

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-241850

(P2008-241850A)

(43) 公開日 平成20年10月9日(2008.10.9)

(51) Int.Cl.

F I

テーマコード (参考)

G 1 O L 11/00 (2006.01)

G 1 O L 11/00 4 O 2 G

G 1 O L 19/00 (2006.01)

G 1 O L 19/00 3 1 2 Z

審査請求 未請求 請求項の数 4 O L (全 11 頁)

(21) 出願番号 特願2007-78956 (P2007-78956)
 (22) 出願日 平成19年3月26日 (2007. 3. 26)

(71) 出願人 000001889
 三洋電機株式会社
 大阪府守口市京阪本通2丁目5番5号
 (74) 代理人 100131071
 弁理士 ▲角▼谷 浩
 (72) 発明者 松本 悟
 大阪府守口市京阪本通2丁目5番5号 三
 洋電機株式会社内
 (72) 発明者 山本 友二
 大阪府守口市京阪本通2丁目5番5号 三
 洋電機株式会社内
 (72) 発明者 古賀 達雄
 大阪府守口市京阪本通2丁目5番5号 三
 洋電機株式会社内

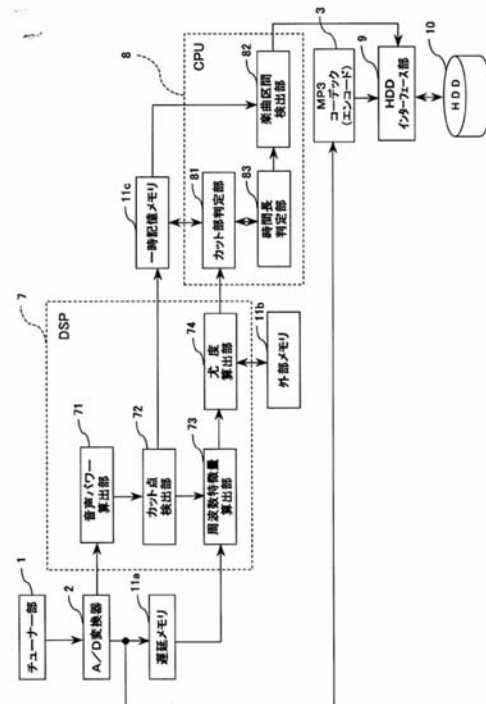
(54) 【発明の名称】 録音または再生装置

(57) 【要約】

【課題】軽い演算処理により、楽曲と話し声から成る音声から楽曲を識別することができる録音または再生装置を提供する。

【解決手段】 音声信号レベルの変化量が所定値以上である時点のカット点として検出するカット点検出部と、カット点近傍においてのみ前記音声信号の周波数領域における特徴量を算出する周波数特徴量算出部と、算出された周波数特徴量に基づき、カット点の属性を判定するカット点判定部と、前記属性とサンプリング点との間隔に基づき、楽曲区間の始点と終点とを検出する楽曲区間検出部とを備える。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

音声信号中の楽曲区間を検出する録音または再生装置であって、
音声信号レベル又はその変化量が所定値以上である時点をカット点として検出するカット点検出部と、
カット点近傍においてのみ前記音声信号の周波数領域における特徴量を算出する周波数特徴量算出部と、
算出された周波数特徴量に基づき、カット点の属性を判定するカット点判定部と、
前記属性とサンプリング点との間隔に基づき、楽曲区間の始点と終点とを検出する楽曲区間検出部を備えることを特徴とする、録音または再生装置。

10

【請求項 2】

前記カット点判定部は、算出された周波数特徴量に基づいてカット点近傍の音声信号が楽曲部分のものか、非楽曲部分のものを判別し、
隣り合う 2 つの非楽曲部分の間隔が所定時間以上であるとき、これらの非楽曲部分の間にある音声信号を楽曲区間として推定することを特徴とする、請求項 1 に記載の録音または再生装置。

【請求項 3】

前記カット点判定部は、算出された周波数特徴量に基づいてカット点近傍の音声信号が楽曲部分のものか、非楽曲部分のものを判別し、
隣り合う 2 つの非楽曲部分のカット点の間隔が所定時間以上であるとき、これらの非楽曲部分のカット点の間にある音声信号を楽曲区間として推定することを特徴とする、請求項 1 に記載の録音または再生装置。

20

【請求項 4】

音声信号中の楽曲区間を検出する録音または再生装置であって、
音声信号レベル又はその変化量が所定値以上である時点をカット点として検出するカット点検出部と、
前記音声信号の周波数領域における特徴量を算出する周波数特徴量算出部と、
算出された周波数特徴量と、検出されたカット点の情報に基づいて楽曲区間の始点と終点を検出する楽曲区間検出部を備えることを特徴とする、録音または再生装置。

30

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、話し声 (s p e a c h) の区間や音楽 (m u s i c) の区間が混在する音声 (a u d i o) の中から音楽 (楽曲) の区間を検出する録音または再生装置に関する。

【背景技術】**【0002】**

通常の放送される音声にはアナウンサーの話し声の区間や音楽の区間が混在している。

40

【0003】

使用者が希望する音楽の録音しようとする場合、音楽の開始タイミングに合わせて録音を手動で開始し、音楽の終了タイミングに合わせて録音を手動で終了するという厄介な操作を行わねばならなかった。しかし、ラジオ等の放送を聴いているときに、好みの音楽を録音しようとしても通常は音楽の開始タイミングまで遡って確実に録音することはできない。

【0004】

そこで、放送を録音しておき、好みの音楽だけを後で編集作業によって切出すことが有効になる。しかし、少なくとも編集に際して、できるだけ音楽区間だけを識別して再生する方が編集作業も容易となる。

50

【 0 0 0 5 】

そこで、音楽が持つ特徴と話し声が持つ特徴とを分析し、両者を自動的に識別する技術が必要となる。

【 0 0 0 6 】

特許文献 1 によれば、M F C C (メルケール周波数ケプストラム計数)などの周波数特徴量を用いて、音楽と話し声の識別を行っている。

【特許文献 1】特開 2 0 0 4 - 2 5 8 5 6 9 号公報

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 7 】

10

しかし、特許文献 1 の技術は、音声信号の周波数領域の特徴量を算出する処理が複雑で大掛かりになり、処理のための負荷が大きくなるという問題点がある。

【 0 0 0 8 】

本発明は、斯かる問題に鑑みてなされたものであり、軽い演算処理により、楽曲と話し声から成る音声から楽曲を識別することができる録音または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】

【 0 0 0 9 】

本願発明に係る録音または再生装置のある態様は、音声信号中の楽曲区間を検出する録音または再生装置であって、音声信号レベルの変化量が所定値以上である時点をカット点として検出するカット点検出部と、カット点近傍においてのみ前記音声信号の周波数領域における特徴量を算出する周波数特徴量算出部と、算出された周波数特徴量に基づき、カット点の属性を判定するカット点判定部と、前記属性とサンプリング点との間隔に基づき、楽曲区間の始点と終点とを検出する楽曲区間検出部とを備えることを特徴とする。

20

【 0 0 1 0 】

好ましくは、前記カット点判定部は、算出された周波数特徴量に基づいてカット点近傍の音声信号が楽曲部分のものか、非楽曲部分のものを判別し、隣り合う 2 つの非楽曲部分の間隔が所定時間以上であるとき、これらの非楽曲部分の間にある音声信号を楽曲区間として推定する。

【 0 0 1 1 】

30

或いは、隣り合う 2 つの非楽曲部分のカット点の間隔が所定時間以上であるとき、これらの非楽曲部分のカット点の間にある音声信号を楽曲区間として推定するものでもよい。

【 0 0 1 2 】

本願発明に係る録音または再生装置の他の態様は、音声信号中の楽曲区間を検出する録音または再生装置であって、音声信号レベル又はその変化量が所定値以上である時点をカット点として検出するカット点検出部と、前記音声信号の周波数領域における特徴量を算出する周波数特徴量算出部と、算出された周波数特徴量と、検出されたカット点の情報に基づいて楽曲区間の始点と終点とを検出する楽曲区間検出部を備えることを特徴とする。

【発明の効果】

【 0 0 1 3 】

40

本発明の録音または再生装置によれば、軽い演算処理により、楽曲と話し声から成る音声から楽曲を識別することができる。

【発明を実施するための最良の形態】

【 0 0 1 4 】

以下、本発明をその実施の形態を示す図面に基づいて説明する。図 1 は、本発明の実施の形態に係る録音または再生装置中の楽曲検出機能を説明するための構成図である。図 1 に示すように、本実施の形態に係る録音または再生装置は、受信したテレビジョン放送、ラジオ放送等の放送信号を、選局して受信して音声信号に復調する。A / D 変換器 2 は、チューナ部 1 により選択されたアナログの音声信号をデジタル信号に変換する。

【 0 0 1 5 】

50

MP3 (MPEG Audio Layer-3) コーデック 3 は、デジタル音声データを符号化し、データ圧縮した符号化データを生成して時刻情報とペアで出力するエンコーダ機能と、符号化データを復号するデコーダ機能とを有する。D/A 変換器 4 は、MP3 コーデック 3 にて復号されたデジタル音声データをアナログ信号データに変換する。そして、このアナログ信号データは、図時省略したアンプを介してスピーカ 5 に入力される。

【0016】

DSP 7 は、音声信号に基づいて、音声信号レベルを検出するため音声信号の振幅値を二乗した音声パワーを算出する。また、DSP 7 は、音声信号レベルの変化量を検出するため音声パワーの変化量を算出する。また、DSP 7 は、音声パワーの変化量が所定値以上となるタイミングをカット点と定義して検出する。また、DSP 7 は、カット点近傍においてのみの周波数領域の特徴量、例えば MFCC を算出し、サンプル音声信号に基づいて算出された MFCC との尤度を算出する。

10

【0017】

CPU 8 は、バス 6 を介して本実施の形態に係る録音または再生装置全体の動作を制御する。また、カット点が楽曲の始点または終点であるか否かについて推定処理等を実行する。HDD 10 は、例えば ATA インタフェースを実現する HDD インタフェース部 9 を介して符号化データと時刻情報とを記録する大容量記録装置である。メモリ 11 は、実行プログラムを展開し、演算処理により生じるデータを一時記憶すると共に、AD 変換直後の音声データを一定期間遅延する機能を持っている。なお、MP3 コーデック 3、DSP 7、CPU 8、HDD インタフェース部 9 及びメモリ 11 は、バス 6 を介して各種のデータを互いに授受する。

20

【0018】

図 2 は、本実施の形態に係る録音または再生装置の一部を示す機能ブロック図である。図 2 に示すように、本実施の形態に係る録音または再生装置は、チューナ 1 で選局された音声信号を、A/D 変換器 2 に入力してデジタルに変換した後、時刻情報とともに MP3 コーデック 3 に入力し MP3 データに入力し MP3 データに圧縮符号化して時刻情報とペアで録音期間中、HDD インタフェース部 9 を介して HDD 10 に継続的に記録する。

【0019】

A/D 変換器 2 からのデジタル音声データを DSP 7 の処理に要する時間分だけ遅延する遅延メモリ 11a に記憶するとともに、DSP 7 内の音声パワー算出部 71 にて音声信号レベルに相当する音声パワー即ち、音声信号の振幅の二乗値を算出する。

30

【0020】

DSP 7 内のカット点検出部 72 では、音声信号レベルの変化が大きいタイミング、即ち音声パワー値の変化量が所定値より大きいタイミングをカット点として検出して、検出出力を発する。同時にカット点における時刻情報とその変化量を一時記憶メモリ 11c に記憶する。

【0021】

図 3 は、カット点検出部 72 の動作を説明するための波形図である。図 3 (a) は、音声パワー値の変化を、図 3 (b) は変化量 (微分値) の変化を、それぞれ示している。図 3 に示すように、カット点検出部 72 は音声パワー算出部 71 にて算出された音声パワー値に基づいて、その微分値が所定の閾値より大きい極大点となる時刻 T_m 、 $T_m + 1$ をカット点として検出する。そして、その検出結果を、周波数特徴量算出部 73 に入力する。

40

【0022】

周波数特徴量算出部 73 では、遅延メモリ 11a から所定時間遅延されて出力される音声データを、カット点検出部 72 からの出力に同期して、カット点に僅か先行するタイミングからカット点より僅か遅れたタイミングにおける微小期間において一時的に MFCC 等の周波数の特徴量を算出して尤度算出部 74 に入力する。

【0023】

本実施例では、楽曲と話し声とは周波数特徴量が異なることに着目して、典型的な楽曲の周波数特徴量と話し声の周波数特徴量を、比較に際して基準データとして外部メモリ 1

50

1 b に予め記憶している。従って、DSP 内の尤度検出部 7 4 は、周波数特徴量算出部 7 3 から入力されるカット点前後の特徴量算出出力と、基準データとの尤度を算出し、その尤度算出出力を、CPU 8 内のカット点判定部 8 1 に入力する。

【0024】

なお、尤度算出に際しては、基準データと比較により尤度を求める前述の方式に代えて、周波数特徴量をあらかじめ設定した評価関数に代入して尤度を求めることにより、楽曲である可能性を算出する方法もあり、必ずしも基準データと比較する必要はない。

【0025】

次に、カット点判定部 8 1 は、尤度算出出力に基づいてカット点における音声信号が音楽か話し声かを判定する。カット点検出部 7 2 より得た時刻情報と変化量が記憶されている一時記憶メモリ 1 1 C には、更に判定結果が関連付けて記憶される。

10

【0026】

図 4 は、上記のような判定結果が関連付けて記憶されている一時記憶メモリ 1 1 C のテーブルを示している。

【0027】

時間長判定部 8 3 は、楽曲が所定時間たとえば 100 秒以上継続することを経験的に見出しており、話し声のサンプリング点間隔が 100 秒未満の場合、その間のサンプリング点が楽曲と判定された場合でも、当該区間は楽曲とはみなさないことにしており、話し声即ち楽曲以外と判定されたサンプリング点間の間隔を測定し、100 秒以上の区間を楽曲区間として判定している。

20

【0028】

時間長判定部 8 3 から得られる判定出力を入力する楽曲区間検出部 8 2 は、一時記憶メモリ 1 1 C のテーブルを書換え、楽曲毎のテーブル（最終テーブル）に変更する。

【0029】

図 5 は、一時記憶メモリ 1 1 C で書き換えられた最終のテーブルを示す図である。この最終のテーブルでは、T 6 が一旦楽曲と判定されたが話し声と判定された前後のサンプリング点 T 5 と T 7 の間隔が短いことを理由に、楽曲とは看做されなかったので、テーブルから削除されたということが示されている。

【0030】

この最終のテーブルは、録音動作が終了した時点で、楽曲区間検出部 8 2 を経由して、HDD インタフェース部 9 へ供給され、更に HDD 10 に記憶される。

30

【0031】

尚、最終テーブルは、楽曲の始点と終了点以外に、中間のカット点や、変化量を残したまま、HDD 10 に記録されているが、再生に際してサビ部分を再生する目的で利用される。

【0032】

HDD 10 に記録されている符号化データは、編集再生操作に応じて最終テーブルに特定された楽曲区間にのみ対応する符号化データを順次読み出し、MP3 コーデック 3 に入力される。MP3 コーデック 3 は、符号化データを復号し、D/A 変換器 4 にて音声信号に変換された後、スピーカ 5 から出力される。これにより、会話等を含んだ音声信号から、楽曲だけを検出し、楽曲を抽出して再生することができる。

40

【0033】

以上の実施形態によれば、カット点の情報と、周波数特徴量の両方の情報を用いて楽曲区間の検出を行うから、楽曲を正確に検出することができる。

【0034】

更に、カット点近傍においてのみ音声信号の周波数領域における特徴量を算出して、楽曲区間の検出を行うから、軽い演算処理により楽曲区間を検出することもできる。

【0035】

なお、本実施の形態では、DSP 7 と CPU 8 とで機能分散しているが、特にこれに限定されるものではなく、CPU 8 のみで両機能を実現しても良いし、A/D 変換器 2、M

50

P 3 コーデック 3、及び D / A 変換器 4 の機能も含めて、全て C P U 8 によりソフトウェア処理する構成であっても良い。また、本実施例では、遅延メモリ 1 1 a、外部メモリ 1 1 b、一時記憶メモリ 1 1 c を個別に表したが、これらのメモリは図 1 中のメモリ 1 1 内に形成されるものである。

【 0 0 3 6 】

本実施例では、録音期間中に楽曲区間を検出して最終テーブルを作成して記録したが、一旦録音期間が終了した空き時間に H D D より録音したデジタル音声データを順次再生しながら楽曲区間を検出して最終テーブルを作成してもよい。また、再生操作に連動して本実施例の動作をすべて実行させるような回路構成にしてもよく、これらの構成が本発明に含まれることは言うまでもない。

10

【 0 0 3 7 】

また、本実施例では、音声信号のレベルを音声信号の振幅を二乗値として検出したが、振幅の絶対値に置き換えても等価である。

【 0 0 3 8 】

更に、本実施例のカット点は、音声信号レベルが大きく変化するタイミングであり、楽曲の正確な開始点や終了点に対応するものではないが、楽曲の再生開始点と終了点として利用するには十分である。

【 0 0 3 9 】

本実施例は、録音後の編集に際して、録音した全ての楽曲の一部を再生して、操作者が要否を判断し必要な楽曲のみを後で改めてライブラリとして残という利用方法として有益な構成であり、高い精度等を問わない使い方を追求するものである。

20

【 0 0 4 0 】

(変形例)

以下の手順により楽曲区間を検出してもよい。

(1) まず、音声信号の周波数特徴量を算出し、算出した周波数特徴量から楽曲に対する尤度を算出する。

(2) 次に、尤度の値が所定値を超える時点を楽曲区間の仮の開始点とし、尤度の値が所定値を下回る時点を楽曲区間の仮の終了点であると判断する。

(3) 次に、仮の開始点の近傍にあるカット点を楽曲区間の真の開始点であると、前記仮の終了点の近傍にあるカット点を楽曲区間の真の終了点であると判断する。

30

(4) そして、真の開始点から前記真の終了点までを楽曲区間として推定する。

【 0 0 4 1 】

上記変形例のような検出によれば、周波数特徴量をだけを使って楽曲区間を検出する特許文献 1 の技術と比較して、楽曲区間の検出精度が向上する。

【 0 0 4 2 】

また、本発明は上記実施例に限定されるものではなく、特許請求の範囲内の記載であれば多種の変形、置換等が可能であることは言うまでもない。

【 図面の簡単な説明 】

【 0 0 4 3 】

【 図 1 】 本発明の実施の形態に係る録音または再生装置中の楽曲検出機能を説明するための構成図である。

40

【 図 2 】 本実施の形態に係る録音または再生装置の一部を示す機能ブロック図である。

【 図 3 】 カット点検出部 7 2 の動作を説明するための波形図である。

【 図 4 】 一時記憶メモリ 1 1 c に記憶されているテーブルを示している。

【 図 5 】 一時記憶メモリ 1 1 c で書き換えられた最終のテーブルを示す図である。

【 符号の説明 】

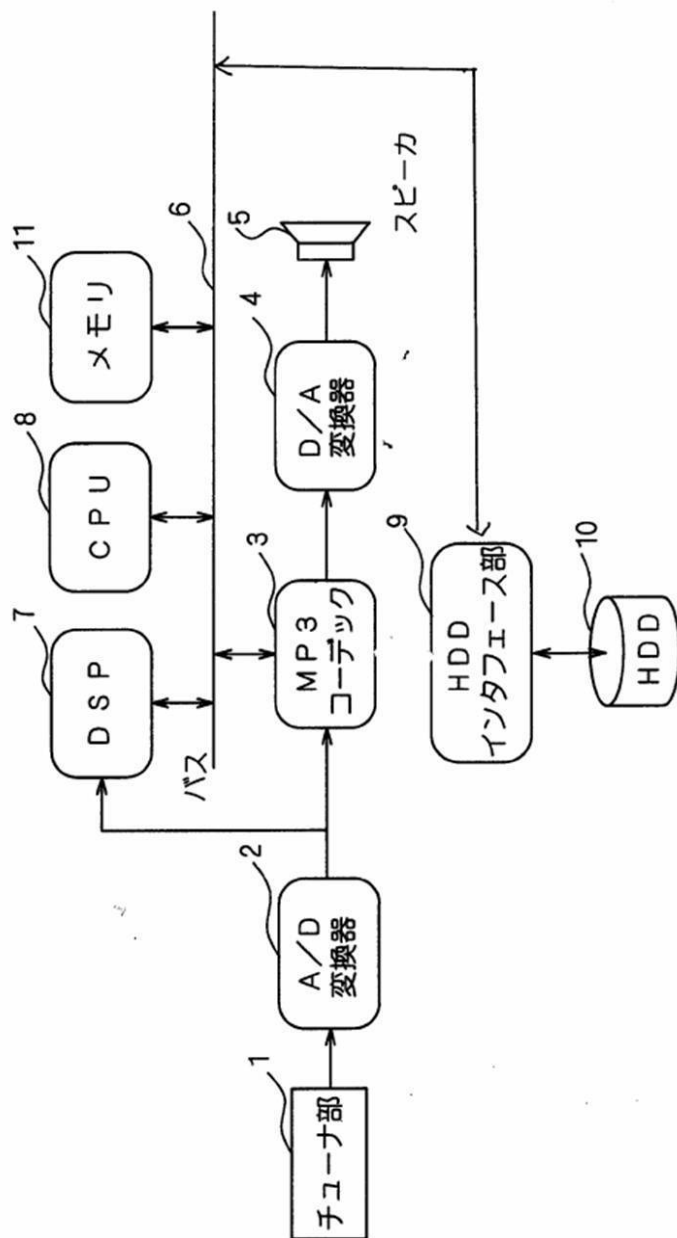
【 0 0 4 4 】

- 1 チューナ部
- 2 A / D 変換器
- 3 M P 3 コーデック

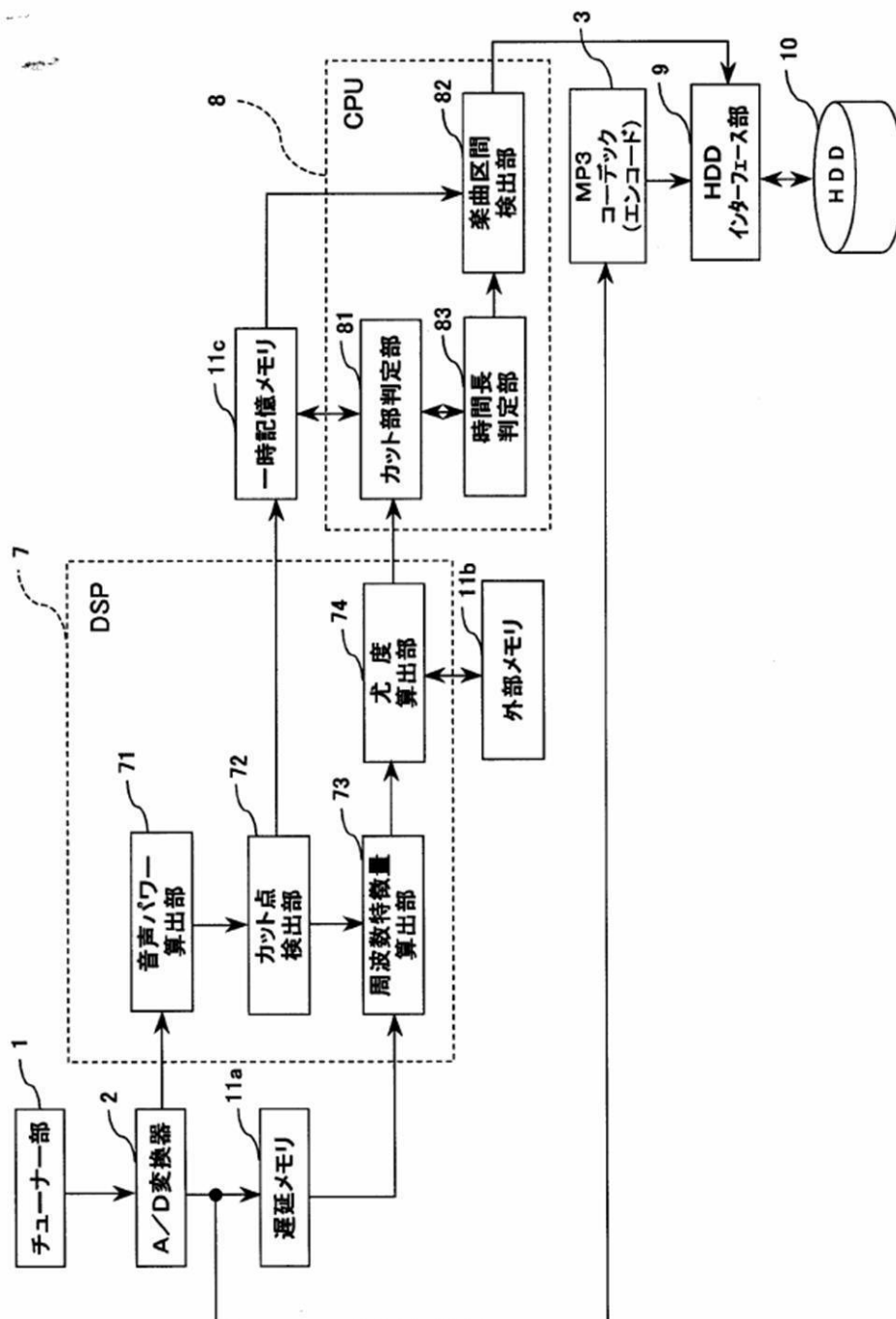
50

- 4 D / A 変換器
- 5 スピーカ
- 6 バス
- 7 D S P
- 8 C P U
- 9 H D D インタフェース部
- 1 0 H D D
- 1 1 メモリ

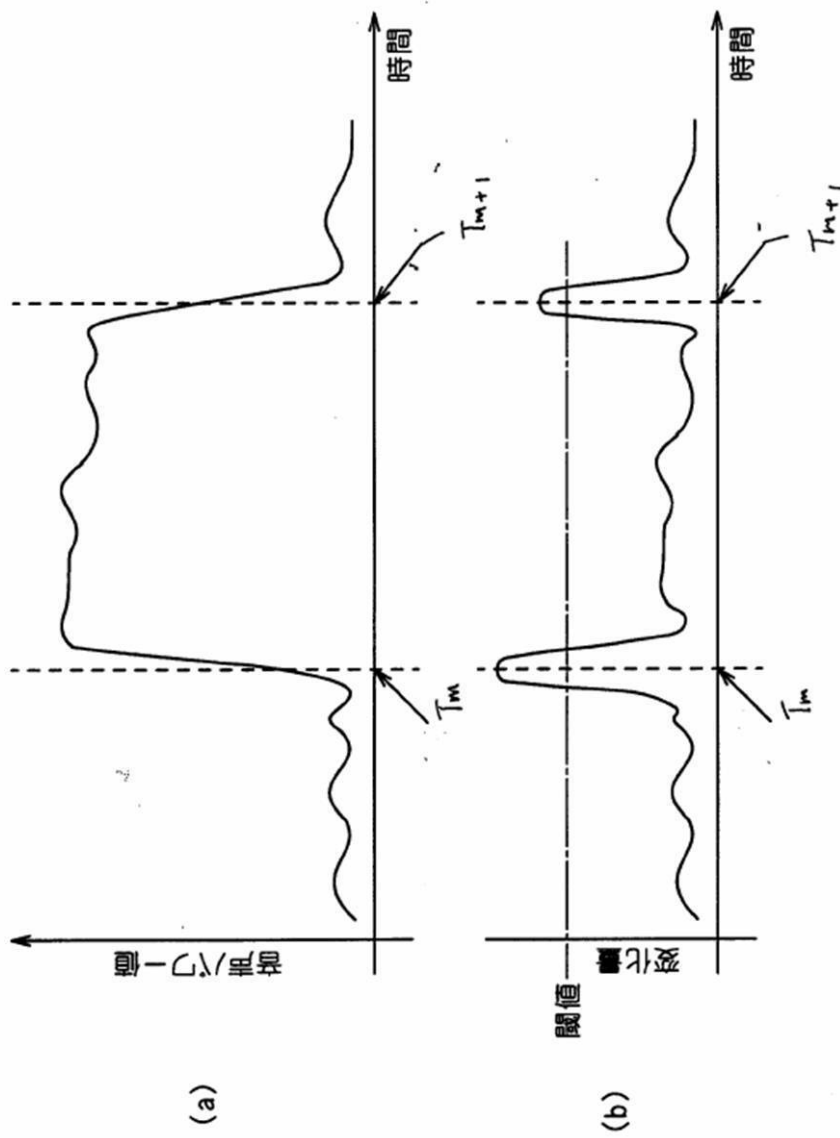
【図 1】



【図 2】



【図 3】



【 図 4 】

時 刻	変 化 量	属 性
T 1	XXXX	楽曲以外
T 2	XXXX	楽 曲
T 3	XXXX	楽 曲
T 4	XXXX	楽 曲
T 5	XXXX	楽曲以外
T 6	XXXX	楽 曲
T 7	XXXX	楽曲以外
T 8	XXXX	楽 曲
T 9	XXXX	楽 曲
T 1 0	XXXX	楽曲以外
T 1 1	XXXX	楽曲以外
⋮	⋮	⋮

【 図 5 】

楽曲 A

時 刻	変 化 量	属 性
T 1	XXXX	楽曲以外
T 2	XXXX	楽 曲
T 3	XXXX	楽 曲
T 4	XXXX	楽 曲
T 5	XXXX	楽曲以外

楽曲 B

時 刻	変 化 量	属 性
T 7	XXXX	楽曲以外
T 8	XXXX	楽 曲
T 9	XXXX	楽 曲
T 1 0	XXXX	楽曲以外