

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5598516号
(P5598516)

(45) 発行日 平成26年10月1日(2014.10.1)

(24) 登録日 平成26年8月22日(2014.8.22)

(51) Int.Cl.			F I		
G 1 0 K	15/04	(2006.01)	G 1 0 K	15/04	3 0 2 D
G 1 0 L	13/00	(2006.01)	G 1 0 L	13/00	1 0 0 Y
G 1 0 L	13/033	(2013.01)	G 1 0 L	13/033	1 0 2 B
G 0 9 B	15/00	(2006.01)	G 0 9 B	15/00	D

請求項の数 5 (全 24 頁)

(21) 出願番号	特願2012-191440 (P2012-191440)	(73) 特許権者	000005267
(22) 出願日	平成24年8月31日 (2012.8.31)		ブラザー工業株式会社
(65) 公開番号	特開2014-48472 (P2014-48472A)		愛知県名古屋市瑞穂区苗代町15番1号
(43) 公開日	平成26年3月17日 (2014.3.17)	(74) 代理人	110000578
審査請求日	平成25年3月13日 (2013.3.13)		名古屋国際特許業務法人
		(72) 発明者	上村 晃弘
			愛知県名古屋市瑞穂区苗代町15番1号
			ブラザー工業株式会社内
		(72) 発明者	幡田 久美
			愛知県名古屋市瑞穂区苗代町15番1号
			ブラザー工業株式会社内
		(72) 発明者	阿瀬見 典昭
			愛知県名古屋市瑞穂区苗代町15番1号
			ブラザー工業株式会社内

最終頁に続く

(54) 【発明の名称】 カラオケ用音声合成システム、及びパラメータ抽出装置

(57) 【特許請求の範囲】

【請求項1】

楽曲を識別する楽曲ID、当該楽曲IDによって識別される楽曲を構成する楽音を表す演奏情報、当該楽曲IDによって識別される楽曲の歌詞を表す歌詞情報、当該歌詞情報によって表される歌詞の発声開始タイミングを示す発声タイミング情報を含む楽曲データから、前記楽曲ID、前記演奏情報、前記歌詞情報、前記発声タイミング情報を取得して前記楽曲データに基づく楽曲を再生する発声情報取得手段と、

前記楽曲データに基づく楽曲の再生中に、前記歌詞の発声開始タイミングで、入力された音声波形を取得する波形取得手段と、

前記音声波形から、前記歌詞を形成する各音節に対する音声波形である音節波形を抽出すると共に、その抽出した各音節波形から、予め規定された少なくとも一つの特徴量である音声パラメータを導出するパラメータ導出手段と、

前記パラメータ導出手段で導出された前記音節ごとの音声パラメータを、当該楽曲を歌唱したユーザを識別するユーザIDと対応付けて、第一記憶装置に記憶するパラメータ登録手段と、

各楽曲についての理想的な歌唱音声に基づく、少なくとも一つの音声パラメータを含む模範音声データを、前記楽曲IDと対応付けて第二記憶装置に記憶する模範記憶手段と、

を有したパラメータ抽出装置と、

指定された楽曲ID、及びユーザIDを取得する識別情報取得手段と、

前記識別情報取得手段で取得した前記楽曲IDと対応付けられた模範歌声データを前記

第二記憶装置から取得する模範取得手段と、

前記第一記憶装置に記憶されている音声パラメータの中から、前記識別情報取得手段で取得した前記ユーザIDと対応付けられた音声パラメータであって、前記模範取得手段で取得した模範音声データに含まれる楽曲IDと対応付けられた音声パラメータに最も類似する音声パラメータを取得するパラメータ取得手段と、

前記パラメータ取得手段で取得した音声パラメータを、前記楽曲IDと対応付けられた音声パラメータに一致するように調整し、その調整された音声パラメータに基づいて音声合成する音声合成手段と、

前記音声合成手段にて音声合成することで生成された合成音を出力する出力手段と、
を有した合成音出力装置と

を備え、

前記発声情報取得手段は、

楽曲の一つである対象楽曲の楽譜を表し、音源モジュールから出力される個々の出力音について、少なくとも音高及び演奏開始タイミングが規定された楽譜データを前記演奏情報として取得し、

前記楽譜データは、

前記対象楽曲の曲中において転調していれば、時間軸に沿って前記対象楽曲が転調した時刻を表す転調フラグを含み、

前記パラメータ登録手段は、

取得した楽譜データに基づいて、前記対象楽曲において同一の調が継続される各区間である調同一区間を特定する区間特定手段と、

前記区間特定手段にて特定した各調同一区間に含まれ、それぞれの調同一区間における時間軸に沿った最後の出力音を主音として特定する主音特定手段と、

前記区間特定手段にて特定した調同一区間に含まれる同一音名の出力音の頻度を表す登場音名頻度を、前記主音特定手段で特定した主音の音名を起点として前記調同一区間ごとに導出する音名頻度導出手段と、

各調にて利用可能な音名の分布を表すテンプレートとして調ごとに予め用意した調テンプレートに、前記音名頻度導出手段で導出した各登場音名頻度を照合した結果、最も相関が高い調それぞれを、メタデータとして推定する調推定手段と、

を備え、

前記調推定手段にて推定したメタデータを、前記音声パラメータと対応付けて前記第一記憶装置に記憶することを特徴とする音声合成システム。

【請求項2】

前記発声情報取得手段は、

前記対象楽曲の歌詞を構成する歌詞構成文字の文字列を、前記歌詞情報として取得し、前記歌詞構成文字の少なくとも1つに対する出力タイミングが、当該歌詞構成文字に対応する前記出力音の演奏開始タイミングと対応付けられた歌詞出力タイミングを、前記発声タイミング情報として取得し、

前記波形取得手段は、

前記楽譜データに基づく前記対象楽曲の演奏中に入力された音声が入力された時間軸に沿って推移した波形を、前記音声波形として取得し、

前記パラメータ導出手段は、

前記音声波形において、個々の出力音に対応する区間での音声波形を、前記音節波形として抽出する

ことを特徴とする請求項1に記載の音声合成システム。

【請求項3】

楽曲を識別する楽曲ID、当該楽曲IDによって識別される楽曲を構成する楽音を表す演奏情報、当該楽曲IDによって識別される楽曲の歌詞を表す歌詞情報、当該歌詞情報によって表される歌詞の発声開始タイミングを示す発声タイミング情報を含む楽曲データから、前記楽曲ID、前記演奏情報、前記歌詞情報、前記発声タイミング情報を取得して前

10

20

30

40

50

記楽曲データに基づく楽曲を再生する発声情報取得手段と、

前記楽曲データに基づく楽曲の再生中に、前記歌詞の発声開始タイミングで、入力された音声波形を取得する波形取得手段と、

前記音声波形から、前記歌詞を形成する各音節に対する音声波形である音節波形を抽出すると共に、その抽出した各音節波形から、予め規定された少なくとも一つの特徴量である音声パラメータを導出するパラメータ導出手段と、

前記パラメータ導出手段で導出された前記音節ごとの音声パラメータを、当該楽曲を歌唱したユーザを識別するユーザIDと対応付けて、第一記憶装置に記憶するパラメータ登録手段と、

各楽曲についての理想的な歌唱音声に基づく、少なくとも一つの音声パラメータを含む模範音声データを、前記楽曲IDと対応付けて第二記憶装置に記憶する模範記憶手段と、

を有したパラメータ抽出装置と、

指定された楽曲ID、及びユーザIDを取得する識別情報取得手段と、

前記識別情報取得手段で取得した前記楽曲IDと対応付けられた模範歌声データを前記第二記憶装置から取得する模範取得手段と、

前記第一記憶装置に記憶されている音声パラメータの中から、前記識別情報取得手段で取得した前記ユーザIDと対応付けられた音声パラメータであって、前記模範取得手段で取得した模範音声データに含まれる楽曲IDと対応付けられた音声パラメータに最も類似する音声パラメータを取得するパラメータ取得手段と、

前記パラメータ取得手段で取得した音声パラメータを、前記楽曲IDと対応付けられた音声パラメータに一致するように調整し、その調整された音声パラメータに基づいて音声合成する音声合成手段と、

前記音声合成手段にて音声合成することで生成された合成音を出力する出力手段と、

を有した合成音出力装置と

を備え、

前記パラメータ登録手段は、

前記発声情報取得手段で取得した歌詞情報によって表される文字列を、単語を構成する文字列である単語文字ごとに分割する単語分割手段と、

各単語の性質を表す性質情報を、当該単語の識別情報と対応付けた単語性質テーブルを予め用意し、前記単語分割手段で分割された各単語文字によって表される単語に対応する前記性質情報をメタデータとして、前記単語性質テーブルから抽出するメタデータ抽出手段と

を備え、

前記メタデータ抽出手段にて抽出したメタデータを、前記音声パラメータを対応付けて前記第一記憶装置に記憶することを特徴とする音声合成システム。

【請求項4】

楽曲を識別する楽曲ID、当該楽曲IDによって識別される楽曲を構成する楽音を表す演奏情報、当該楽曲IDによって識別される楽曲の歌詞を表す歌詞情報、当該歌詞情報によって表される歌詞の発声開始タイミングを示す発声タイミング情報を含む楽曲データから、前記楽曲ID、前記演奏情報、前記歌詞情報、前記発声タイミング情報を取得して前記楽曲データに基づく楽曲を再生する発声情報取得手段と、

前記楽曲データに基づく楽曲の再生中に、前記歌詞の発声開始タイミングで、入力された音声波形を取得する波形取得手段と、

前記音声波形から、前記歌詞を形成する各音節に対する音声波形である音節波形を抽出すると共に、その抽出した各音節波形から、予め規定された少なくとも一つの特徴量である音声パラメータを導出するパラメータ導出手段と、

前記パラメータ導出手段で導出された前記音節ごとの音声パラメータを、当該楽曲を歌唱したユーザを識別するユーザIDと対応付けて、第一記憶装置に記憶するパラメータ登録手段と、

各楽曲についての理想的な歌唱音声に基づく、少なくとも一つの音声パラメータを含む

10

20

30

40

50

模範音声データを、前記楽曲IDと対応付けて第二記憶装置に記憶する模範記憶手段と、
を備え、

前記発声情報取得手段は、

楽曲の一つである対象楽曲の楽譜を表し、音源モジュールから出力される個々の出力音について、少なくとも音高及び演奏開始タイミングが規定された楽譜データを前記演奏情報として取得し、

前記楽譜データは、

前記対象楽曲の曲中において転調していれば、時間軸に沿って前記対象楽曲が転調した時刻を表す転調フラグを含み、

前記パラメータ登録手段は、

取得した楽譜データに基づいて、前記対象楽曲において同一の調が継続される各区間である調同一区間を特定する区間特定手段と、

前記区間特定手段にて特定した各調同一区間に含まれ、それぞれの調同一区間における時間軸に沿った最後の出力音を主音として特定する主音特定手段と、

前記区間特定手段にて特定した調同一区間に含まれる同一音名の出力音の頻度を表す登場音名頻度を、前記主音特定手段で特定した主音の音名を起点として前記調同一区間ごとに導出する音名頻度導出手段と、

各調にて利用可能な音名の分布を表すテンプレートとして調ごとに予め用意した調テンプレートに、前記音名頻度導出手段で導出した各登場音名頻度を照合した結果、最も相関が高い調それぞれを、メタデータとして推定する調推定手段と、

を備え、

前記調推定手段にて推定したメタデータを、前記音声パラメータと対応付けて前記第一記憶装置に記憶することを特徴とするパラメータ抽出装置。

【請求項5】

楽曲を識別する楽曲ID、当該楽曲IDによって識別される楽曲を構成する楽音を表す演奏情報、当該楽曲IDによって識別される楽曲の歌詞を表す歌詞情報、当該歌詞情報によって表される歌詞の発声開始タイミングを示す発声タイミング情報を含む楽曲データから、前記楽曲ID、前記演奏情報、前記歌詞情報、前記発声タイミング情報を取得して前記楽曲データに基づく楽曲を再生する発声情報取得手段と、

前記楽曲データに基づく楽曲の再生中に、前記歌詞の発声開始タイミングで、入力された音声波形を取得する波形取得手段と、

前記音声波形から、前記歌詞を形成する各音節に対する音声波形である音節波形を抽出すると共に、その抽出した各音節波形から、予め規定された少なくとも一つの特徴量である音声パラメータを導出するパラメータ導出手段と、

前記パラメータ導出手段で導出された前記音節ごとの音声パラメータを、当該楽曲を歌唱したユーザを識別するユーザIDと対応付けて、第一記憶装置に記憶するパラメータ登録手段と、

各楽曲についての理想的な歌唱音声に基づく、少なくとも1つの音声パラメータを含む模範音声データを、前記楽曲IDと対応付けて第二記憶装置に記憶する模範記憶手段と、

を備え、

前記パラメータ登録手段は、

前記発声情報取得手段で取得した歌詞情報によって表される文字列を、単語を構成する文字列である単語文字ごとに分割する単語分割手段と、

各単語の性質を表す性質情報を、当該単語の識別情報と対応付けた単語性質テーブルを予め用意し、前記単語分割手段で分割された各単語文字によって表される単語に対応する前記性質情報をメタデータとして、前記単語性質テーブルから抽出するメタデータ抽出手段と

を備え、

前記メタデータ抽出手段にて抽出したメタデータを、前記音声パラメータを対応付けて前記第一記憶装置に記憶することを特徴とするパラメータ抽出装置。

10

20

30

40

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、カラオケで音声合成を実行する音声合成システム、及び音声合成に必要な音声パラメータを音声から抽出するパラメータ抽出装置に関する。

【背景技術】

【0002】

従来、ユーザが歌唱することで入力された歌声音声の音高ピッチを補正して出力するカラオケ装置が知られている（特許文献1参照）。

この特許文献1に記載されたカラオケ装置では、ユーザが歌唱することで入力された歌声音声の音高ピッチを、楽曲中の楽音に用いられている音高の中で最も近い音高に合致させるように補正している。

10

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2003-167587号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

この特許文献1に記載されたカラオケ装置では、音高が補正された歌声音声は、ユーザが発声して当該カラオケ装置に音声を入力している期間中に出力されるものの、音声を入力していない期間には出力されない。

20

【0005】

一般的に、ユーザ自身が歌唱している最中は、補正前の発声音もユーザ自身の耳にて聴取される。このため、演奏音に沿って2種類の歌唱音声を同時に聞き分けて歌唱練習することは困難であり、一般的な歌唱練習は、一旦歌唱した後で、自分の発声音声を該当曲の演奏音とともに聴いて行われる。

【0006】

ところが、特許文献1に記載されたカラオケ装置では、ユーザ自身の歌声音声を当該楽曲の適切な音高で聴取するのは、ユーザ自身が楽曲を歌唱している最中であるため、当該楽曲の歌唱後や、ましてや、別の曲に対しては、再び歌唱しなければ歌唱練習することが困難であるという問題があった。

30

【0007】

そこで、本発明は、ユーザ自身が楽曲を歌唱している期間以外にも、ユーザ自身の歌声音声を当該楽曲の適切な音高で聴取可能とすることを目的とする。

【課題を解決するための手段】

【0008】

上記目的を達成するためになされた本発明は、パラメータ抽出装置と、合成音出力装置とを備えた音声合成システムに関する。

本発明の音声合成システムを構成するパラメータ抽出装置は、発声情報取得手段と、波形取得手段と、パラメータ導出手段と、パラメータ登録手段と、模範記憶手段とを備えている。

40

【0009】

このうち、発声情報取得手段は、楽曲を識別する楽曲ID、当該楽曲IDによって識別される楽曲を構成する楽音を表す演奏情報、当該楽曲IDによって識別される楽曲の歌詞を表す歌詞情報、当該歌詞情報によって表される歌詞の発声開始タイミングを示す発声タイミング情報を含む楽曲データから、楽曲ID、演奏情報、歌詞情報、発声タイミング情報を取得して楽曲データに基づく楽曲を再生する。

【0010】

そして、波形取得手段が、楽曲データに基づく楽曲の再生中に、歌詞の発声開始タイミ

50

ングで、入力された音声波形を取得し、パラメータ導出手段が、音声波形から、歌詞を形成する各音節に対する音声波形である音節波形を抽出すると共に、その抽出した各音節波形から、予め規定された少なくとも一つの特徴量である音声パラメータを導出する。

【0011】

パラメータ登録手段は、パラメータ導出手段で導出された音節ごとの音声パラメータを、当該楽曲を歌唱したユーザを識別するユーザIDと対応付けて、第一記憶装置に記憶する。さらに、模範記憶手段が、各楽曲についての理想的な歌唱音声に基づく、少なくとも1つの音声パラメータを含む模範音声データを、前記楽曲IDと対応付けて第二記憶装置に記憶する。

【0012】

一方、音声合成システムを構成する合成音出力装置は、識別情報取得手段と、パラメータ取得手段と、模範取得手段と、音声合成手段と、出力手段とを備えている。

この合成音出力装置では、識別情報取得手段が、指定された楽曲ID、及びユーザIDを取得する。その取得した楽曲IDと対応付けられた模範歌声データを、模範取得手段が第二記憶装置から取得する。

さらに、パラメータ取得手段は、第一記憶装置に記憶されている音声パラメータの中から、識別情報取得手段で取得した前記ユーザIDと対応付けられた音声パラメータであって、模範取得手段で取得した模範音声データに含まれる楽曲IDと対応付けられた音声パラメータに最も類似する音声パラメータを取得する。

【0013】

音声合成手段は、パラメータ取得手段で取得した音声パラメータを、楽曲IDと対応付けられた音声パラメータに一致するように調整し、その調整された音声パラメータに基づいて音声合成する。すると、その生成された合成音を、出力手段が出力する。

さらに、発声情報取得手段は、楽曲の一つである対象楽曲の楽譜を表し、音源モジュールから出力される個々の出力音について、少なくとも音高及び演奏開始タイミングが規定された楽譜データを前記演奏情報として取得する。楽譜データには、対象楽曲の曲中において転調していれば、時間軸に沿って前記対象楽曲が転調した時刻を表す転調フラグが含まれている。

そして、本発明のパラメータ登録手段は、区間特定手段と、主音特定手段と、音名頻度導出手段と、調推定手段とを備えている。

このうち、区間特定手段は、取得した楽譜データに基づいて、対象楽曲において同一の調が継続される各区間である調同一区間を特定する。主音特定手段は、区間特定手段にて特定した各調同一区間に含まれ、それぞれの調同一区間における時間軸に沿った最後の出力音を主音として特定する。そして、音名頻度導出手段は、区間特定手段にて特定した調同一区間に含まれる同一音名の出力音の頻度を表す登場音名頻度を、主音特定手段で特定した主音の音名を起点として前記調同一区間ごとに導出する。さらに、調推定手段は、各調にて利用可能な音名の分布を表すテンプレートとして調ごとに予め用意した調テンプレートに、音名頻度導出手段で導出した各登場音名頻度を照合した結果、最も相関が高い調それぞれを、メタデータとして推定する。

なお、本発明のパラメータ登録手段は、調推定手段にて推定したメタデータを、音声パラメータと対応付けて第一記憶装置に記憶する。

【0014】

このような音声合成システムによれば、一つの楽曲をユーザが歌唱後に、模範歌唱状態のユーザ音声を聴くことができる。

つまり、ユーザの歌声から導出した音声パラメータを用いて、楽曲における発声開始タイミングかつ発声音高にて音声出力されるように、ユーザ自身の声による音声合成を実施できる。

【0015】

しかも、本発明の音声合成システムにおいては、一つの楽曲についての音声パラメータ及び模範歌声データを生成した後は、他の楽曲について、再び音声パラメータ及び模範歌

10

20

30

40

50

さらに、本発明がパラメータ抽出装置としてなされている場合には、パラメータ登録手段は、区間特定手段と、主音特定手段と、音名頻度導出手段と、調推定手段と、を備え、調推定手段にて推定したメタデータを、音声パラメータと対応付けて第一記憶装置に記憶しても良い。

また、本発明がパラメータ抽出装置としてなされている場合には、パラメータ登録手段は、単語分割手段と、メタデータ抽出手段とを備え、そのメタデータ抽出手段にて抽出したメタデータを、音声パラメータを対応付けて第一記憶装置に記憶しても良い。

【0028】

このようなパラメータ抽出装置によれば、一つの楽曲をユーザが最初に歌唱した際に、音声パラメータを導出すると共に、模範歌声データを生成することができる。

10

そして、本発明のパラメータ抽出装置によれば、第一記憶装置に記憶された音声パラメータ、及び第二記憶装置に記憶された模範歌声データに基づいて、楽曲における発声開始タイミングかつ発声音高にて音声が出力されるように、音声合成装置に音声合成を実施させることができる。

【図面の簡単な説明】

【0029】

【図1】本発明が適用された音声合成システムの全体構成を示すブロック図である。

【図2】パラメータ登録処理の処理手順を示すフローチャートである。

【図3】パラメータデータ登録処理の処理手順を示すフローチャートである。

【図4】第一実施形態におけるメタデータ推定処理の処理手順を示すフローチャートである。

20

【図5】メタデータ推定処理の処理内容を示す図である。

【図6】メタデータ推定処理の処理内容を示す図である。

【図7】メタデータ推定処理の処理内容を示す図である。

【図8】第一実施形態におけるパラメータデータ登録処理の概要を示す図である。

【図9】パラメータデータ及び模範歌声データの概要を示す図である。

【図10】模範歌声データ登録処理の処理手順を示すフローチャートである。

【図11】音声合成処理の処理手順を示すフローチャートである。

【図12】第二実施形態におけるメタデータ推定処理の処理手順を示すフローチャートである。

30

【図13】第二実施形態におけるパラメータデータ登録処理の概要を示す図である。

【発明を実施するための形態】

【0030】

以下に本発明の実施形態を図面と共に説明する。

[第一実施形態]

音声合成システム

図1に示すように、音声合成システム1は、ユーザ(歌唱者)が指定した楽曲(以下、指定楽曲と称す)における理想的な歌唱音声を、当該ユーザ自身の音声にて出力するように、音声合成を実施するシステムである。

【0031】

これを実現するために、音声合成システム1は、音声入力装置10と、情報格納サーバ25と、情報処理装置30と、データ格納サーバ50と、音声出力端末60とを備えている。

40

【0032】

音声入力装置10は、カラオケの用途に用いられる音楽データMDに基づいて楽曲を演奏すると共に、その楽曲の演奏中に音声の入力を受け付ける。情報格納サーバ25は、楽曲ごとに用意された音楽データMDそれぞれを格納すると共に、音声入力装置10を介して入力された音声それぞれのデータである音声波形データSVを格納する。

【0033】

さらに、情報処理装置30は、情報格納サーバ25に格納されている音声波形データS

50

V及び音楽データMDに基づいて、パラメータデータPM及び模範歌声データEDを生成する。

【0034】

なお、ここで言うパラメータデータPMとは、詳しくは後述するが、いわゆるフォルマント合成に用いる、少なくとも一つの音声パラメータを含むデータである。また、模範歌声データEDとは、各楽曲についての理想的な歌唱音声に基づく、少なくとも一つの音声パラメータを含むデータである。この少なくとも一つの音声パラメータには、例えば、発声音声における各音節での基本周波数、メル周波数ケプストラム(MFCC)、パワー、及びそれらの時間差分を含む。

【0035】

データ格納サーバ50は、情報処理装置30にて生成されたパラメータデータPM及び模範歌声データEDを格納する。また、音声出力端末60は、データ格納サーバ50に格納されているパラメータデータPM及び模範歌声データEDに基づいて、指定楽曲を歌唱した場合の理想的な歌唱音声となるように、ユーザ自身の音声により音声合成した合成音を出力する。なお、本実施形態における音声合成システム1は、音声出力端末60を複数台備えていても良い。

情報格納サーバ

情報格納サーバ25は、記憶内容を読み書き可能に構成された記憶装置を中心に構成された装置であり、通信網を介して、音声入力装置10に接続されている。

【0036】

この情報格納サーバ25に格納される音楽データMDは、楽曲MIDIデータDMと、歌詞データ群DLと、ガイドボーカルデータGDとを有し、それぞれ対応する楽曲ごとに対応付けられている。さらに、楽曲がデュエット曲の場合、音楽データMDには、歌詞の男女パートを区別する情報が含まれていても良い。

【0037】

楽曲MIDIデータDMは、周知のMIDI(Musical Instrument Digital Interface)規格によって、一つの楽曲の楽譜を表すデータである。この楽曲MIDIデータDMの各々は、楽曲を区別するデータである識別データと、当該楽曲にて用いられる楽器ごとの楽譜を表す楽譜トラックと、当該楽曲において調が変化する時刻を表す変調フラグとを少なくとも有している。

【0038】

そして、楽譜トラックには、MIDI音源から出力される個々の出力音について、少なくとも、音高(いわゆるノートナンバー)と、音源モジュールが出力音を出力する期間(以下、音符長)とが規定されている。ただし、楽譜トラックの音符長は、当該出力音の出力を開始するまでの当該楽曲の演奏開始からの時間を表す演奏開始タイミング(いわゆるノートオンタイミング)と、当該出力音の出力を終了するまでの当該楽曲の演奏開始からの時間を表す演奏終了タイミング(いわゆるノートオフタイミング)とによって規定されている。

【0039】

なお、楽譜トラックは、楽器種類ごとに用意されている。

一方、歌詞データ群DLは、周知のカラオケ装置を構成する表示装置に表示される歌詞に関するデータであり、楽曲の歌詞を構成する文字(以下、歌詞構成文字とする)を表す歌詞テロップデータDTと、歌詞構成文字の出力タイミングである歌詞出力タイミングを、楽曲MIDIデータDMの演奏と対応付けるタイミング対応関係が規定された歌詞出力データDOとを備えている。

【0040】

具体的に、本実施形態におけるタイミング対応関係は、楽曲MIDIデータDMの演奏を開始するタイミングに、歌詞テロップデータDTの出力を開始するタイミングが対応付けられた上で、特定楽曲の時間軸に沿った各歌詞構成文字の歌詞出力タイミングが、楽曲MIDIデータDMの演奏開始からの経過時間によって規定されている。なお、ここで言

10

20

30

40

50

う経過時間とは、例えば、表示された歌詞構成文字の色替えを実行するタイミングを表す時間であり、色替えの速度によって規定されている。また、ここで言う歌詞構成文字は、歌詞を構成する文字の各々であっても良いし、その文字の各々を時間軸に沿った特定の規則に従って一群とした文節やフレーズであっても良い。

【 0 0 4 1 】

ガイドボーカルデータ G D は、当該楽曲についての理想的な歌唱音声として予め用意された音声データである。ここで言う理想的な歌唱音声とは、当該楽曲を構成する楽音の通りに歌唱したと考えられる音声であり、カラオケ装置に周知の採点機能にて採点した場合に満点近くとなる音声である。

【 0 0 4 2 】

一般的なカラオケ装置においては、ガイドボーカルデータ G D に基づく理想的な歌唱音声は、楽曲 M I D I データ D M に基づく楽器演奏とともに再生される。

音声入力装置

次に、音声入力装置 1 0 について説明する。

【 0 0 4 3 】

音声入力装置 1 0 は、通信部 1 1 と、入力受付部 1 2 と、表示部 1 3 と、音声入力部 1 4 と、音声出力部 1 5 と、音源モジュール 1 6 と、記憶部 1 7 と、制御部 2 0 とを備えている。すなわち、本実施形態における音声入力装置 1 0 は、いわゆる周知のカラオケ装置として構成されている。

【 0 0 4 4 】

このうち、通信部 1 1 は、通信網を介して、音声入力装置 1 0 が外部との間で通信を行う。本実施形態における通信網とは、例えば、公衆無線通信網やネットワーク回線である。

【 0 0 4 5 】

入力受付部 1 2 は、外部からの操作に従って情報や指令の入力を受け付ける入力機器である。本実施形態における入力機器とは、例えば、キーやスイッチ、リモコンの受付部などである。

【 0 0 4 6 】

表示部 1 3 は、少なくとも、文字コードで示される情報を含む画像を表示する表示装置である。本実施形態における表示装置とは、例えば、液晶ディスプレイや C R T などである。また、音声入力部 1 4 は、音を電気信号に変換して制御部 2 0 に入力する装置（いわゆるマイクロホン）である。音声出力部 1 5 は、制御部 2 0 からの電気信号を音に変換して出力する装置（いわゆるスピーカ）である。

【 0 0 4 7 】

さらに、音源モジュール 1 6 は、楽曲 M I D I データ D M に基づいて、音源からの音を模擬した音（即ち、出力音）を出力する装置、例えば、M I D I 音源である。

記憶部 1 7 は、記憶内容を読み書き可能に構成された不揮発性の記憶装置である。本実施形態における記憶装置とは、例えば、ハードディスク装置や、フラッシュメモリなどである。

【 0 0 4 8 】

また、制御部 2 0 は、電源が切断されても記憶内容を保持する必要がある処理プログラムやデータを格納する R O M 2 1 と、処理プログラムやデータを一時的に格納する R A M 2 2 と、R O M 2 1 や R A M 2 2 に記憶された処理プログラムに従って各処理（各種演算）を実行する C P U 2 3 とを少なくとも有した周知のコンピュータを中心に構成されている。

【 0 0 4 9 】

そして、R O M 2 1 には、ユーザによって指定された楽曲（以下、対象楽曲と称す）を演奏するカラオケ演奏処理を制御部 2 0 が実行する処理プログラムが格納されている。本実施形態におけるカラオケ演奏処理では、対象楽曲が演奏されている期間中に、音声入力部 1 4 を介して入力された音声を音声波形データ S V として、当該対象楽曲を識別する楽

10

20

30

40

50

曲ID及び対象楽曲を歌唱したユーザ音声を識別するユーザIDと対応付けて、情報格納サーバ25に格納する。

【0050】

つまり、音声入力装置10では、カラオケ演奏処理に従って、対象楽曲に対応する楽曲MIDIデータDMに基づいて、楽曲を演奏すると共に、対象楽曲に対応する歌詞データ群DLに基づいて歌詞を表示部13に表示する。そして、カラオケ演奏処理の実行中に、音声入力部14を介して入力された音声を音声波形データSVとして、楽曲ID及びユーザIDと対応付けて、情報格納サーバ25に格納する。

【0051】

なお、情報格納サーバ25に格納される音声波形データSVには、発声者の特徴を表す発声者特徴情報も対応付けられる。この発声者特徴情報には、例えば、少なくとも、ユーザの性別を含む。

10

【0052】

データ格納サーバ50は、記憶内容を読み書き可能に構成された記憶装置を中心に構成された装置であり、通信網を介して情報処理装置30に接続されている。

情報処理装置

情報処理装置30は、通信部31と、入力受付部32と、表示部33と、記憶部34と、制御部40とを備えている。

【0053】

このうち、通信部31は、通信網を介して外部との間で通信を行う。入力受付部32は、外部からの操作に従って情報や指令の入力を受け付ける入力機器である。表示部33は、画像を表示する表示装置である。

20

【0054】

記憶部34は、記憶内容を読み書き可能に構成された不揮発性の記憶装置である。制御部40は、ROM41、RAM42、CPU43を少なくとも有した周知のコンピュータを中心に構成されている。

【0055】

さらに、情報処理装置30のROM41には、情報格納サーバ25に格納された音声波形データSVの中で、指定された楽曲ID及びユーザIDが対応付けられた音声波形データSVを取得して、パラメータデータPM、及び模範歌声データEDを生成してデータ格納サーバ50に格納するパラメータ登録処理を制御部40が実行する処理プログラムが格納されている。

30

【0056】

このパラメータ登録処理にてデータ格納サーバ50に格納されるパラメータデータPM及び模範歌声データEDには、メタデータが含まれる。本実施形態におけるメタデータとは、当該音声が発声されたときの感情を含む音声の性質を表すものである。また、ここで言う感情には、例えば、情緒や、雰囲気などを含む。さらに、音声の性質には、感情を推定するために必要な情報を含んでも良い。

パラメータ登録処理

情報処理装置30が実行するパラメータ登録処理は、起動されると、図2に示すように、まず、入力受付部32を介して、ユーザIDを取得する(S110)。すなわち、S110では、ユーザが情報処理装置30へのログインを行う。

40

【0057】

続いて、入力受付部32を介して楽曲IDを取得する(S120)。すなわち、S120では、ユーザが楽曲を指定する。

続いて、S110にて取得したユーザID、及びS120にて取得した楽曲IDに対応する音声波形データSVに基づいて、パラメータデータPMを生成して登録するパラメータデータ登録処理を実行する(S130)。さらに、S120にて取得した楽曲IDに対応する楽曲の模範歌声データEDを生成して登録する模範歌声データ登録処理を実行する(S140)。なお、パラメータデータ登録処理及び模範歌声データ登録処理の詳細につ

50

いては、後述する。

【 0 0 5 8 】

その後、本パラメータ登録処理を終了する。

パラメータデータ登録処理

パラメータデータ登録処理は、図 3 に示すように、起動されると、先の S 1 2 0 にて取得した楽曲 ID に対応する楽曲 M I D I データ DM を取得する (S 2 1 0)。続いて、先の S 1 2 0 にて取得した楽曲 ID に対応する歌詞データ群 DL を取得し (S 2 2 0)、当該楽曲 ID に対応し、かつ先の S 1 1 0 にて指定されたユーザ ID に対応する一つの音声波形データ SV を取得する (S 2 3 0)。

【 0 0 5 9 】

さらに、S 2 3 0 で取得した音声波形データ SV において、当該音声波形データ SV の発声内容に含まれる音節それぞれに対応する区間での音声波形 (以下、音節波形と称す) を特定する (S 2 4 0)。

【 0 0 6 0 】

具体的に、本実施形態の S 2 4 0 では、S 2 1 0 で取得した楽曲 M I D I データ DM のうち、歌唱旋律を表す楽譜トラック (以下、メロディトラックと称す) に規定された各出力音の演奏開始タイミング及び演奏終了タイミングを抽出すると共に、各出力音に対応付けられた歌詞構成文字の音節を特定する。そして、音声波形データ SV において、各出力音の演奏開始タイミングから演奏終了タイミングまでの区間それぞれに対応する区間での音声波形を音節波形として特定する。なお、本実施形態の S 2 4 0 にて特定される音節波形それぞれは、当該音節波形にて発声した音節の内容と対応付けられたものである。

【 0 0 6 1 】

さらに、音節波形それぞれから音声パラメータを導出する (S 2 5 0)。本実施形態の S 2 5 0 にて導出する音声パラメータには、少なくとも、基本周波数、メル周波数ケプストラム (M F C C)、パワー、それらの時間差分を含む。これらの基本周波数、M F C C、パワーの導出方法は、周知であるため、ここでの詳しい説明は省略するが、例えば、基本周波数であれば、音節波形の時間軸に沿った自己相関、音節波形の周波数スペクトルの自己相関、またはケプストラム法などの手法を用いて導出すれば良い。また、M F C C であれば、音節波形に対して時間分析窓を適用して、時間分析窓ごとに周波数解析 (例えば、F F T) をした結果について、周波数ごとの大きさを対数化した結果を、さらに、周波数解析することで導出すれば良い。パワーについては、音節波形に対して時間分析窓を適用して振幅の二乗した結果を時間方向に積分することで導出すれば良い。

【 0 0 6 2 】

続いて、S 2 4 0 にて特定した各音節波形に対するメタデータを推定するメタデータ推定処理を実行する (S 2 6 0)。

メタデータ推定処理

本実施形態におけるメタデータ推定処理は、図 4 に示すように、起動されると、まず、先の S 2 1 0 にて取得した楽曲 M I D I データに基づいて、当該楽曲において同一の調が継続される各区間である調同一区間を特定する (S 3 1 0)。具体的に、本実施形態の S 3 1 0 では、図 5 に示すように、楽曲 M I D I データに含まれる転調フラグに基づき、時間軸に沿って互いに隣接する転調フラグの間の区間を、調同一区間として特定する。

【 0 0 6 3 】

続いて、S 3 1 0 にて特定した調同一区間における主音を特定する (S 3 2 0)。具体的に、本実施形態の S 3 2 0 では、図 6 に示すように、調同一区間において、時間軸に沿って最後の出力音を、当該調同一区間における主音として特定する。本実施形態では、S 3 1 0 にて特定した調同一区間のそれぞれについて、主音を特定する。

【 0 0 6 4 】

そして、S 3 2 0 にて特定した主音の音名を起点とし、当該主音が特定された調同一区間に含まれる出力音それぞれの音名を階級とし、各音名の登場回数を度数としたヒストグラム (以下、登場音名頻度と称す) を導出する (S 3 3 0)。具体的に、本実施形態の S

10

20

30

40

50

330にて導出する登場音名頻度は、図7(A)に示すように、調同一区間に含まれる同一音名の出力音の登場回数(登場頻度)を集計したものである。そして、本実施形態においては、オクターブが異なる出力音であっても、音名が同一であれば、同一音名の出力音として集計する。なお、本実施形態では、各調同一区間について、登場音名頻度を導出する。

【0065】

続いて、S330にて導出した登場音名頻度を、各調にて利用可能な音名の分布を表すテンプレートとして調ごとに予め用意した調テンプレートに照合した結果に基づいて、当該調同一区間における調を特定する(S340)。具体的に、本実施形態のS340では、長調の楽曲にて利用可能な音名の分布を表す長調テンプレート(図7(B)参照)と、短調の楽曲にて利用可能な音名の分布を表す短調テンプレート(図7(C)参照)とを予め用意し、それぞれの調テンプレートにS330にて導出した登場音名頻度を照合する。その結果、最も高い相関を示す調テンプレートに対応する調を、当該調同一区間における調として特定する。なお、本実施形態のS340では、調同一区間のそれぞれについての調を特定する。

10

【0066】

さらに、S340で特定した調同一区間における楽曲の調に対応する音声の性質を、メタデータとして特定する(S350)。具体的に、本実施形態のS350では、調同一区間における調が長調であれば、当該調同一区間での歌詞(即ち、発声内容)が「明るい」という感情を表す音声の性質をメタデータとして特定する。また、調同一区間における調が短調であれば、当該調同一区間での歌詞が「暗い」という感情を表す音声の性質をメタデータとして特定する。なお、本実施形態においては、調同一区間に含まれる全ての音節について、当該調同一区間に対応するメタデータを割り当てている。

20

【0067】

その後、パラメータデータ登録処理のS270へと戻り、S250にて導出した音声パラメータと、S260にて推定したメタデータとを対応する音節ごとに対応付けることで、パラメータデータPMを生成しデータ格納サーバ50に格納するパラメータデータ登録を実行する(S270)。なお、本実施形態のS270にてデータ格納サーバ50に格納されるパラメータデータPMには、音声パラメータ及びメタデータに加えて、発声した音節の内容(種類)や、ユーザID、発声者特徴情報を含む。

30

【0068】

その後、本パラメータデータ登録処理を終了する。

以上説明したように、図8に示すように、本実施形態のパラメータデータ登録処理では、楽曲の演奏期間中に入力された音声波形を処理対象とする。そして、その音声波形に基づく音声波形データSVを、当該楽曲のメロディラインを構成する各出力音の演奏期間に対応する区間(即ち、発声内容に含まれる各音節)ごとに分割して音節波形を生成すると共に、各音節波形から音声パラメータを導出する。

【0069】

これと共に、パラメータデータ登録処理では、当該楽曲において同一の調が継続する期間(即ち、調同一区間)それぞれを特定し、各調同一区間における調(調性)を特定する。そして、その特定した調からイメージされる感情として予め規定された音声の性質をメタデータとして特定する。

40

【0070】

その上で、パラメータデータ登録処理では、対応する音節ごとに、音声パラメータとメタデータとを対応付けてパラメータデータPMを生成し、データ格納サーバ50に格納する。すなわち、データ格納サーバ50に格納されるパラメータデータPMは、図9(A)に示すように、当該楽曲における歌詞構成文字の時間軸に沿った音節の登場順に、メタデータと対応付けられた音声パラメータを含むものである。

【0071】

なお、本実施形態におけるパラメータデータ登録処理は、ユーザごと、かつ楽曲ごとに

50

実施される。したがって、パラメータデータPMは、同一のユーザ、かつ同一の楽曲であっても、当該楽曲を歌唱するごとに、異なるパラメータデータPMとして、データ格納サーバに格納される。

模範歌声データ登録処理

模範歌声データ登録処理は、図10に示すように、起動されると、先のS120にて取得された楽曲IDに対応する楽曲MIDIデータDMを取得する(S810)。続いて、先のS120にて取得した楽曲IDに対応する歌詞データ群DLを取得する(S820)。さらに、当該楽曲IDに対応するガイドボーカルデータGDを取得する(S830)。

【0072】

その取得したガイドボーカルデータGDにおいて、当該ガイドボーカルデータGDの発声内容に含まれる音節それぞれに対応する区間での音声波形を特定する(S840)。本実施形態におけるS840は、パラメータデータ登録処理におけるS240と同様に実施すれば良い。

【0073】

さらに、S840にて特定した音声波形のそれぞれから音声パラメータを導出する(S850)。続いて、上述したメタデータ推定処理を実行する(S860)。

その後、S850にて導出した音声パラメータを、S860にて推定したメタデータと対応付けたデータを模範歌声データEDとして生成し、データ格納サーバ50に格納する模範歌声データ登録を実行する(S870)。なお、本実施形態のS870にて生成される模範歌声データEDは、音声パラメータ及びメタデータに加えて、発声した音節の内容(種類)や、楽曲IDを含む。

【0074】

その後、本模範歌声データ登録処理を終了する。

すなわち、データ格納サーバ50に格納される模範歌声データEDは、図9(B)に示すように、当該楽曲における歌詞構成文字の時間軸に沿った音節の登場順に、メタデータと対応付けられた音声パラメータを含むものである。

【0075】

通常、模範歌声データEDのメタデータ(図9(B))と、歌唱者のメタデータ(図9(A))とは、同じ曲で、曲調が一致していれば、メタデータの種類(明るい、暗い)は一致していることが多い。

【0076】

歌唱者がメジャー調な曲を歌唱すれば、メタデータは「明るい」、マイナー調の曲を歌唱すれば、メタデータは「暗い」となる可能性が高い。

しかし、ユーザが歌唱した曲の曲調と、模範音声として聴きたい曲の曲調と、が異なる場合もある。

【0077】

このため、更に、適切な音高で聴くためには、ガイドボーカルをメジャー調(明るい)、マイナー調(暗い)の2種類を用意して、2種類のメタデータと音声パラメータとを作成すればよい。

【0078】

CPU23は、楽曲IDに対応付けられたガイドボーカル種類を特定し、パラメータ登録処理を、ガイドボーカルの種類毎に行うことで、ガイドボーカルの種類数に応じたパラメータデータ登録が行える。

【0079】

本実施形態における模範歌声データ登録処理は、楽曲ごとに実施される。したがって、模範歌声データEDは、一つの楽曲に対して一つ生成され、データ格納サーバ50に格納される。

【0080】

本実施例において、ガイドボーカルデータGDのメタデータ、音声パラメータは、カラオケ装置側で音楽データMDを少なくとも1回(ガイドボーカルがメジャー調、マイナー

10

20

30

40

50

調の２種類なら、２回）演奏することによって作成している。

【 0 0 8 1 】

予め、音楽データMDに歌詞の発生音高（ガイドボーカル）とともに、ガイドボーカルの模範音声データをも、演奏の進行に同期した歌詞の出力タイミングに対応付けて組み込んでおくならば、ガイドボーカルの模範音声データ登録処理は、行わなくてもよい。

音声出力端末

この音声出力端末60は、図1に示すように、情報受付部61と、表示部62と、音出力部63と、通信部64と、記憶部65と、制御部67とを備えている。本実施形態における音声出力端末60として、例えば、携帯電話や携帯情報端末などの周知の携帯端末や、いわゆるパーソナルコンピュータといった周知の情報処理装置を想定しても良い。

10

【 0 0 8 2 】

このうち、情報受付部61は、入力装置（図示せず）を介して入力された情報を受け付ける。表示部62は、制御部67からの指令に基づいて画像を表示する。音出力部63は、音を出力する周知の装置であり、例えば、PCM音源と、スピーカとを備えている。

【 0 0 8 3 】

通信部64は、通信網を介して音声出力端末60が外部との間で情報通信を行うものである。記憶部65は、記憶内容を読み書き可能に構成された不揮発性の記憶装置である。記憶部65には、各種処理プログラムや各種データが記憶される。

【 0 0 8 4 】

また、制御部67は、ROM、RAM、CPUを少なくとも有した周知のコンピュータ

20

を中心に構成されている。制御部67のROMには、パラメータデータPM及び模範歌声データEDに基づいて、楽曲についての理想的な歌唱音声をユーザ自身の音声により実現するように音声合成した合成音を出力する音声合成処理を、制御部67が実行するための処理プログラムが格納されている。

音声合成処理について

音声合成処理は、音声出力端末60の情報受付部61を介して起動指令が入力されると起動される。

【 0 0 8 5 】

そして、音声合成処理では、図11に示すように、情報受付部61を介してユーザID

30

【 0 0 8 6 】

そして、S520にて取得した楽曲IDに対応する模範歌声データEDを取得する（S530）。さらに、S510にて取得したユーザIDと対応付けられたパラメータデータPMの中で、S530にて取得した模範歌声データEDに含まれる音声パラメータに最も類似するパラメータデータPMを取得する（S540）。なお、S540において、「最も類似」しているものと判定する条件は、例えば、音声パラメータとしての特徴量ごとに導出した相関値の演算結果を、パラメータデータPMごとに導出し、その導出した演算結果の中で最も大きい演算結果に対応する音声パラメータを含むパラメータデータPMを「

40

【 0 0 8 7 】

また、メタデータが複数種類（明るい、暗い）存在する、つまり、楽曲IDに対応する模範歌声データEDが複数あるならば、メタデータが一致している模範歌声データEDを選択すればよい。一致するメタデータ側の音声パラメータを採用することで、更に最適な音声合成が行われる。

【 0 0 8 8 】

続いて、S540にて取得したパラメータデータPMに含まれる音声パラメータを、S530にて取得した模範歌声データEDに含まれる音声パラメータに一致するように調整する（S550）。その調整された音声パラメータに基づいて、音声合成する（S560

50

)。このS560における音声合成は、フォルマント合成による周知の音声合成の手法を用いれば良い。なお、楽曲がデュエット曲である場合には、発声者特徴情報からユーザの性別を特定し、デュエット曲の男女パートに一致する歌詞部分を音声合成しても良い。

【0089】

さらに、S560にて音声合成することによって生成された合成音を音出力部63から出力する(S570)。

本実施形態においては、S530からS570の各ステップを、楽曲での時間軸に沿って順次実行する。具体的には、当該楽曲の楽曲MIDIデータDMに基づいて、各楽音の出力タイミングにて、各音節に対する合成音出力されるように音声合成を実行する。

【0090】

なお、S520にて取得した楽曲IDに対応する楽曲の演奏が終了するまで(S580:NO)、S530からS570を繰り返す。そして、当該楽曲の演奏が終了すると(S580:YES)、本音声合成処理を終了する。

[第一実施形態の効果]

以上説明したように、音声合成システム1によれば、一つの楽曲をユーザが歌唱後に、模範歌唱状態のユーザ音声聴くことができる。つまり、音声合成システム1によれば、ユーザの歌声から導出した音声パラメータを用いて、楽曲における発声開始タイミングかつ発声音高にて音声出力されるように、ユーザ自身の声による音声合成を実施できる。

【0091】

しかも、音声合成システム1においては、一つの楽曲について模範歌声データEDを生成した後は、他の楽曲については、該当する楽曲データMDを演奏するだけで、ユーザ側は毎回歌唱して模範歌声データEDを生成する必要がない。

【0092】

以上のことから、音声合成システム1によれば、ユーザ自身が楽曲を歌唱している期間以外にも、ユーザ自身の歌声音声を当該楽曲も含め、様々な楽曲をユーザ自身の声で、かつ適切な音高で聴取可能とすることができる。

【0093】

なお、本実施形態のメタデータ推定処理では、対象楽曲における各調同一区間の調によって表される可能性が高い歌唱者の感情をメタデータとしている。すなわち、本実施形態のメタデータ推定処理によれば、各調同一区間に対応する歌詞を発声したときの発声者の感情をメタデータとすることができ、しかも、各調同一区間における調を確実に特定することができる。

[第二実施形態]

第二実施形態の音声合成システムは、第一実施形態の音声合成システム1とは、主として、メタデータ推定処理の処理内容が異なる。このため、本実施形態においては、第一実施形態と同様の構成及び処理には、同一の符号を付して説明を省略し、第一実施形態とは異なるメタデータ推定処理を中心に説明する。

メタデータ推定処理

本実施形態のメタデータ推定処理は、図12に示すように、パラメータデータ登録処理のS260にて起動されると、先のS210にて取得した歌詞データ群DLに含まれている歌詞テロップデータDTによって表される歌詞を形態素解析する(S710)。すなわち、本実施形態のS710では、形態素解析を実行することで、歌詞を構成する文字列を、当該歌詞中の単語を構成する文字列である単語文字ごとに分割する。なお、S710にて実行する形態素解析は、周知の処理であるため、ここでの詳しい説明は省略する。

【0094】

続いて、予め用意された単語性質テーブルが格納された単語メタデータデータベース(図中DB)100から、S710の形態素解析した結果である単語ごとに単語性質情報を取得する(S720)。ただし、ここで言う単語性質テーブルとは、各単語の性質を表す単語性質情報を当該単語の識別情報と対応付けたテーブルである。また、ここで言う単語の性質とは、当該単語の意味や、当該単語によって表される感情を含む。

10

20

30

40

50

【 0 0 9 5 】

そして、S 7 2 0にて取得した単語性質情報をメタデータとして、当該単語を発声した区間に割り当てる(S 7 3 0)。

その後、本メタデータ推定処理を終了して、パラメータデータ登録処理へと戻る。

【 0 0 9 6 】

以上説明したように、本実施形態のメタデータ推定処理では、図 1 3 に示すように、対象楽曲の歌詞に対して形態素解析を実行し、対象楽曲の歌詞を、単語を構成する文字列である単語文字ごとに分割する。その上で、予め用意された単語メタデータデータベース 1 0 0 に格納されている単語性質テーブルに含まれる単語性質情報の中から、各単語に対応する単語性質情報を取得し、当該単語性質情報のそれぞれを、対応する音節の音声パラメータに対するメタデータとしている。

10

[第二実施形態の効果]

以上説明したように、本実施形態のメタデータ推定処理によれば、発声者が発声した単語の意味や、当該単語によって表される感情などを、メタデータとすることができる。

[その他の実施形態]

以上、本発明の実施形態について説明したが、本発明は上記実施形態に限定されるものではなく、本発明の要旨を逸脱しない範囲において、様々な態様にて実施することが可能である。

【 0 0 9 7 】

例えば、音声合成システム 1 においては、情報格納サーバ 2 5 が設けられていたが、本発明の音声合成システムにおいては、情報格納サーバ 2 5 は設けられていなくとも良い。この場合、音楽データ M D や音声波形データ S V は、音声入力装置 1 0 の記憶部 1 7 に格納されても良いし、データ格納サーバ 5 0 に格納されても良いし、さらには、情報処理装置 3 0 の記憶部 3 4 に格納されても良い。

20

【 0 0 9 8 】

同様に、音声合成システム 1 においては、データ格納サーバ 5 0 が設けられていたが、音声合成システム 1 においては、データ格納サーバ 5 0 は設けられていなくとも良い。この場合、パラメータデータ P M や模範歌声データ E D は、情報処理装置 3 0 の記憶部 3 4 に格納されても良いし、音声入力装置 1 0 の記憶部 1 7 に格納されても良いし、さらには、情報格納サーバ 2 5 に格納されても良い。

30

【 0 0 9 9 】

なお、上記実施形態における模範歌声データ E D は、模範歌声データ登録処理を実行することで生成されていたが、模範歌声データ E D は、予め音楽データ M D に組み込まれていても良い。この場合、楽曲 I D に対応した理想的な音声波形データ S V r , 及びその音声波形データ S V r に基づく音声パラメータを含むデータが模範歌声データ E D として、組み込まれていても良い。音楽データ M D に組み込む場合には、当然のことながら、模範歌声データ E D を含む、1 つの音楽データ M D を記憶すればよい。

【 0 1 0 0 】

また、音声合成システム 1 は、音声入力装置 1 0 と、情報処理装置 3 0 と、音声出力端末 6 0 とを別個に備えていたが、これらの音声入力装置 1 0 と、情報処理装置 3 0 と、音声出力端末 6 0 とは共通した装置であっても良い。すなわち、音声入力装置 1 0 (即ち、カラオケ装置) は、情報処理装置 3 0 、及び音声出力端末 6 0 の少なくとも一方を兼ねても良い。

40

[実施形態と特許請求の範囲との対応関係]

最後に、上記実施形態の記載と、特許請求の範囲の記載との関係を説明する。

【 0 1 0 1 】

上記実施形態におけるカラオケ演奏処理が、本発明における発声情報取得手段に相当し、パラメータデータ登録処理における S 2 3 0 が、本発明における波形取得手段に相当し、S 2 5 0 が、本発明におけるパラメータ導出手段に相当し、S 2 7 0 が、本発明におけるパラメータ登録手段に相当し、パラメータ登録処理における S 1 4 0 が、本発明にお

50

る模範記憶手段に相当する。

【0102】

さらに、上記実施形態における音声合成処理におけるS510、S520が、本発明における識別情報取得手段に相当し、S540が、本発明におけるパラメータ取得手段に相当し、S530が、本発明における模範取得手段に相当する。そして、音声合成処理におけるS550、S560が、本発明における音声合成手段に相当し、S570が、本発明における出力手段に相当する。

【0103】

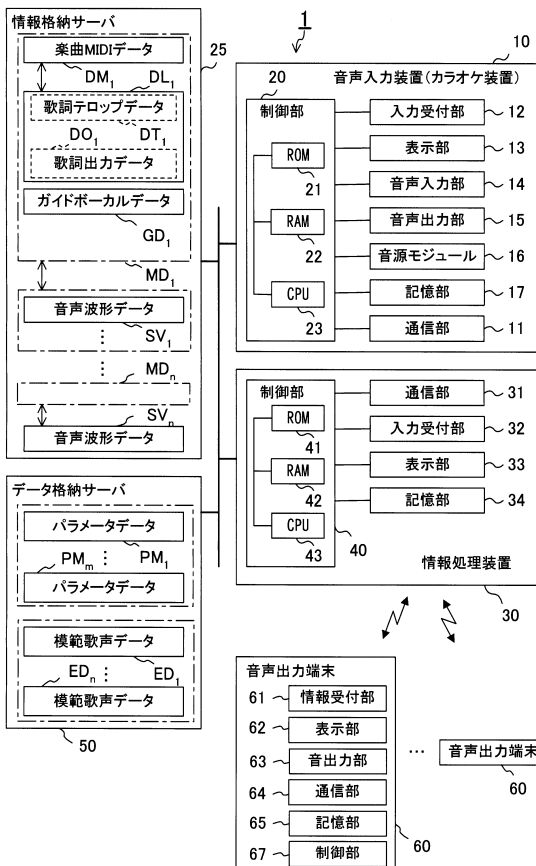
また、上記第一実施形態のメタデータ推定処理におけるS310が、特許請求の範囲の記載における区間特定手段に相当し、S320が、特許請求の範囲の記載における主音特定手段に相当し、S330が、特許請求の範囲の記載における音名頻度導出手段に相当し、S340、S350が、調推定手段に相当する。さらに、上記第二実施形態のメタデータ推定処理におけるS710が、特許請求の範囲の記載における単語分割手段に相当し、S720、S720が、メタデータ抽出手段に相当する。

【符号の説明】

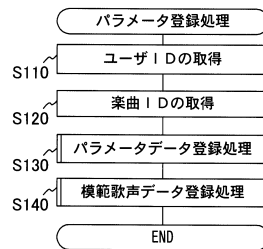
【0104】

- 1 ... 音声合成システム
- 10 ... 音声入力装置
- 11 ... 通信部
- 12 ... 入力受付部
- 13 ... 表示部
- 14 ... 音声入力部
- 15 ... 音声出力部
- 16 ... 音源モジュール
- 17 ... 記憶部
- 20 ... 制御部
- 21, 41 ... ROM
- 22, 42 ... RAM
- 23, 43 ... CPU
- 25 ... MIDI格納サーバ
- 30 ... 情報処理装置
- 31 ... 通信部
- 32 ... 入力受付部
- 33 ... 表示部
- 34 ... 記憶部
- 40 ... 制御部
- 50 ... データ格納サーバ
- 60 ... 音声出力端末
- 61 ... 情報受付部
- 62 ... 表示部
- 63 ... 音出力部
- 64 ... 通信部
- 65 ... 記憶部
- 67 ... 制御部
- 100 ... 単語メタデータデータベース

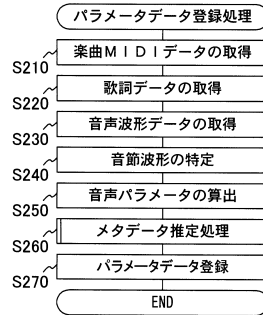
【図1】



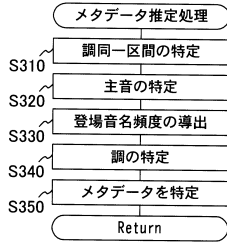
【図2】



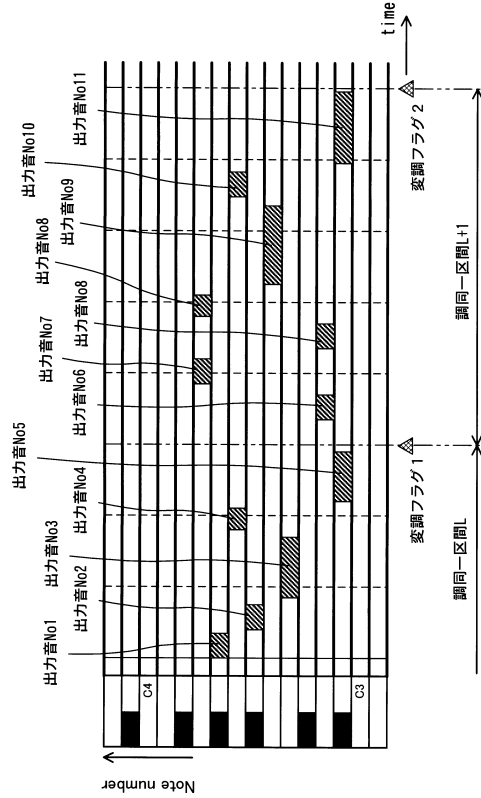
【図3】



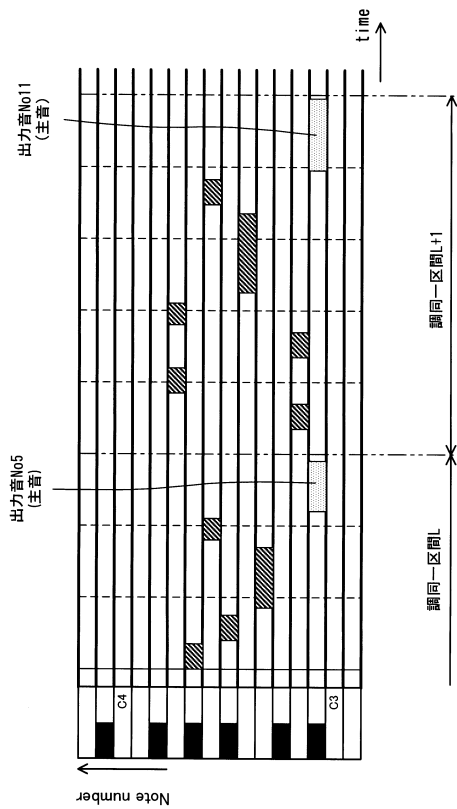
【図4】



【図5】



【図6】



【図9】

ユーザ ID	メタデータ	音節	音声パラメータ						
39	明るい	/a/	1.3	2.9	0.78	3.39	5.08	
	⋮	/u/	1.6	3.2	0.72	4.91	1.72	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
40	明るい	/a/	1.5	3.1	0.82	3.45	4.85	
	⋮	/u/	1.7	2.8	0.68	5.11	1.69	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

(A)

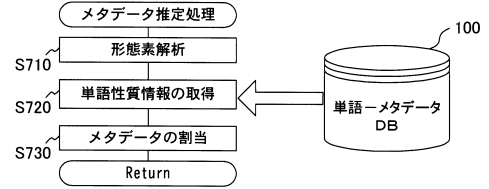
楽曲 ID	メタデータ	音節	音声パラメータ						
12345	明るい	/a/	1.3	2.9	0.78	3.50	6.08	
	⋮	/u/	1.6	3.2	0.72	4.50	1.70	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
12346	明るい	/a/	1.5	3.1	0.82	3.50	5.05	
	⋮	/u/	1.7	2.8	0.70	5.50	1.80	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

(B)

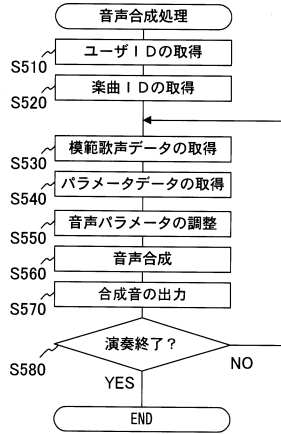
【図10】



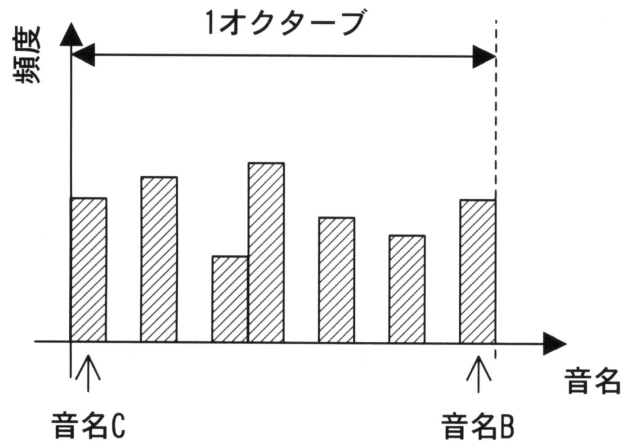
【図12】



【図11】

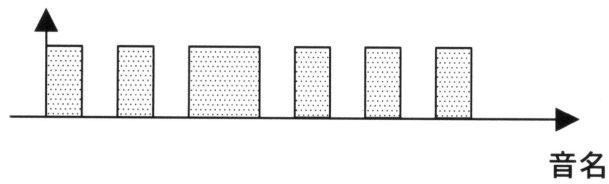


【図7】



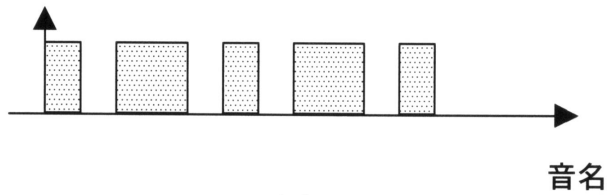
(A)

長調テンプレート



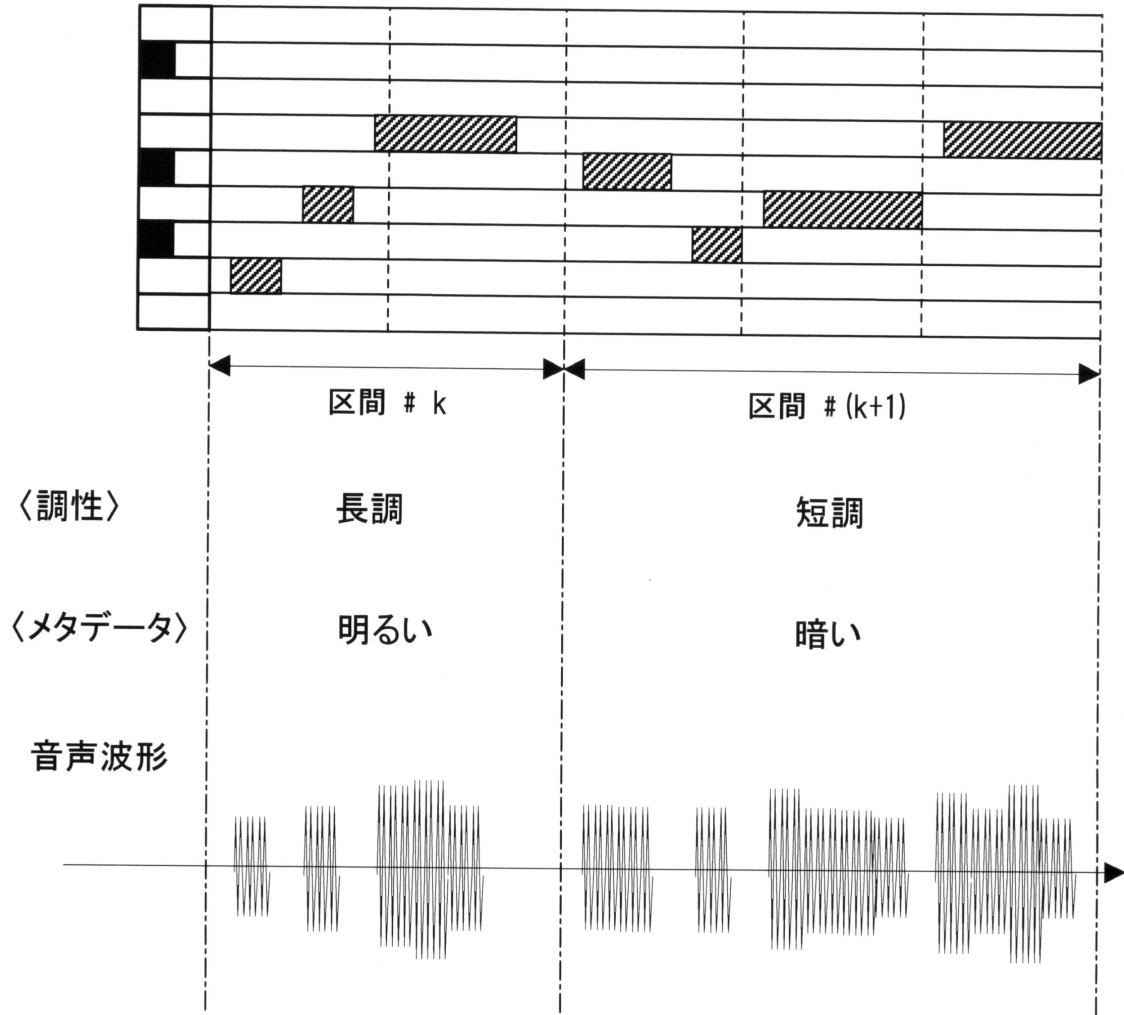
(B)

短調テンプレート

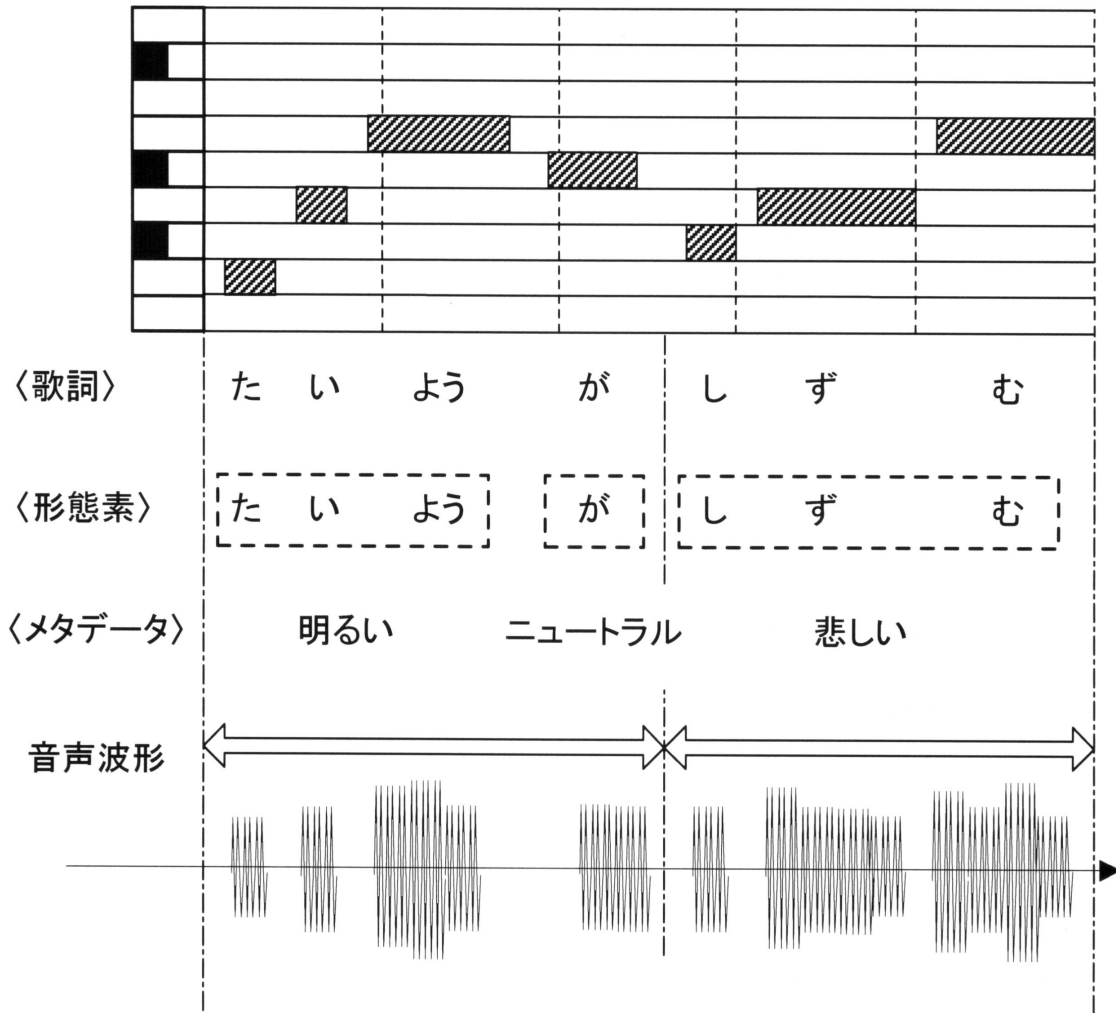


(C)

【図8】



【図13】



フロントページの続き

(72)発明者 久野 琢磨

愛知県名古屋市瑞穂区苗代町15番1号 ブラザー工業株式会社内

審査官 富澤 直樹

(56)参考文献 特開2009-244789(JP,A)

特開2007-193151(JP,A)

特開2005-004107(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10K 15/04

G09B 15/00

G10L 13/00

G10L 13/033