(19)日本国特許庁(JP)　　(12)公開特許公報(A)　　(11)公開番号
特開2022-17561
(P2022-17561A)
(43)公開日　令和4年1月25日(2022.1.25)

(51)国際特許分類　　FI
G10L 13/00 (2006.01)　　G10L 13/00 100Y

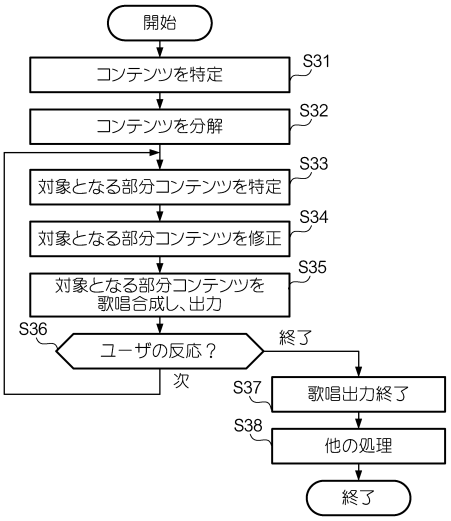
審査請求　有　請求項の数　9　OL　（全28頁）

(21)出願番号　特願2021-183657(P2021-183657)
(22)出願日　令和3年11月10日(2021.11.10)
(62)分割の表示　特願2017-116831(P2017-116831)の分割
原出願日　平成29年6月14日(2017.6.14)

(71)出願人　000004075
ヤマハ株式会社
静岡県浜松市中区中沢町１０番１号
(74)代理人　110000752
特許業務法人朝日特許事務所
(72)発明者　倉光 大樹
静岡県浜松市中区中沢町１０番１号 ヤマハ株式会社内
(72)発明者　奈良 頌子
静岡県浜松市中区中沢町１０番１号 ヤマハ株式会社内
(72)発明者　宮木 強
静岡県浜松市中区中沢町１０番１号 ヤマハ株式会社内
(72)発明者　椎原 浩雅
最終頁に続く

(54)【発明の名称】　情報処理装置、歌唱音声の出力方法、及びプログラム

(57)【要約】
【課題】ユーザとのインタラクションに応じて歌唱音声を出力する。
【解決手段】歌唱音声の出力方法は、コンテンツに含まれる文字列を分解して得られた複数の部分コンテンツの中から第１の部分コンテンツを特定するステップ（Ｓ３１）と、第１の部分コンテンツに含まれる文字列を用いて合成された歌唱音声を出力するステップ（Ｓ３５）と、歌唱音声に対するユーザの反応を受け付けるステップ（Ｓ３６）と、反応に応じて、第１の部分コンテンツに続く第２の部分コンテンツに含まれる文字列を用いて合成された歌唱音声を出力するステップ（Ｓ３５）とを有する。
【選択図】図１３



【特許請求の範囲】
【請求項１】
コンテンツに含まれる文字列を分解して得られた複数の部分コンテンツの中から第１の部分コンテンツを特定するステップと、
前記第１の部分コンテンツに含まれる文字列を用いて合成された歌唱音声を出力するステップと、
前記歌唱音声に対するユーザの反応を受け付けるステップと、
前記反応に応じて、前記第１の部分コンテンツに続く第２の部分コンテンツに含まれる文字列を用いて合成された歌唱音声を出力するステップと
を有する歌唱音声の出力方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ユーザの入力に対し歌唱を含む音声を用いて応答する技術に関する。
【背景技術】
【０００２】
ユーザの指示に応じて楽曲を出力する技術が知られている。例えば特許文献１は、ユーザの状況や嗜好に応じて楽曲の雰囲気を変える技術を開示している。特許文献２は、運動体の状態に応じた楽音を出力する装置において、飽きの来ない独特な選曲をする技術を開示している。
【先行技術文献】
【特許文献】
【０００３】
【特許文献１】特開２００６－８５０４５号公報
【特許文献２】特許第４４９６９９３号公報
【発明の開示】
【発明が解決しようとする課題】
【０００４】
特許文献１及び２はいずれも、ユーザとのインタラクションに応じて歌唱音声を出力するものではなかった。
これに対し本発明は、ユーザとのインタラクションに応じて歌唱音声を出力する技術を提供する。
【課題を解決するための手段】
【０００５】
本発明は、コンテンツに含まれる文字列を分解して得られた複数の部分コンテンツの中から第１の部分コンテンツを特定するステップと、前記第１の部分コンテンツに含まれる文字列を用いて合成された歌唱音声を出力するステップと、前記歌唱音声に対するユーザの反応を受け付けるステップと、前記反応に応じて、前記第１の部分コンテンツに続く第２の部分コンテンツに含まれる文字列を用いて合成された歌唱音声を出力するステップとを有する歌唱音声の出力方法を提供する。
【０００６】
この歌唱音声の出力方法は、前記反応に応じて、前記第２の部分コンテンツに含まれる文字列を用いた歌唱合成に用いられる要素を決定するステップを有してもよい。
【０００７】
前記要素は、前記歌唱合成のパラメータ、メロディ、若しくはテンポ、又は前記歌唱音声における伴奏のアレンジを含んでもよい。
【０００８】
前記歌唱音声の合成は、複数のデータベースの中から選択された少なくとも１つのデータベースに記録された素片を用いて行われ、この歌唱音声の出力方法は、前記反応に応じて、前記第２の部分コンテンツに含まれる文字列を用いた歌唱合成の際に用いられるデータ

ベースを選択するステップを有してもよい。
【０００９】
前記歌唱音声の合成は、複数のデータベースの中から選択された複数のデータベースに記録された素片を用いて行われ、前記データベースを選択するステップにおいて、複数のデータベースが選択され、この歌唱音声の出力方法は、前記複数のデータベースの利用比率を、前記反応に応じて決定するステップを有してもよい。
【００１０】
この歌唱音声の出力方法は、前記第１の部分コンテンツに含まれる文字列の一部を他の文字列に置換するステップを有し、前記歌唱音声を出力するステップにおいて、一部が前記他の文字列に置換された前記第１の部分コンテンツに含まれる文字列を用いて合成された歌唱音声が出力されてもよい。
【００１１】
前記他の文字列と前記置換の対象となる文字列とは、音節数又はモーラ数が同じであってもよい。
【００１２】
この歌唱音声の出力方法は、前記反応に応じて、前記第２の部分コンテンツの一部を他の文字列に置換するステップを有し、前記歌唱音声を出力するステップにおいて、一部が前記他の文字列に置換された前記第２の部分コンテンツに含まれる文字列を用いて合成された歌唱音声が出力されてもよい。
【００１３】
この歌唱音声の出力方法は、前記第１の部分コンテンツに含まれる文字列が示す事項に応じた時間長となるよう合成された歌唱音声を、前記第１の部分コンテンツの歌唱音声と前記第２の部分コンテンツの歌唱音声との間に出力するステップを有してもよい。
【００１４】
この歌唱音声の出力方法は、前記第１の部分コンテンツに含まれる第１文字列が示す事項に応じた第２文字列を用いて合成された歌唱音声を、当該第１の部分コンテンツの歌唱音声の出力後、当該第１文字列が示す事項に応じた時間長に応じたタイミングで出力するステップを有してもよい。
【００１５】
また、本発明は、コンテンツに含まれる文字列を分解して得られた複数の部分コンテンツの中から第１の部分コンテンツを特定する特定部と、前記第１の部分コンテンツに含まれる文字列を用いて合成された歌唱音声を出力する出力部と、前記歌唱音声に対するユーザの反応を受け付ける受け付け部とを有し、前記出力部は、前記反応に応じて、前記第１の部分コンテンツに続く第２の部分コンテンツに含まれる文字列を用いて合成された歌唱音声を出力する情報処理システムを提供する。
【発明の効果】
【００１６】
本発明によれば、ユーザとのインタラクションに応じて歌唱音声を出力することができる。
【図面の簡単な説明】
【００１７】
【図１】一実施形態に係る音声応答システム１の概要を示す図。
【図２】音声応答システム１の機能の概要を例示する図。
【図３】入出力装置１０のハードウェア構成を例示する図。
【図４】応答エンジン２０及び歌唱合成エンジン３０のハードウェア構成を例示する図。
【図５】学習機能５１に係る機能構成を例示する図。
【図６】学習機能５１に係る動作の概要を示すフローチャート。
【図７】学習機能５１に係る動作を例示するシーケンスチャート。
【図８】分類テーブル５１６１を例示する図。
【図９】歌唱合成機能５２に係る機能構成を例示する図。

【図１０】歌唱合成機能５２に係る動作の概要を示すフローチャート。
【図１１】歌唱合成機能５２に係る動作を例示するシーケンスチャート。
【図１２】応答機能５３に係る機能構成を例示する図。
【図１３】応答機能５３に係る動作を例示するフローチャート。
【図１４】音声応答システム１の動作例１を示す図。
【図１５】音声応答システム１の動作例２を示す図。
【図１６】音声応答システム１の動作例３を示す図。
【図１７】音声応答システム１の動作例４を示す図。
【図１８】音声応答システム１の動作例５を示す図。
【図１９】音声応答システム１の動作例６を示す図。
【図２０】音声応答システム１の動作例７を示す図。
【図２１】音声応答システム１の動作例８を示す図。
【図２２】音声応答システム１の動作例９を示す図。
【図２３】音声応答システム１の動作例１０を示す図。
【図２４】音声応答システム１の動作例１１を示す図。
【発明を実施するための形態】
【００１８】
１．システム概要
図１は、一実施形態に係る音声応答システム１の概要を示す図である。音声応答システム１は、ユーザが声によって入力（又は指示）を行うと、それに対し自動的に音声による応答を出力するシステムであり、いわゆるＡＩ（Artificial Intelligence）音声アシスタントである。以下、ユーザから音声応答システム１に入力される音声を「入力音声」といい、入力音声に対し音声応答システム１から出力される音声を「応答音声」という。特にこの例において、音声応答は歌唱を含む。すなわち、音声応答システム１は、歌唱合成システムの一例である。例えば、音声応答システム１に対しユーザが「何か歌って」と話しかけると、音声応答システム１は自動的に歌唱を合成し、合成された歌唱を出力する。
【００１９】
音声応答システム１は、入出力装置１０、応答エンジン２０、及び歌唱合成エンジン３０を含む。入出力装置１０は、マンマシンインターフェースを提供する装置であり、ユーザからの入力音声を受け付け、その入力音声に対する応答音声を出力する装置である。応答エンジン２０は、入出力装置１０により受け付けられた入力音声を分析し、応答音声を生成する。この応答音声は、少なくとも一部に歌唱音声を含む。歌唱合成エンジン３０は、応答音声に用いられる歌唱音声を合成する。
【００２０】
図２は、音声応答システム１の機能の概要を例示する図である。音声応答システム１は、学習機能５１、歌唱合成機能５２、及び応答機能５３を有する。応答機能５３は、ユーザの入力音声を分析し、分析結果に基づいて応答音声を提供する機能であり、入出力装置１０及び応答エンジン２０により提供される。学習機能５１は、ユーザの入力音声からユーザの嗜好を学習する機能であり、歌唱合成エンジン３０により提供される。歌唱合成機能５２は、応答音声に用いられる歌唱音声を合成する機能であり、歌唱合成エンジン３０により提供される。学習機能５１、歌唱合成機能５２、及び応答機能５３の関係は以下のとおりである。学習機能５１は、応答機能５３により得られた分析結果を用いてユーザの嗜好を学習する。歌唱合成機能５２は、学習機能５１によって行われた学習に基づいて歌唱音声を合成する。応答機能５３は、歌唱合成機能５２により合成された歌唱音声を用いた応答をする。各機能の詳細は後述する。
【００２１】
図３は、入出力装置１０のハードウェア構成を例示する図である。入出力装置１０は、マイクロフォン１０１、入力信号処理部１０２、出力信号処理部１０３、スピーカ１０４、ＣＰＵ（Central Processing Unit）１０５、センサー１０６、モータ１０７、及びネットワークＩＦ１０８を有する。マイクロフォン１０１はユーザの音声を電気信号（入

力音信号）に変換する装置である。入力信号処理部１０２は、入力音信号に対しアナログ／デジタル変換等の処理を行い、入力音声を示すデータ（以下「入力音声データ」という）を出力する装置である。出力信号処理部１０３は、応答音声を示すデータ（以下「応答音声データ」という）に対しデジタル／アナログ変換等の処理を行い、出力音信号を出力する装置である。スピーカ１０４は、出力音信号を音に変換する（出力音信号に基づいて音を出力する）装置である。ＣＰＵ１０５は、入出力装置１０の他の要素を制御する装置であり、メモリー（図示略）からプログラムを読み出して実行する。センサー１０６は、ユーザの位置（入出力装置１０から見たユーザの方向）を検知するセンサーであり、一例としては赤外線センサー又は超音波センサーである。モータ１０７は、ユーザのいる方向に向くように、マイクロフォン１０１及びスピーカ１０４の少なくとも一方の向きを変化させる。一例において、マイクロフォン１０１がマイクロフォンアレイであり、ＣＰＵ１０５が、マイクロフォンアレイにより収音された音に基づいてユーザのいる方向を検知してもよい。ネットワークＩＦ１０８は、ネットワーク（例えばインターネット）を介した通信を行うためのインターフェースであり、例えば、所定の無線通信規格（例えばいわゆるＷｉＦｉ（登録商標））に従った通信を行うためのアンテナ及びチップセットを含む。

【００２２】

図４は、応答エンジン２０及び歌唱合成エンジン３０のハードウェア構成を例示する図である。応答エンジン２０は、ＣＰＵ２０１、メモリー２０２、ストレージ２０３、及び通信ＩＦ２０４を有するコンピュータ装置である。ＣＰＵ２０１は、プログラムに従って各種の演算を行い、コンピュータ装置の他の要素を制御する。メモリー２０２は、ＣＰＵ２０１がプログラムを実行する際のワークエリアとして機能する主記憶装置であり、例えばＲＡＭ（Random Access Memory）を含む。ストレージ２０３は、各種のプログラム及びデータを記憶する不揮発性の補助記憶装置であり、例えばＨＤＤ（Hard Disk Drive）又はＳＳＤ（Solid State Drive）を含む。通信ＩＦ２０４は、所定の通信規格（例えばEthernet）に従った通信を行うためのコネクタ及びチップセットを含む。この例において、ストレージ２０３は、コンピュータ装置を音声応答システム１における応答エンジン２０として機能させるためのプログラム（以下「応答プログラム」という）を記憶している。ＣＰＵ２０１が応答プログラムを実行することにより、コンピュータ装置は応答エンジン２０として機能する。応答エンジン２０は、例えばいわゆるＡＩである。

【００２３】

歌唱合成エンジン３０は、ＣＰＵ３０１、メモリー３０２、ストレージ３０３、及び通信ＩＦ３０４を有するコンピュータ装置である。各要素の詳細は応答エンジン２０と同様である。この例において、ストレージ３０３は、コンピュータ装置を音声応答システム１における歌唱合成エンジン３０として機能させるためのプログラム（以下「歌唱合成プログラム」という）を記憶している。ＣＰＵ３０１が歌唱合成プログラムを実行することにより、コンピュータ装置は歌唱合成エンジン３０として機能する。

【００２４】

この例において、応答エンジン２０及び歌唱合成エンジン３０は、インターネット上において、いわゆるクラウドサービスとして提供される。なお、応答エンジン２０及び歌唱合成エンジン３０は、クラウドコンピューティングによらないサービスであってもよい。以下、学習機能５１、歌唱合成機能５２、及び応答機能５３のそれぞれについて、その機能の詳細及び動作を説明する。

【００２５】

２．学習機能

２－１．構成

図５は、学習機能５１に係る機能構成を例示する図である。学習機能５１に係る機能要素として、音声応答システム１は、音声分析部５１１、感情推定部５１２、楽曲解析部５１３、歌詞抽出部５１４、嗜好分析部５１５、記憶部５１６、及び処理部５１０を有する。また、入出力装置１０は、ユーザの入力音声を受け付ける受け付け部、及び応答音声を出力する出力部として機能する。

【００２６】
音声分析部５１１は、入力音声を分析する。ここでいう分析は、応答音声を生成するために用いられる情報を入力音声から取得する処理をいい、具体的には、入力音声をテキスト化（すなわち文字列に変換）する処理、得られたテキストからユーザの要求を判断する処理、ユーザの要求に対してコンテンツを提供するコンテンツ提供部６０を特定する処理、特定されたコンテンツ提供部６０に対し指示を行う処理、コンテンツ提供部６０からデータを取得する処理、取得したデータを用いて応答を生成する処理を含む。この例において、コンテンツ提供部６０は、音声応答システム１の外部システムである。コンテンツ提供部６０は、少なくとも、楽曲等のコンテンツを音として再生するためのデータ（以下「楽曲データ」という）を出力するサービス（例えば、楽曲のストリーミングサービス又はネットラジオ）を提供するコンピュータリソースであり、例えば、音声応答システム１の外部サーバである。
【００２７】
楽曲解析部５１３は、コンテンツ提供部６０から出力される楽曲データを解析する。楽曲データの解析とは、楽曲の特徴を抽出する処理をいう。楽曲の特徴は、例えば、曲調、リズム、コード進行、テンポ、及びアレンジの少なくとも１つを含む。特徴の抽出には公知の技術が用いられる。
【００２８】
歌詞抽出部５１４は、コンテンツ提供部６０から出力される楽曲データから歌詞を抽出する。一例において、楽曲データは、音データに加えメタデータを含む。音データは、楽曲の信号波形を示すデータであり、例えば、ＰＣＭ（Pulse Code Modulation）データ等の非圧縮データ、又はＭＰ３データ等の圧縮データを含む。メタデータはその楽曲に関連する情報を含むデータであり、例えば、楽曲タイトル、実演者名、作曲者名、作詞者名、アルバムタイトル、及びジャンル等の楽曲の属性、並びに歌詞等の情報を含む。歌詞抽出部５１４は、楽曲データに含まれるメタデータから、歌詞を抽出する。楽曲データがメタデータを含まない場合、歌詞抽出部５１４は、音データに対し音声認識処理を行い、音声認識により得られたテキストから歌詞を抽出する。
【００２９】
感情推定部５１２は、ユーザの感情を推定する。この例において、感情推定部５１２は、入力音声からユーザの感情を推定する。感情の推定には公知の技術が用いられる。一例において、感情推定部５１２は、音声応答システム１が出力する音声における（平均）音高と、それに対するユーザの応答の音高との関係に基づいてユーザの感情を推定してもよい。あるいは、感情推定部５１２は、音声分析部５１１によりテキスト化された入力音声、又は分析されたユーザの要求に基づいてユーザの感情を推定してもよい。
【００３０】
嗜好分析部５１５は、ユーザが再生を指示した楽曲の再生履歴、解析結果、及び歌詞、並びにその楽曲の再生を指示したときのユーザの感情のうち少なくとも１つを用いて、ユーザの嗜好を示す情報（以下「嗜好情報」という）を生成する。嗜好分析部５１５は、生成された嗜好情報を用いて、記憶部５１６に記憶されている分類テーブル５１６１を更新する。分類テーブル５１６１は、ユーザの嗜好を記録したテーブル（又はデータベース）であり、例えば、ユーザ毎かつ感情毎に、楽曲の特徴（例えば、音色、曲調、リズム、コード進行、及びテンポ）、楽曲の属性（実演者名、作曲者名、作詞者名、及びジャンル）、及び歌詞を記録したものである。記憶部５１６は、歌唱合成に用いるパラメータをユーザと対応付けて記録したテーブルから、トリガを入力したユーザに応じたパラメータを読み出す読み出し部の一例である。なおここで、歌唱合成に用いるパラメータとは、歌唱合成の際に参照されるデータをいい、分類テーブル５１６１の例では、音色、曲調、リズム、コード進行、テンポ、実演者名、作曲者名、作詞者名、ジャンル、及び歌詞を含む概念である。
【００３１】
２－２．動作

図６は、学習機能５１に係る音声応答システム１の動作の概要を示すフローチャートである。ステップＳ１１において、音声応答システム１は、入力音声を分析する。ステップＳ１２において、音声応答システム１は、入力音声により指示された処理を行う。ステップＳ１３において、音声応答システム１は、入力音声が学習の対象となる事項を含むか判断する。入力音声が学習の対象となる事項を含むと判断された場合（Ｓ１３：ＹＥＳ）、音声応答システム１は、処理をステップＳ１４に移行する。入力音声が学習の対象となる事項を含まないと判断された場合（Ｓ１３：ＮＯ）、音声応答システム１は、処理をステップＳ１８に移行する。ステップＳ１４において、音声応答システム１は、ユーザの感情を推定する。ステップＳ１５において、音声応答システム１は、再生が指示された楽曲を解析する。ステップＳ１６において、音声応答システム１は、再生が指示された楽曲の歌詞を取得する。ステップＳ１７において、音声応答システム１は、ステップＳ１４～Ｓ１６において得られた情報を用いて、分類テーブルを更新する。

【００３２】

ステップＳ１８以降の処理は学習機能５１すなわち分類テーブルの更新と直接は関係ないが、分類テーブルを用いる処理を含むので説明する。ステップＳ１８において、音声応答システム１は、入力音声に対する応答音声を生成する。このとき、必要に応じて分類テーブルが参照される。ステップＳ１９において、音声応答システム１は、応答音声を出力する。以下、学習機能５１に係る音声応答システム１の動作をより詳細に説明する。

【００３３】

図７は、学習機能５１に係る音声応答システム１の動作を例示するシーケンスチャートである。ユーザは、例えば音声応答システム１の加入時又は初回起動時に、音声応答システム１に対しユーザ登録を行う。ユーザ登録は、例えば、ユーザ名（又はログインＩＤ）及びパスワードの設定を含む。図７のシーケンスの開始時点において入出力装置１０は起動しており、ユーザのログイン処理が完了している。すなわち、音声応答システム１において、入出力装置１０を使用しているユーザが特定されている。また、入出力装置１０は、ユーザの音声入力（発声）を待ち受けている状態である。なお、音声応答システム１がユーザを特定する方法はログイン処理に限定されない。例えば、音声応答システム１は、入力音声に基づいてユーザを特定してもよい。

【００３４】

ステップＳ１０１において、入出力装置１０は、入力音声を受け付ける。入出力装置１０は、入力音声をデータ化し、音声データを生成する。音声データは、入力音声の信号波形を示す音データ及びヘッダを含む。ヘッダには、入力音声の属性を示す情報が含まれる。入力音声の属性は、例えば、入出力装置１０を特定するための識別子、その音声を発したユーザのユーザ識別子（例えば、ユーザ名又はログインＩＤ）、及びその音声を発した時刻を示すタイムスタンプを含む。ステップＳ１０２において、入出力装置１０は、入力音声を示す音声データを音声分析部５１１に出力する。

【００３５】

ステップＳ１０３において、音声分析部５１１は、音声データを用いて入力音声を分析する。この分析において、音声分析部５１１は、入力音声が学習の対象となる事項を含むか判断する。この例において学習の対象となる事項とは、楽曲を特定する事項をいい、具体的には楽曲の再生指示である。

【００３６】

ステップＳ１０４において、処理部５１０は、入力音声により指示された処理を行う。処理部５１０が行う処理は、例えば楽曲のストリーミング再生である。この場合、コンテンツ提供部６０は複数の楽曲データが記録された楽曲データベースを有する。処理部５１０は、指示された楽曲の楽曲データを楽曲データベースから読み出す。処理部５１０は、読み出した楽曲データを、入力音声の送信元の入出力装置１０に送信する。別の例において、処理部５１０が行う処理は、ネットラジオの再生である。この場合、コンテンツ提供部６０は、ラジオ音声のストリーミング放送を行う。処理部５１０は、コンテンツ提供部６０から受信したストリーミングデータを、入力音声の送信元の入出力装置１０に送信する

。
【００３７】
ステップＳ１０３において入力音声が学習の対象となる事項を含むと判断された場合、処理部５１０はさらに、分類テーブルを更新するための処理を行う（ステップＳ１０５）。この例において、分類テーブルを更新するための処理には、感情推定部５１２に対する感情推定の要求（ステップＳ１０５１）、楽曲解析部５１３に対する楽曲解析の要求（ステップＳ１０５２）、及び歌詞抽出部５１４に対する歌詞抽出の要求（ステップＳ１０５３）を含む。
【００３８】
感情推定が要求されると、感情推定部５１２は、ユーザの感情を推定し（ステップＳ１０６）、推定した感情を示す情報（以下「感情情報」という）を、要求元である処理部５１０に出力する（ステップＳ１０７）。この例において、感情推定部５１２は、入力音声を用いてユーザの感情を推定する。感情推定部５１２は、例えば、テキスト化された入力音声に基づいて感情を推定する。一例において、感情を示すキーワードがあらかじめ定義されており、テキスト化された入力音声がこのキーワードを含んでいた場合、感情推定部５１２は、ユーザがその感情であると判断する（例えば、「クソッ」というキーワードが含まれていた場合、ユーザの感情が「怒り」であると判断する）。別の例において、感情推定部５１２は、入力音声の音高、音量、速度又はこれらの時間変化に基づいて感情を推定する。一例において、入力音声の平均音高がしきい値よりも低い場合、感情推定部５１２はユーザの感情が「悲しい」であると判断する。別の例において、感情推定部５１２は、音声応答システム１が出力する音声における（平均）音高と、それに対するユーザの応答の音高との関係に基づいてユーザの感情を推定してもよい。具体的には、音声応答システム１が出力する音声の音高が高いにもかかわらず、ユーザが応答した音声の音高が低い場合、感情推定部５１２はユーザの感情が「悲しい」であると判断する。さらに別の例において、感情推定部５１２は、音声における語尾の音高と、それに対するユーザの応答の音高との関係に基づいてユーザの感情を推定してもよい。あるいは、感情推定部５１２は、これら複数の要素を複合的に考慮してユーザの感情を推定してもよい。
【００３９】
別の例において、感情推定部５１２は、音声以外の入力を用いてユーザの感情を推定してもよい。音声以外の入力としては、例えば、カメラにより撮影されたユーザの顔の映像、又は温度センサーにより検知されたユーザの体温、若しくはこれらの組み合わせが用いられる。具体的には、感情推定部５１２は、ユーザの表情からユーザの感情が「楽しい」、「怒り」、「悲しい」のいずれであるかを判断する。また、感情推定部５１２は、ユーザの顔の動画において、表情の変化に基づいてユーザの感情を判断してもよい。あるいは、感情推定部５１２は、ユーザの体温が高いと「怒り」、低いと「悲しい」と判断してもよい。
【００４０】
楽曲解析が要求されると、楽曲解析部５１３は、ユーザの指示により再生される楽曲を解析し（ステップＳ１０８）、解析結果を示す情報（以下「楽曲情報」という）を、要求元である処理部５１０に出力する（ステップＳ１０９）。
【００４１】
歌詞抽出が要求されると、歌詞抽出部５１４は、ユーザの指示により再生される楽曲の歌詞を取得し（ステップＳ１１０）、取得した歌詞を示す情報（以下「歌詞情報」という）を、要求元である処理部５１０に出力する（ステップＳ１１１）。
【００４２】
ステップＳ１１２において、処理部５１０は、感情推定部５１２、楽曲解析部５１３、及び歌詞抽出部５１４からそれぞれ取得した感情情報、楽曲情報、及び歌詞情報の組を、嗜好分析部５１５に出力する。
【００４３】
ステップＳ１１３において、嗜好分析部５１５は、複数組の情報を分析し、ユーザの嗜好

を示す情報を得る。この分析のため、嗜好分析部５１５は、過去のある期間（例えば、システムの稼働開始から現時点までの期間）に渡って、これらの情報の組を複数、記録する。一例において、嗜好分析部５１５は、楽曲情報を統計処理し、統計的な代表値（例えば、平均値、最頻値、又は中央値）を計算する。この統計処理により、例えば、テンポの平均値、並びに音色、曲調、リズム、コード進行、作曲者名、作詞者名、及び実演者名の最頻値が得られる。また、嗜好分析部５１５は、形態素解析等の技術を用いて歌詞情報により示される歌詞を単語レベルに分解したうえで各単語の品詞を特定し、特定の品詞（例えば名詞）の単語についてヒストグラムを作成し、登場頻度が所定の範囲（例えば上位５％）にある単語を特定する。さらに、嗜好分析部５１５は、特定された単語を含み、構文上の所定の区切り（例えば、分、節、又は句）に相当する単語群を歌詞情報から抽出する。例えば、「好き」という語の登場頻度が高い場合、この語を含む「そんな君が好き」、「とても好きだから」等の単語群が歌詞情報から抽出される。これらの平均値、最頻値、及び単語群は、ユーザの嗜好を示す情報（パラメータ）の一例である。あるいは、嗜好分析部５１５は、単なる統計処理とは異なる所定のアルゴリズムに従って複数組の情報を分析し、ユーザの嗜好を示す情報を得てもよい。あるいは、嗜好分析部５１５は、ユーザからフィードバックを受け付け、これらのパラメータの重みをフィードバックに応じて調整してもよい。ステップＳ１１４において、嗜好分析部５１５は、ステップＳ１１３により得られた情報を用いて、分類テーブル５１６１を更新する。

【００４４】

図８は、分類テーブル５１６１を例示する図である。この図では、ユーザ名が「山田太郎」であるユーザの分類テーブル５１６１を示している。分類テーブル５１６１において、楽曲の特徴、属性、及び歌詞が、ユーザの感情と対応付けて記録されている。分類テーブル５１６１を参照すれば、例えば、ユーザ「山田太郎」が「嬉しい」という感情を抱いているときには、「恋」、「愛」、及び「love」という語を歌詞に含み、テンポが約６０であり、「I→V→VIm→IIIm→IV→I→IV→V」というコード進行を有し、ピアノの音色が主である楽曲を好むことが示される。本実施形態によれば、ユーザの嗜好を示す情報を自動的に得ることができる。分類テーブル５１６１に記録される嗜好情報は、学習が進むにつれ、すなわち音声応答システム１の累積使用時間が増えるにつれ、蓄積され、よりユーザの嗜好を反映したものとなる。この例によれば、ユーザの嗜好を反映した情報を自動的に得ることができる。

【００４５】

なお、嗜好分析部５１５は、分類テーブル５１６１の初期値をユーザ登録時又は初回ログイン時等、所定のタイミングにおいて設定してもよい。この場合において、音声応答システム１は、システム上でユーザを表すキャラクタ（例えばいわゆるアバター）をユーザに選択させ、選択されたキャラクタに応じた初期値を有する分類テーブル５１６１を、そのユーザに対応する分類テーブルとして設定してもよい。

【００４６】

この実施形態において説明した分類テーブル５１６１に記録されるデータはあくまで例示である。例えば、分類テーブル５１６１にはユーザの感情が記録されず、少なくとも、歌詞が記録されていればよい。あるいは、分類テーブル５１６１には歌詞が記録されず、少なくとも、ユーザの感情と楽曲解析の結果とが記録されていればよい。

【００４７】

３．歌唱合成機能

３－１．構成

図９は、歌唱合成機能５２に係る機能構成を例示する図である。歌唱合成機能５２に係る機能要素として、音声応答システム１は、音声分析部５１１、感情推定部５１２、記憶部５１６、検知部５２１、歌唱生成部５２２、伴奏生成部５２３、及び合成部５２４を有する。歌唱生成部５２２は、メロディ生成部５２２１及び歌詞生成部５２２２を有する。以下において、学習機能５１と共通する要素については説明を省略する。

【００４８】

歌唱合成機能５２に関し、記憶部５１６は、素片データベース５１６２を記憶する。素片データベースは、歌唱合成において用いられる音声素片データを記録したデータベースである。音声素片データは、１又は複数の音素をデータ化したものである。音素とは、言語上の意味の区別の最小単位（例えば母音や子音）に相当するものであり、ある言語の実際の調音と音韻体系全体を考慮して設定される、その言語の音韻論上の最小単位である。音声素片は、特定の発声者によって発声された入力音声のうち所望の音素や音素連鎖に相当する区間が切り出されたものである。本実施形態における音声素片データは、音声素片の周波数スペクトルを示すデータである。以下の説明では、「音声素片」の語は、単一の音素（例えばモノフォン）や、音素連鎖（例えばダイフォンやトライフォン）を含む。

【００４９】

記憶部５１６は、素片データベース５１６２を複数、記憶してもよい。複数の素片データベース５１６２は、例えば、それぞれ異なる歌手（又は話者）により発音された音素を記録したものを含んでもよい。あるいは、複数の素片データベース５１６２は、単一の歌手（又は話者）により、それぞれ異なる歌い方又は声色で発音された音素を記録したものを含んでもよい。

【００５０】

歌唱生成部５２２は、歌唱音声を生成する、すなわち歌唱合成する。歌唱音声とは、与えられた歌詞を与えられたメロディに従って発した音声をいう。メロディ生成部５２２１は、歌唱合成に用いられるメロディを生成する。歌詞生成部５２２２は、歌唱合成に用いられる歌詞を生成する。メロディ生成部５２２１及び歌詞生成部５２２２は、分類テーブル５１６１に記録されている情報を用いてメロディ及び歌詞を生成してもよい。歌唱生成部５２２は、メロディ生成部５２２１により生成されたメロディ及び歌詞生成部５２２２により生成された歌詞を用いて歌唱音声を生成する。伴奏生成部５２３は、歌唱音声に対する伴奏を生成する。合成部５１９は、歌唱生成部５２２により生成された歌唱音声、伴奏生成部５２３により生成された伴奏、及び素片データベース５１６２に記録されている音声素片を用いて歌唱音声を合成する。

【００５１】

３－２．動作

図１０は、歌唱合成機能５２に係る音声応答システム１の動作（歌唱合成方法）の概要を示すフローチャートである。ステップＳ２１において、音声応答システム１は、歌唱合成をトリガするイベントが発生したか判断する。すなわち、音声応答システム１は、歌唱合成をトリガするイベントを検知する。歌唱合成をトリガするイベントは、例えば、ユーザから音声入力が行われたというイベント、カレンダーに登録されたイベント（例えば、アラーム又はユーザの誕生日）、ユーザから音声以外の手法（例えば入出力装置１０に無線接続されたスマートフォン（図示略）への操作）により歌唱合成の指示が入力されたというイベント、及びランダムに発生するイベントのうち少なくとも１つを含む。歌唱合成をトリガするイベントが発生したと判断された場合（Ｓ２１：ＹＥＳ）、音声応答システム１は、処理をステップＳ２２に移行する。歌唱合成をトリガするイベントが発生していないと判断された場合（Ｓ２１：ＮＯ）、音声応答システム１は、歌唱合成をトリガするイベントが発生するまで待機する。

【００５２】

ステップＳ２２において、音声応答システム１は、歌唱合成パラメータを読み出す。ステップＳ２３において、音声応答システム１は、歌詞を生成する。ステップＳ２４において、音声応答システム１は、メロディを生成する。ステップＳ２５において、音声応答システム１は、生成した歌詞及びメロディの一方を他方に合わせて修正する。ステップＳ２６において、音声応答システム１は、使用する素片データベースを選択する。ステップＳ２７において、音声応答システム１は、ステップＳ２３、Ｓ２６、及びＳ２７において得られた、メロディ、歌詞、及び素片データベースを用いて歌唱合成を行う。ステップＳ２８において、音声応答システム１は、伴奏を生成する。ステップＳ２９において、音声応答システム１は、歌唱音声と伴奏とを合成する。ステップＳ２３～Ｓ２９の処理は、図６

のフローにおけるステップＳ１８の処理の一部である。以下、歌唱合成機能５２に係る音声応答システム１の動作をより詳細に説明する。

【００５３】

図１１は、歌唱合成機能５２に係る音声応答システム１の動作を例示するシーケンスチャートである。歌唱合成をトリガするイベントを検知すると、検知部５２１は歌唱生成部５２２に対し歌唱合成を要求する（ステップＳ２０１）。歌唱合成の要求はユーザの識別子を含む。歌唱合成を要求されると、歌唱生成部５２２は、記憶部５１６に対しユーザの嗜好を問い合わせる（ステップＳ２０２）。この問い合わせはユーザ識別子を含む。問い合わせを受けると、記憶部５１６は、分類テーブル５１６１の中から、問い合わせに含まれるユーザ識別子と対応する嗜好情報を読み出し、読み出した嗜好情報を歌唱生成部５２２に出力する（ステップＳ２０３）。さらに歌唱生成部５２２は、感情推定部５１２に対しユーザの感情を問い合わせる（ステップＳ２０４）。この問い合わせはユーザ識別子を含む。問い合わせを受けると、感情推定部５１２は、そのユーザの感情情報を歌唱生成部５２２に出力する（ステップＳ２０５）。

【００５４】

ステップＳ２０６において、歌唱生成部５２２は、歌詞のソースを選択する。歌詞のソースは入力音声に応じて決められる。歌詞のソースは、大きくは、処理部５１０及び分類テーブル５１６１のいずれかである。処理部５１０から歌唱生成部５２２に出力される歌唱合成の要求は、歌詞（又は歌詞素材）を含んでいる場合と、歌詞を含んでいない場合とがある。歌詞素材とは、それ単独では歌詞を形成することができず、他の歌詞素材と組み合わせることによって歌詞を形成する文字列をいう。歌唱合成の要求が歌詞を含んでいる場合とは、例えば、ＡＩによる応答そのもの（「明日の天気は晴れです」等）にメロディを付けて応答音声を出力する場合をいう。歌唱合成の要求は処理部５１０によって生成されることから、歌詞のソースは処理部５１０であるということもできる。さらに、処理部５１０は、コンテンツ提供部６０からコンテンツを取得する場合があるので、歌詞のソースはコンテンツ提供部６０であるということもできる。コンテンツ提供部６０は、例えば、ニュースを提供するサーバ又は気象情報を提供するサーバである。あるいは、コンテンツ提供部６０は、既存の楽曲の歌詞を記録したデータベースを有するサーバである。図ではコンテンツ提供部６０は１台のみ示しているが、複数のコンテンツ提供部６０が存在してもよい。歌唱合成の要求に歌詞が含まれている場合、歌唱生成部５２２は、歌唱合成の要求を歌詞のソースとして選択する。歌唱合成の要求に歌詞が含まれていない場合（例えば、入力音声による指示が「何か歌って」のように歌詞の内容を特に指定しないものである場合）、歌唱生成部５２２は、分類テーブル５１６１を歌詞のソースとして選択する。

【００５５】

ステップＳ２０７において、歌唱生成部５２２は、選択されたソースに対し歌詞素材の提供を要求する。ここでは、分類テーブル５１６１すなわち記憶部５１６がソースとして選択された例を示している。この場合、この要求はユーザ識別子及びそのユーザの感情情報を含む。歌詞素材提供の要求を受けると、記憶部５１６は、要求に含まれるユーザ識別子及び感情情報に対応する歌詞素材を分類テーブル５１６１から抽出する（ステップＳ２０８）。記憶部５１６は、抽出した歌詞素材を歌唱生成部５２２に出力する（ステップＳ２０９）。

【００５６】

歌詞素材を取得すると、歌唱生成部５２２は、歌詞生成部５２２２に対し歌詞の生成を要求する（ステップＳ２１０）。この要求は、ソースから取得した歌詞素材を含む。歌詞の生成が要求されると、歌詞生成部５２２２は、歌詞素材を用いて歌詞を生成する（ステップＳ２１１）。歌詞生成部５２２２は、例えば、歌詞素材を複数、組み合わせることにより歌詞を生成する。あるいは、各ソースは１曲全体分の歌詞を記憶していてもよく、この場合、歌詞生成部５２２２は、ソースが記憶している歌詞の中から、歌唱合成に用いる１曲分の歌詞を選択してもよい。歌詞生成部５２２２は、生成した歌詞を歌唱生成部５２２に出力する（ステップＳ２１２）。

【００５７】
ステップＳ２１３において、歌唱生成部５２２は、メロディ生成部５２２１に対しメロディの生成を要求する。この要求は、ユーザの嗜好情報及び歌詞の音数を特定する情報を含む。歌詞の音数を特定する情報は、生成された歌詞の文字数、モーラ数、又は音節数である。メロディの生成が要求されると、メロディ生成部５２２１は、要求に含まれる嗜好情報に応じてメロディを生成する（ステップＳ２１４）。具体的には例えば以下のとおりである。メロディ生成部５２２１は、メロディの素材（例えば、２小節又は４小節程度の長さを有する音符列、又は音符列をリズムや音高の変化といった音楽的な要素に細分化した情報列）のデータベース（以下「メロディデータベース」という。図示略）にアクセスすることができる。メロディデータベースは、例えば記憶部５１６に記憶される。メロディデータベースには、メロディの属性が記録されている。メロディの属性は、例えば、適合する曲調又は歌詞、作曲者名等の楽曲情報を含む。メロディ生成部５２２１は、メロディデータベースに記録されている素材の中から、要求に含まれる嗜好情報に適合する１又は複数の素材を選択し、選択された素材を組み合わせて所望の長さのメロディを得る。歌唱生成部５２２は、生成したメロディを特定する情報（例えばＭＩＤＩ等のシーケンスデータ）を歌唱生成部５２２に出力する（ステップＳ２１５）。
【００５８】
ステップＳ２１６において、歌唱生成部５２２は、メロディ生成部５２２１に対しメロディの修正、又は歌詞生成部５２２２に対し歌詞の生成を要求する。この修正の目的の一つは、歌詞の音数（例えばモーラ数）とメロディの音数とを一致させることである。例えば、歌詞のモーラ数がメロディの音数よりも少ない場合（字足らずの場合）、歌唱生成部５２２は、歌詞の文字数を増やすよう、歌詞生成部５２２２に要求する。あるいは、歌詞のモーラ数がメロディの音数よりも多い場合（字余りの場合）、歌唱生成部５２２は、メロディの音数を増やすよう、メロディ生成部５２２１に要求する。この図では、歌詞を修正する例を説明する。ステップＳ２１７において、歌詞生成部５２２２は、修正の要求に応じて歌詞を修正する。メロディの修正をする場合、メロディ生成部５２２１は、例えば音符を分割して音符数を増やすことによりメロディを修正する。歌詞生成部５２２２又はメロディ生成部５２２１は、歌詞の文節の区切りの部分とメロディのフレーズの区切り部分とを一致させるよう調整してもよい。歌詞生成部５２２２は、修正した歌詞を歌唱生成部５２２に出力する（ステップＳ２１８）。
【００５９】
歌詞を受けると、歌唱生成部５２２は、歌唱合成に用いられる素片データベース５１６２を選択する（ステップＳ２１９）。素片データベース５１６２は、例えば、歌唱合成をトリガしたイベントに関するユーザの属性に応じて選択される。あるいは、素片データベース５１６２は、歌唱合成をトリガしたイベントの内容に応じて選択されてもよい。さらにあるいは、素片データベース５１６２は、分類テーブル５１６１に記録されているユーザの嗜好情報に応じて選択されてもよい。歌唱生成部５２２は、これまでの処理で得られた歌詞及びメロディに従って、選択された素片データベース５１６２から抽出された音声素片を合成し、合成歌唱のデータを得る（ステップＳ２２０）。なお、分類テーブル５１６１には、歌唱における声色の変更、タメ、しゃくり、ビブラート等の歌唱の奏法に関するユーザの嗜好を示す情報が記録されてもよく、歌唱生成部５２２は、これらの情報を参照して、ユーザの嗜好に応じた奏法を反映した歌唱を合成してもよい。歌唱生成部５２２は、生成された合成歌唱のデータを合成部５２４に出力する（ステップＳ２２２１）。
【００６０】
さらに、歌唱生成部５２２は、伴奏生成部５２３に対し伴奏の生成を要求する（Ｓ２２２）。この要求は、歌唱合成におけるメロディを示す情報を含む。伴奏生成部５２３は、要求に含まれるメロディに応じて伴奏を生成する（ステップＳ２２３）。メロディに対し自動的に伴奏を付ける技術としては、周知の技術が用いられる。メロディデータベースにおいてメロディのコード進行を示すデータ（以下「コード進行データ」）が記録されている場合、伴奏生成部５２３は、このコード進行データを用いて伴奏を生成してもよい。あ

るいは、メロディデータベースにおいてメロディに対する伴奏用のコード進行データが記録されている場合、伴奏生成部５２３は、このコード進行データを用いて伴奏を生成してもよい。さらにあるいは、伴奏生成部５２３は、伴奏のオーディオデータをあらかじめ複数、記憶しておき、その中からメロディのコード進行に合ったものを読み出してもよい。また、伴奏生成部５２３は、例えば伴奏の曲調を決定するために分類テーブル５１６１を参照し、ユーザの嗜好に応じた伴奏を生成してもよい。伴奏生成部５２３は、生成された伴奏のデータを合成部５２４に出力する（ステップＳ２２４）。

【００６１】

合成歌唱及び伴奏のデータを受けると、合成部５２４は、合成歌唱及び伴奏を合成する（ステップＳ２２５）。合成に際しては、演奏の開始位置やテンポを合わせることによって、歌唱と伴奏とが同期するように合成される。こうして伴奏付きの合成歌唱のデータが得られる。合成部５２４は、合成歌唱のデータを出力する。

【００６２】

ここでは、最初に歌詞が生成され、その後、歌詞に合わせてメロディを生成する例を説明した。しかし、音声応答システム１は、先にメロディを生成し、その後、メロディに合わせて歌詞を生成してもよい。また、ここでは歌唱と伴奏とが合成された後に出力される例を説明したが、伴奏が生成されず、歌唱のみが出力されてもよい（すなわちアカペラでもよい）。また、ここでは、まず歌唱が合成された後に歌唱に合わせて伴奏が生成される例を説明したが、まず伴奏が生成され、伴奏に合わせて歌唱が合成されてもよい。

【００６３】

４．応答機能

図１２は、応答機能５３に係る音声応答システム１の機能構成を例示する図である。応答機能５３に係る機能要素として、音声応答システム１は、音声分析部５１１、感情推定部５１２、及びコンテンツ分解部５３１を有する。以下において、学習機能５１及び歌唱合成機能５２と共通する要素については説明を省略する。コンテンツ分解部５３１は、一のコンテンツを複数の部分コンテンツに分解する。この例においてコンテンツとは、応答音声として出力される情報の内容をいい、具体的には、例えば、楽曲、ニュース、レシピ、又は教材（スポーツ教習、楽器教習、学習ドリル、クイズ）をいう。

【００６４】

図１３は、応答機能５３に係る音声応答システム１の動作を例示するフローチャートである。ステップＳ３１において、音声分析部５１１は、再生するコンテンツを特定する。再生するコンテンツは、例えばユーザの入力音声に応じて特定される。具体的には、音声分析部５１１が入力音声を解析し、入力音声により再生が指示されたコンテンツを特定する。一例において、「ハンバーグのレシピ教えて」という入力音声が与えられると、音声分析部１１は、「ハンバーグのレシピ」を提供するよう、処理部５１０に指示する。処理部５１０は、コンテンツ提供部６０にアクセスし、「ハンバーグのレシピ」を説明したテキストデータを取得する。こうして取得されたデータが、再生されるコンテンツとして特定される。処理部５１０は、特定されたコンテンツをコンテンツ分解部５３１に通知する。

【００６５】

ステップＳ３２において、コンテンツ分解部５３１は、コンテンツを複数の部分コンテンツに分解する。一例において、「ハンバーグのレシピ」は複数のステップ（材料を切る、材料を混ぜる、成形する、焼く等）から構成されるところ、コンテンツ分解部５３１は、「ハンバーグのレシピ」のテキストを、「材料を切るステップ」、「材料を混ぜるステップ」、「成形するステップ」、及び「焼くステップ」の４つの部分コンテンツに分解する。コンテンツの分解位置は、例えばＡＩにより自動的に判断される。あるいは、コンテンツに区切りを示すマーカーをあらかじめ埋め込んでおき、そのマーカーの位置でコンテンツが分解されてもよい。

【００６６】

ステップＳ３３において、コンテンツ分解部５３１は、複数の部分コンテンツのうち対象となる一の部分コンテンツを特定する（特定部の一例）。対象となる部分コンテンツは再

生される部分コンテンツであり、元のコンテンツにおけるその部分コンテンツの位置関係に応じて決められる。「ハンバーグのレシピ」の例では、コンテンツ分解部５３１は、まず、「材料を切るステップ」を対象となる部分コンテンツとして特定する。次にステップＳ３３の処理が行われるとき、コンテンツ分解部５３１は、「材料を混ぜるステップ」を対象となる部分コンテンツとして特定する。コンテンツ分解部５３１は、特定した部分コンテンツをコンテンツ修正部５３２に通知する。

【００６７】

ステップＳ３４において、コンテンツ修正部５３２は、対象となる部分コンテンツを修正する。具体的修正の方法は、コンテンツに応じて定義される。例えば、ニュース、気象情報、及びレシピといったコンテンツに対して、コンテンツ修正部５３２は修正を行わない。例えば、教材又はクイズのコンテンツに対して、コンテンツ修正部５３２は、問題として隠しておきたい部分を他の音（例えばハミング、「ラララ」、ビープ音等）に置換する。このとき、コンテンツ修正部５３２は、置換前の文字列とモーラ数又は音節数が同一の文字列を用いて置換する。コンテンツ修正部５３２は、修正された部分コンテンツを歌唱生成部５２２に出力する。

【００６８】

ステップＳ３５において、歌唱生成部５２２は、修正された部分コンテンツを歌唱合成する。歌唱生成部５２２により生成された歌唱音声は、最終的に、入出力装置１０から応答音声として出力される。応答音声を出力すると、音声応答システム１はユーザの応答待ち状態となる（ステップＳ３６）。ステップＳ３６において、音声応答システム１は、ユーザの応答を促す歌唱又は音声（例えば「できましたか？」等）を出力してもよい。音声分析部５１１は、ユーザの応答に応じて次の処理を決定する。次の部分コンテンツの再生を促す応答が入力された場合（Ｓ３６：次）、音声分析部５１１は、処理をステップＳ３３に移行する。次の部分コンテンツの再生を促す応答は、例えば、「次のステップへ」、「できた」、「終わった」等の音声である。次の部分コンテンツの再生を促す応答以外の応答が入力された場合（Ｓ３６：終了）、音声分析部５１１は、音声の出力を停止するよう処理部５１０に指示する。

【００６９】

ステップＳ３７において、処理部５１０は、部分コンテンツの合成音声の出力を、少なくとも一時的に停止する。ステップＳ３８において、処理部５１０は、ユーザの入力音声に応じた処理を行う。ステップＳ３８における処理には、例えば、現在のコンテンツの再生中止、ユーザから指示されたキーワード検索、及び別のコンテンツの再生開始が含まれる。例えば、「歌を止めて欲しい」、「もう終わり」、又は「おしまい」等の応答が入力された場合、処理部５１０は、現在のコンテンツの再生を中止する。例えば、「短冊切りってどうやるの？」又は「アーリオオーリオって何？」等、質問型の応答が入力された場合、処理部５１０は、ユーザの質問に回答するための情報をコンテンツ提供部６０から取得する。処理部５１０は、ユーザの質問に対する回答の音声を出力する。この回答は歌唱ではなく、話声であってもよい。「○○の曲かけて」等、別のコンテンツの再生を指示する応答が入力された場合、処理部５１０は、指示されたコンテンツをコンテンツ提供部６０から取得し、再生する。

【００７０】

なおここではコンテンツが複数の部分コンテンツに分解され、部分コンテンツ毎にユーザの反応に応じて次の処理を決定する例を説明した。しかし、応答機能５３が応答音声を出力する方法はこれに限定さない。例えば、コンテンツは部分コンテンツに分解されず、そのまま話声として、又はそのコンテンツを歌詞として用いた歌唱音声として出力されてもよい。音声応答システム１は、ユーザの入力音声に応じて、又は出力されるコンテンツに応じて、部分コンテンツに分解するか、分解せずそのまま出力するか判断してもよい。

【００７１】

５．動作例

以下、具体的な動作例をいくつか説明する。各動作例において特に明示はしないが、各動

作例は、それぞれ、上記の学習機能、歌唱合成機能、及び応答機能の少なくとも１つ以上に基づくものである。なお以下の動作例はすべて日本語が使用される例を説明するが、使用される言語は日本語に限定されず、どのような言語でもよい。

【００７２】

５－１．動作例１

図１４は、音声応答システム１の動作例１を示す図である。この例において、ユーザは「佐藤一太郎（実演者名）の『さくらさくら』（楽曲名）をかけて」という入力音声により、楽曲の再生を要求する。音声応答システム１は、この入力音声に従って楽曲データベースを検索し、要求された楽曲を再生する。このとき、音声応答システム１は、この入力音声を入力したときのユーザの感情及びこの楽曲の解析結果を用いて、分類テーブルを更新する。分類テーブルは、楽曲の再生が要求される度に分類テーブルを更新する。分類テーブルは、ユーザが音声応答システム１に対し楽曲の再生を要求する回数が増えるにつれ（すなわち、音声応答システム１の累積使用時間が増えるにつれ）、よりそのユーザの嗜好を反映したものになっていく。

【００７３】

５－２．動作例２

図１５は、音声応答システム１の動作例２を示す図である。この例において、ユーザは「何か楽しい曲歌って」という入力音声により、歌唱合成を要求する。音声応答システム１は、この入力音声に従って歌唱合成を行う。歌唱合成に際し、音声応答システム１は、分類テーブルを参照する。分類テーブルに記録されている情報を用いて、歌詞及びメロディを生成する。したがって、ユーザの嗜好を反映した楽曲を自動的に作成することができる。

【００７４】

５－３．動作例３

図１６は、音声応答システム１の動作例３を示す図である。この例において、ユーザは「今日の天気は？」という入力音声により、気象情報の提供を要求する。この場合、処理部５１０はこの要求に対する回答として、コンテンツ提供部６０のうち気象情報を提供するサーバにアクセスし、今日の天気を示すテキスト（例えば「今日は一日快晴」）を取得する。処理部５１０は、取得したテキストを含む、歌唱合成の要求を歌唱生成部５２２に出力する。歌唱生成部５２２は、この要求に含まれるテキストを歌詞として用いて、歌唱合成を行う。音声応答システム１は、入力音声に対する回答として「今日は一日快晴」にメロディ及び伴奏を付けた歌唱音声を出力する。

【００７５】

５－４．動作例４

図１７は、音声応答システム１の動作例４を示す図である。この例において、図示された応答が開始される前に、ユーザは音声応答システム１を２週間、使用し、恋愛の歌をよく再生していた。そのため、分類テーブルには、そのユーザが恋愛の歌が好きであることを示す情報が記録される。音声応答システム１は、「出会いの場所はどこがいい？」や、「季節はいつがいいかな？」など、歌詞生成のヒントとなる情報を得るためにユーザに質問をする。音声応答システム１は、これらの質問に対するユーザの回答を用いて歌詞を生成する。なおこの例において、使用期間がまだ２週間と短いため、音声応答システム１の分類テーブルは、まだユーザの嗜好を十分に反映できておらず、感情との対応付けも十分ではない。そのため、本当はユーザはバラード調の曲が好みであるにも関わらず、それとは異なるロック調の曲を生成したりする。

【００７６】

５－５．動作例５

図１８は、音声応答システム１の動作例５を示す図である。この例は、動作例３からさらに音声応答システム１の使用を続け、累積使用期間が１月半となった例を示している。動作例３と比較すると分類テーブルはユーザの嗜好をより反映したものとなっており、合成される歌唱はユーザの嗜好に沿ったものになっている。ユーザは、最初は不完全だった音

声応答システム１の反応が徐々に自分の嗜好に合うように変化していく体験をすることができる。

【００７７】

５－６．動作例６

図１９は、音声応答システム１の動作例６を示す図である。この例において、ユーザは、「ハンバーグのレシピを教えてくれる？」という入力音声により、「ハンバーグ」の「レシピ」のコンテンツの提供を要求する。音声応答システム１は、「レシピ」というコンテンツが、あるステップが終了してから次のステップに進むべきものである点を踏まえ、コンテンツを部分コンテンツに分解し、ユーザの反応に応じて次の処理を決定する態様で再生することを決定する。

【００７８】

「ハンバーグ」の「レシピ」はステップ毎に分解され、各ステップの歌唱を出力する度に、音声応答システム１は「できましたか？」、「終わりましたか？」等、ユーザの応答を促す音声を出力する。ユーザが「できたよ」、「次は？」等、次のステップの歌唱を指示する入力音声を発すると、音声応答システム１は、それに応答して次のステップの歌唱を出力する。ユーザが「タマネギのみじん切りってどうやるの？」と質問する入力音声を発すると、音声応答システム１は、それに応答して「タマネギのみじん切り」の歌唱を出力する。「タマネギのみじん切り」の歌唱を終えると、音声応答システム１は、「ハンバーグ」の「レシピ」の続きから歌唱を開始する。

【００７９】

音声応答システム１は、第１の部分コンテンツの歌唱音声と、それに続く第２の部分コンテンツの歌唱音声との間に、別のコンテンツの歌唱音声を出力してもよい。音声応答システム１は、例えば、第１の部分コンテンツに含まれる文字列が示す事項に応じた時間長となるよう合成された歌唱音声を、第１の部分コンテンツの歌唱音声と第２の部分コンテンツの歌唱音声との間に出力する。具体的には、第１の部分コンテンツが「ここで材料を２０分、煮込みましょう」というように、待ち時間が２０分発生することを示していた場合、音声応答システム１は、材料を煮込んでいる間に流す２０分の歌唱を合成し、出力する。

【００８０】

また、音声応答システム１は、第１の部分コンテンツに含まれる第１文字列が示す事項に応じた第２文字列を用いて合成された歌唱音声を、第１の部分コンテンツの歌唱音声の出力後、第１文字列が示す事項に応じた時間長に応じたタイミングで出力してもよい。具体的には、第１の部分コンテンツが「ここで材料を２０分、煮込みましょう」というように、待ち時間が２０分発生することを示していた場合、音声応答システム１は、「煮込み終了です」（第２文字列の一例）という歌唱音声を、第１の部分コンテンツを出力してから２０分後に出力してもよい。あるいは、第１の部分コンテンツが「ここで材料を２０分、煮込みましょう」である例において、待ち時間の半分（１０分）経過したときに、「煮込み終了まであと１０分です」などとラップ風に歌唱してもよい。

【００８１】

５－７．動作例７

図２０は、音声応答システム１の動作例７を示す図である。この例において、ユーザは、「世界史の年号の暗記問題出してくれる？」という入力音声により、「世界史」の「暗記問題」のコンテンツの提供を要求する。音声応答システム１は、「暗記問題」というコンテンツが、ユーザの記憶を確認するためのものである点を踏まえ、コンテンツを部分コンテンツに分解し、ユーザの反応に応じて次の処理を決定する態様で再生することを決定する。

【００８２】

例えば、音声応答システム１は、「卑弥呼にサンキュー（２３９）魏の皇帝」という年号暗記文を、音声応答システム１は、「卑弥呼に」及び「サンキュー魏の皇帝」という２つの部分コンテンツに分解する。音声応答システム１は、「卑弥呼に」という歌唱を出力す

るとユーザの反応を待つ。ユーザが何か音声を発すると、音声応答システム１は、ユーザが発した音声が正解であるか判断し、その判断結果に応じた音声を出力する。例えば、ユーザが「サンキュー魏の皇帝」という正解の音声を発した場合、音声応答システム１は、「正解です」等の音声を出力する。あるいは、ユーザが「わかりません」等、正解ではない音声を発した場合、音声応答システム１は、「卑弥呼にサンキュー魏の皇帝」という正解の歌唱を出力する。

【００８３】

５－８．動作例８

図２１は、音声応答システム１の動作例８を示す図である。動作例７と同様、ユーザは、「世界史」の「暗記問題」のコンテンツの提供を要求する。音声応答システム１は、「暗記問題」というコンテンツが、ユーザの記憶を確認するためのものである点を踏まえ、このコンテンツの一部を隠して出力する。隠すべき部分は、例えばコンテンツにおいて定義されていてもよいし、処理部５１０すなわちＡＩが形態素解析等の結果に基づいて判断してもよい。

【００８４】

例えば、音声応答システム１は、「卑弥呼にサンキュー（２３９）魏の皇帝」という年号暗記文のうち、「にサンキュー」の部分を隠して歌唱する。具体的には、音声応答システム１は、隠す部分を他の音又は文字列（例えばハミング、「ラララ」、ビープ音等）に置換する。置換に用いられる音又は文字列は、置換前とモーラ数又は音節数が同一である音又は文字列である。一例において、音声応答システム１は、「卑弥呼・ラ・ラ・ラ・ラ・ラ・魏の皇帝」という歌唱を出力する。音声応答システム１は、この歌唱を出力するとユーザの反応を待つ。ユーザが何か音声を発すると、音声応答システム１は、ユーザが発した音声が正解であるか判断し、その判断結果に応じた音声を出力する。例えば、ユーザが「卑弥呼にサンキュー魏の皇帝」という音声を発した場合、音声応答システム１は、「正解です」等の音声を出力する。あるいは、ユーザが「わかりません」という音声を発した場合、音声応答システム１は、「卑弥呼にサンキュー魏の皇帝」という正解の歌唱を出力する。

【００８５】

また、音声応答システム１は、第１の部分コンテンツに対するユーザの反応に応じて、それに続く第２の部分コンテンツの一部又は全部を他の文字列に置換してもよい。例えば、問題集やクイズのコンテンツにおいて、第１問（第１の部分コンテンツの一例）に正解した場合と不正解だった場合とで、第２問（第２の部分コンテンツの一例）において他の文字列に置換する文字数を変化させてもよい（例えば、第１問が正解だった場合には第２問はより多くの文字を隠し、第１問が不正解だった場合には第２問はより少ない文字を隠す）。

【００８６】

５－９．動作例９

図２２は、音声応答システム１の動作例９を示す図である。この例において、ユーザは、「工場における工程の手順書を読み上げてくれる？」という入力音声により、「手順書」のコンテンツの提供を要求する。音声応答システム１は、「手順書」というコンテンツが、ユーザの記憶を確認するためのものである点を踏まえ、コンテンツを部分コンテンツに分解し、ユーザの反応に応じて次の処理を決定する態様で再生することを決定する。

【００８７】

例えば、音声応答システム１は、手順書をランダムな位置で区切り、複数の部分コンテンツに分解する。音声応答システム１は、一の部分コンテンツの歌唱を出力すると、ユーザの反応を待つ。例えば「スイッチＡを押した後、メータＢの値が１０以下となったところでスイッチＢを押す」という手順のコンテンツにつき、音声応答システム１が「スイッチＡを押した後」という部分を歌唱し、ユーザの反応を待つ。ユーザが何か音声を発すると、音声応答システム１は、次の部分コンテンツの歌唱を出力する。あるいはこのとき、ユーザが次の部分コンテンツを正しく言えたか否かに応じて、次の部分コンテンツの歌唱

のスピードを変更してもよい。具体的には、ユーザが次の部分コンテンツを正しく言えた場合、音声応答システム１は、次の部分コンテンツの歌唱のスピードを上げる。あるいは、ユーザが次の部分コンテンツを正しく言えなかった場合、音声応答システム１は、次の部分コンテンツの歌唱のスピードを下げる。

【００８８】

５－１０．動作例１０

図２３は、音声応答システム１の動作例１０を示す図である。動作例１０は、高齢者の認知症対策の動作例である。この例において、ユーザが高齢者であることはあらかじめユーザ登録等により設定されている。音声応答システム１は、例えばユーザの指示に応じて既存の歌を歌い始める。音声応答システム１は、ランダムな位置、又は所定の位置（例えばサビの手前）において歌唱を一時停止する。その際、「うーん分からない」、「忘れちゃった」等のメッセージを発し、あたかも歌詞を忘れたかのように振る舞う。音声応答システム１は、この状態でユーザの応答を待つ。ユーザが何か音声を発すると、音声応答システム１は、ユーザが発した言葉（の一部）を正解の歌詞として、その言葉の続きから歌唱を出力する。なお、ユーザが何か言葉を発した場合、音声応答システム１は「ありがとう」等の応答を出力してもよい。ユーザの応答待ちの状態で所定時間が経過したときは、音声応答システム１は、「思い出した」等の話声を出力し、一時停止した部分の続きから歌唱を再開してもよい。

【００８９】

５－１１．動作例１１

図２４は、音声応答システム１の動作例１１を示す図である。この例において、ユーザは「何か楽しい曲歌って」という入力音声により、歌唱合成を要求する。音声応答システム１は、この入力音声に従って歌唱合成を行う。歌唱合成の際に用いる素片データベースは、例えばユーザ登録時に選択されたキャラクタに応じて選択される（例えば、男性キャラクタが選択された場合、男性歌手による素片データベースが用いられる）。ユーザは、歌の途中で「女性の声に変えて」等、素片データベースの変更を指示する入力音声を発する。音声応答システム１は、ユーザの入力音声に応じて、歌唱合成に用いる素片データベースを切り替える。素片データベースの切り替えは、音声応答システム１が歌唱音声を出力しているときに行われてもよいし、動作例７～１０のように音声応答システム１がユーザの応答待ちの状態のときに行われてもよい。

【００９０】

既に説明したように、音声応答システム１は、単一の歌手（又は話者）により、それぞれ異なる歌い方又は声色で発音された音素を記録した複数の素片データベースを有してもよい。このような場合において、音声応答システム１は、ある音素について、複数の素片データベースから抽出した複数の素片を、ある比率（利用比率）で組み合わせて、すなわち加算して用いてもよい。さらに、音声応答システム１は、この利用比率を、ユーザの反応に応じて決めてもよい。具体的には、ある歌手について、通常の声と甘い声とで２つの素片データベースが記録されているときに、ユーザが「もっと甘い声で」という入力音声を発すると甘い声の素片データベースの利用比率を高め、「もっともっと甘い声で」という入力音声を発すると甘い声の素片データベースの利用比率をさらい高める。

【００９１】

６．変形例

本発明は上述の実施形態に限定されるものではなく、種々の変形実施が可能である。以下、変形例をいくつか説明する。以下の変形例のうち２つ以上のものが組み合わせて用いられてもよい。

【００９２】

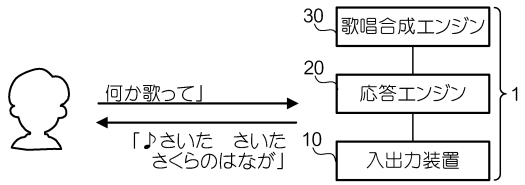
本稿において歌唱音声とは、少なくともその一部に歌唱を含む音声をいい、歌唱を含まない伴奏のみの部分、又は話声のみの部分を含んでいてもよい。例えば、コンテンツを複数の部分コンテンツに分解する例において、少なくとも１つの部分コンテンツは、歌唱を含んでいなくてもよい。また、歌唱は、ラップ、又は詩の朗読を含んでもよい。

【００９３】
実施形態においては、学習機能５１、歌唱合成機能５２、及び応答機能５３が相互に関連している例を説明したが、これらの機能は、それぞれ単独で提供されてもよい。例えば、学習機能５１により得られた分類テーブルが、例えば楽曲を配信する楽曲配信システムにおいてユーザの嗜好を知るために用いられてもよい。あるいは、歌唱合成機能５２は、学習機能５１により生成された分類テーブルではなく、ユーザが手入力した分類テーブルを用いて歌唱合成を行ってもよい。また、音声応答システム１の機能要素の少なくとも一部は省略されてもよい。例えば、音声応答システム１は、感情推定部５１２を有していなくてもよい。
【００９４】
入出力装置１０、応答エンジン２０、及び歌唱合成エンジン３０に対する機能の割り当ては、実施形態において例示されたものに限定されない。例えば、音声分析部５１１及び感情推定部５１２が入出力装置に実装されてもよい。また、入出力装置１０、応答エンジン２０、及び歌唱合成エンジン３０の相対的な配置は、実施形態において例示されたものに限定されない。例えば、歌唱合成エンジン３０は入出力装置１０と応答エンジン２０との間に配置され、応答エンジン２０から出力される応答のうち歌唱合成が必要と判断される応答について、歌唱合成を行ってもよい。また、音声応答システム１において用いられるコンテンツは、コンテンツ提供部６０から提供されるもの、すなわちネットワーク又はクラウド上に存在するものに限定されない。音声応答システム１において用いられるコンテンツは、入出力装置１０又は入出力装置１０と通信可能な装置等の、ローカルな装置に記憶されていてもよい。
【００９５】
入出力装置１０、応答エンジン２０、及び歌唱合成エンジン３０のハードウェア構成は実施形態において例示されたものに限定されない。例えば、入出力装置１０は、タッチスクリーン及びディスプレイを有するコンピュータ装置、例えばスマートフォン又はタブレット端末であってもよい。これに関連し、音声応答システム１に対するユーザの入力は音声を介するものに限定されず、タッチスクリーン、キーボード、又はポインティングデバイスを介して入力されるものであってもよい。また、入出力装置１０は、人感センサーを有してもよい。この場合において、音声応答システム１は、この人感センサーを用いて、ユーザが近くにいるかいないかに応じて、動作を制御してもよい。例えば、ユーザが入出力装置１０の近くにいないと判断される場合、音声応答システム１は、音声を出力しない（対話を返さない）という動作をしてもよい。ただし、音声応答システム１が出力する音声の内容によっては、ユーザが入出力装置１０の近くにいるいないにかかわらず、音声応答システム１はその音声を出力してもよい。例えば、動作例６の後半で説明したような、残りの待ち時間を案内する音声については、音声応答システム１は、ユーザが入出力装置１０の近くにいるいないにかかわらず出力してもよい。なお、ユーザが入出力装置１０の近くにいるかいないかの検出については、ユーザに動きがあまりない場合の対応を考え、カメラや温度センサーなど、人感センサー以外のセンサーを用いたり、複数のセンサーを併用したりしてもよい。
【００９６】
実施形態において例示したフローチャート及びシーケンスチャートはあくまで例示であり、音声応答システム１の動作はこれに限定されない。実施形態で例示したフローチャート又はシーケンスチャートにおいて、処理の順序が入れ替えられたり、一部の処理が省略されたり、新たな処理が追加されたりしてもよい。
【００９７】
入出力装置１０、応答エンジン２０、及び歌唱合成エンジン３０において実行されるプログラムは、ＣＤ－ＲＯＭ又は半導体メモリー等の記録媒体に記憶された状態で提供されてもよいし、インターネット等のネットワークを介したダウンロードにより提供されてもよい。
【符号の説明】

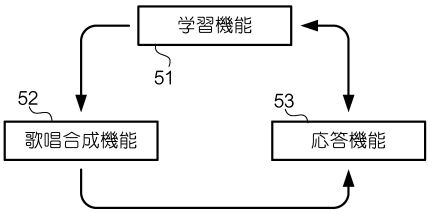
【００９８】
１…音声応答システム、１０…入出力装置、２０…応答エンジン、３０…歌唱合成エンジン、５１…学習機能、５２…歌唱合成機能、５３…応答機能、６０…コンテンツ提供部、１０１…マイクロフォン、１０２…入力信号処理部、１０３…出力信号処理部、１０４…スピーカ、１０５…ＣＰＵ、１０６…センサー、１０７…モータ、１０８…ネットワークＩＦ、２０１…ＣＰＵ、２０２…メモリー、２０３…ストレージ、２０４…通信ＩＦ、３０１…ＣＰＵ、３０２…メモリー、３０３…ストレージ、３０４…通信ＩＦ、５１０…処理部、５１１…音声分析部、５１２…感情推定部、５１３…楽曲解析部、５１４…歌詞抽出部、５１５…嗜好分析部、５１６…記憶部、５２１…検知部、５２２…歌唱生成部、５２３…伴奏生成部、５２４…合成部、５２２１…メロディ生成部、５２２２…歌詞生成部、５３１…コンテンツ分解部、５３２…コンテンツ修正部

【図面】

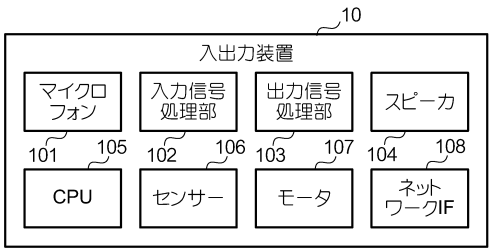
【図１】



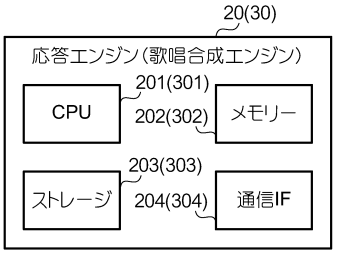
【図２】



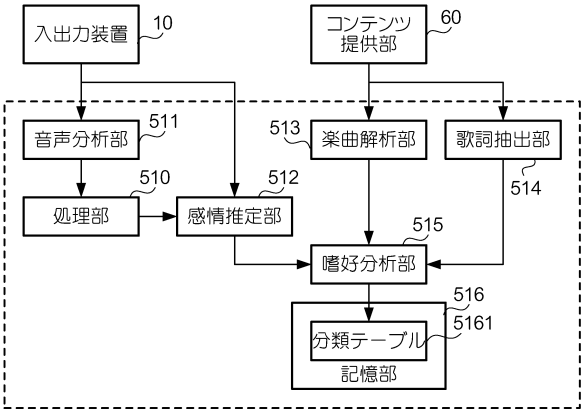
【図３】



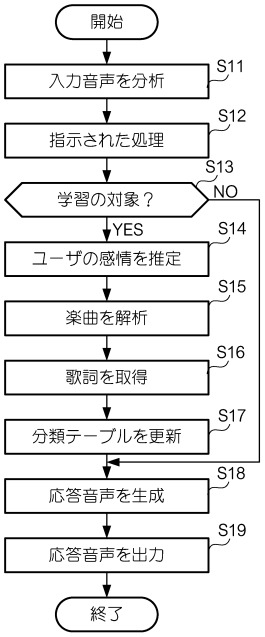
【図４】



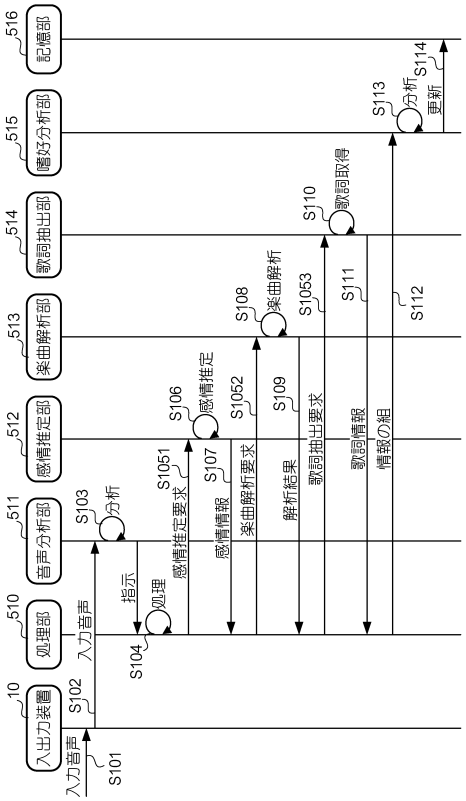
【図５】



【図６】



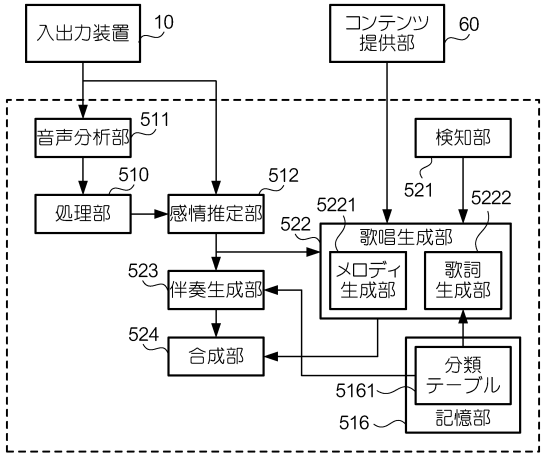
【図７】



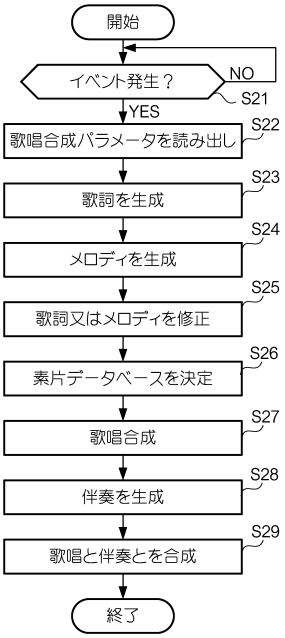
【図８】

ユーザ名	山田太郎				
感情	歌詞素材	テンポ	音色	コード進行	…
嬉しい	そんな君が好き とても好きだから 恋 愛 love	60	ピアノ	Ⅰ→Ⅴ→Ⅵm→ Ⅲm→Ⅳ→Ⅰ→ Ⅳ→Ⅴ	…
怒り	…	…	…	…	…
悲しい	…	…	…	…	…
楽しい	…	…	…	…	…
…	…	…	…	…	…

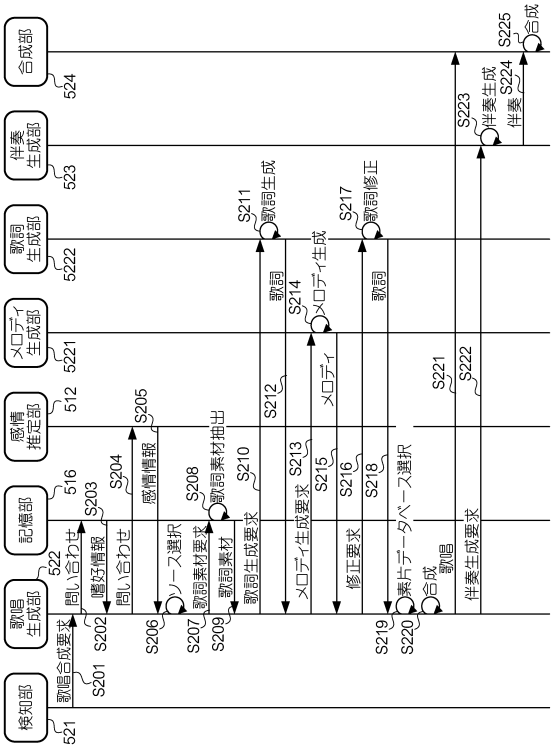
【図９】



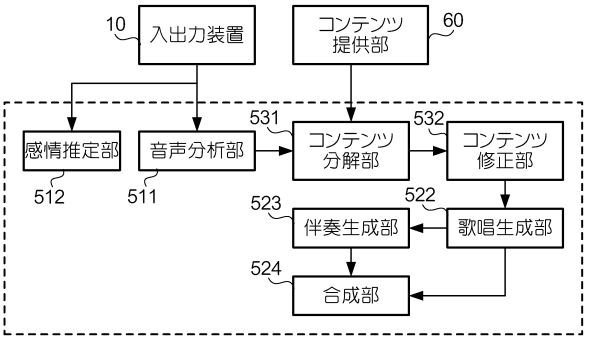
【図１０】



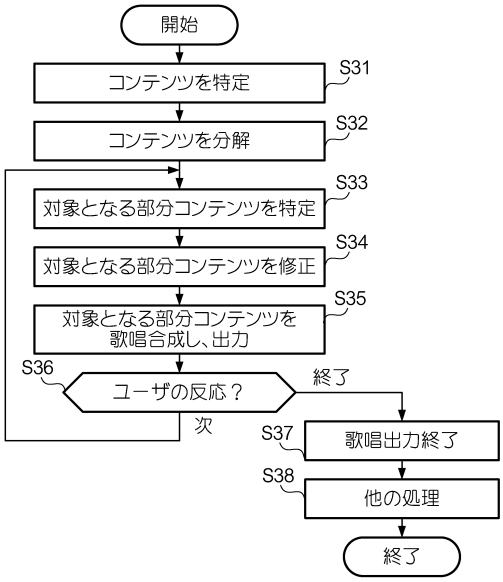
【図１１】



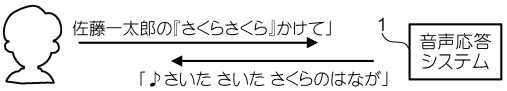
【図１２】



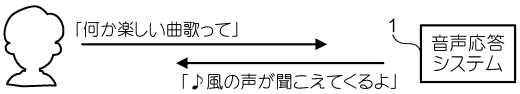
【図１３】



【図１４】

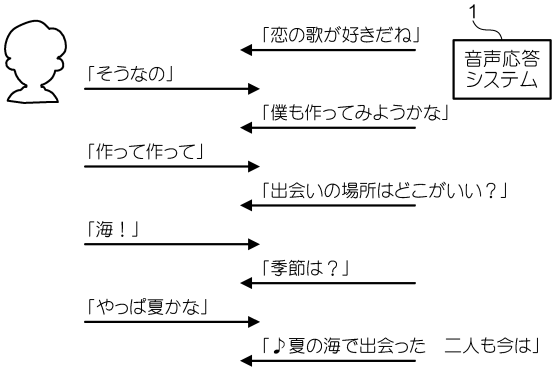


【図１５】

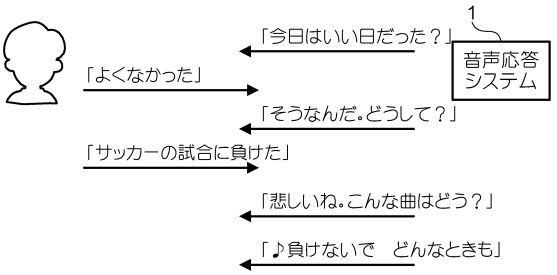


【図１６】

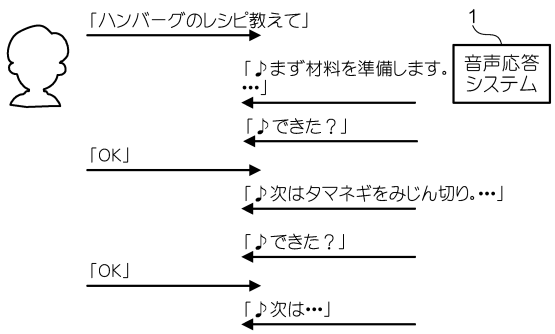
【図１７】



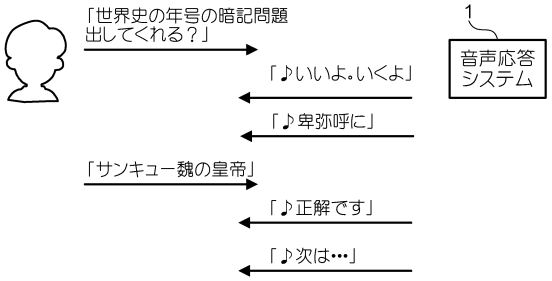
【図１８】



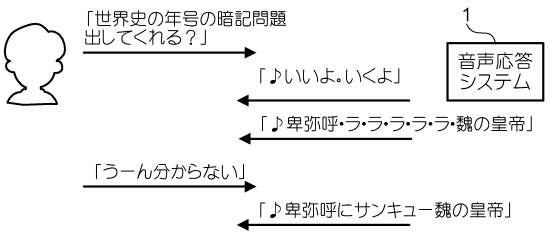
【図１９】



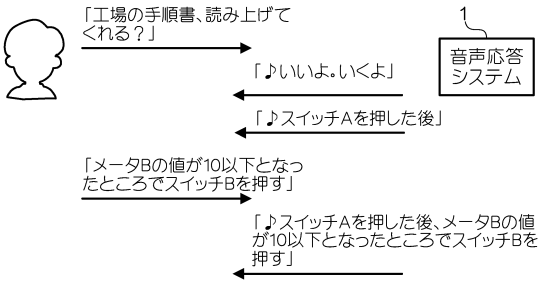
【図２０】



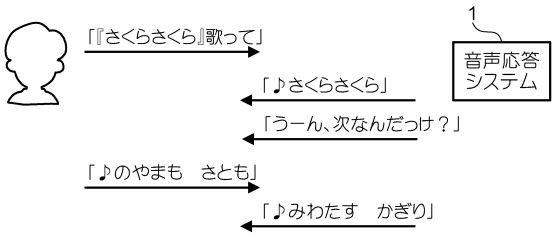
【図２１】



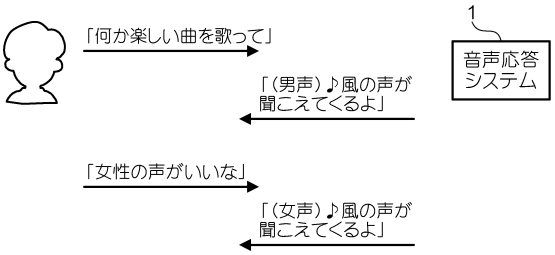
【図２２】



【図２３】


1
音声応答システム
「『さくらさくら』歌って」
「♪さくらさくら」
「うーん、次なんだっけ？」
「♪のやまも　さとも」
「♪みわたす　かぎり」

【図２４】


1
音声応答システム
「何か楽しい曲を歌って」
「（男声）♪風の声が聞こえてくるよ」
「女性の声がいいな」
「（女声）♪風の声が聞こえてくるよ」

【手続補正書】
【提出日】令和3年11月11日(2021.11.11)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
入出力装置を介して入力された、ユーザの要求を示す入力音声を取得する第１取得手段と、
前記入力音声に基づく要求に対する回答を示すテキストをサーバから取得する第２取得手段と、
前記テキストを歌詞として歌唱音声を合成する歌唱合成手段と、
前記歌唱音声を出力する出力手段と
を有する情報処理装置。
【請求項２】
前記第２取得手段は、前記入力音声を検索キーとして得られた検索結果を前記テキストとして取得する
請求項１に記載の情報処理装置。
【請求項３】
前記テキストを、第１の部分コンテンツ及び第２の部分コンテンツを含む複数の部分コンテンツに分解する分解手段を有し、
前記出力手段は、前記第１の部分コンテンツの歌唱音声を出力した後、前記ユーザの反応を待って前記第２の部分コンテンツの歌唱音声を出力する
請求項２に記載の情報処理装置。
【請求項４】
前記歌唱合成手段は、前記第１の部分コンテンツの歌唱音声を合成した後、当該第１の部分コンテンツに対して前記ユーザの応答を促す歌唱音声を合成し、
前記出力手段は、前記第１の部分コンテンツの歌唱音声を出力した後、前記ユーザの応答を促す歌唱音声を出力する
請求項３に記載の情報処理装置。
【請求項５】
前記歌唱合成手段は、前記歌唱音声を合成する際、歌詞及びメロディの一方を他方に合わせて修正をする
請求項１乃至４のいずれか一項に記載の情報処理装置。
【請求項６】
前記歌唱合成手段は、歌詞の音数とメロディの音数とを一致させるよう、前記修正をする
請求項５に記載の情報処理装置。
【請求項７】
前記ユーザの嗜好を示す嗜好情報を記憶する記憶手段と、
前記第２取得手段は、前記嗜好情報に応じて特定されたジャンルの歌の歌詞となる前記テキストを取得する
請求項１乃至６のいずれか一項に記載の情報処理装置。
【請求項８】
入出力装置を介して入力された、ユーザの要求を示す入力音声を取得し、
前記入力音声に基づく要求に対する回答を示すテキストを他のサーバから取得し、
前記テキストを歌詞として歌唱音声を合成し、
前記歌唱音声を出力する
歌唱音声の出力方法。

【請求項９】
コンピュータに、
入出力装置を介して入力された、ユーザの要求を示す入力音声を取得し、
前記入力音声に基づく要求に対する回答を示すテキストを他のサーバから取得し、
前記テキストを歌詞として歌唱音声を合成し、
前記歌唱音声を出力する
処理を実行させるためのプログラム。

フロントページの続き

静岡県浜松市中区中沢町１０番１号 ヤマハ株式会社内
(72)発明者 山内 健一
静岡県浜松市中区中沢町１０番１号 ヤマハ株式会社内
(72)発明者 山中 晋
静岡県浜松市中区中沢町１０番１号 ヤマハ株式会社内