

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4750010号
(P4750010)

(45) 発行日 平成23年8月17日(2011.8.17)

(24) 登録日 平成23年5月27日(2011.5.27)

(51) Int.Cl.		F I			
H04S	5/02	(2006.01)	H04S	5/02	J
G10L	11/00	(2006.01)	G10L	11/00	402G
			G10L	11/00	402K
			G10L	11/00	402L
			G10L	11/00	101D
請求項の数 6 (全 12 頁) 最終頁に続く					

(21) 出願番号	特願2006-352297 (P2006-352297)	(73) 特許権者	000005049
(22) 出願日	平成18年12月27日(2006.12.27)		シャープ株式会社
(65) 公開番号	特開2008-166976 (P2008-166976A)		大阪府大阪市阿倍野区長池町22番22号
(43) 公開日	平成20年7月17日(2008.7.17)	(74) 代理人	100097113
審査請求日	平成21年2月18日(2009.2.18)		弁理士 堀 城之
		(74) 代理人	100124316
			弁理士 塩田 康弘
		(72) 発明者	藤井 修
			大阪府大阪市阿倍野区長池町22番22号
			シャープ株式会社内
		審査官	菊池 充
最終頁に続く			

(54) 【発明の名称】 音響音声再生装置

(57) 【特許請求の範囲】

【請求項1】

デジタル音響音声データのスペクトル成分からホルマント成分を抽出し、前記スペクトル成分の前記ホルマント成分のパワー又はエネルギーと、前記スペクトル成分から前記ホルマント成分を除いた成分のパワー又はエネルギーとのパワー比又はエネルギー比を算出するホルマント周波数検出部と、

前記パワー比又はエネルギー比を累積加算するパワー比又はエネルギー比加算部と、

前記パワー比又はエネルギー比の累積加算値に応じてスピーカ構成を決定するスピーカ構成決定部と

を有することを特徴とする音響音声再生装置。

10

【請求項2】

前記スピーカ構成決定部は、前記パワー比又はエネルギー比の累積加算値が、前記ホルマント成分のみであり、前記ホルマント成分を除いた成分がないことを示す場合には、前記スピーカ構成をセンターチャンネルのスピーカの1台と決定することを特徴とする請求項1記載の音響音声再生装置。

【請求項3】

前記スピーカ構成決定部が前記スピーカ構成をセンターチャンネルのスピーカの1台と決定した場合には、前記デジタル音響音声データの時間領域のデータである左チャンネルの音響音声データ及び右チャンネルの音響音声データからモノラルの音声データを生成する再生データ生成部をさらに有することを特徴とする請求項2記載の音響音声再生装置。

20

【請求項 4】

前記スピーカ構成決定部は、前記パワー比又はエネルギー比の累積加算値が、前記ホルマント成分と、少ない前記ホルマント成分を除いた成分とからなることを示す場合は、前記スピーカ構成を左チャンネルのスピーカ、前記センターチャンネルのスピーカ及び右チャンネルのスピーカの3台と決定することを特徴とする請求項1乃至3のいずれかに記載の音響音声再生装置。

【請求項 5】

前記再生データ生成部は、前記スピーカ構成決定部が前記スピーカ構成を前記左チャンネルのスピーカ、前記センターチャンネルのスピーカ及び前記右チャンネルのスピーカの3台と決定した場合には、前記左チャンネルの音響音声データ及び前記右チャンネルの音響音声データの同一成分から台詞成分の音声データを抽出することを特徴とする請求項4記載の音響音声再生装置。

10

【請求項 6】

前記スピーカ構成決定部は、前記パワー比又はエネルギー比の累積加算値が、前記ホルマント成分と、多くの前記ホルマント成分を除いた成分とからなることを示す場合又は、多くの前記ホルマント成分を除いた成分のみであり、前記ホルマント成分がないことを示す場合は、前記スピーカ構成を左チャンネルのスピーカ及び右チャンネルのスピーカの2台と決定することを特徴とする請求項1乃至5のいずれかに記載の音響音声再生装置。

【発明の詳細な説明】

【技術分野】

20

【0001】

本発明は、符号化されたデジタルデータからコンテンツを分類して原音響音声信号を再生する音響音声再生装置に関する。

【背景技術】

【0002】

従来の音声再生装置には、入力される音声信号の周波数スペクトルによる母音長と、高レベル音の長さ、低レベル音の長さにより入力信号を検出する検出手段と、この検出手段の検出結果について、予め設定したメンバーシップ関係に基づき入力信号ソースモードをファジィ推論を使って判断する判断手段と、この判断手段の出力により音声出力モードを選択する自動切替手段とを備えているものがある（例えば、特許文献1参照。）。 30

【0003】

【特許文献1】特開平7-115606号公報（請求項1，[0008]～[0019]、図1）

【発明の開示】

【発明が解決しようとする課題】

【0004】

上記した従来の音声再生装置では、多数の帯域通過フィルタやアナログ/デジタル変換器（ADC）が必要であるため、装置が高価になるとともに、消費電力が大きいという課題があった。また、ADCを用いているため、量子化雑音に起因して音質の劣化が生じるという課題があった。 40

【0005】

本発明は、上述した事情に鑑みてなされたものであり、上述のような課題を解決することができる音響音声再生装置を提供することを目的とする。

【課題を解決するための手段】

【0006】

上記課題を解決するために、請求項1記載の発明に係る音響音声再生装置は、デジタル音響音声データのスペクトル成分からホルマント成分を抽出し、前記スペクトル成分の前記ホルマント成分のパワー又はエネルギーと、前記スペクトル成分から前記ホルマント成分を除いた成分のパワー又はエネルギーとのパワー比又はエネルギー比を算出するホルマント周波数検出部と、前記パワー比又はエネルギー比を累積加算するパワー比又はエネルギー比加算 50

部と、前記パワー比又はエネルギー比の累積加算値に応じてスピーカ構成を決定するスピーカ構成決定部とを有することを特徴としている。

【0007】

また、請求項2記載の発明は、請求項1記載の音響音声再生装置に係り、前記スピーカ構成決定部は、前記パワー比又はエネルギー比の累積加算値が、前記ホルマント成分のみであり、前記ホルマント成分を除いた成分がないことを示す場合には、前記スピーカ構成をセンターチャンネルのスピーカの1台と決定することを特徴としている。

【0008】

また、請求項3記載の発明は、請求項2記載の音響音声再生装置に係り、前記スピーカ構成決定部が前記スピーカ構成をセンターチャンネルのスピーカの1台と決定した場合には、前記デジタル音響音声データの時間領域のデータである左チャンネルの音響音声データ及び右チャンネルの音響音声データからモノラルの音声データを生成する再生データ生成部をさらに有することを特徴としている。

10

【0009】

また、請求項4記載の発明は、請求項1乃至3のいずれかに記載の音響音声再生装置に係り、前記スピーカ構成決定部は、前記パワー比又はエネルギー比の累積加算値が、前記ホルマント成分と、少ない前記ホルマント成分を除いた成分とからなることを示す場合は、前記スピーカ構成を左チャンネルのスピーカ、前記センターチャンネルのスピーカ及び右チャンネルのスピーカの3台と決定することを特徴としている。

【0010】

20

また、請求項5記載の発明は、請求項4記載の音響音声再生装置に係り、前記再生データ生成部は、前記スピーカ構成決定部が前記スピーカ構成を前記左チャンネルのスピーカ、前記センターチャンネルのスピーカ及び前記右チャンネルのスピーカの3台と決定した場合には、前記左チャンネルの音響音声データ及び前記右チャンネルの音響音声データの同一成分から台詞成分の音声データを抽出することを特徴としている。

【0011】

また、請求項6記載の発明は、請求項1乃至5のいずれかに記載の音響音声再生装置に係り、前記スピーカ構成決定部は、前記パワー比又はエネルギー比の累積加算値が、前記ホルマント成分と、多くの前記ホルマント成分を除いた成分とからなることを示す場合又は、多くの前記ホルマント成分を除いた成分のみであり、前記ホルマント成分がないことを示す場合は、前記スピーカ構成を左チャンネルのスピーカ及び右チャンネルのスピーカの2台と決定することを特徴としている。

30

【発明の効果】

【0012】

本発明によれば、ソースのコンテンツが異なっても聴取者が聴きやすい音響音声再生装置を安価に構成できるとともに、消費電力を大幅に削減することができる。また、量子化雑音に起因する音質劣化を防止することができる。

【発明を実施するための最良の形態】

【0013】

実施の形態1 .

40

図1は、本発明の実施の形態1に係る音響音声再生装置の構成を示すブロック図である。この例の音響音声再生装置は、地上デジタルテレビジョン放送、BSデジタルテレビジョン放送やCSデジタルテレビジョン放送等で採用されている音声圧縮符号化技術の1つであるAAC (Advanced Audio Coding) 方式が適用されている。この例の音響音声再生装置は、アンパッキング・逆量子化部1と、ブロックサイズ判定部2と、周波数・時間変換部3と、ホルマント周波数検出部4と、パワー比加算部5と、スピーカ構成決定部6と、再生データ生成部7と、アンプ8と、L (左) チャンネルスピーカ9と、C (センター) チャンネルスピーカ10と、R (右) チャンネルスピーカ11とから構成されている。

【0014】

アンパッキング・逆量子化部1は、圧縮符号化されたデジタル音響音声データC D A D

50

(ビットストリーム(Bitstream))からヘッダ、スケールファクタ、ブロックサイズ等の符号化データを展開する。また、アンパッキング・逆量子化部1は、展開したヘッダ、スケールファクタ、ブロックサイズ等の符号化データを逆量子化するとともに、これらに基づいて、逆量子化したブロックサイズごとの周波数成分の逆修正離散余弦変換(IMDCT; Inverse Modified Discrete Cosine Transform)係数(スペクトルデータ)を算出する。

【0015】

ブロックサイズ判定部2は、アンパッキング・逆量子化部1で展開されたブロックサイズに基づいて、1フレームをロングブロック(例えば、2048サンプルブロック)又はショートブロック(例えば、256サンプルブロック)のいずれかで、周波数・時間変換部3に周波数・時間変換処理を行わせるかを判定する。この判定は、例えば、定常音の場合にはロングブロックで周波数・時間変換処理を行うと判定し、過渡音の場合にはショートブロックで周波数・時間変換処理を行うと判定する。

10

【0016】

このような判定を行うのは、過渡音をショートブロックで周波数・時間変換処理を行った場合には、経時マスキング効果に起因する量子化誤差の影響を低減することができるからである。ここで、経時マスキング効果とは、複数の音圧レベルの音が時間的に接近して発生している場合、ある大きな音圧レベルの音によって、その直前及び直後に発生した音圧レベルの低い別の音がマスキングされ、人には聴こえない又は聴こえにくくなるという聴感効果をいう。

20

【0017】

周波数・時間変換部3は、アンパッキング・逆量子化部1で算出された逆量子化後のスペクトルデータについて、ブロックサイズ判定部2で判定されたブロックサイズごとに、公知のIMDCT処理を施すことにより、対応する時間領域のデータ、即ち、Lチャンネルの音響音声データ及びRチャンネルの音響音声データに直交変換する。

【0018】

ホルマント周波数検出部4は、アンパッキング・逆量子化部1で算出された逆量子化後のスペクトルデータについて、以下に示す手法によりホルマントを分析する。ここで、ホルマントとは、声道の伝達関数の極と零により音声波の周波数スペクトル上に生じる数個の共振の山をいう。第n番目(nは自然数)のホルマントピーク近傍のスペクトル成分の周波数を f_i (iは自然数)、各スペクトル成分に対する加重係数を W_i とした場合、ホルマント周波数 F_n は、近似的には、式(1)により算出される。

30

【0019】

【数1】

$$F_n = \frac{\sum_i W_i \times f_i}{\sum_i f_i} \dots (1)$$

40

式(1)において、加重係数 W_i は、各スペクトル成分の振幅 A_i と、着目される周波数領域内でのスペクトル成分の最大振幅 A_{max} との比 A_i / A_{max} で与えられる。

【0020】

従って、ホルマント周波数検出部4は、ホルマントの着目される周波数領域を、例えば、基本周波数 F_0 について100Hz~300Hz、ホルマント周波数 F_1 について250Hz~1100Hz、ホルマント周波数 F_2 について800Hz~3500Hz、ホルマント周波数 F_3 について1600Hz~3800Hzとそれぞれ設定し、上記式(1)を用いてホルマント周波数 $F_0 \sim F_3$ を算出する。

【0021】

次に、ホルマント周波数検出部4は、算出したホルマント周波数 $F_0 \sim F_3$ の値が図2

50

に示すホルマント判定表のいずれかの母音と相関があるか否かを判定し、相関がある場合には母音が存在すると判断し、図2に示すホルマント振幅の重み係数 $L_1 \sim L_3$ を乗算することによりホルマントスペクトルを抽出する。図2は、三浦 種敏監修、勝木 保次外7名著、社団法人電子情報通信学会編著、新版「聴覚と音声」、第7版、社団法人電子情報通信学会、平成6年6月10日、p355に記載されている、バーネイ・ピーターソン(Barney Peterson)氏によって測定された米語母音のホルマントの平均の基本周波数 F_0 と、平均ホルマント周波数 $F_1 \sim F_3$ と、それぞれの平均ホルマント振幅とを示している。この場合、抽出されるホルマントスペクトルの帯域幅は、正弦波のスペクトルデータについてMDC T処理を施した場合に統計的分布より得られるので、予めROMやフラッシュメモリ等の記憶部に保存しておくことが好ましい。

10

【0022】

また、抽出したホルマントスペクトル、即ち、ホルマント成分のスペクトルを2乗してホルマント成分のスペクトルパワーを求めるとともに、全帯域のスペクトルを2乗して全帯域のスペクトルパワーを求める。そして、ホルマント成分のスペクトルパワーの総和、即ち、エネルギーを $P_{ow}(F_{sum})$ 、さらに全帯域のスペクトルパワーの総和を $P_{ow}(S_{sum})$ 、同様にスペクトル成分からホルマント成分を除いた成分のスペクトルパワーの総和を $P_{ow}(N_{sum})$ とすると、ホルマント成分を除いた成分のエネルギー $P_{ow}(N_{sum})$ は、式(2)で算出される。

$$P_{ow}(N_{sum}) = P_{ow}(S_{sum}) - P_{ow}(F_{sum}) \cdots (2)$$

従って、ホルマント周波数検出部4は、ホルマント成分のパワーとホルマント成分を除いた成分のパワー又はエネルギーの比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ を算出することができる。これ以降、パワー比又はエネルギー比を総称して、「パワー比」と称する。

20

【0023】

パワー比加算部5は、内部にRAM等の記憶部を有し、ホルマント周波数検出部4で算出された1フレーム分のパワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ を、例えば、数十秒間記憶部に保存し、例えば、数十秒間経過した後、パワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ を累積加算する。

【0024】

スピーカ構成決定部6は、パワー比加算部5で算出されたパワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ の累積加算値に応じて、例えば、以下に示すようにスピーカ構成を決定する。

30

(1) パワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ の累積加算値が、ソースがホルマント成分のみであり、ホルマント成分を除いた成分がないことを示す場合(例えば、日本語の場合、50%以上)は、ソースがニュース、対談又は討論等の音声の主であるコンテンツであると判断し、スピーカ構成をCチャンネルスピーカ10の1台と決定する。

【0025】

このように、ソースがニュース、対談又は討論等の音声の主であるコンテンツである場合にスピーカ構成をCチャンネルスピーカ10の1台と決定するのは以下に示す理由による。即ち、音声をCチャンネルスピーカからモノラルで放射した場合には、聴取者がその聴取位置を変更しても違和感なく同じように聴くことができる。これに対し、音声をLチャンネルスピーカ及びRチャンネルスピーカからステレオで放射した場合には、聴取者がその聴取位置を変更した場合、聴取位置によっては違和感を持つ場合がある。

40

【0026】

(2) パワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ の累積加算値が、ソースがホルマント成分と、少ないホルマント成分を除いた成分とからなることを示す場合は、ソースがドラマ等の音声の主であるがテーマ音楽や主題歌等も随時挿入されるコンテンツであると判断し、スピーカ構成をLチャンネルスピーカ9、Cチャンネルスピーカ10及びRチャンネルスピーカ11の3台と決定する。なお、ソースが映画の場合には、予めシネマモード(例えば、5.1チャンネル)が設定されているので、このシネマモードに基づいてスピーカ構成を決定すれば良い。

50

【 0 0 2 7 】

(3) パワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ の累積加算値が、ソースがホルマント成分と、多くのホルマント成分を除いた成分とからなることを示す場合は、ソースが楽器の演奏をバックに歌手が歌を歌っている等の楽曲と音声とが混在しているコンテンツである判断し、スピーカ構成をLチャンネルスピーカ9及びRチャンネルスピーカ11の2台と決定する。

【 0 0 2 8 】

(4) パワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ の累積加算値が、ソースが多くのホルマント成分を除いた成分のみであり、ホルマント成分がないことを示す場合は、ソースが楽器の演奏のみ等の楽曲が主であるコンテンツであると判断し、スピーカ構成をLチャンネルスピーカ9及びRチャンネルスピーカ11の2台と決定する。

10

【 0 0 2 9 】

(5) パワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ の累積加算値が、ソースが少ないホルマント成分を除いた成分のみであり、ホルマント成分がないことを示す場合は、自然な騒音や台詞と台詞の間、曲間等が主であるコンテンツと判断し、スピーカ構成をLチャンネルスピーカ9、Cチャンネルスピーカ10及びRチャンネルスピーカ11の3台又はLチャンネルスピーカ9及びRチャンネルスピーカ11の2台と決定する。

【 0 0 3 0 】

再生データ生成部7は、スピーカ構成決定部6の決定結果が(1)の場合、周波数・時間変換部3から供給されるLチャンネルの音響音声データ及びRチャンネルの音響音声データからCチャンネルスピーカ10に供給すべきモノラルの音声データを生成し、Lチャンネルの音響音声データ、Rチャンネルの音響音声データ及びモノラルの音声データをアンプ8に供給する。

20

【 0 0 3 1 】

また、再生データ生成部7は、スピーカ構成決定部6の決定結果が(2)の場合、周波数・時間変換部3から供給されるLチャンネルの音響音声データ及びRチャンネルの音響音声データの同一成分からCチャンネルスピーカ10に供給すべき台詞成分の音声データを抽出し、Lチャンネルの音響音声データ、Rチャンネルの音響音声データ及び台詞成分の音声データをアンプ8に供給する。

【 0 0 3 2 】

なお、台詞成分の音声データの抽出手法については、例えば、ドルビープロロジックII(ドルビーは登録商標)等、通常のLチャンネルの音響音声データ及びRチャンネルの音響音声データを3チャンネル以上のサラウンドサウンドとして拡張する手法を用いても良い。

30

【 0 0 3 3 】

さらに、再生データ生成部7は、スピーカ構成決定部6の決定結果が(3)及び(4)の場合、周波数・時間変換部3から供給されるLチャンネルの音響音声データ及びRチャンネルの音響音声データをそのままアンプ8に供給する。

【 0 0 3 4 】

また、再生データ生成部7は、スピーカ構成決定部6の決定結果が(5)の場合、周波数・時間変換部3から供給されるLチャンネルの音響音声データ及びRチャンネルの音響音声データをそのままアンプ8に供給したり、Lチャンネルの音響音声データ及びRチャンネルの音響音声データの同一成分からCチャンネルスピーカ10に供給すべき音響音声データを抽出し、Lチャンネルの音響音声データ、Rチャンネルの音響音声データ及び抽出した音響音声データをアンプ8に供給したりする。

40

【 0 0 3 5 】

アンプ8は、デジタル/アナログ変換器(DAC)、前置増幅器(プリアンプ)、主増幅器(メインアンプ)等からなり、Lチャンネルの音響音声データ、Rチャンネルの音響音声データ、モノラルの音声データあるいは台詞成分等の抽出された音声データのアナログの音響音声信号又は音声信号への変換、音質の調整、音量の調整、電力増幅等を行う。

50

【 0 0 3 6 】

Lチャンネルスピーカ9、Cチャンネルスピーカ10及びRチャンネルスピーカ11は、いずれも全音域を出力可能なスピーカであり、例えば、居間等において聴取者の聴取位置に向かって前方左側、前方面、前方右側にそれぞれ配置され、アンプ8から供給されるアナログの音響音声信号等に基づいて楽音又は音声を放射する。

【 0 0 3 7 】

次に、上記構成の音響音声再生装置の動作について説明する。アンパッキング・逆量子化部1は、外部から供給されたデジタル音響音声データC D A Dから符号化データを展開した後、展開した符号化データを逆量子化するとともに、これらに基づいて、逆量子化したブロックサイズごとの周波数成分のスペクトルデータを算出する。

10

【 0 0 3 8 】

これにより、ブロックサイズ判定部2は、上記展開されたブロックサイズに基づいて、1フレームをロングブロック又はショートブロックのいずれかで、周波数・時間変換部3に周波数・時間変換処理を行わせるかを判定する。次に、周波数・時間変換部3は、上記逆量子化後のスペクトルデータについて、ブロックサイズ判定部2で判定されたブロックサイズごとに、公知のIMDCT処理を施すことにより、Lチャンネルの音響音声データ及びRチャンネルの音響音声データに直交変換する。

【 0 0 3 9 】

一方、ホルマント周波数検出部4は、アンパッキング・逆量子化部1で算出された逆量子化後のスペクトルデータについて、上記式(1)を用いてホルマント周波数 $F_0 \sim F_3$ を算出した後、算出したホルマント周波数 $F_0 \sim F_3$ の値が図2に示すホルマント判定表のいずれかの母音と相関があるか否かを判定し、相関がある場合には母音が存在すると判断し、図2に示すホルマント振幅の重み係数 $L_1 \sim L_3$ を乗算することによりホルマントスペクトルを抽出する。次に、ホルマント周波数検出部4は、抽出したホルマントスペクトルに基づいて、ホルマント成分のパワーとホルマント成分を除いた成分のパワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ を算出する。

20

【 0 0 4 0 】

これにより、パワー比加算部5は、上記1フレーム分のパワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ を数十秒間記憶部に保存し、数十秒間経過した後、パワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ を累積加算する。次に、スピーカ構成決定部6は、上記パワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ の累積加算値に応じて、上記(1)～(5)に示すようにスピーカ構成を決定する。例えば、ソースがニュース、対談又は討論等の音声が主であるコンテンツである場合にはスピーカ構成がCチャンネルスピーカ10の1台と決定され、ソースがドラマ等の音声が主であるがテーマ音楽や主題歌等も随時挿入されるコンテンツである場合には、スピーカ構成がLチャンネルスピーカ9、Cチャンネルスピーカ10及びRチャンネルスピーカ11の3台と決定される。

30

【 0 0 4 1 】

一方、ソースが楽器の演奏をバックに歌手が歌を歌っている等の楽曲と音声とが混在しているコンテンツである場合には、スピーカ構成がLチャンネルスピーカ9及びRチャンネルスピーカ11の2台と決定され、ソースが楽器の演奏のみ等の楽曲が主であるコンテンツである場合には、スピーカ構成がLチャンネルスピーカ9及びRチャンネルスピーカ11の2台と決定される。さらに、ソースが自然な騒音や台詞と台詞の間、曲間等が主であるコンテンツである場合には、スピーカ構成がLチャンネルスピーカ9、Cチャンネルスピーカ10及びRチャンネルスピーカ11の3台又はLチャンネルスピーカ9及びRチャンネルスピーカ11の2台と決定される。

40

【 0 0 4 2 】

これにより、再生データ生成部7は、スピーカ構成決定部6の決定結果が(1)の場合、上記Lチャンネルの音響音声データ及びRチャンネルの音響音声データからCチャンネルスピーカ10に供給すべきモノラルの音声データを生成し、Lチャンネルの音響音声データ、Rチャンネルの音響音声データ及びモノラルの音声データをアンプ8に供給する。

50

一方、スピーカ構成決定部 6 の決定結果が (2) の場合、再生データ生成部 7 は、上記 L チャンネルの音響音声データ及び R チャンネルの音響音声データの同一成分から C チャンネルスピーカ 1 0 に供給すべき台詞成分の音声データを抽出し、L チャンネルの音響音声データ、R チャンネルの音響音声データ及び台詞成分の音声データをアンプ 8 に供給する。

【 0 0 4 3 】

また、再生データ生成部 7 は、スピーカ構成決定部 6 の決定結果が (3) 及び (4) の場合、上記 L チャンネルの音響音声データ及び R チャンネルの音響音声データをそのままアンプ 8 に供給し、スピーカ構成決定部 6 の決定結果が (5) の場合、周波数・時間変換部 3 から供給される L チャンネルの音響音声データ及び R チャンネルの音響音声データをそのままアンプ 8 に供給したり、L チャンネルの音響音声データ及び R チャンネルの音響音声データの同一成分から C チャンネルスピーカ 1 0 に供給すべき音響音声データを抽出し、L チャンネルの音響音声データ、R チャンネルの音響音声データ及び抽出した音響音声データをアンプ 8 に供給したりする。

【 0 0 4 4 】

これにより、アンプ 8 は、上記 L チャンネルの音響音声データ、上記 R チャンネルの音響音声データ、上記モノラルの音声データあるいは上記台詞成分等の抽出された音声データをアナログの音響音声信号又は音声信号へ変換した後、聴取者の操作に応じて、音質の調整、音量の調整、電力増幅等を行い、L チャンネルスピーカ 9、C チャンネルスピーカ 1 0 又は R チャンネルスピーカ 1 1 に供給する。従って、スピーカ構成決定部 6 の決定結果が (1) の場合には、例えば、C チャンネルスピーカ 1 0 からニュースを読むアナウンサーの音声が発射される。また、スピーカ構成決定部 6 の決定結果が (2) の場合には、例えば、L チャンネルスピーカ 9 及び R チャンネルスピーカ 1 1 からドラマで使用されている楽音が放射されるとともに、C チャンネルスピーカ 1 0 から俳優の音声が発射される。

【 0 0 4 5 】

また、スピーカ構成決定部 6 の決定結果が (3) の場合には、例えば、L チャンネルスピーカ 9 及び R チャンネルスピーカ 1 1 から楽器の演奏をバックに歌手が歌を歌っている等の楽曲と音声とが発射される。さらに、スピーカ構成決定部 6 の決定結果が (4) の場合には、例えば、L チャンネルスピーカ 9 及び R チャンネルスピーカ 1 1 から楽器の演奏の楽音が放射される。また、スピーカ構成決定部 6 の決定結果が (5) の場合には、L チャンネルスピーカ 9 及び R チャンネルスピーカ 1 1 又は、L チャンネルスピーカ 9、C チャンネルスピーカ 1 0 及び R チャンネルスピーカ 1 1 から自然な騒音等が発射される。

【 0 0 4 6 】

このように、本発明の実施の形態 1 によれば、圧縮された符号化データからホルマント成分を抽出し、このホルマント成分とホルマント成分を除いた成分とからソースを自動的に判断し、この判断結果に基づいてスピーカ構成を決定している。従って、ソースがニュース、対談又は討論等の音声の主であるコンテンツである場合や、ドラマ等の音声の主であるがテーマ音楽や主題歌等も随時挿入されるコンテンツである場合には、聴取者の聴取範囲を広げることができる。一方、ソースが楽器の演奏をバックに歌手が歌を歌っている等の楽曲と音声とが混在しているコンテンツである場合や、楽器の演奏のみ等の楽曲が主であるコンテンツである場合には、当該コンテンツの制作者 (例えば、ミキシングエンジニア) の意図通りにステレオで再生することができる。

【 0 0 4 7 】

また、本発明の実施の形態 1 によれば、デジタルの I M D C T 係数 (スペクトルデータ) からホルマント成分を抽出しているため、上記した従来の音響音声再生装置のように、アナログの音響音声信号を、アナログ / デジタル変換及び高速フーリエ変換 (F F T : Fast Fourier Transform) した後にホルマント成分を抽出する場合と比べて、装置を安価に構成できるとともに、消費電力を大幅に削減することができる。また、A D C が不要であるため、量子化雑音に起因する音質劣化を防止することができる。

【 0 0 4 8 】

実施の形態 2 .

上述の実施の形態 1 においては、ホルマント周波数検出部 4 によりホルマントを分析し、パワー比加算部 5 によりパワー比 $P_{ow}(F_{sum}) / P_{ow}(N_{sum})$ を累積加算する例を示したが、これに限定されない。例えば、ホルマント周波数検出部 4 及びパワー比加算部 5 に換えて、一定時間のフレームごとに音響音声信号の音響パワーを算出するパワー算出部と、算出された音響パワー値を予め設定したしきい値と比較してそのフレームの有音無音を判定する有音無音判定部と、フレームごとに音響信号の波形の零交差回数を算出する零交差算出部と、算出された零交差回数を予め設定したしきい値と比較してそのフレームの子音性を判定する子音性判定部と、連続する所定複数フレーム区間におけるパワー値の最大値と最小値を検出し、その差分値を算出する定常性判定部と、複数フレームにおいて無音と判定されたフレームの存在比率と、子音性が高いと判定されたフレームの存在比率と、差分値とがそれぞれに予め設定したしきい値よりもすべて大きい場合にその複数フレーム区間における音響信号は音声と判定し、音声と判定されない場合であって、複数フレームにおいて無音と判定されたフレームの存在比率と、差分値とがそれぞれに予めしきい値以下に設定されたしきい値よりも小さい場合にはその複数フレーム区間の音響信号は非音声と判定し、それ以外を不定と判定し、判定結果を複数フレームごとに出力する音声判定部とを設けても良い。そして、スピーカ構成決定部 6 は、音声判定部の判定結果に基づいて、スピーカ構成を決定しても良い。この構成によれば、上記した実施の形態 1 と略同様の効果が得られる。なお、上記パワー算出部、上記有音無音判定部、上記零交差算出部、上記子音性判定部、上記定常性判定部及び上記音声判定部については、例えば、特許第 2 8 3 5 4 8 3 号公報を参照されたい。

10

20

【 0 0 4 9 】

実施の形態 3 .

上述の実施の形態 1 においては、ホルマント周波数検出部 4、パワー比加算部 5、スピーカ構成決定部 6 及び再生データ生成部 7 が処理を行うタイミングについては特に言及していないが、そのタイミングは任意で良い。例えば、ソースのコンテンツが切り替わった直後にスピーカ構成決定部 6 が上記 (1) ~ (5) の判断を行っても良いし、所定時間経過後に上記判断を行っても良い。また、今まで (3) ~ (5) の判断を行っていたが、コンテンツが切り替わることにより (1) 及び (2) の判断を行った場合でも、直ちに C チャンネルスピーカ 1 0 にモノラルの音声データ又は台詞成分の音声データのすべてを供給するように制御するのではなく、C チャンネルスピーカ 1 0 に供給する音声データの割合を時間の経過とともに増加させるように構成しても良い。逆に、今まで (1) 又は (2) の判断を行っていたが、コンテンツが切り替わることにより (3) ~ (5) の判断を行った場合でも、直ちに C チャンネルスピーカ 1 0 へのモノラルの音声データ又は台詞成分の音声データの供給を中止するように制御するのではなく、C チャンネルスピーカ 1 0 に供給する音声データの割合を時間の経過とともに減少させるように構成しても良い。このように構成すれば、聴取者に違和感を与えることが少ない。

30

【 0 0 5 0 】

実施の形態 4 .

上述の実施の形態 1 においては、音響音声再生装置をハードウェアで構成した例を示したが、これに限定されない。即ち、上記音響音声再生装置のうち、アンパッキング・逆量子化部 1、ブロックサイズ判定部 2、周波数・時間変換部 3、ホルマント周波数検出部 4、パワー比加算部 5、スピーカ構成決定部 6、再生データ生成部 7 を、CPU (中央処理装置) と、ROM や RAM 等の内部記憶装置と、FD ドライブ、HD ドライブ、MO ディスクドライブ、CD / DVD ドライブ等の外部記憶装置と、出力手段と、入力手段とを有するコンピュータによって構成しても良い。そして、CPU の機能が音響音声再生プログラムとして、ROM 等の半導体メモリや、FD、HD や CD - ROM 等の記憶媒体に記憶されていると構成しても良い。この場合、音響音声再生プログラムは、記憶媒体から CPU に各々読み込まれ、CPU の動作を制御する。音響音声再生プログラムが起動されると

40

50

、CPUは上記アンパッキング・逆量子化部1、上記ブロックサイズ判定部2、上記周波数・時間変換部3、上記ホルマント周波数検出部4、上記パワー比加算部5、上記スピーカ構成決定部6、上記再生データ生成部7として機能し、音響音声再生プログラムの制御により、上記した処理を実行するのである。

【0051】

以上、本発明の実施の形態について図面を参照して詳述してきたが、具体的な構成はこれらの実施の形態に限られるものではなく、本発明の要旨を逸脱しない範囲の設計の変更等があっても本発明に含まれる。

例えば、上述の各実施の形態では、本発明は、AAC方式が適用された音響音声再生装置に適用する例を示したが、これに限定されない。本発明は、例えば、MP3 (MPEG Audio Layer-3) 方式、AC-3 (Audio Code number 3) 方式、あるいはWMA (Windows Media Audio) (Windowsは登録商標) 方式等の音声圧縮符号化技術が適用された音響音声再生装置にも適用することができる。

10

また、上述の各実施の形態では、符号化デジタルデータを生成するための変換方法としては、MDCTを挙げたが、DCTといった直交変換する変換方法であれば本発明が適用可能である。

【0052】

また、上述の各実施の形態では、それぞれLチャンネルスピーカ9、Cチャンネルスピーカ10及びRチャンネルスピーカ11を聴取位置に向かって前方左側、前方正面及び前方右側にそれぞれ配置する例を示したが、これに限定されない。例えば、上記3台のスピーカに加えて、低音出力用のサブウーファースピーカーを聴取位置に向かって前方正面に配置しても良い。また、上記4台のスピーカに加えて、それぞれLチャンネルスピーカ及びRチャンネルスピーカを聴取位置に向かって後方左側及び後方右側にそれぞれ配置しても良い。

20

【0053】

また、上述の各実施の形態では、本発明を地上デジタルテレビジョン放送等のテレビジョン信号を構成する圧縮されたデジタル音響音声データを復号化する場合に適用する例を示したが、これに限定されない。本発明は、例えば、ミニディスク(MD)、コンパクトディスク(CD)、DVD (Digital Versatile Disk)、ハードディスク(HD)、あるいはフラッシュメモリなどの半導体メモリ等の記録媒体に記録された圧縮されたデジタル音響音声データを復号化する場合にも適用することができる。

30

【図面の簡単な説明】

【0054】

【図1】本発明の実施の形態1に係る音響音声再生装置の構成を示すブロック図である。

【図2】ホルマント判定表の一例を示す図である。

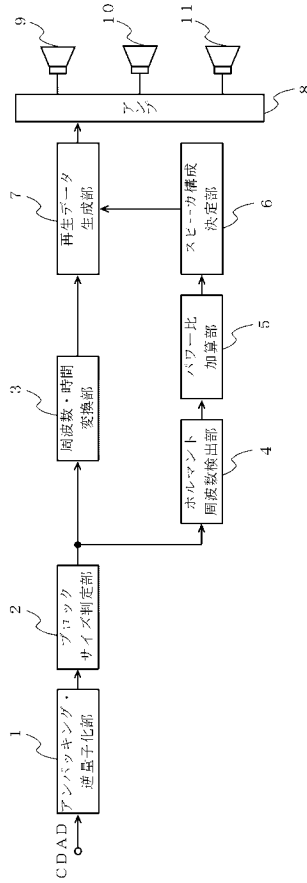
【符号の説明】

【0055】

- 1 アンパッキング・逆量子化部
- 2 ブロックサイズ判定部
- 3 周波数・時間変換部
- 4 ホルマント周波数検出部
- 5 パワー比加算部 (パワー比又はエネルギー比加算部)
- 6 スピーカ構成決定部
- 7 再生データ生成部
- 8 アンブ
- 9 Lチャンネルスピーカ (左チャンネルのスピーカ)
- 10 Cチャンネルスピーカ (センターチャンネルのスピーカ)
- 11 Rチャンネルスピーカ (右チャンネルのスピーカ)

40

【図1】



【図2】

		i	ɪ	e	æ	a	ɔ	u	ɯ	ʌ	ɔ̃
基本周波数 F ₀ (Hz)	M	136	135	130	127	124	128	137	141	130	133
	W	235	232	223	210	212	216	232	231	221	218
	Ch	272	269	260	251	256	263	276	274	261	261
ホルマント 周波数F ₁ (Hz)	M	270	390	530	660	730	570	440	300	840	490
	W	310	430	610	880	850	590	470	370	790	500
	Ch	370	530	690	1010	1030	680	560	430	850	560
ホルマント 周波数F ₂ (Hz)	M	2290	1950	1840	1720	1090	840	1020	870	1190	1350
	W	2790	2450	2330	2060	1220	920	1180	850	1400	1640
	Ch	3200	2750	2610	2320	1370	1060	1410	1170	1590	1820
ホルマント 周波数F ₃ (Hz)	M	3010	2590	2480	2410	2440	2410	2240	2240	2390	1690
	W	3310	3070	2980	2850	2610	2710	2680	2670	2780	1860
	Ch	3730	3600	3570	3320	3170	3180	3310	3260	2360	2160
ホルマント 振幅 (dB)	L ₁	-4	-3	-2	-1	-1	0	-1	-3	-1	-5
	L ₂	-24	-23	-17	-12	-5	-7	-12	-19	-10	-15
	L ₃	-28	-27	-24	-22	-28	-34	-34	-43	-27	-20

(M: 男性, W: 女性, Ch: 子供)

フロントページの続き

(51)Int.Cl. F I
G 1 0 L 11/00 1 0 1 F
H 0 4 S 5/02 X

(56)参考文献 特開平04 - 249484 (JP, A)
特開2006 - 254187 (JP, A)
特開平07 - 115606 (JP, A)
特開2004 - 118108 (JP, A)
特開平10 - 207491 (JP, A)
特開平03 - 285500 (JP, A)
実開平05 - 060100 (JP, U)
特開平03 - 236691 (JP, A)
特開2002 - 095095 (JP, A)
特開平06 - 253398 (JP, A)

(58)調査した分野(Int.Cl., DB名)

H 0 4 S 1 / 0 0 - 7 / 0 0
G 1 0 L 1 1 / 0 0 - 1 3 / 0 8
G 1 0 L 1 9 / 0 0 - 2 1 / 0 6