

(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2006-195385

(P2006-195385A)

(43) 公開日 平成18年7月27日(2006.7.27)

(51) Int.C1.	F 1	テーマコード (参考)
G10K 15/04 (2006.01)	G10K 15/04	302D 5D015
G10L 15/10 (2006.01)	G10L 3/00	531Z 5D108
G10L 15/00 (2006.01)	G10L 3/00	551B 551G

審査請求 未請求 請求項の数 5 O L (全 12 頁)

(21) 出願番号	特願2005-9563 (P2005-9563)	(71) 出願人	000005821 松下電器産業株式会社 大阪府門真市大字門真1006番地
(22) 出願日	平成17年1月17日 (2005.1.17)	(74) 代理人	100098291 弁理士 小笠原 史朗
		(72) 発明者	飯阪 篤 大阪府門真市大字門真1006番地 松下電器産業株式会社内
		(72) 発明者	山本 章裕 大阪府門真市大字門真1006番地 松下電器産業株式会社内
		(72) 発明者	開発 勇治 大阪府門真市大字門真1006番地 松下電器産業株式会社内

最終頁に続く

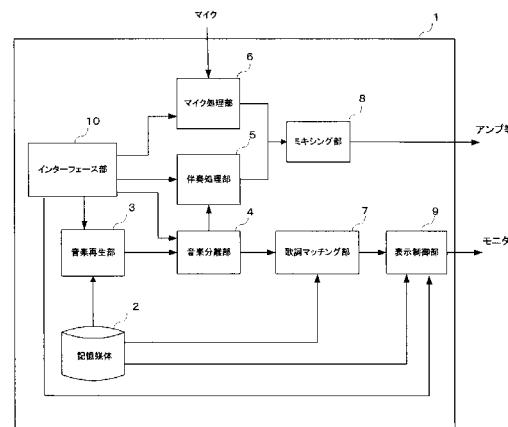
(54) 【発明の名称】 音楽再生装置および音楽再生プログラム

(57) 【要約】

【課題】 カラオケ専用のデータを入手せずとも、市販の音楽CD等を用いて正確な歌詞表示がされるカラオケが楽しめる音楽再生装置を提供すること。

【解決手段】 例えば市販の音楽CDを音楽データとして記憶媒体2に記憶する。また、各音楽データに対応する歌詞データも、インターネット等から入手して記憶媒体2に記憶しておく。音楽データを再生する際に、再生対象の音楽データをボーカル信号と伴奏信号に分離し、分離したボーカル信号を母音認識して、母音列を抽出する。次に、当該抽出した母音列と、当該音楽データに対応する歌詞データの母音列と比較して歌唱位置を特定する。そして、伴奏信号の出力と共に、特定した歌唱位置に対応する歌詞を上記歌詞データから読み出して表示を行う。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

音楽データおよび歌詞データが記録された媒体から当該音楽データを再生すると共に歌詞を画面に表示する音楽再生装置であって、

前記音楽データを音楽信号に再生する再生部と、

前記音楽信号をボーカル信号とボーカル以外の信号である伴奏信号とに分離する音楽分離部と、

前記ボーカル信号を音声認識して文字データを生成する音声認識部と、

前記文字データと前記歌詞データとの比較を行うことにより、前記歌詞データ上における歌詞の位置を特定する情報である歌唱位置特定情報を生成する歌詞比較部と、

音声出力するための前記伴奏信号を出力する伴奏出力部と、

前記歌唱位置特定情報に基づいて前記歌詞データから画面に表示すべき歌詞を抽出して前記画面に表示する表示制御部とを有する音楽再生装置。

【請求項 2】

前記音声認識部は、前記ボーカル信号を母音認識して第1の母音文字データを生成し、

前記歌詞比較部は、前記歌詞データを母音列に変換した第2の母音文字データを生成し、前記第1の母音文字データから、複数の文字数で設定される比較単位で第1の文字列を抽出し、当該第1の文字列と一致した前記第2の母音文字データ上の文字列に対応する前記歌詞データ上の文字列の位置を前記歌唱位置特定情報として生成する、請求項1記載の音楽再生装置。

【請求項 3】

前記歌詞比較部は、前記第1の文字列と一致する文字列が前記第2の母音文字データ上に複数存在するとき、

前記音声認識部は、前記ボーカル信号を子音も含む音声認識して子音文字データを生成し、

前記歌詞比較部は、前記子音文字データから前記比較単位で第2の文字列を抽出し、当該第2の文字列と一致した前記母音列への変換前の歌詞データ上における文字列の位置を前記歌唱位置特定情報として生成する、請求項2記載の音楽再生装置。

【請求項 4】

前記表示制御部は、前記伴奏出力部が出力する伴奏信号に併せて前記歌唱位置特定情報によって示される歌詞を前記画面に表示する、請求項1記載の音楽再生装置。

【請求項 5】

音楽データおよび歌詞データが記録された媒体から当該音楽データを再生すると共に歌詞を画面に表示する音楽再生装置のコンピュータに、

前記音楽データを音楽信号に再生する再生ステップと、

前記音楽信号をボーカル信号とボーカル以外の信号である伴奏信号とに分離する音楽分離する音楽分離ステップと、

前記ボーカル信号を音声認識して文字データを生成する音声認識ステップと、

前記文字データと前記歌詞データとの比較を行うことにより、前記歌詞データ上における歌詞の位置を特定する情報である歌唱位置特定情報を生成する歌詞比較ステップと、

音声出力するための前記伴奏信号を出力する伴奏出力ステップと、

前記歌唱位置特定情報に基づいて前記歌詞データから画面に表示すべき歌詞を抽出して画面に表示する表示制御ステップとを実行させるための音楽再生プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音楽再生装置および音楽再生プログラムに関し、より特定的には、音楽を再生する際に歌詞表示を行う音楽再生装置および音楽再生プログラムに関するものである。

【背景技術】

10

20

30

40

50

【0002】

従来、音楽再生に合わせて再生している音楽の歌詞をディスプレイ等に表示する音楽再生装置として、カラオケ装置がある。カラオケ装置では、再生する音楽の歌詞の文字情報である歌詞データと、音楽の伴奏のみが含まれている伴奏データと、音楽内の位置を特定するための情報であるタイムコードとが含まれているカラオケ用データを、ネットワークからダウンロード、あるいは記録媒体等から読み込んでいる。そして、タイムコードに基づいて歌詞データを表示することで、伴奏に合わせて歌詞が表示される。

【0003】

また、上記のようなカラオケ用データを使わずに、例えば市販の音楽CDを利用してカラオケを楽しむ音楽データ処理装置も開示されている（例えば、特許文献1参照）。特許文献1に開示されている音楽データ処理装置は、ショックブルーフ手段と、音声認識手段と、出力手段とを備えている。ショックブルーフ手段は、例えば市販の音楽CDを再生する際に、再生の為の通常の読み取り速度よりも高速で間欠的に音楽CDから音楽データを読み取る。音声認識手段は、ショックブルーフ手段により読み取られた音楽データに含まれているボーカル情報を音声認識して文字情報を取得する。そして、出力手段で、音声認識手段により取得された文字情報を出力する。これによって、カラオケ用データが無くともカラオケを楽しむことが出来る。

【特許文献1】特開2001-312288号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

上述のように、従来のカラオケ装置では、音楽再生時に歌詞表示をするためには、カラオケ用データが必要となる。従って、利用者が市販の音楽CDを保有していても、別途同じ音楽のカラオケ用データを購入しなければならないという問題点がある。一方、上記特許文献1に開示された音楽データ処理装置は、カラオケ用データを購入せずとも、市販の音楽CDさえ保有していれば、音楽再生時に歌詞表示を行うことが可能ではある。しかし、音楽に含まれているボーカル（人間の音声）には、例えば叫び声や、不明瞭な発音等、非常に多種多様なものが含まれている。そのため、現在の音声認識技術では多種多様なボーカルを文字（テキスト）データとして正確に抽出することが出来ず、その結果、歌詞表示がおかしくなるという問題点がある。

【0005】

それ故に、本発明の目的は、カラオケ用データ入手せずとも、正確な歌詞表示がされるカラオケが楽しめる音楽再生装置および音楽再生プログラムを提供することである。

【課題を解決するための手段】

【0006】

第1の発明は、音楽データおよび歌詞データが記録された媒体から当該音楽データを再生すると共に歌詞を画面に表示する音楽再生装置であって、音楽データを音楽信号に再生する再生部と、音楽信号をボーカル信号とボーカル以外の信号である伴奏信号とに分離する音楽分離部と、ボーカル信号を音声認識して文字データを生成する音声認識部と、文字データと歌詞データとの比較を行うことにより、歌詞データ上における歌詞の位置を特定する情報である歌唱位置特定情報を生成する歌詞比較部と、音声出力するための伴奏信号を出力する伴奏出力部と、歌唱位置特定情報に基づいて歌詞データから画面に表示すべき歌詞を抽出して画面に表示する表示制御部とを有する音楽再生装置である。

【0007】

第2の発明は、上記第1の発明において、音声認識部は、ボーカル信号を母音認識して第1の母音文字データを生成し、歌詞比較部は、歌詞データを母音列に変換した第2の母音文字データを生成し、第1の母音文字データから、複数の文字数で設定される比較単位で第1の文字列を抽出し、当該第1の文字列と一致した第2の母音文字データ上の文字列に対応する歌詞データ上の文字列の位置を歌唱位置特定情報として生成する。

【0008】

10

20

30

40

50

第3の発明は、上記第2の発明において、歌詞比較部は、第1の文字列と一致する文字列が第2の母音文字データ上に複数存在するとき、音声認識部は、ボーカル信号を子音も含む音声認識して子音文字データを生成し、歌詞比較部は、子音文字データから比較単位で第2の文字列を抽出し、当該第2の文字列と一致した母音列への変換前の歌詞データ上における文字列の位置を歌唱位置特定情報として生成する。

【0009】

第4の発明は、上記第1の発明において、表示制御部は、伴奏出力部が出力する伴奏信号に併せて歌唱位置特定情報によって示される歌詞を画面に表示する。

【0010】

第5の発明は、音楽データおよび歌詞データが記録された媒体から当該音楽データを再生すると共に歌詞を画面に表示する音楽再生装置のコンピュータに、音楽データを音楽信号に再生する再生ステップと、音楽信号をボーカル信号とボーカル以外の信号である伴奏信号とに分離する音楽分離する音楽分離ステップと、ボーカル信号を音声認識して文字データを生成する音声認識ステップと、文字データと歌詞データとの比較を行うことにより、歌詞データ上における歌詞の位置を特定する情報である歌唱位置特定情報を生成する歌詞比較ステップと、音声出力するための伴奏信号を出力する伴奏出力ステップと、歌唱位置特定情報に基づいて歌詞データから画面に表示すべき歌詞を抽出して画面に表示する表示制御ステップとを実行させるための音楽再生プログラムである。

【発明の効果】

【0011】

上記第1の発明によれば、例えば市販のCDと歌詞データとがあれば、別途カラオケ用のデータを入手せずとも正確な歌詞表示ができるため、手軽にカラオケを楽しむことができる。

【0012】

上記第2の発明によれば、歌詞マッチング部は、ボーカル信号を母音認識するため、子音認識に比べ、より確実な認識ができる。そして、母音文字列を用いて歌詞のマッチング処理を行うため、より正確な歌詞表示ができる。

【0013】

上記第3の発明によれば、母音認識だけで歌詞が特定できないときは子音認識を併用するため、母音認識のみ使用するときに比べ、より正確な歌詞表示ができる。

【0014】

上記第4の発明によれば、伴奏とその伴奏に合った歌詞とが同期されて出力されるため、利用者にとって歌うタイミングが把握しやすくなる。

【0015】

上記第5の発明によれば、上記第1の発明と同様の効果が得られる。

【発明を実施するための最良の形態】

【0016】

以下、本発明の実施の形態について、図面を参照して説明する。尚、この実施例により本発明が限定されるものではない。

【0017】

図1は、本発明の実施形態にかかる音楽再生装置（以下、カラオケ装置と称す）の構成を示したブロック図である。図1において、カラオケ装置1は、記憶媒体2と、音楽再生部3と、音楽分離部4と、伴奏処理部5と、マイク処理部6と、歌詞マッチング部7と、ミキシング部8と、表示制御部9と、インターフェース部10とで構成される。

【0018】

記憶媒体2は、音楽データ、歌詞データ、映像データ等を記憶する媒体である。例えば、ハードディスクやフラッシュメモリやCD-ROMやDVD-ROM等が記憶媒体2に相当する。また、音楽データは、例えば、MP3(MPEG Audio Layer-3)形式やogg Vorbis形式の圧縮オーディオデータや、WAV形式の非圧縮のオーディオデータが相当する。歌詞データは、具体的にはテキストデータであり、音楽デ

10

20

30

40

50

ータと同一名称で「txt」という拡張子が付されたファイル名のテキストファイルを、インターネットからダウンロード等することで記憶媒体2に保存される。映像データは、例えば、MPEG4(Moving Picture Experts Group phase 4)形式の動画データやJPEG形式の静止画データである。

【0019】

音楽再生部3は、インターフェース部10から出力される、ユーザの操作内容を示す操作信号等に従って、記憶媒体2に記憶された音楽データを読み込む。また、音楽再生部3は、読み込んだ音楽データを再生し、L、Rの2チャンネルの音楽信号として音楽分離部4に出力する。なお、音楽再生部3は、L、Rの2チャンネルの音楽信号が出力できればどのような形態でもよく、例えばラジオ放送を受信するチューナーであってもよい。また、音楽再生部3は、音楽信号を所定秒分、例えば3秒分毎に区切って出力する。すなわち、音楽信号を3秒分溜めて出力し、また3秒分溜めて出力するという動作を繰り返す。音楽分離部4は、音楽再生部3から出力された音楽信号を、ボーカルのみが含まれるボーカル信号とボーカルを除いた音楽のみが含まれる伴奏信号とに分離する。また、音楽分離部4は、上記ボーカル信号を歌詞マッチング部7へ出力し、上記伴奏信号を伴奏処理部5へ出力する。伴奏処理部5は、歌詞マッチング部7における処理時間を確保するために、音楽分離部4から出力された伴奏信号を所定の時間遅延させてからミキシング部8へ出力する。マイク処理部6は、マイクなどによって入力されたユーザの声信号の入力レベル調整やエコーの付加などの処理を行い、当該ユーザの声信号をミキシング部8へ出力する。歌詞マッチング部7は、音楽分離部4から入力されたボーカル信号と記憶媒体2に記憶されている歌詞データとを用いて、歌詞表示のための制御情報である歌唱位置特定情報を生成し、表示制御部9へ出力する。ミキシング部8は、マイク処理部6から出力されたユーザの声信号と伴奏処理部5から出力された伴奏信号とをミキシングしてアンプ等へ出力する。なお、アンプについては、カラオケ装置1に内蔵されていてもよい。表示制御部9は、記憶媒体2から歌詞データを読み込み、当該動画等にオーバーラップさせる形で歌詞マッチング部7から出力される歌唱位置特定情報に基づいた歌詞表示をモニタに行う。なお、表示制御部9は、記憶媒体2から動画・静止画データ等を読み込み、歌詞表示に併せて動画・静止画等をモニタに表示してもよい。インターフェース部10は、タッチパネルやリモコンなどが相当し、例えばユーザによる再生操作やキー(曲の調)の変更操作等を受け付け、ユーザの操作内容を示す操作信号を音楽再生部3等に出力する。

【0020】

図2は、本発明の実施の形態における歌詞マッチング部7の詳細を示すブロック図である。図2において、歌詞マッチング部7は、メモリー部71と、音声認識部72と、音データベース73と、歌詞比較部74とで構成される。メモリー部71は、音楽分離部4から出力されたボーカル信号を保持する。音声認識部72は、上記ボーカル信号を母音認識し、第1の母音文字データを生成して歌詞比較部74へ出力する。音データベース73は、あらゆる年齢層の男女の声と母音とが対応付けられたデータベースであり、音声認識部72による母音認識処理の際に用いられる。歌詞比較部74は、記憶媒体2から歌詞データを読み込み、母音文字列に変換して第2の母音文字データを生成する。そして、歌詞比較部74は、音声認識部72から出力された第1の母音文字データと、当該第2の母音文字データを比較し、現在の再生(演奏)位置に対応する歌詞データ上の位置(例えば、歌詞の先頭から何文字目か、等)を特定する。また、歌詞比較部74は、特定した位置についての情報を歌唱位置特定情報として、表示制御部9に出力する。

【0021】

本実施形態は、上記のような構成において、まず、例えば市販の音楽CDをリッピング等することで、MP3形式等の音楽データとして記憶媒体2に記憶しておく。また、各音楽データに対応する歌詞データをインターネット等からダウンロードして、記憶媒体2に記憶しておく。そして、上記の音楽データを再生する際に、再生対象の音楽データをボーカル信号と伴奏信号に分離する。更に、分離したボーカル信号を母音認識して、母音列を抽出し、当該母音列と再生対象の音楽データに対応する歌詞データの母音列と比較するこ

10

20

30

40

50

とで歌唱位置を特定する。そして、伴奏信号の出力と共に、特定した歌唱位置に対応する歌詞を上記歌詞データから読み出して表示を行うものである。なお、音声認識として母音認識を用いているのは、音声認識では子音の認識は困難であるが、母音の認識は確実性が高いいためである。

【0022】

以下、図3～図4を用いて、カラオケ装置1が行うカラオケ再生処理の詳細動作を説明する。図3は、カラオケ装置1が行うカラオケ再生処理の詳細を示すフローチャートである。本実施形態のカラオケ装置1では、曲再生中に利用者がカラオケモードに切り替えると、その時点で再生中の曲をカラオケするかどうか問合せる。その結果、利用者がカラオケすることを選択した場合は、再生中の曲をカラオケモードで最初から再生します。一方、再生中の曲をカラオケしない場合は、次の曲（利用者がカラオケしたい曲）が利用者により選択されるのを待つ。また、曲が再生されてないときに利用者がカラオケモードに切り替えたときは、その後に利用者が選んだ曲をカラオケモードで再生する。

10

【0023】

図3において、まず、インターフェース部10、例えばモニタ（タッチパネル）上に表示されるカラオケモード移行ボタンが、利用者により押下される（ステップS101）。これにより、カラオケ装置1は、通常の再生モードからカラオケモードに移行し、以下のカラオケ再生処理を開始する。

【0024】

カラオケモードに移行後、音楽再生部3は、曲が再生中であるか否かの判定を行う（ステップS102）。ステップS102における判定の結果、再生中でない場合（ステップS102でNO）、音楽再生部3は、利用者の曲選択を待つ。そして、利用者が曲を選択すれば、音楽再生部3は、当該選択された曲の音楽データを記憶媒体2から読み込む。そして、音楽再生部3は、当該音楽を再生し、再生された音楽信号を音楽分離部4へ出力する。次に、音楽分離部4は、音楽再生部3から出力された音楽信号をボーカル信号と伴奏信号とに分離する。そして、音楽分離部4は、ボーカル信号を歌詞マッチング部7に、伴奏信号を伴奏処理部5に出力する（ステップS114）。その後、カラオケ装置1は、後述するステップS106の処理へと進む。

20

【0025】

ここで、音楽分離部4が行う音楽信号の分離手法としては、次のような手法が考えられる。例えばステレオ音声の音楽データである場合、ボーカル音声が通常L Rのステレオ音声のセンターに定位していることを利用し、Lチャンネルの音声信号とRチャンネルの音声信号の減算を行うことで、センター定位成分、つまりLチャンネル音声信号とRチャンネル音声信号で同一レベル、同一周波数、同位相の成分であるボーカルをキャンセルすることができる。そのため、当該減算処理の対象周波数帯域をボーカル情報の含まれている帯域にすることで、ベースなどセンターに定位しているボーカル以外の音が消えないようすることができ、その結果、伴奏信号を作成することができる。また、伴奏信号が作成出来れば、LまたはRチャンネルの音声信号と、上記作成した伴奏信号との減算を行うことにより、ボーカル信号を作成することができる。

30

【0026】

図3に戻り、ステップS102における判定の結果、楽曲が再生中である場合（ステップS102でYES）、表示制御部9は、現在再生中の曲についてカラオケを行うか否かの確認を行うために、例えば、「この曲を歌いますか？」という表示を、モニタに出力する。また、当該表示と共に、表示制御部9は、タッチパネル上に「YES」および「NO」のボタンを表示する（ステップS103）。利用者は、カラオケを行う場合には、YESを選び、行わない場合は、NOを選ぶ。

40

【0027】

次に、音楽再生部3は、YESおよびNOのいずれのボタンが選ばれたかを判定する（ステップS104）。ステップS104の判定の結果、NOのボタンが選ばれた場合（ステップS104でNO）は、音楽再生部3は、現在の再生曲が終了するのを待ち、次の曲

50

の再生および上記ステップ S 114 と同様の処理で音楽信号を分離し、ボーカル信号の歌詞マッチング部 7 への出力および伴奏信号の伴奏処理部 5 への出力を開始する（ステップ S 115）。そして、カラオケ装置 1 は、後述のステップ S 106 の処理へ進む。一方、ステップ S 104 の判定の結果、YES のボタンが選ばれた場合（ステップ S 104 で YES）は、音楽再生部 3 は、現在再生中の曲の再生を停止する。そして、当該曲の先頭から再生し直し、上記ステップ S 113 と同様の処理で音楽信号を分離し、ボーカル信号の歌詞マッチング部 7 への出力および伴奏信号の伴奏処理部 5 への出力をを行う（ステップ S 105）。そして、カラオケ装置 1 は、次に説明するステップ 106 の処理へと進む。ここで、上記ステップ S 105、S 114、S 115 のいずれかにおいて伴奏信号を出力された伴奏処理部 5 は、所定の時間（本実施形態では 3 秒とする）待機してから、上記伴奏信号をミキシング部 8 へ出力する。これは、以下に説明するステップ S 106～S 109 の処理が完了するまで、伴奏信号の出力を遅延させ、表示制御部 9 の歌詞表示処理との同期を取るためである。10

【0028】

音楽再生部 3 による曲の再生が開始されると、表示制御部 9 は、再生中の曲名、歌手名等を音楽データから読み込み、モニタに表示する。更に、表示制御部 9 は、モニタ上にカラオケモード終了ボタンを表示する（ステップ S 106）。

【0029】

次に、歌詞マッチング部 2 は、再生中の楽曲に対応した歌詞データを記憶媒体 2 から読み込む（ステップ S 107）。なお、歌詞データが記憶媒体 2 に存在しない場合は、カラオケ装置 1 は、通信部（図示せず）を用いて所定のネットワークにアクセスし、歌詞データを取得すると同時に記憶媒体 2 に記憶してもかまわない。また、歌詞データが存在しない旨の表示を行い、利用者の操作を受けて CD-ROM 等の外部媒体から読み込ませてもかまわない。20

【0030】

次に、歌詞マッチング部 7 は、ステップ S 107 で読み込んだ歌詞データの中から後述する歌詞マッチング処理（ステップ S 109）で処理対象とする部分の絞り込みを行う（ステップ S 108）。すなわち、後述する歌詞マッチング処理において、処理の高速化の観点から、歌詞データの全てを処理対象とはせずに、歌詞データのうちの一部分のみを処理対象とするためである。上記絞り込みの方法として、後述するステップ S 110 の処理で記憶部（図示せず）に記憶される表示済位置情報あるいはタイムカウンターの値を使うことが考えられる。表示済位置情報とは、表示制御部 9 が表示した歌詞の歌詞データ上の位置を示す情報である（例えば、表示した歌詞の最後の文字が歌詞データ上で先頭から何文字目にあたるか等）。すなわち、表示済位置情報の示す歌詞データ上の位置は、直前に表示制御部 9 が表示した歌詞の位置を示しているため、当該表示した歌詞の続きの歌詞が存在する当該位置の周辺を処理対象として絞り込めばよいことになる。また、タイムカウンター値とは、オーディオ装置がカウントするタイムカウンターの値である。そのため、当該タイムカウンター値を用いれば、表示すべき歌詞が存在する歌詞データ上のおよその位置を推測し、絞り込むことが可能となる。例えば、演奏時間 4 分の曲において、タイムカウンターの値が 2 分のときは、歌詞データ上の中間地点前後を絞り込むことが考えられる。30

【0031】

ステップ S 108 で歌詞データの絞込みを行った後、歌詞マッチング部 7 は、後述する歌詞マッチング処理（ステップ S 109；図 4）を行う。当該歌詞マッチング処理の結果、表示制御部 9 が歌詞を表示するときに用いる制御情報である歌唱位置特定情報が歌詞マッチング部 7 より表示制御部 9 に出力される。40

【0032】

次に、表示制御部 9 は、上記歌唱位置特定情報に基づき、歌詞データから歌詞（文字列）を予め設定された長さ分読み込む。そして、表示制御部 9 は、現在歌唱すべき位置が分かるように表示文字を所定の色で変化させていく等の処理を施して、読み込んだ歌詞をモ50

ニタに表示する。これに合わせて、上述した待機時間を経た伴奏処理部5から上記伴奏信号がミキシング部8に出力される。ミキシング部8は、当該伴奏信号とマイク処理部から出力されるマイク信号（すなわち、利用者の歌声）とをミキシングし、アンプに出力する（ステップS110）。これによって、伴奏信号の出力に合わせて、歌詞データから読みこまれた歌詞がモニタに表示されることになる。また、このとき、表示制御部9は、歌詞データ上の現在の歌唱位置を上述した表示済位置情報として、RAM等の記憶部（図示せず）に記憶しておく。更に、オーディオ装置がカウントするタイムカウンターの値も同様に記憶部（図示せず）に記憶しておく。これは、上述したようにステップ108において歌詞データから歌詞を絞り込むときに用いるためである。

【0033】

10

次に、音楽再生部3は、カラオケモード終了ボタンが押されたかどうかの判定を行う（ステップS111）。ステップS111の判定の結果、終了ボタンが押されていない場合は（ステップS111でNO）、音楽再生部3は、曲の再生が終了したか否かの判定を行う（ステップS112）。ステップS112の判定の結果、曲が再生中である場合（ステップS112でNO）、カラオケ装置1は、ステップ108に戻って、ステップS108～S111の処理を繰り返す。一方、ステップS112の判定の結果、曲の再生が終了した場合（ステップS112でYES）は、カラオケ装置1は、自動的に次の曲の再生を開始するか、あるいは、ユーザ操作による次の曲の再生開始を待つ（ステップ113）。曲の再生が開始すれば、カラオケ装置1は、上記ステップS106に戻って、ステップS106～S113の処理を繰り返す。

【0034】

20

一方、ステップS111の判定の結果、終了ボタンが押された場合（ステップS111でYES）は、カラオケ装置1は、通常の音楽再生モードに切り替わる（ステップS116）。以上で、本実施形態にかかるカラオケ再生処理は終了する。

【0035】

30

次に、図4を用いて、上記ステップS109における歌詞マッチング処理の詳細について説明する。図4は、上記ステップS109で示した歌詞マッチング処理の詳細を示すフローチャートである。図4において、まず、音声認識部72は、音楽分離部4から出力されたボーカル信号の中でボーカル音声が存在する区間検出を行う（ステップS71）。ここで、区間検出の手法については、例えば、音量レベルを確認し、所定の閾値よりもレベルが高い部分はボーカル音声が存在すると判断する、等の手法が考えられる。

【0036】

40

次に、音声認識部72は、ステップS71のボーカル区間の検出結果からボーカル区間の有無を判定する（ステップS72）。ステップS72の判定の結果、ボーカル区間が存在しない場合（ステップS72でNO）、後述する歌唱位置特定情報の出力処理（ステップS76）へ進む。一方、ステップS72の判定の結果、ボーカル区間が存在する場合（ステップS72でYES）、音声認識部72は、ステップS71で調べたボーカルの存在する区間について、ステップS108において絞り込んだ歌詞を母音に変換する（ステップS73）。そして、変換した母音列を辞書として、音データベース73を用いた母音認識を行う（ステップ74）。母音認識の手法としては、例えば、ケプストラム法を用いることが考えられる。また、このとき、母音認識のマッチング度合いを示すパラメータを母音認識の結果としてRAM等の記憶部（図示せず）に一時的に記憶しておく。次に、ステップ75において、上記母音認識のマッチング度合いを示すパラメータが所定の閾値以上である歌詞が1つしか存在しなければ（すなわち、一致する歌詞が1つだけならば）、当該歌詞を歌唱位置として特定する。一方、所定の閾値以上である歌詞が2つ以上存在する場合は、当該所定の閾値以上である歌詞の元の歌詞（つまり、子音を含む文字列）を辞書として通常の音声認識を行い、歌唱位置を特定する。すなわち、一致する母音列が複数存在した場合は、当該複数の母音列に対応する歌詞を候補として、子音を含む通常の音声認識を行うことで歌唱位置を1つに絞る（ステップS75）。上記のような処理の結果、母音列（あるいは子音列）が一致した位置に対応する歌詞データ上の位置（例えば先頭から

50

150文字目、等)が歌唱位置であると特定することができる。

【0037】

図4に戻り、ステップS72の処理あるいはS75の処理の次に、歌詞比較部74は、表示制御部9に歌唱位置特定情報を出力する(ステップS76)。歌唱位置特定情報とは、表示制御部9が歌詞を表示するときに用いる制御情報である。歌詞比較部74は、ステップS72の判定の結果、ボーカル区間が存在しない場合は、ボーカル無しの旨の歌唱位置特定情報を出力する。一方、ボーカル区間が存在する場合は、ステップS75で特定した歌唱位置の情報(例えば先頭から150文字目、等)を出力する。以上で、歌詞マッチング処理が終了する。この後、表示制御部9は、上記歌唱位置特定情報に基づいて上述したステップS110(図3)の処理を行うが、このときボーカル無しの旨の歌唱位置特定情報が出力された場合は、歌詞のモニタへの表示は行わない。一方、歌唱位置の情報が含まれる歌唱位置特定情報が出力された場合は、表示制御部9は、歌詞データを記憶媒体2から読み込み、例えば先頭から150文字目の文字から所定の長さ分の文字列をモニタに表示する。このとき、上述したように待機時間(本実施形態では3秒)を経た伴奏信号もミキシング部8から出力される。その結果、伴奏信号と歌詞表示との出力の同期が取られることになる。

【0038】

また、上述した「さわやかな」と「はなたばは」の例のように、歌詞において同じ母音が連続するときは、子音を含めた通常の音声認識を行い、その結果を用いて歌唱位置特定情報の出力を行うが、上記通常の音声認識の結果、子音の認識もできなかつた場合は、等速度で歌唱位置を変化させる表示を行う旨の歌唱位置特定情報を出力する。つまり、「あああああ」という叫び声に相当するような歌詞の部分は子音が含まれていないため、通常の音声認識を用いても子音の認識ができないこととなる。そのため、このような場合は、歌詞比較部74は、等速度で歌唱位置を変化させる旨の歌唱位置特定情報を表示制御部9に出力する。その結果、表示制御部9は、当該歌唱位置特定情報に基づき、「あああああ」の歌詞表示を行い、左から右方向に一定の速度で文字色を変化させていく。

【0039】

このように、本実施形態では、音楽データを再生する際に、音楽データのボーカル信号とデキストファイルである歌詞データとのマッチングを行うことで、歌詞表示位置を特定する。これにより、別途カラオケ用のデータを購入しなくとも、正確な歌詞表示が行われるカラオケを楽しむことができる。更に、伴奏信号の出力についても、所定の時間(歌詞マッチング処理に要する時間)待機させ、歌詞表示に併せて出力するようにしている。そのため、伴奏に合わせて歌詞が表示され、利用者にとって歌うべきタイミングが把握しやすくなる。

【0040】

なお、本実施形態におけるカラオケ装置1は、一般のオーディオ装置に限らず、車載用のオーディオ装置やナビゲーション装置と一体型のオーディオ装置として構成することもできる。特に、車内環境は比較的大音量で音楽を楽しめ、他人に聞かれない空間であるため、車載用としての利用は、より一層効果的である。

【0041】

また、上述した音楽データについても、ステレオ音声の音楽データに限らず、例えばモノラル音声や5.1chの音楽データでもよい。

【0042】

また、上述した実施形態は、カラオケ装置のコンピュータに実行させるプログラムの形態で実施されてもよい。この場合は、例えばカラオケ装置1の記憶部(図示せず)に図3、4に示すような処理が実行できるプログラムを格納し、制御部(図示せず)が図3、4に示すような処理を実行すればよい。

【産業上の利用可能性】

【0043】

本発明にかかる音楽再生装置は、音楽データを音声認識し、テキストデータである歌詞

データとマッチングすることで正確な歌詞表示を行うことができ、家庭用オーディオ装置や車載用のオーディオ装置等の用途に有用である。

【図面の簡単な説明】

【0044】

【図1】本発明の実施形態にかかる音楽再生装置を示す機能ブロック図

【図2】図1の歌詞マッチング部7の機能構成を示すブロック図

【図3】本発明の実施形態に係るカラオケ再生処理を示すフロー・チャート

【図4】図3のステップS109で示した歌詞マッチング処理の詳細を示すフロー・チャート

【符号の説明】

【0045】

1 音楽再生装置

2 記憶媒体

3 音楽再生部

4 音楽分離部

5 伴奏処理部

6 マイク処理部

7 歌詞マッチング部

8 ミキシング部

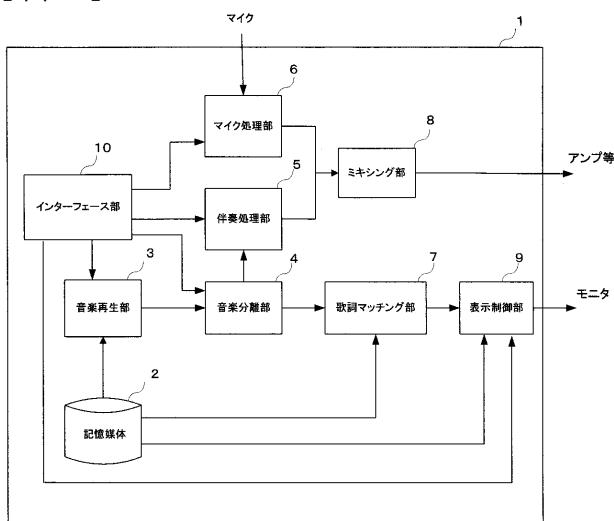
9 表示制御部

10 インターフェース部

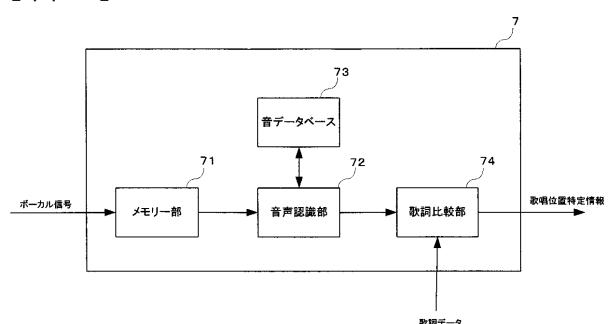
10

20

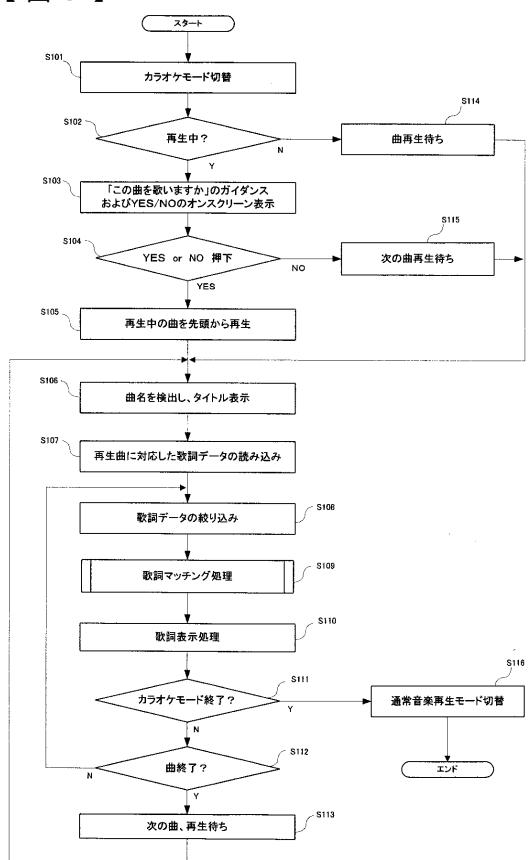
【図1】



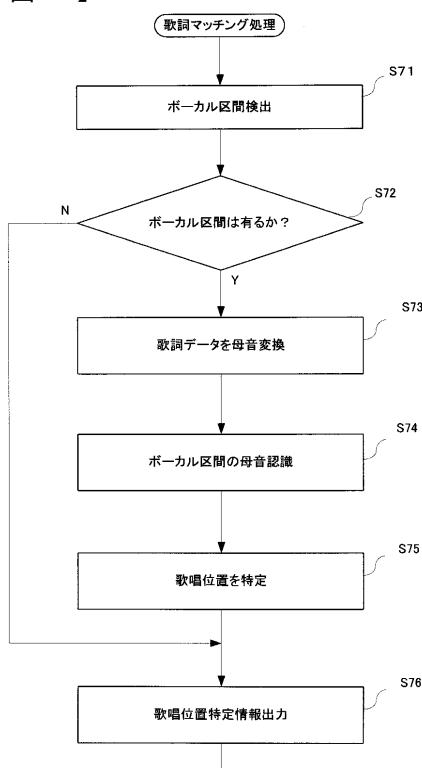
【図2】



【図3】



【図4】



フロントページの続き

(72)発明者 中北 学
大阪府門真市大字門真1006番地 松下電器産業株式会社内

(72)発明者 森 敏昭
大阪府門真市大字門真1006番地 松下電器産業株式会社内

F ターム(参考) 5D015 AA06 KK02
5D108 BA22 BA35