; and
 - means for switching to the use of short block encoding if any of the short block perceptual entropy values exceeds a third threshold. 50
11. The apparatus of claim 10, further comprising
- means for switching to short block encoding of the succeeding block if none of the short block perceptual entropy values exceeds the third threshold. 55
12. The apparatus of claims 10 or 11, wherein the third threshold is dependent on the difference between long block perceptual entropy of the present and preceding frames.

Fig.1.

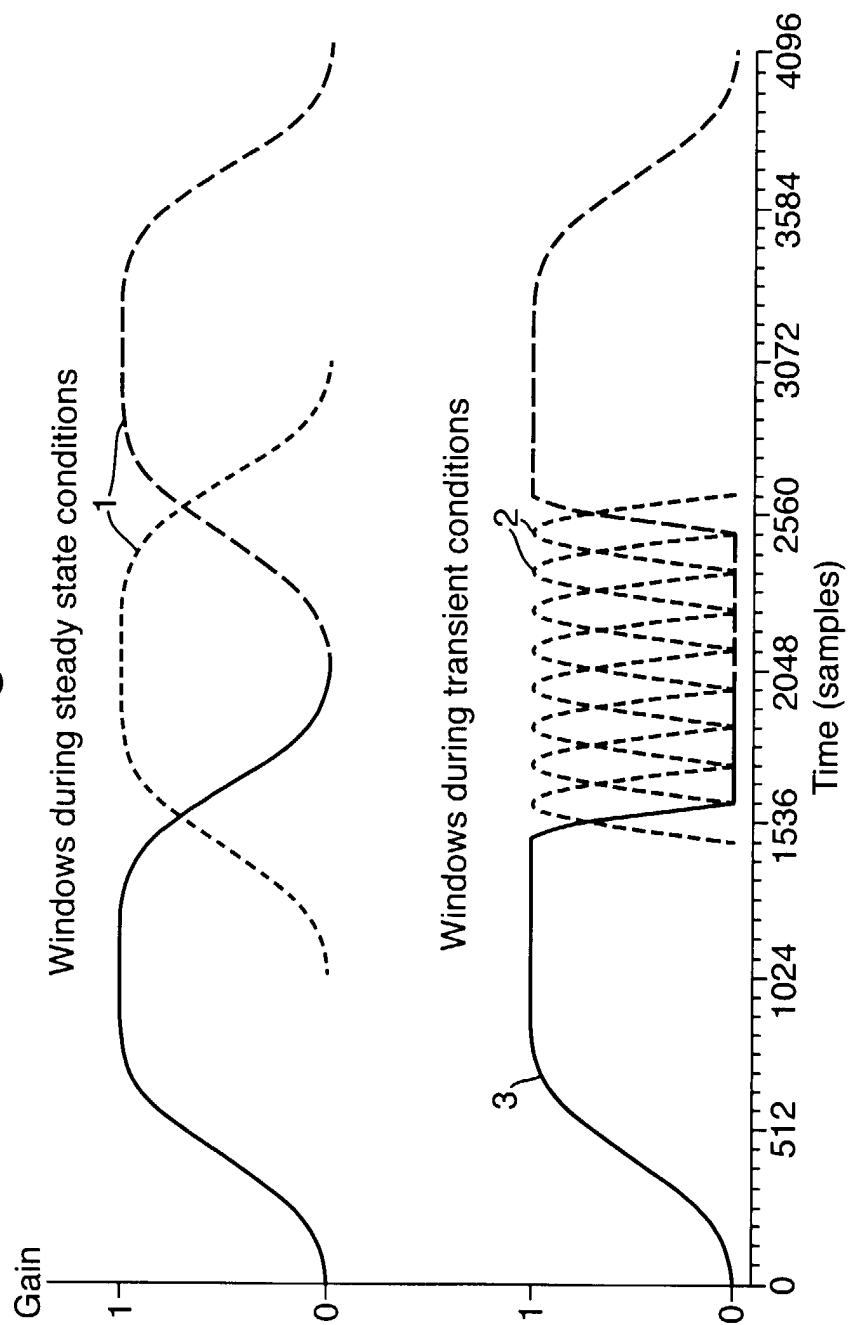


Fig.2.

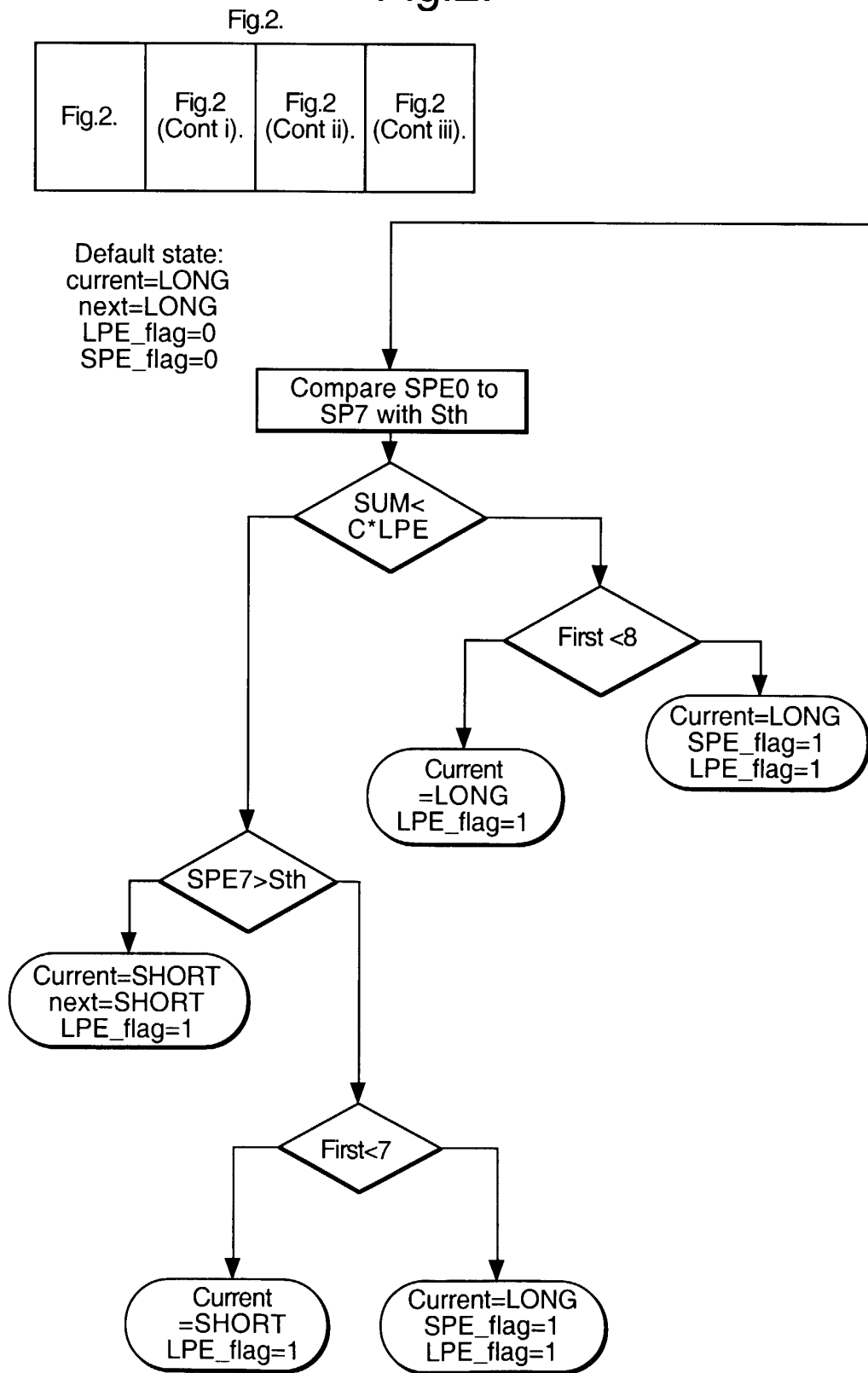


Fig.2 (Cont i).

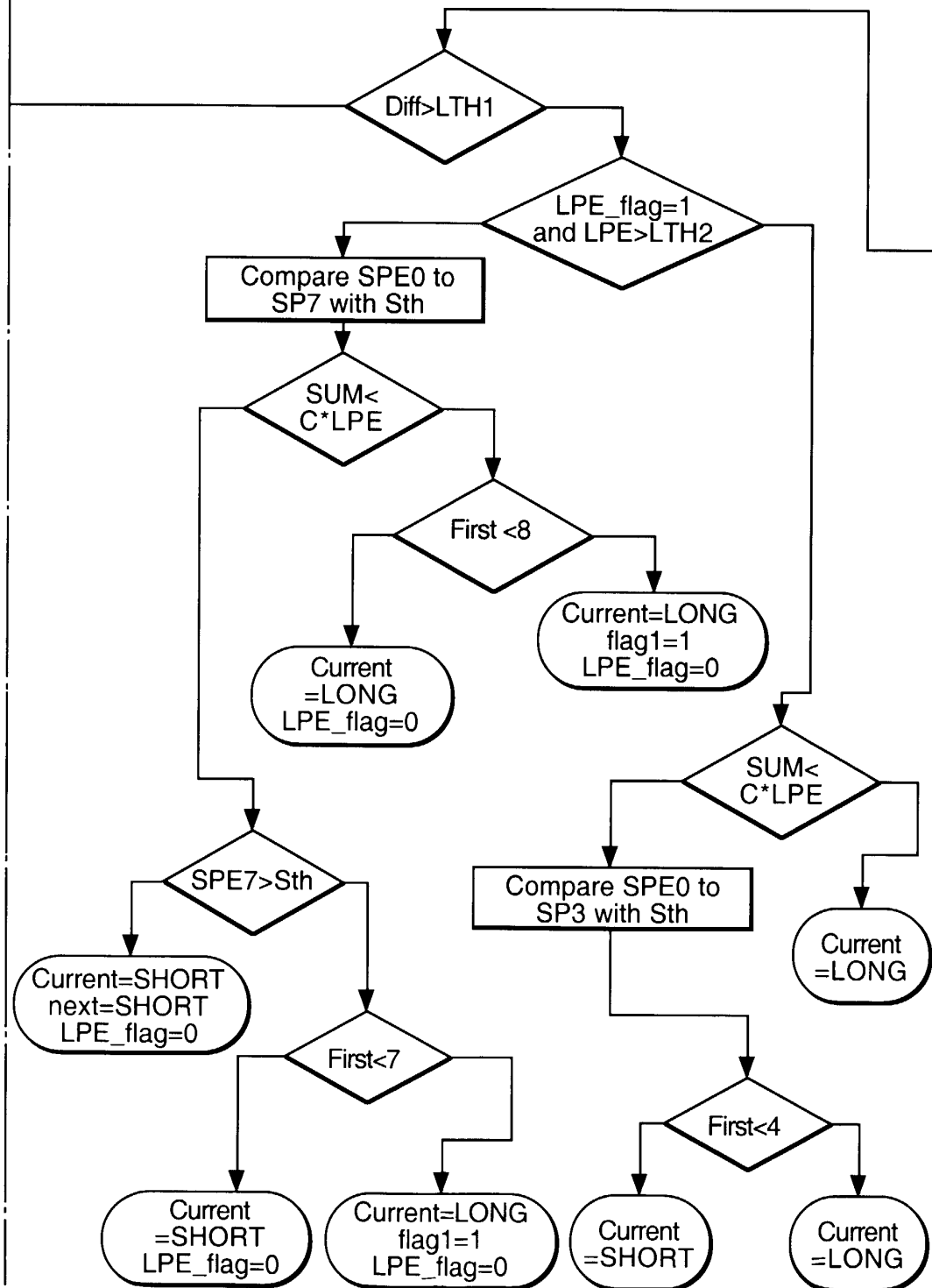


Fig.2 (Cont ii).

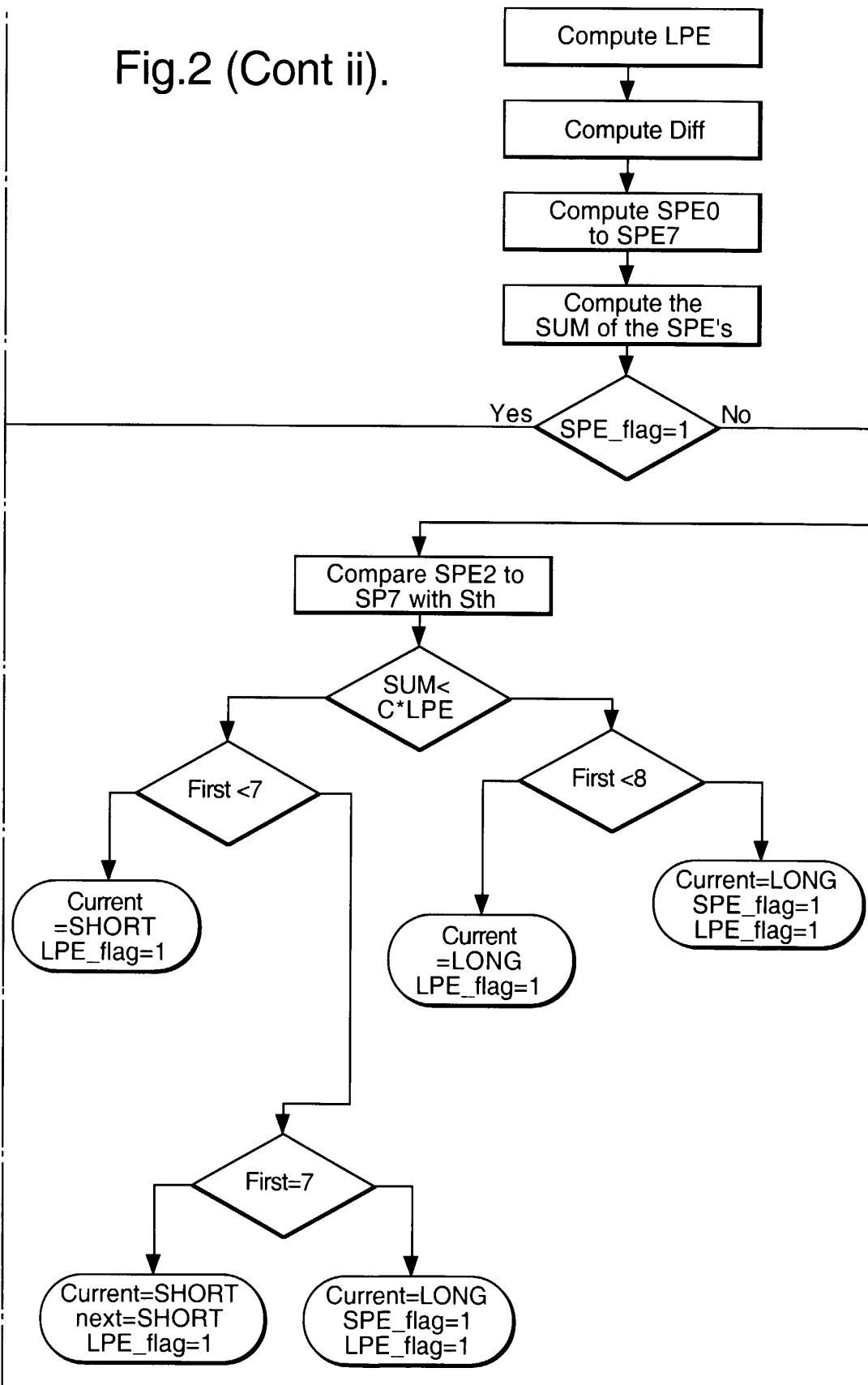


Fig.2 (Cont iii).

