



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 1 175 030 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
23.01.2002 Bulletin 2002/04

(51) Int Cl.7: H04H 5/00

(21) Application number: 01305191.7

(22) Date of filing: 14.06.2001

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

(71) Applicant: NOKIA MOBILE PHONES LTD.
02150 Espoo (FI)

(72) Inventor: Wang, Ye
33720 Tampere (FI)

(30) Priority: 07.07.2000 US 612207

(74) Representative: Read, Matthew Charles et al
Venner Shipley & Co. 20 Little Britain
London EC1A 7DH (GB)

(54) **Method and system for multichannel perceptual audio coding using the cascaded discrete cosine transform or modified discrete cosine transform**

(57) A method and apparatus for coding audio signals having M sound channels in order to reduce the amount of audio data for transmission or storage. A comparison device is used to compare the coding efficiency in intra-channel signal redundancy reduction and the coding efficiency in inter-channel signal redundancy reduction in order to select a more efficient coding process.

In particular, the modified discrete cosine transform (MDCT) process is used to compute the intra-channel coding efficiency and the cascade discrete cosine transform of the MDCT coefficients is used to compute the inter-channel coding efficiency. The efficiencies can be evaluated by computing a gain or the cross-channel correlation coefficients in the audio signals of the M sound channels.

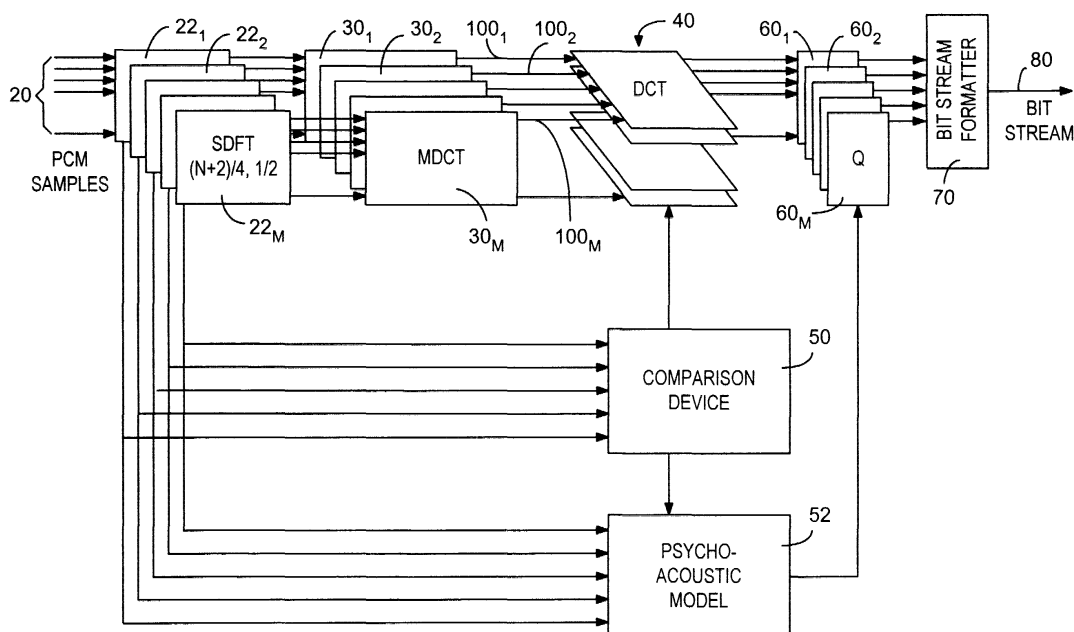


FIG. 3

EP 1 175 030 A2

DescriptionField of the Invention

[0001] The present invention relates generally to audio coding and, in particular, to the coding technique used in a multiple channel surround sound system.

Background of the Invention

[0002] As it is well known in the art, the International Organization for Standardization (IOS) founded the Moving Pictures Expert Group (MPEG) with the intention to develop and standardize compression algorithms for video and audio signals. One of the most efficient audio coding techniques since 1997 is the MPEG-2 Advanced Audio Coding (AAC) algorithm.

The driving force to develop the AAC algorithm has been the quest for an efficient coding method for surround sound signals, such as 5-channel signals including left (L), right (R), center (C), left-surround (LS) and right-surround (RS) signals. MPEG-2 AAC basically makes use of the signal masking properties of the human ear in order to reduce the amount of data. Generally, an N-channel surround sound system, running with a bit rate of M bps/ch does not necessarily have a total bit rate of MxN bps, but rather an overall bit rate significantly less than MxN bps due to cross channel (inter-channel) redundancy. To exploit the inter-channel redundancy, two methods have been used in MPEG-2 AAC standards: Mid-Side (MS) Stereo Coding and Intensity Stereo Coding. Both the MS Stereo and Intensity Stereo coding methods operate on channel pairs, as shown in Figure 1. As shown in Figure 1, the signals in one channel pairs are denoted by (100_L , 100_R) and (100_{LS} , 100_{RS}). The rationale behind the application of stereo audio coding is based on the fact that the human auditory system as well as a stereo recording system use two audio signal detectors. While a human being has two ears, a stereo recording system has two microphones. With these two audio signal detectors, the human auditory system or the stereo recording system receives and records an audio signal from the same source twice, once through each audio signal detector. The two sets of recorded data of the audio signal from the same source contain time and signal level differences caused mainly by the positions of the detectors in relation to the source.

[0003] It is believed that the human auditory system itself is able to detect and discard the inter-channel redundancy, thereby avoiding extra processing. At low frequencies, the human auditory system locates sound sources mainly based on the inter-aural time difference (ITD) of the arrived signals. At high frequencies, the difference in signal strength or intensity level at both ears, or inter-aural level difference (ILD), is the major cue. In order to remove the redundancy in the received signals in a stereo sound system, the psychoacoustic model analyzes the received signals with consecutive time blocks and determines for each block the spectral components of the received audio signal in the frequency domain in order to remove certain spectral components, thereby mimicking the masking properties of the human auditory system. Like any perceptual audio coder, the MPEG audio coder does not attempt to retain the input signal exactly after encoding and decoding, rather its goal is to reduce the amount of audio data yet maintaining the output signals similar to what the human auditory system might perceive. Thus, the MS Stereo coding technique applies a matrix to the signals of the (L,R) or (LS, RS) pair in order to compute the sum and difference of the two original signals, dealing mainly with the spectral image at the mid-frequency range. Intensity Stereo coding replaces the left and the right signals by a single representative signal plus directional information. The replacement of signals in the Intensity Stereo coding scheme is psychoacoustically justified in the higher frequency range at around 2kHz.

[0004] While conventional audio coding techniques can reduce a significant amount of channel redundancy in channel pairs (L/R or LS/RS) based on the dual channel correlation, they may not be efficient in coding audio signals when a large number of channels are used in a surround sound system.

[0005] It is advantageous and desirable to provide a more efficient encoding system and method in order to further reduce the redundancy in the stereo sound signals. In particular, the method can be advantageously applied to a surround sound system having a large number of sound channels (6 or more, for example). Such system and method can also be used in audio streaming over Internet Protocol (IP) for personal computer (PC) users, mobile IP and third-generation (3G) systems for mobile laptop users, digital radio, digital television, and digital archives of movie sound tracks and the like.

Summary of the Invention

[0006] The primary objective of the present invention is to improve the efficiency in encoding audio signals in a sound system in order to reduce the amount of audio data for transmission or storage.

[0007] Accordingly, the first aspect of the present invention is a method of coding audio signals in a sound system having a plurality of sound channels for providing M sets of audio signals, wherein M is a positive integer greater than 2. The method comprises the steps of computing a first value representative of coding efficiency in intra-channel signal

redundancy reduction in said audio signals; computing a second value representative of coding efficiency in inter-channel signal redundancy reduction in said audio signals; comparing the first value to the second value in order to select a more efficient coding process; and encoding the audio signals according to the selected process.

[0008] Preferably, the intra-channel signal redundancy reduction is carried out in accordance with a modified discrete cosine transform process, and the inter-channel signal redundancy reduction is carried out in accordance with a cascaded discrete cosine transform process.

[0009] Preferably, the inter-channel signal redundancy reduction is carried out in order to reduce redundancy in the audio signals among L channels, wherein L is a positive integer greater than 2 but smaller than M+1.

[0010] Preferably, the encoding step includes a signal masking process according to a psychoacoustic model simulating a human auditory system.

[0011] Preferably, the method further includes the step of converting the encoded signals into a bit stream.

[0012] The second aspect of the present invention is an encoder for a sound system having a plurality of sound channels for providing M sets of audio signals, wherein M is a positive integer greater than 2. The encoder comprises a first mechanism, responsive to said audio signals, for providing a first reduced audio signal having a first magnitude indicative of a first data amount by removing intra-channel signal redundancy in said audio signals; a second mechanism, responsive to the first reduced audio signal, for providing a second reduced audio signal having a second magnitude indicative of a second data amount by removing inter-channel signal redundancy in said audio signals; and a third mechanism, responsive to the first audio signal and the second audio signal, for comparing the first and second data amounts for providing a third signal indicative of the reduced audio signal having a magnitude corresponding to the lesser data amount.

[0013] Preferably, the first mechanism removes the intra-channel signal redundancy by a modified discrete cosine transform process, and the second mechanism removes the inter-channel signal redundancy by a cascaded discrete cosine transform process in L of the M sets of audio signals, wherein L is a positive integer greater than 2 but smaller than M+1.

[0014] Preferably, the encoder also includes a mechanism for masking the audio signals according to a psychoacoustic model simulating a human auditory system.

[0015] Preferably, the encoder also includes a quantizer for quantizing the third signal into an encoded signal and a bit-stream formatter for converting the encoded signal into a bit-stream.

[0016] The present invention will become apparent upon reading the description taken in conjunction with Figures 2a to 3.

Brief Description of the Drawings

[0017]

Figure 1 is a diagrammatic representation illustrating a conventional audio coding method for a surround sound system.

Figure 2a is a diagrammatic representation illustrating an audio coding method using an M channel cascaded discrete cosine transform in an M channel sound system.

Figure 2b is a diagrammatic representation illustrating an audio coding method using an L channel cascaded discrete cosine transform in an M channel sound system, where $L < M$.

Figure 3 is a block diagram illustrating a system for audio coding, according to the present invention.

Detailed Description

[0018] The present invention improves the coding efficiency in audio coding for a sound system having M sound channels for sound reproduction, wherein M is greater than 2. In the encoder of the present invention, the individual or intra-channel masking thresholds for each of the sound channels are calculated in a fashion similar to a basic Advanced Audio Coding (AAC) encoder. This method is herein referred to as the intra-channel signal redundancy reduction method. Unlike the convention coding method, however, it also relies on the inter-channel discrete cosine transform (DCT) of the modified discrete cosine transform coefficients. This method is herein referred to as the cascaded MDCT-DCT coding method for inter-channel signal redundancy reduction. The MDCT-DCT coefficients should be quantized according to the highest threshold, taking into account the inter-channel masking effect, known as the masking level difference (MLD). This is characterized by a decreasing masking threshold when the masking mechanism is spatially separated from the source being masked.

[0019] As shown in Figures 2a and 2b, one of the audio coding steps of the present coding method is to perform an inter-channel DCT of multiple channel MDCT coefficients in a cascaded manner in order to reduce the inter-channel redundancy in an M channel sound system, wherein M is greater than 2. Figures 2a and 2b diagrammatically illustrate

M sound channels, and a group of DCT units **40** are used to perform inter-channel DCT from audio signals **100**₁, **100**₂, **100**₃, ..., **100**_{M+1}, and **100**_M. When a block of N samples (the transform length) are used to compute a series of MDCT coefficients, the maximum number of DCT units **40** used to perform the inter-channel DCT is equal to the number of MDCT coefficients. The MDCT transform length N is determined by transform gain, computational complexity and the pre-echo problem; and the number of MDCT coefficients is N/2. Typically, the MDCT transform length N is between 256 and 2048 samples. Accordingly, the number of DCT units required to perform the inter-channel DCT is between 128 and 1024. In practice, however, the number of DCT units needed for performing the inter-channel DCT is much less.

[0020] As shown in Figure 2a, the cascaded MDCT-DCT is carried out with M DCT units **40**. It is also possible, however, to perform the inter-channel DCT of the MDCT coefficients of L channels, wherein L is a subset of M with L being greater than 2 and smaller than M+1. For example, in a 5-channel sound system consisting of left (L), right (R), center (C), left-surround (LS) and right-surround (RS) channels, it is possible to perform the cascaded inter-channel DCT of the MDCT coefficients involving only 4 channels, namely, L, R, LS and RS. Likewise, in a 12-channel sound system, it is possible to perform an inter-channel DCT of only 5 or 6 channel MDCT coefficients. As shown in Figure 2b, the cascaded MDCT-DCT is carried out with M-3 DCT units **40** in order to compute the cross correlation among audio signals **100**₃, ..., and **100**_{M+1}.

[0021] In some surround sound recording and reproduction cases, the correlation in the audio signals among L (>2) channels is strong. Accordingly, the efficiency of audio coding using the cascaded MDCT-DCT method is higher than the efficiency of the intra-channel MDCT method alone. However, if the correlation in the audio signals among the L channels is weak, it is possible that this inter-channel DCT technique may not be as efficient as the intra-channel signal redundancy reduction using the MDCT coding method. Thus, it is advantageous to provide a comparison device to compare the coding efficiency of the two methods for each sampling block or a group of sampling blocks and select the more efficient method.

[0022] The efficiency of the intra-channel MDCT coding method is represented by Equations 1 and 2 below. In a block of N samples with each block having a series of sound amplitude values of a(k)'s, the MDCT coefficients in the frequency domain are given by:

$$(\alpha_r)_m = (1/N) \sum_{k=0}^{N-1} a(k)_m \cos[(\pi/4N)(4k + 2 + N)(2r + 1)] \quad (1)$$

$$bits_per_coef1 = (1/MN) \sum_{r=0}^{(N/2)-1} \sum_{m=1}^M \log_2[abs(\alpha_r)_m] \quad (2)$$

In the above equations, m represents a channel number and M represents the number of sound channels involved.

[0023] In particular, if it is desirable to determine the cross correlation among all M channels, then a cascaded inter-channel DCT of the M sets of MDCT coefficients should be performed, as given in Equations 3 and 4 below:

$$\beta_{r,s} = (1/MN) \sum_{k=0}^{N-1} \sum_{j=0}^{M-1} a(k,j) \cos[(\pi/4N)(4k + 2 + N)(2r + 1)] \cos[(\pi/M)(2j+1)s] \quad (3)$$

$$bits_per_coef2 = (1/MN) \sum_{r=0}^{(N/2)-1} \sum_{s=1}^M \log_2[abs(\beta_{r,s})] \quad (4)$$

It should be noted that the coefficient a(k) in Equation 1 and the coefficient a(k,j) in Equation 3 may include a modified function of sin(πk/N).

[0024] In order to ensure that the efficiency of the cascaded MDCT-DCT process is higher than that of the intra-channel MDCT process, it is possible to compute the gain according to Equation 5 as follows:

$$G = \sum_{i=0}^{L-1} \{bits_per_coef1 - bits_per_coef2\} / bits_per_coef1 \quad (5)$$

where L is the number of frames of the test signal used to calculate the average gain G. If G is positive, then the efficiency of the cascaded inter-channel DCT process is higher than the efficiency of the intra-channel MDCT process. Accordingly, the cascaded inter-channel DCT should be used for audio coding in order to reduce the amount of encoded data.

[0025] Alternatively, the efficiency in the inter-channel signal redundancy reduction using the cascaded MDCT-DCT process can be evaluated using a cross-channel correlation method. The normalized cross-channel correlation coefficient between any two channels p and q is represented by the following equation:

$$C_{pq} = \sum_{r=0}^{(N/2)-1} (\alpha_r)_p (\alpha_r)_q / \sqrt{\sum_{r=0}^{(N/2)-1} (\alpha_r)_p^2 \sum_{r=0}^{(N/2)-1} (\alpha_r)_q^2} \quad (6)$$

The absolute value of C_{pq} can be used to set a threshold over which the cascaded MDCT-DCT process should be used. In an M channel system, it is possible to calculate $M(M-1)/2$ normalized cross-channel correlation coefficients. For example, in a three channel system having channels 1, 2 and 3, it is possible to calculate the normalized cross-channel correlation coefficients C_{12} , C_{13} , and C_{23} . The sum of the absolute values of these normalized cross-channel correlation coefficients can be used to compare the efficiency of the intra-channel MDCT method to the inter-channel cascaded MDCT-DCT method.

[0026] Accordingly, the present invention provides a system for efficient audio coding to reduce redundancy in an M channel sound system, as shown in Figure 3. As shown, the pulsed code modulation (PCM) samples **20** in the M channels are first conveyed to a set of M Shifted Discrete Fourier Transform (SDFT) devices **22₁**, **22₂**, ..., **22_M** so that the real parts of the SDFT coefficients form a group of M MDCT coefficients in a group of M MDCT units **30₁**, **30₂**, ..., **30_M**, respectively. The devices **22₁**, **22₂**, ..., **22_M** and the MDCT units **30₁**, **30₂**, ..., **30_M** together perform an intra-channel decorrelation.

[0027] For a set of signal sequences $\{a(k)_m\}$, the Shifted Discrete Fourier Transform coefficient is defined as follows:

$$(\alpha_r)_m = (1/N) \sum_{k=0}^{N-1} a(k)_m \exp \{i2\pi(k+u)(r+v)/N\} \quad (7)$$

where $u=(N+2)/4$ and $v=1/2$, being the shift in the time domain and the shift in the frequency domain, respectively. Thus, the relationship between the MDCT coefficients (Eq.1) and the SDFT coefficients (Eq.7) is as follows:

$$\begin{aligned} (\alpha_r)_m &= (1/N) \sum_{k=0}^{N-1} a(k)_m \cos[(\pi/4N)(4k+2+N)(2r+1)] \\ &= (1/2N) \sum_{k=0}^{N-1} \tilde{a}(k)_m \exp \{i2\pi(k+u)(r+v)/N\} \end{aligned} \quad (8)$$

where $\tilde{a}(k)_m = a(k)_m - a(N/2-1-k)_m$ for $k=0, \dots, (N/2)-1$; and $\tilde{a}(k)_m = a(k)_m - a(3N/2-1-k)_m$ for $k=(N/2), \dots, (N-1)$ with N being an even number. Accordingly, the right-hand side of Eq.8 is $\text{SDFT}_{u,v}(\tilde{a}(k)_m/2)$ or $\text{real}\{\text{SDFT}_{u,v}(a(k)_m/2)\}$.

[0028] As shown in Figure 3, a number of DCT units **40** are used to compute the inter-channel signal redundancy reduction in these M sets of MDCT coefficients. The number of DCT units **40** can be equal to or less than the number of MDCT coefficients in each of the M channels, as discussed earlier in conjunction with Equation 3. A comparison device **50** is used to compute the gain G (Equation 5) or the threshold from the cross-channel correlation coefficients

C_{pq} (Equation 6) to ensure that the coding according to the cascaded inter-channel DCT of the MDCT coefficients is more efficient than the intra-channel decorrelation by the MDCT units $30_1, 30_2, \dots, 30_M$. If the gain G is negative or the cross-channel correlation is lower than a pre-determined threshold, it can cause the DCT units 40 to turn off. A masking mechanism 52 , based on a so-called psychoacoustic model, is used to remove the audio data believed not to be used by a human auditory system. As shown in Figure 3, the masking mechanism is also operatively connected to the comparison device 50 so that the masking is carried out according to the intra-channel MDCT manner or the inter-channel MDCT-DCT manner. Finally, the 2-D spectral image is quantized by a group of quantizers $60_1, 60_2, \dots, 60_M$ according to the masking threshold calculated by the psychoacoustic model and the quantized data is further processed by a bit stream formatter 70 into a bit stream 80 for transmission or storage.

[0029] The efficiency of the cascaded MDCT-DCT coding process in removing cross-channel redundancy, in general, increases with the number of sound channels involved. For example, if a sound system consists of 6 or more surround sound speakers, then the reduction in cross-channel redundancy using the cascaded MDCT-DCT processing is usually significant. However, if the number of channels to be used in the cascaded MDCT-DCT processing is 2, then the efficiency may not be improved at all. It should be noted that, like any perceptual audio coder, the goal of the cascaded MDCT-DCT processing is to reduce the audio data for transmission or storage. While the processing method is intended to produce signal outputs similar to what a human auditory system might perceive, its goal is not to replicate the input signals.

[0030] It should be noted that the so-called psychoacoustic model may consist of a certain perceptual model and a certain band mapping model. The surround sound encoding system may consist of components such as an AAC gain control and a certain long-term prediction model. However, these components are well-known in the art and they can be modified, replaced or omitted. Thus, although the invention has been described with respect to a preferred embodiment thereof, it will be understood by those skilled in the art that the foregoing and various other changes, omissions and deviations in the form and detail thereof may be made without departing from the spirit and scope of this invention.

Claims

1. A method of coding audio signals in a sound system having a plurality of sound channels for providing M sets of audio signals, wherein M is a positive integer greater than 2, said method comprising the steps of:

computing a first value representative of coding efficiency in intra-channel signal redundancy reduction in said audio signals;
 computing a second value representative of coding efficiency in inter-channel signal redundancy reduction in said audio signals;
 comparing the first value to the second value in order to select a more efficient coding process; and
 encoding the audio signals according to the selected process.

2. The method of claim 1, wherein the intra-channel signal redundancy reduction is carried out in accordance with a modified discrete cosine transform process, and the inter-channel signal redundancy reduction is carried out in accordance with a cascaded discrete cosine transform process.

3. The method of claim 1, wherein the inter-channel signal redundancy reduction is carried out in order to reduce redundancy in the audio signals among L channels, wherein L is a positive integer greater than 2 but smaller than $M+1$.

4. The method of claim 1, wherein the encoding step includes a signal masking process according to a psychoacoustic model simulating a human auditory system.

5. The method of claim 1, further comprising the step of converting the encoded signals into a bit stream.

6. An encoding apparatus for a sound system having a plurality of sound channels for providing M sets of audio signals, wherein M is a positive integer greater than 2, said encoding apparatus comprising:

first means, responsive to said audio signals, for providing a first reduced audio signal have a first magnitude indicative of a first data amount by removing intra-channel signal redundancy in said audio signals;
 second means, responsive to the first reduced audio signal, for providing a second reduced audio signal have a second magnitude indicative of a second data amount by removing inter-channel signal redundancy in said audio signals; and

third means, responsive to the first audio signal and the second audio signal, for comparing the first and second data amounts for providing a third signal indicative of the reduced audio signal having a magnitude corresponding to the lesser data amount.

- 5 **7.** The encoding apparatus of claim 6, wherein the first means removes the intra-channel signal redundancy by a modified discrete cosine transform process, and the second means removes the inter-channel signal redundancy by a cascaded discrete cosine transform process.
- 10 **8.** The encoding apparatus of claim 6, wherein the second means removes the inter-channel signal redundancy in L of the M sets of audio signals and wherein L is a positive integer greater than 2 but smaller than M+1.
- 9.** The encoding apparatus of claim 6, further comprising a mechanism for masking the audio signals according to a psychoacoustic model simulating a human auditory system.
- 15 **10.** The encoding apparatus of claim 6, further comprising a mechanism for quantizing the third signal into an encoded signal.
- 11.** The encoding apparatus of claim 10, further comprising a mechanism for converting the encoded signal into a bit stream.
- 20 **12.** The encoding apparatus of claim 6, wherein the third means is capable of computing a value indicative of cross-channel correlation coefficients among the M sets of audio signals and comparing said value to a pre-determined threshold in order to compare the first and second data amounts.

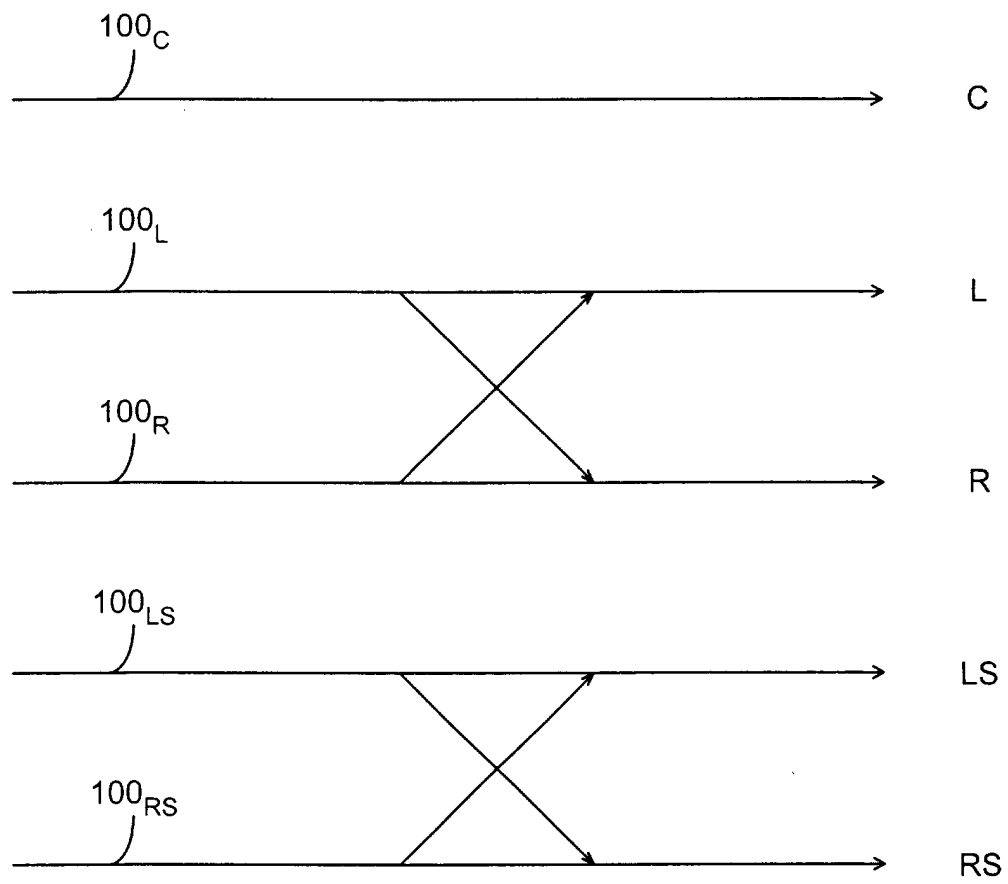


FIG. 1
(PRIOR ART)

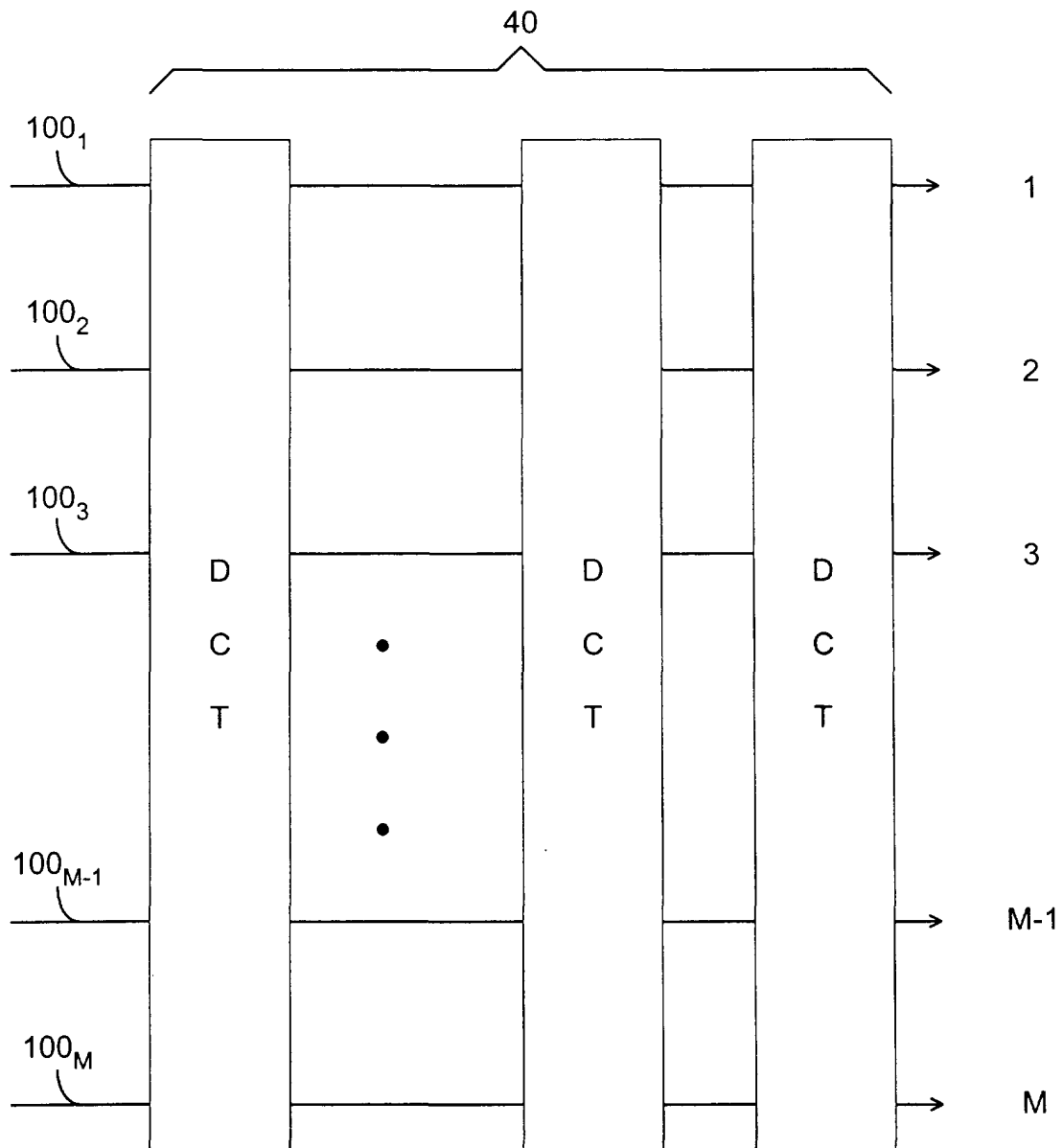


FIG. 2a

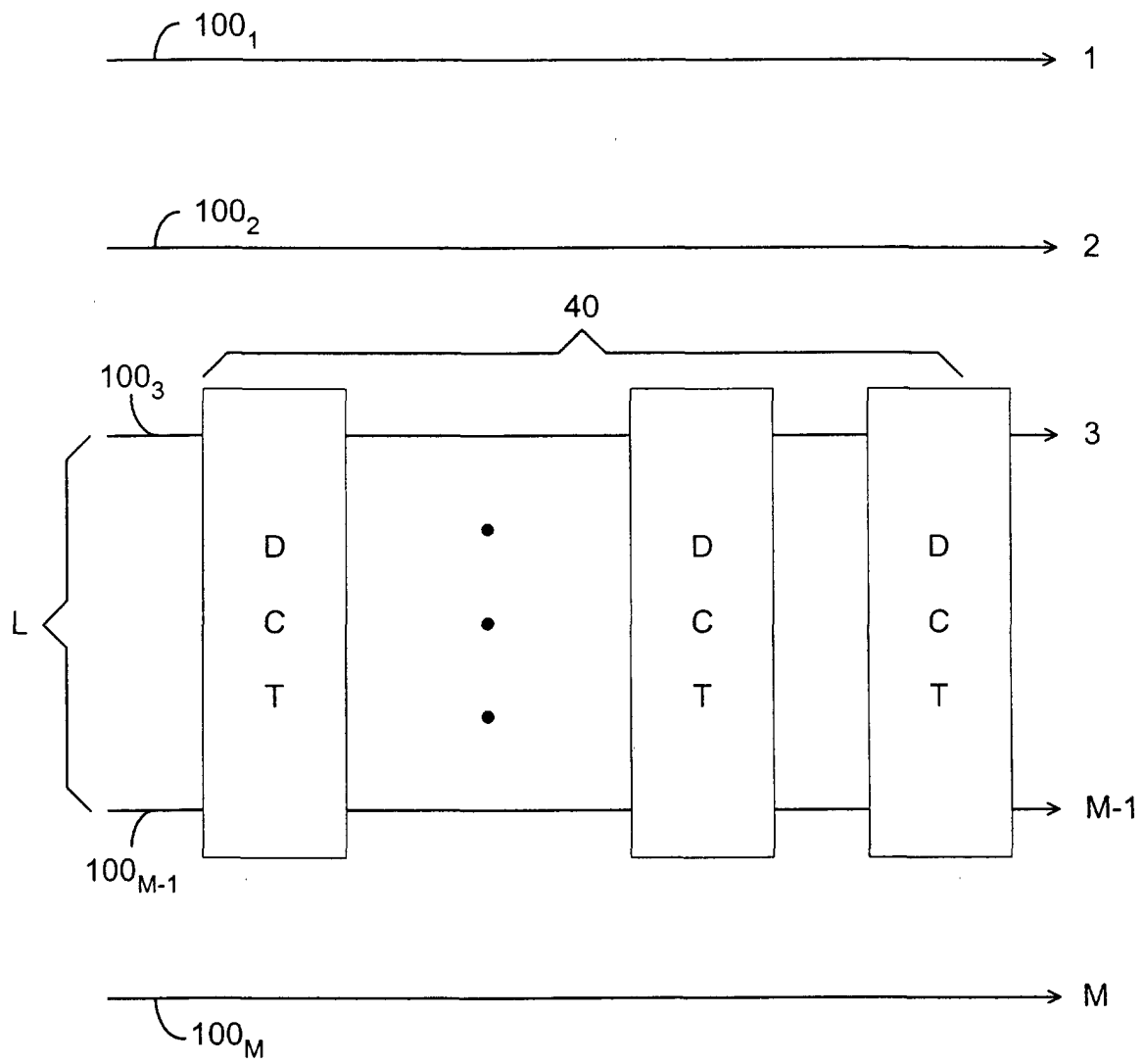


FIG. 2b

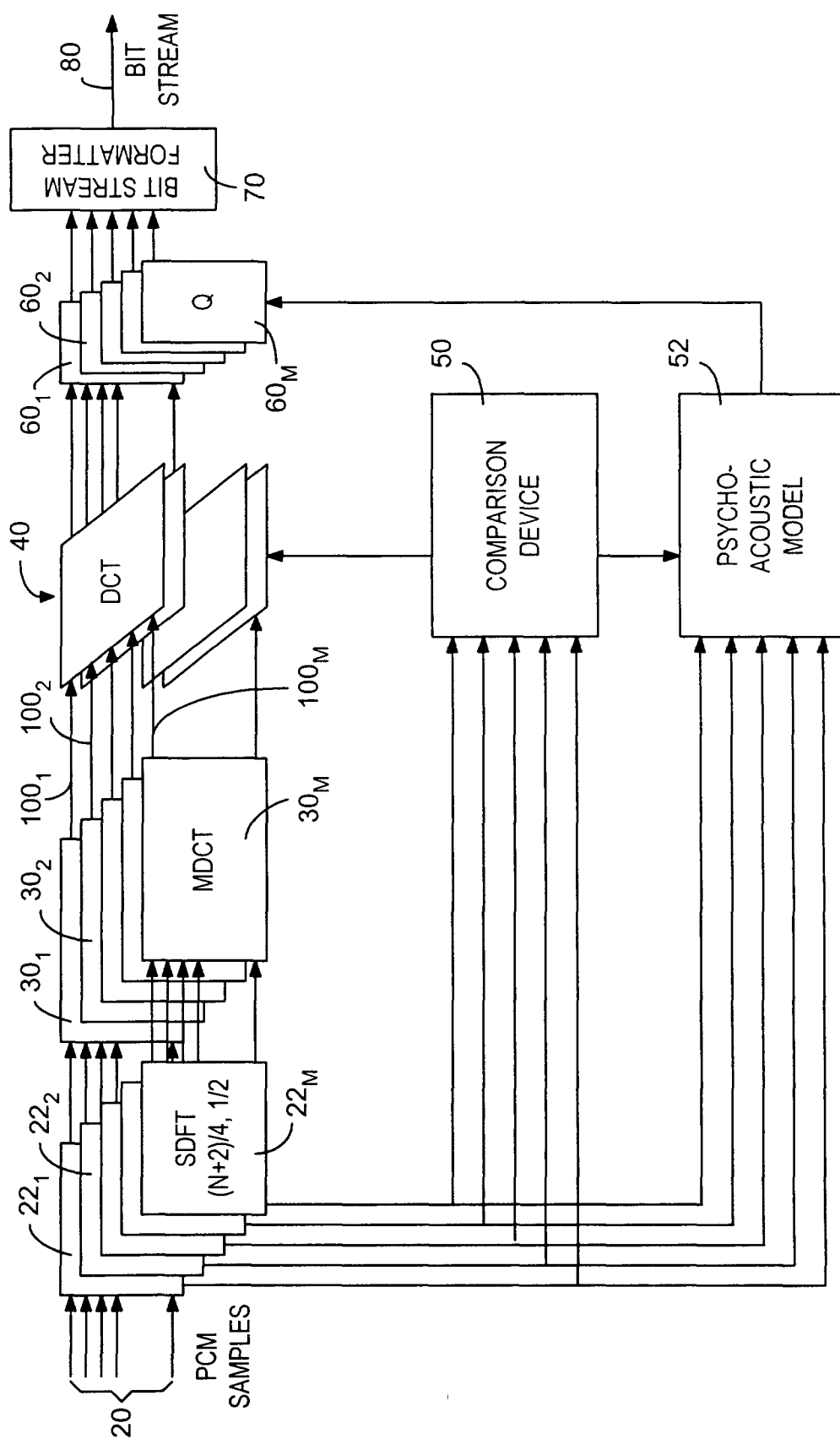


FIG. 3