

(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2016-157088

(P2016-157088A)

(43) 公開日 平成28年9月1日(2016.9.1)

(51) Int. Cl.		F I			テーマコード (参考)
G 1 0 K	15/04	(2006.01)	G 1 0 K	15/04	3 0 2 D
G 1 0 K	15/02	(2006.01)	G 1 0 K	15/02	5 D 1 0 8
G 1 0 L	13/00	(2006.01)	G 1 0 L	13/00	1 0 0 Y

審査請求 未請求 請求項の数 6 O L (全 21 頁)

(21) 出願番号	特願2015-36705 (P2015-36705)	(71) 出願人	000005016
(22) 出願日	平成27年2月26日 (2015.2.26)		パイオニア株式会社
			神奈川県川崎市幸区新小倉1番1号
		(74) 代理人	100107331
			弁理士 中村 聡延
		(74) 代理人	100104765
			弁理士 江上 達夫
		(72) 発明者	石岡 克仁
			神奈川県川崎市幸区新小倉1番1号 パイ
			オニア株式会社内
		(72) 発明者	菅原 啓太郎
			神奈川県川崎市幸区新小倉1番1号 パイ
			オニア株式会社内
		Fターム(参考)	5D108 BD02 BD14

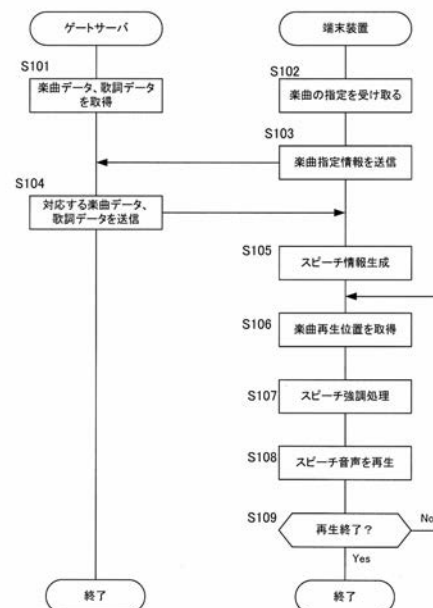
(54) 【発明の名称】 楽曲再生システム、端末装置、楽曲データ提供方法、及び、プログラム

(57) 【要約】

【課題】歌詞を含む音楽を再生している際に、ユーザがその曲を歌うための歌詞音声を取り取り易く提供する。

【解決手段】楽曲再生システムは、サーバと端末装置とを備える。再生すべき楽曲である再生楽曲をユーザが選択すると、端末装置は、選択された再生楽曲を指定する情報をサーバに送信する。サーバは、楽曲の歌詞データを取得し、端末装置へ送信する。端末装置は、再生楽曲に対応する歌詞データをサーバから受信し、歌詞データに基づいて歌詞音声データを生成する。そして、端末装置は、楽曲中の歌詞部分に先行するように歌詞音声データを楽曲データに加算して歌詞音声付楽曲データを生成し、音声再生装置に送信する。こうして、ユーザが選択した楽曲についての歌詞音声付楽曲データが音声再生装置により再生される。

【選択図】図12



【特許請求の範囲】**【請求項 1】**

サーバと、端末装置とを備える楽曲再生システムであって、
前記サーバは、
楽曲の歌詞データを取得する取得手段と、
前記歌詞データを前記端末装置へ送信する送信手段と、を備え、
前記端末装置は、
再生すべき楽曲である再生楽曲を選択するための入力手段と、
選択された再生楽曲の楽曲データを取得する楽曲データ取得手段と、
選択された再生楽曲を指定する情報を前記サーバに送信し、当該再生楽曲に対応する歌詞データを前記サーバから受信する第 1 の通信手段と、
前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、
前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段と、
前記歌詞音声付楽曲データを音声再生装置に送信する第 2 の通信手段と、
を備えることを特徴とする楽曲再生システム。

【請求項 2】

前記楽曲データ取得手段は、前記サーバから前記楽曲データを取得することを特徴とする請求項 1 に記載の楽曲再生システム。

【請求項 3】

再生すべき楽曲である再生楽曲を選択するための入力手段と、
選択された再生楽曲の楽曲データを取得する楽曲データ取得手段と、
選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得手段と、
前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、
前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段と、
前記歌詞音声付楽曲データを音声再生装置に送信する通信手段と、
を備えることを特徴とする端末装置。

【請求項 4】

端末装置によって実行される楽曲データ提供方法であって、
再生すべき楽曲である再生楽曲を選択するための入力工程と、
選択された再生楽曲の楽曲データを取得する楽曲データ取得工程と、
選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得工程と、
前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、
前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成工程と、
前記歌詞音声付楽曲データを音声再生装置に送信する通信工程と、
を備えることを特徴とする楽曲データ提供方法。

【請求項 5】

コンピュータを備える端末装置によって実行されるプログラムであって、
再生すべき楽曲である再生楽曲を選択するための入力手段、
選択された再生楽曲の楽曲データを取得する楽曲データ取得手段、
選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得手段、
前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段、
前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段、
前記歌詞音声付楽曲データを音声再生装置に送信する通信手段、
として前記コンピュータを機能させることを特徴とするプログラム。

【請求項 6】

請求項 5 に記載のプログラムを記憶したことを特徴とする記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、楽曲の再生に伴って歌詞の情報を出力する手法に関する。

【背景技術】

【0002】

カラオケの演奏曲に先行して歌詞データを音声合成して出力するカラオケ装置が知られている（例えば、特許文献1、2）。

【先行技術文献】

【特許文献】

10

【0003】

【特許文献1】特開4 - 67467号公報

【特許文献2】特開10 - 63274号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

カラオケ装置の場合、再生される楽曲に歌詞が含まれないため、先行技術により出力される歌詞音声聞き取りにくくなることはない。しかし、カラオケではなく通常の音楽を再生して聞いているような場合には、先行技術の手法により歌詞を音声出力すると、出力された歌詞音声は元の音楽に含まれる歌詞の部分と重なって聞き取りにくくなってしまふことがある。また、例えば車両の運転中に音楽を聞いている場合には、先行技術の手法により出力される歌詞音声は車載用ナビゲーション装置による道案内の音声メッセージなどと重なって聞き取りにくくなってしまふこともある。

20

【0005】

本発明の解決しようとする課題としては、上記のものが一例として挙げられる。本発明は、歌詞を含む音楽を再生している際に、ユーザがその曲を歌うための歌詞音声を聞き取り易く提供することを目的とする。

【課題を解決するための手段】

【0006】

請求項1に記載の発明は、サーバと、端末装置とを備える楽曲再生システムであって、前記サーバは、楽曲の歌詞データを取得する取得手段と、前記歌詞データを前記端末装置へ送信する送信手段と、を備え、前記端末装置は、再生すべき楽曲である再生楽曲を選択するための入力手段と、選択された再生楽曲の楽曲データを取得する楽曲データ取得手段と、選択された再生楽曲を指定する情報を前記サーバに送信し、当該再生楽曲に対応する歌詞データを前記サーバから受信する第1の通信手段と、前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段と、前記歌詞音声付楽曲データを音声再生装置に送信する第2の通信手段と、を備えることを特徴とする。

30

【0007】

請求項3に記載の発明は、端末装置であって、再生すべき楽曲である再生楽曲を選択するための入力手段と、選択された再生楽曲の楽曲データを取得する楽曲データ取得手段と、選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得手段と、前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段と、前記歌詞音声付楽曲データを音声再生装置に送信する通信手段と、を備えることを特徴とする。

40

【0008】

請求項4に記載の発明は、端末装置によって実行される楽曲データ提供方法であって、再生すべき楽曲である再生楽曲を選択するための入力工程と、選択された再生楽曲の楽曲

50

データを取得する楽曲データ取得工程と、選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得工程と、前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成工程と、前記歌詞音声付楽曲データを音声再生装置に送信する通信工程と、を備えることを特徴とする。

【0009】

請求項5に記載の発明は、コンピュータを備える端末装置によって実行されるプログラムであって、再生すべき楽曲である再生楽曲を選択するための入力手段、選択された再生楽曲の楽曲データを取得する楽曲データ取得手段、選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得手段、前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段、前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段、前記歌詞音声付楽曲データを音声再生装置に送信する通信手段、として前記コンピュータを機能させることを特徴とする。

【図面の簡単な説明】

【0010】

【図1】アシストボーカルの概念を示す図である。

【図2】アシストボーカル処理のフローチャートである。

【図3】スピーチ情報生成処理のフローチャートである。

【図4】スピーチ情報生成処理の概要を示す。

【図5】歌詞ブロック化の例を示す。

【図6】スピーチ挿入方法の例を示す。

【図7】スピーチ強調処理の例を示す。

【図8】スピーチ強調処理の他の例に係る構成を示す。

【図9】スピーチ強調処理の他の例に係る構成を示す。

【図10】楽曲再生システムの全体構成を示すブロック図である。

【図11】端末装置の内部構成例を示すブロック図である。

【図12】第1実施例の楽曲再生システムによるアシストボーカル処理のフローチャートである。

【図13】第2実施例の楽曲再生システムによるアシストボーカル処理のフローチャートである。

【図14】スピーチのみを再生するアシストボーカル処理のフローチャートである。

【図15】外部ソースにより再生されている楽曲の特定方法を説明する図である。

【発明を実施するための形態】

【0011】

本発明の好適な実施形態では、サーバと、端末装置とを備える楽曲再生システムにおいて、前記サーバは、楽曲の歌詞データを取得する取得手段と、前記歌詞データを前記端末装置へ送信する送信手段と、を備え、前記端末装置は、再生すべき楽曲である再生楽曲を選択するための入力手段と、選択された再生楽曲の楽曲データを取得する楽曲データ取得手段と、選択された再生楽曲を指定する情報を前記サーバに送信し、当該再生楽曲に対応する歌詞データを前記サーバから受信する第1の通信手段と、前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段と、前記歌詞音声付楽曲データを音声再生装置に送信する第2の通信手段と、を備える。

【0012】

上記の楽曲再生システムは、サーバと端末装置とを備える。再生すべき楽曲である再生楽曲をユーザが選択すると、端末装置は、選択された再生楽曲を指定する情報をサーバに送信する。サーバは、楽曲の歌詞データを取得し、端末装置へ送信する。端末装置は、再

10

20

30

40

50

生楽曲に対応する歌詞データをサーバから受信し、歌詞データに基づいて歌詞音声データを生成する。そして、端末装置は、楽曲中の歌詞部分に先行するように歌詞音声データを楽曲データに加算して歌詞音声付楽曲データを生成し、音声再生装置に送信する。こうして、ユーザが選択した楽曲についての歌詞音声付楽曲データが音声再生装置により再生される。

【0013】

上記の楽曲再生システムの一態様では、前記楽曲データ取得手段は、前記サーバから前記楽曲データを取得する。

【0014】

本発明の他の好適な実施形態では、端末装置は、再生すべき楽曲である再生楽曲を選択するための入力手段と、選択された再生楽曲の楽曲データを取得する楽曲データ取得手段と、選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得手段と、前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段と、前記歌詞音声付楽曲データを音声再生装置に送信する通信手段と、を備える。この端末装置によれば、歌詞音声付楽曲データを生成することにより、楽曲とともに歌詞音声を再生することができる。

【0015】

本発明の他の好適な実施形態では、端末装置によって実行される楽曲データ提供方法は、再生すべき楽曲である再生楽曲を選択するための入力工程と、選択された再生楽曲の楽曲データを取得する楽曲データ取得工程と、選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得工程と、前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段と、前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成工程と、前記歌詞音声付楽曲データを音声再生装置に送信する通信工程と、を備える。この方法によれば、歌詞音声付楽曲データを生成することにより、楽曲とともに歌詞音声を再生することができる。

【0016】

本発明の他の好適な実施形態では、コンピュータを備える端末装置によって実行されるプログラムは、再生すべき楽曲である再生楽曲を選択するための入力手段、選択された再生楽曲の楽曲データを取得する楽曲データ取得手段、選択された再生楽曲に対応する歌詞データを取得する歌詞データ取得手段、前記歌詞データに基づいて、歌詞音声データを生成する歌詞音声データ生成手段、前記楽曲中の歌詞部分に先行するように、前記歌詞音声データを前記楽曲データに加算して歌詞音声付楽曲データを生成する歌詞音声付楽曲データ生成手段、前記歌詞音声付楽曲データを音声再生装置に送信する通信手段、として前記コンピュータを機能させる。このプログラムをコンピュータで実行することにより、上述の端末装置を実現することができる。このプログラムは記憶媒体に記憶して取り扱うことができる。

【実施例】

【0017】

以下、図面を参照して本発明の好適な実施例について説明する。

【0018】

[1] アシストボーカル

[1.1] アシストボーカルの概念

車両を運転しているユーザが車内で音楽を再生して聞いている際、聞いている曲を歌いたくなることがある。しかし、運転中は歌詞の情報を見ることができないため、ユーザはその曲の歌詞を記憶していないと歌うことはできない。

【0019】

本実施例では、歌詞を含む楽曲を再生している際に、その楽曲に含まれる歌詞を音声信号として出力し、ユーザに教える。具体的には、端末装置のメモリなどに記憶されている

10

20

30

40

50

楽曲を再生している際に、その楽曲に含まれる歌詞を、その歌詞がその楽曲中で再生される前に、音声として出力してユーザに伝える。これにより、ユーザは、運転中であっても、再生中の曲を歌うことができる。また、運転手以外のユーザも、歌詞集などを見ることなく曲を歌うことができる。

【 0 0 2 0 】

このように、楽曲中で歌詞が再生されるタイミングに先行して、その歌詞の内容を音声出力してユーザに伝える機能を「アシストボーカル」と呼ぶ。なお、本実施例では、再生される楽曲はカラオケではなく、歌詞を含む通常の曲であるものとする。

【 0 0 2 1 】

図 1 は、アシストボーカルの概念を示す。図 1 は、1つの楽曲を模式的に示したものである。図 1 の横軸は時間を示す。1つの楽曲中には、複数のブロックに分けて歌詞の部分が含まれている。再生される楽曲に含まれる歌詞の部分を「ボーカル」と呼ぶ。また、楽曲中において、ボーカル以外の部分を「間奏」と呼ぶ。よって、通常 1つの楽曲は、複数の間奏と複数のボーカルとにより構成される。

【 0 0 2 2 】

図 1 の例では、楽曲は、3つのボーカル 1～3 と、複数の間奏とにより構成されている。ボーカル 1 の内容（歌詞）は「あいうえお」であり、ボーカル 2 の内容は「かきくけこ」であり、ボーカル 3 の内容は「さしすせそ」であるものとする。

【 0 0 2 3 】

このような楽曲が再生されている状況において、本実施例では、楽曲中のボーカル 1 が再生されるタイミングに先行して、ボーカル 1 に対応する歌詞「あいうえお」が音声出力される。なお、本明細書では、アシストボーカルにより音声出力される歌詞音声を「スピーチ」と呼んで、楽曲中に含まれる「ボーカル」と区別する。

【 0 0 2 4 】

図 1 の例では、ボーカル 1 に先行して、ボーカル 1 に対応するスピーチ 1 が出力される。同様に、ボーカル 2 に先行してスピーチ 2 が出力され、ボーカル 3 に先行してスピーチ 3 が出力される。

【 0 0 2 5 】

スピーチは、曲に含まれるボーカルの歌詞のみを音声信号として出力するものであり、基本的に音程やリズムなどの要素を含まない。また、後述するように、スピーチは基本的に対応するボーカルの前の間奏に挿入されるので、必要に応じてその長さが調整され、通常は楽曲の再生中にボーカルとして再生される場合よりも短い時間とされる。典型的な例では、スピーチは対応するボーカルの歌詞を早口で話した音声となる。

【 0 0 2 6 】

[1 . 2] アシストボーカル処理

次に、スピーチを出力するためのアシストボーカル処理について説明する。図 2 は、アシストボーカル処理のフローチャートである。なお、この処理は、車両に搭載された端末装置、典型的にはスマートフォンなどの携帯端末などにより実行されるが、その詳細については後述する。以下の説明では、端末装置が処理を実行するものとして説明する。

【 0 0 2 7 】

まず、端末装置は、アシストボーカルがオンになっているか否かを判定する（ステップ S 1 ）。ここで、アシストボーカルのオン/オフは、ユーザが手動で行う場合と、自動で行う場合とがある。手動で行う場合、ユーザはアシストボーカルによりスピーチの再生を行いたいときに所定のボタンなどを操作してアシストボーカルをオンに設定し、端末装置はこれを検出する。一方、自動で行う場合、端末装置は例えばマイクなどを利用してユーザの声を判定し、ユーザが曲を歌っている又は歌唱に準ずる行為を行っている場合に、自動的にアシストボーカルをオンに設定する。なお、アシストボーカルの自動設定方法についてはさらに後述する。

【 0 0 2 8 】

アシストボーカルがオンに設定されていない場合（ステップ S 1 : N o ）、処理は終了

10

20

30

40

50

する。一方、アシストボーカルがオンに設定されている場合（ステップS1：Yes）、端末装置は、再生中の楽曲を特定する（ステップS2）。この場合に、車内で再生されている楽曲は、サーバからダウンロードされるなどして端末装置の内部に記憶されている楽曲、CDや車載器のメモリなどの記憶媒体に記憶されている楽曲、ラジオなどから再生されている楽曲などを含む。端末装置の内部に記憶されている楽曲を再生している場合、端末装置はその再生中の楽曲を容易に特定することができる。一方、CDなどの記憶媒体に記憶されている楽曲が再生されている場合やラジオから楽曲が再生されている場合には、端末装置は、車内のスピーカから再生されている楽曲をマイクで集音し、そのオーディオデータを外部の音楽検索サーバに送信する。音楽検索サーバは、多数の楽曲のデータをデータベース化して記憶しており、端末装置から受信したオーディオデータと一致する楽曲を特定してその楽曲を示す情報（例えば、曲名、アーティスト名など、以下、「楽曲特定情報」と呼ぶ。）を端末装置に送信する。こうして、端末装置は、現在再生されている楽曲の楽曲特定情報を取得する。

10

【0029】

こうして、再生中の楽曲が特定されると、端末装置は、スピーチ情報生成処理を実行する（ステップS3）。図3は、スピーチ情報生成処理のフローチャートである。また、図4は、スピーチ情報生成処理の概要を示す。

【0030】

図3において、端末装置は、ステップS2で特定された楽曲の歌詞データを外部サーバなどから取得する（ステップS31）。ここで、「歌詞データ」とは、その楽曲において、どのタイミングにどのような歌詞が再生されるかを規定する情報であり、具体的には、楽曲に含まれる歌詞を示す歌詞テキストデータと、その歌詞が再生される再生時刻（曲の開始時刻からの経過時間）を示す再生時刻データとを対応付けた情報である。

20

【0031】

次に、端末装置は、楽曲解析データを取得する（ステップS32）。楽曲解析データとは、その楽曲における拍位置、小節位置などの音楽的特徴を示す情報であり、再生された楽曲のオーディオデータに基づいて生成される。具体的には、端末装置は内部に楽曲解析アプリケーションを内蔵しておき、車両のスピーカから再生された楽曲をマイクで集音してオーディオデータを取得し、そのオーディオデータを解析することにより拍位置などの楽曲解析データを取得する。なお、端末装置に楽曲解析アプリケーションを内蔵する代わりに、外部の楽曲解析装置やサーバなどを利用して楽曲解析データを取得してもよい。

30

【0032】

次に、端末装置は、歌詞ブロック化を行う（ステップS33）。歌詞ブロック化とは、ステップS31で取得した歌詞データに含まれる歌詞テキストデータをブロック化する処理であり、1つのブロックは、1つのスピーチに対応する。即ち、歌詞ブロック化は、歌詞テキストデータを、スピーチの単位に分割する処理である。

【0033】

図4の例では、端末装置は、歌詞テキストデータとして「あいうえおかきくけこさしすせそ」を取得しており、端末装置は、これを「あいうえお」、「かきくけこ」、「さしすせそ」の3つのブロックに分割してブロック歌詞データを生成する。

40

【0034】

図5は、歌詞ブロック化の例を示す。図5（A）に第1の方法を示す。この方法では、楽曲に含まれる間奏と間奏との間を1つのブロックとする。なお、「間奏」は、楽曲のうち「ボーカル」以外の部分である。具体的には、端末装置は、ボーカル以外の区間（非ボーカル区間）の長さ I_t が所定長さ t_1 よりも長い場合に、その区間を間奏と判定する。

【0035】

但し、例外的に、間奏の長さとの関係で複数のブロックを1つのブロックにまとめる場合がある。図5（B）に示す例のように、ボーカル3の長さ V_{t3} に対して、その直前の間奏2の長さ I_{t2} が非常に短い（ $I_{t2} < 1 \cdot V_{t3}$ ；1は任意の係数）場合、間奏2の間にボーカル3のスピーチを出力することは難しい。このような場合に、その1つ

50

の前の間奏 1 の長さ $I t 1$ が所定長より長ければ、端末装置は、ボーカル 2 とボーカル 3 を 1 つのブロックとする。これにより、ボーカル 2 とボーカル 3 に対応するスピーチは間奏 1 においてされる。

【0036】

図 5 (C) に第 2 の方法を示す。この方法では、端末装置は、歌詞データに含まれる区切りに基づいて各ブロックを決定する。即ち、歌詞データに含まれる歌詞テキストデータに予め区切りの情報が含まれている場合には、端末装置はその区切りに従って歌詞テキストデータをブロック化することができる。

【0037】

次に、端末装置は、歌詞スピーチ化を行う (ステップ S 3 4)。歌詞ブロック化により得られたブロック歌詞データはあくまで歌詞を示すテキストデータであり、歌詞スピーチ化はブロック歌詞データを音声データに変換する処理である。具体的には、端末装置は、テキスト - 音声変換 (TTS: Text To Speech) ソフトウェアを内蔵し、ステップ S 3 3 で得られた各ブロック歌詞データを音声データに変換する。これにより、図 4 に示すように、各ブロック歌詞データから、音声データであるスピーチ 1 ~ 3 が生成される。なお、端末装置に TTS ソフトウェアを内蔵する代わりに、外部サーバなどによる TTS 変換を利用してもよい。

【0038】

次に、端末装置は、スピーチ長変更を行う (ステップ S 3 5)。スピーチ長変更とは、歌詞スピーチ化により得られた各スピーチの時間的な長さを短縮して、短い時間で再生できるようにする処理である。既に述べたように、各スピーチは対応するボーカルに先行する間奏において再生されるが、間奏の時間的な長さには制限があるので、スピーチを短くして再生する必要がある。このため、スピーチ長変更が行われる。

【0039】

基本的には、人間により聞き取り可能な範囲で、各スピーチの再生時間を短く (再生速度を速く) する。例えば、ステップ S 3 4 で得られた各スピーチの時間的な長さ (「オリジナルスピーチ長」と呼ぶ。) を「 $S t$ 」とし、スピーチ長変換係数を「 2 」とすると、スピーチ長変更による変更後の長さ「 $S t v$ 」は、

$$S t v = S t \cdot 2 \quad (2 < 1.0) \quad (1)$$

で与えられる。例えば、 $2 = 0.7$ とすれば、スピーチ長変更により各スピーチは元の 3 割増しの速さで再生されることになる。

【0040】

また、上記のような一括変更に加えて、各スピーチ毎に対応する間奏の時間に応じてさらに再生時間を短くしてもよい。なお、この場合、同じ文字数のスピーチ、又は、同じ歌詞の言葉であっても、曲中の位置 (先行する間奏の長さ) に応じて、再生時間が異なることになる。

【0041】

次に、端末装置は、スピーチ挿入タイミングを算出する (ステップ S 3 6)。端末装置は、あるボーカルに対応するスピーチを、そのボーカルの再生タイミングに先行して挿入する。図 4 に示す例では、ボーカル 1 に対応するスピーチ 1 はボーカルの再生タイミングより前に挿入される。同様に、ボーカル 2 に対応するスピーチ 2 はボーカル 2 の再生タイミングより前に挿入され、ボーカル 3 に対応するスピーチ 3 はボーカル 3 の再生タイミングより前に挿入される。

【0042】

スピーチを挿入する方法の具体例を図 6 に示す。図 6 は、ボーカル 2 に対応するスピーチ 2 を挿入するタイミングの例を示す。

【0043】

方法 1 では、スピーチは、対応するボーカルの開始タイミングよりも一定時間前に終了する。具体的に、図 6 に示すように、スピーチ 2 は、ボーカル 2 の再生開始タイミングより一定時間 $T 2$ 前に終了するように挿入される。即ち、スピーチ 2 はボーカル 2 の再生開

10

20

30

40

50

始より一定時間 T 2 前に終了する。この場合、スピーチ 2 の再生開始タイミングはスピーチ 2 の長さに応じて決まる。方法 1 では、スピーチの再生が終了してから、対応するボーカルが再生されるまでに一定時間が確保されるので、ユーザは余裕を持ってボーカル部分を歌うことができる。

【 0 0 4 4 】

方法 2 では、スピーチの終了タイミングを楽曲の拍位置と一致させる。具体的に、図 6 の例では、スピーチ 2 は、ボーカル 2 の再生開始タイミングより N 拍前 (N は任意の整数 ; 本例では N = 1) に終了するように挿入される。この場合、スピーチ 2 の再生開始タイミングはスピーチ 2 の長さに応じて決まる。なお、楽曲の拍の位置は、前述の楽曲解析データから取得される。

10

【 0 0 4 5 】

方法 3 では、スピーチの再生開始タイミングと再生終了タイミングの両方を楽曲の拍位置と一致させる。具体的に、図 6 の例では、スピーチ 2 の再生開始タイミング及び再生終了タイミングをとともに 4 拍子の 3 拍目に一致させている。

【 0 0 4 6 】

方法 2、3 のように、スピーチの終了タイミング、又は、開始 / 終了タイミングの両方を楽曲の拍位置と一致させると、スピーチが楽曲と連動するのでユーザが楽曲を歌いやすくなる。

【 0 0 4 7 】

以上のようにして、端末装置は、スピーチの挿入タイミングを決定する。具体的には、各スピーチについて、その再生開始タイミングと再生終了タイミングとを、楽曲の先頭からの経過時間により規定する。各スピーチの再生開始タイミングと再生終了タイミングは、スピーチ情報の一部として記憶される。即ち、スピーチ情報は、各スピーチに対応する音声信号 (以下、「スピーチ信号」とも呼ぶ。) と、各スピーチの再生開始タイミング / 再生終了タイミングとを含む。

20

【 0 0 4 8 】

次に、処理は図 2 に示すメインルーチンに戻り、端末装置は、再生中の楽曲の現在の再生位置を取得する (ステップ S 4)。具体的には、端末装置は、再生中の楽曲の再生開始時刻からの経過時間をカウントすることにより、現在の再生位置を取得する。

【 0 0 4 9 】

次に、端末装置は、スピーチ強調処理を行う (ステップ S 5)。スピーチ強調処理は、楽曲に含まれるボーカルと、スピーチとを区別して聞き取り易くする処理であるが、その詳細は後述する。

30

【 0 0 5 0 】

次に、端末装置は、スピーチ情報に含まれる各スピーチの再生開始タイミング / 再生終了タイミングと、現在の再生位置とに基づいて、スピーチを再生する (ステップ S 6)。具体的には、スピーチの再生開始タイミングでスピーチの再生を開始し、スピーチの再生終了タイミングでスピーチの再生を終了する。これにより、楽曲中のボーカルに先行して、対応するスピーチが再生されることになる。

【 0 0 5 1 】

次に、端末装置は、スピーチの再生を終了すべきか否かを判定する (ステップ S 7)。スピーチの再生を終了すべき場合とは、スピーチ情報が無くなった場合、楽曲の再生自体が終了した場合、ユーザの操作によりアシストボーカルがオフされた場合、などが挙げられる。スピーチの再生を終了すべきでない場合 (ステップ S 7 : N o)、処理はステップ S 4 へ戻り、スピーチの再生を継続する。一方、スピーチの再生を終了すべきである場合 (ステップ S 7 : Y e s)、アシストボーカル処理は終了する。

40

【 0 0 5 2 】

[1 . 3] アシストボーカルの自動オン設定方法

次に、図 2 に示すアシストボーカル処理のステップ S 1 においてアシストボーカルを自動的にオンに設定する方法について説明する。

50

【 0 0 5 3 】

基本的な方法としては、端末装置は、ユーザが発している音声をマイクで集音し、ユーザが楽曲に合わせて歌唱している（歌を歌っている）又は歌唱に準ずる行為を行っているとは判定される場合にアシストボーカルを自動的にオンにする。例えば、マイクにより集音した音声データを解析した結果、鼻歌を歌っている、断片的に曲を歌っている、ハミングしているなどと判定される場合には、アシストボーカルをオンにする。一方、音声データが歌唱しているのではなく、同乗者との会話である場合にはアシストボーカルをオンにしない。音声データが鼻歌を歌っている部分を含んでいるような場合でも、大部分が会話であるような場合にもアシストボーカルをオンにはしない。

【 0 0 5 4 】

なお、音声データに含まれるユーザの音声歌唱であるか否かは、音声データに含まれるリズムや音程の有無に基づいて判断することができる。例えばリズムが規則的である場合や音程の変化が大きい場合には歌唱であると判断し、リズムが不規則である場合は音程の変化が小さい場合に歌唱ではない（会話である）と判断することができる。また、前述の楽曲解析アプリケーションを利用し、音声データから拍や小節が抽出できた場合に歌唱であると判断し、抽出できない場合に歌唱ではないと判断してもよい。また、前述の音楽検索サーバ又は音楽検索機能を利用し、音声データから楽曲が特定できた場合に歌唱であると判断し、楽曲が特定できない場合に歌唱ではないと判断してもよい。

【 0 0 5 5 】

また、端末装置は、集音した音声データと、再生中の楽曲との相関を算出し、一定値以上の相関がある場合に、ユーザが歌唱していると判断してアシストボーカルをオンにしてもよい。また、端末装置が再生中の曲の歌詞データを既に取得している場合には、マイクにより集音した音声データと歌詞データとの相関が一定値以上である場合に、ユーザが歌っていると判断してもよい。また、歌詞データに基づいて、歌詞が存在しないはずの楽曲の間奏位置においてもユーザの音声出力されている場合には、それは会話であると判断してもよい。

【 0 0 5 6 】

また、マイクで集音したリズムの情報を利用してもよい。例えば、ユーザが楽曲のリズムに合わせて手や指でステアリングなどを叩いているとか、足で床を踏んでリズムを取っていると判断される場合には、ユーザが歌唱に準ずる行為を行っているとは判定し、アシストボーカルをオンにしてもよい。この場合、マイクで集音したリズムと再生中の楽曲のリズムとの相関を算出し、相関が一定値以上である場合にアシストボーカルをオンにしてもよい。また、再生中の楽曲のリズムとの相関を算出しなくても、マイクで集音されたリズムが、一定のリズムの繰り返しになっているような場合には、アシストボーカルをオンにしてもよい。

【 0 0 5 7 】

さらには、車内を撮影するカメラでユーザの状態を撮影し、ユーザが楽曲に合わせて首を振っているような場合に、アシストボーカルをオンにしてもよい。また、車内を撮影するカメラにより、助手席や後部座席に同乗者がいるか否かを検出し、同乗者の有無により、ユーザが歌っているのか会話しているのかの判定基準を変化させてもよい。

【 0 0 5 8 】

また、上記の例では、ユーザが歌唱していると判断した場合に、アシストボーカルをオンにする例を説明したが、ユーザが歌唱していても、ユーザが歌詞を知っていてアシストボーカルを再生する必要がないと判断した場合には、アシストボーカルをオンにしなくてもよい。具体的には、例えば集音した音声データと、再生中の楽曲との相関が一定値以上であり、かつ歌詞データとの相関が一定値以上である場合には、ユーザが歌詞を知っていると判断し、歌唱していてもアシストボーカルをオンにしない。

【 0 0 5 9 】

ただしこの場合、ユーザが途中から歌詞が分からなくなる可能性があるため、スピーチ情報を生成し、出力する準備をしておいてもよい。そのあとに、集音した音声データと、

10

20

30

40

50

再生中の楽曲との相関が一定値未満であり、または歌詞データとの相関が一定値未満である場合には、ユーザは歌詞を知らないと判断し、アシストボーカルを出力する。

【 0 0 6 0 】

また、上記の例では、アシストボーカルの自動オン設定の方法について説明したが、アシストボーカルの自動オフ設定も行うことができる。アシストボーカルをオンしている間に、ユーザが楽曲に合わせて歌唱していない（歌を歌っていない）又は歌唱に準ずる行為（鼻歌を歌っている、断片的に曲を歌っている、ハミングをしている等）を行っていないと判定された場合に、アシストボーカルを自動的にオフにしてもよい。同様に、会話が検出されたら、アシストボーカルを自動的にオフにしてもよいし、リズムをとっていないと判断されたり、ユーザが楽曲に合わせて頭を振っていないと判断された場合、アシストボーカルを自動的にオフにしてもよい。

10

【 0 0 6 1 】

また、上記の例では、ユーザが歌唱しているもしくは歌唱に準ずる行為をしているか否かに基づき、アシストボーカルの自動オン設定もしくは自動オフ設定を行うことを説明したが、再生されている楽曲の構成に基づき自動オン設定もしくは自動オフ設定してもよい。例えば、楽曲のサビの部分だけ歌唱したいというユーザに対しては、楽曲のサビの部分を再生する際に、アシストボーカルを自動的にオン設定し、楽曲のサビ以外の部分を再生する際に、アシストボーカルを自動的にオフ設定してもよい。逆に、サビの部分は知っていてサビ以外の部分を練習したいというユーザに対しては、楽曲のサビ以外の部分を再生する際に、アシストボーカルを自動的にオン設定し、楽曲のサビの部分を再生する際に、アシストボーカルを自動的にオフ設定してもよい。

20

【 0 0 6 2 】

[1 . 4] スピーチ強調処理

次に、図 2 に示すアシストボーカル処理のステップ S 5 において実行されるスピーチ強調処理について説明する。スピーチ強調処理は、ユーザがスピーチとボーカルとを区別して聞き取り易くする方法であり、以下のいくつかの方法を示す。

【 0 0 6 3 】

[1 . 4 . 1] スピーチとボーカルが重なる場合の処理

スピーチは基本的に対応するボーカルの直前の間奏中に再生され、ボーカルとは時間的に重ならないことが好ましい。このために前述のスピーチ長変更処理（ステップ S 3 5 ）を行うのであるが、スピーチの長さの間奏の長さによっては、スピーチ長を短縮してもスピーチを間奏中に再生しきれないこともある。即ち、間奏の長さよりも、スピーチの長さの方が長い場合、スピーチとボーカルとが部分的に重なって再生される。このようにスピーチとボーカルとを重ねて再生することに代えて、以下のいずれかの処理を行ってもよい。

30

【 0 0 6 4 】

（ 1 ）ボーカルのレベルを調整する。

【 0 0 6 5 】

スピーチとボーカルとが重なってしまう場合、ボーカルの音量レベルを下げる方法がある。図 7（ A ）は、スピーチの後方部分と、ボーカルの先頭部分とが重なり、重複部分 X が生じる場合を示す。この場合、重複部分 X においてボーカルの音量を調整する。具体的には、ボーカルの音量をスピーチが聞こえる程度まで低下させる、もしくはゼロにする。これにより、重複部分 X では、スピーチの再生が優先され、スピーチが聞き取り易くなる。

40

【 0 0 6 6 】

図 7（ B ）は、逆にスピーチの先頭部分と、1 つ前のボーカルの後方部分とが重なり、重複部分 X が生じる場合を示す。この場合にも、重複部分 X において、ボーカルの音量を調整する。具体的には、ボーカルの音量をスピーチが聞こえる程度まで低下させる、もしくはゼロにする。また、重複部分 X において、急にボーカルの音量レベルを下げるのではなく、ボーカルをフェードアウトさせて徐々に音量レベルを下げるようにしてもよい。こ

50

れにより、重複部分 X では、スピーチの再生が優先され、スピーチが聞き取り易くなる。

【 0 0 6 7 】

具体的に上記のレベル調整は、楽曲信号においてボーカルの成分と楽器などの演奏の成分とが分離している場合には、ボーカルの成分の音量レベルを低下させればよい。一方、ボーカルの部分が楽器などの演奏の部分と合成されており、ボーカルのみの音量を調整できない場合には、楽曲信号全体の音量レベルを低下させてもよいし、又は、楽曲信号のうち一般的にボーカル（人間の声）に相当する周波数帯域の成分のみ音量レベルを低下させるようにしてもよい。

【 0 0 6 8 】

（ 2 ）スピーチのレベルを調整する。

10

【 0 0 6 9 】

スピーチとボーカルとが重なってしまう場合、逆にスピーチの音量レベルを下げる方法もある。図 7（ C ）は、スピーチの後方部分と、ボーカルの先頭部分とが重なり、重複部分 X が生じる場合を示す。この場合、重複部分 X において、スピーチの音量を調整する。具体的には、スピーチの音量を低下させる、もしくはゼロにする。急にスピーチの音量を下げるのではなく、スピーチをフェードアウトさせて徐々に音量を下げるようにしてもよい。この場合、重複部分 X では、スピーチが聞き取れなくなるが、一般的にユーザがある程度知っている楽曲を聞く場合には、歌詞の全てを覚えてはいないものの、歌詞の先頭部分がわかれば、その後は歌詞を思い出して歌うことができるということも多い。よって、図 7（ C ）のように、スピーチの先頭部分が聞き取れば、スピーチの後方部分が聞き取りにくくなくても構わないということも多い。この手法はそのような場合に有効である。

20

【 0 0 7 0 】

〔 1 . 4 . 2 〕スピーチとボーカルを異なる方向から聞かせる処理

人間には、同時に異なる方向から到来する音を聞き分ける能力がある（いわゆるカクテルパーティ効果）。これを利用し、ユーザがスピーチとボーカルとを聞き分けることができるようにする手法が考えられる。なお、この手法は、スピーチとボーカルとが時間的に重なるか否かに拘わらず実行される。

【 0 0 7 1 】

（ 1 ）左右のスピーカで位相を調整する方法

図 8（ A ）は、左右のスピーカから出力されるスピーチの位相を反転させる構成を示す。左（ L ）チャンネルの楽曲信号は加算器 3 2 に供給され、右（ R ）チャンネルの楽曲信号は加算器 3 3 に供給される。一方、スピーチ信号は、そのまま加算器 3 3 に供給されるとともに、位相反転器 3 1 で位相が反転されて加算器 3 2 に供給される。加算器 3 2 の出力は左スピーカ 3 0 L に供給され、加算器 3 3 の出力は右スピーカ 3 0 R に供給される。

30

【 0 0 7 2 】

この構成によれば、ボーカルを含む楽曲の音像は左右スピーカの間に定位するのに対し、スピーチの音像はユーザの耳回りに定位することになり、ユーザはスピーチと楽曲中のボーカルとを聞き分けやすくなる。なお、図 8（ A ）の例では、位相反転器 3 1 により左スピーカ 3 0 L に供給されるスピーチ信号の位相のみを反転しているが、逆に右スピーカ 3 0 R に供給されるスピーチ信号の位相のみを反転させてもよい。また、左右のスピーカに供給されるスピーチ信号の間に一定の位相差があればスピーチの音像位置と楽曲の音像位置とを異ならせることができるので、一方のスピーカに供給されるスピーチ信号を必ずしも反転（ 1 8 0 ° 変化）させる必要はない。即ち、一方のスピーカに供給されるスピーチ信号と、他方のスピーカに供給されるスピーチ信号との間に一定の位相差を与えてやればよい。

40

【 0 0 7 3 】

（ 2 ）音像の定位を制御する方法

図 8（ B ）は、スピーチの音像を任意の位置に設定可能な構成を示す。左（ L ）チャンネルの楽曲信号は加算器 3 2 に供給され、右（ R ）チャンネルの楽曲信号は、加算器 3 3 に供給される。一方、スピーチ信号は、音像定位制御演算部 3 4、クロストークキャンセ

50

ル部 3 5 を経由して加算器 3 2、3 3 に供給される。音像定位制御演算部 3 4 は、目標のスピーカ位置と聴取位置（ユーザの位置）との間の伝達関数をスピーチ信号に畳み込み、クロストークキャンセル部 3 5 は楽曲を出力しているスピーカと聴取位置との間の伝達関数をキャンセルする処理を行う。これにより、楽曲の音像は左右のスピーカ 3 0 L、3 0 R の間に定位させるとともに、スピーチの音像を目標のスピーカ位置に定位させることができるので、ユーザはスピーチとボーカルとを聞き分けやすくなる。

【0074】

（3）ヘッドレストスピーカを利用する方法

車両のスピーカに加えて車両のシートにヘッドレストスピーカが搭載されている場合、車両のスピーカからボーカルを含む楽曲を出力し、ヘッドレストスピーカからスピーチを出力することができる。この場合の構成例を図 9 に示す。

【0075】

左右チャンネルの楽曲信号はそれぞれ車両のスピーカ 3 0 L、3 0 R に供給される。また、スピーチ信号は、そのまま右のヘッドレストスピーカ 3 5 R に供給されるとともに、位相反転器 3 1 で位相が反転されて左のヘッドレストスピーカ 3 5 L に供給される。この場合も、2 つのヘッドレストスピーカ 3 5 L、3 5 R に供給されるスピーチ信号に位相差が与えられているため、スピーチの音像は楽曲の音像と異なる位置に定位し、ユーザはスピーチと楽曲中のボーカルとを聞き分けやすくなる。なお、この例においても、図 8（A）の例と同様に、一方のヘッドレストスピーカに供給されるスピーチ信号と、他方のヘッドレストスピーカに供給されるスピーチ信号との間に一定の位相差を与えてやればよい。

【0076】

ヘッドレストスピーカを利用する場合には、運転席のヘッドレストスピーカの代わりに、助手席のヘッドレストスピーカを利用してスピーチを再生してもよい。また、車両の複数の座席にヘッドレストスピーカが搭載されている場合には、各座席毎にスピーチの再生の可否を選択して設定できるようにしてもよい。こうすると、スピーチを聞いて楽曲を歌いたい搭乗者の座席のヘッドレストスピーカのみからスピーチが再生されるように設定することができる。

【0077】

また、位相差を与えることに代えて、図 8（B）で説明した処理と同様に、音像定位制御演算部 3 4 と、クロストークキャンセル部 3 5 とを用いることで、スピーチの音像を任意の位置に定位させてもよい。これにより、ユーザがスピーチとボーカルとを聞き分けやすくなることができる。

【0078】

〔2〕システム構成

次に、上述のアシストボーカルを実現する楽曲再生システムの構成例を説明する。

【0079】

〔2.1〕第 1 実施例

第 1 実施例では、アシストボーカル処理を主として端末装置側で実行する。第 1 実施例による楽曲再生システムの全体構成を図 10 に示す。第 1 実施例の楽曲再生システムでは、複数の車両 1 と、コンテンツプロバイダ 2 と、ゲートサーバ 3 とがネットワーク 4 を介して通信可能とされる。なお、複数の車両 1 は、無線通信によりネットワーク 4 を介してコンテンツサーバ 2、ゲートサーバ 3 と通信可能となっている。

【0080】

コンテンツプロバイダ 2 は、音楽配信業者などのサーバであり、楽曲データ、楽曲のメタデータ、歌詞データなどを提供する。ゲートサーバ 3 は、本実施例によるアシストボーカルを実現するために機能するサーバであり、コンテンツプロバイダ 2 から必要な楽曲の楽曲データ、メタデータ、歌詞データなどを取得して、図示しないデータベースに記憶している。

【0081】

車両 1 の内部構成の一例を図 11（A）に示す。車両 1 は、端末装置 10 と、音楽再生

装置 20 と、スピーカ 30 とを備える。

【0082】

端末装置 10 は、典型的にはスマートフォンなどの携帯端末であり、通信部 11 と、制御部 12 と、記憶部 13 と、マイク 14 と、操作部 15 とを備える。通信部 11 は、ネットワーク 4 を通じてゲートサーバ 3 と通信する。制御部 12 は、CPU などからなり、端末装置 10 の全体を制御する。

【0083】

記憶部 13 は、ROM、RAM などのメモリであり、制御部 12 が各種の処理を実行するためのプログラムを記憶するとともに、ワークメモリとしても機能する。記憶部 13 に記憶されたプログラムを制御部 12 が実行することにより、アシストボーカル処理を含む処理が実行される。また、記憶部 13 は、ユーザが保存した楽曲の楽曲データを記憶していてもよい。

10

【0084】

マイク 14 は、車内で再生されている楽曲、ユーザによる歌唱、会話などの音声を集音して音声データを生成する。操作部 15 は、典型的にはタッチパネルなどであり、ユーザによる操作、選択の入力を受け付ける。

【0085】

音楽再生装置 20 は、例えばカーオーディオなどであり、アンプなどを含む。スピーカ 30 は、車両に搭載されたスピーカである。音楽再生装置 20 は、端末装置 10 から供給される楽曲データに基づいて楽曲をスピーカ 30 から再生する。

20

【0086】

車両 1 の内部構成の他の例を図 11 (B) に示す。この例では、車両 1 は端末装置 10 x を備える。端末装置 10 x は、図 11 (A) に示す携帯端末などの端末装置 10 とカーオーディオなどの音楽再生装置 20 の機能を併せ持つ装置である。端末装置 10 x は、端末装置 10 と同様に通信部 11、制御部 12、記憶部 13、マイク 14、操作部 15 を備えるとともに、音楽再生装置 20 に相当する音楽再生部 16 を備える。端末装置 10 x はスピーカ 30 に接続され、楽曲データに基づいて楽曲をスピーカ 30 から再生する。

【0087】

次に、第 1 実施例の楽曲再生システムによるアシストボーカル処理について説明する。図 12 は、第 1 実施例に係るアシストボーカル処理のフローチャートである。この処理では、アシストボーカル処理を主として端末装置 10 又は 10 x (以下、代表して単に「端末装置 10」と記す。) により実行する。

30

【0088】

まず、ゲートサーバ 3 は、ネットワーク 4 を介してコンテンツプロバイダ 2 に接続し、複数の楽曲について、楽曲データ及び歌詞データを取得し、内部のデータベースに保存しておく (ステップ S101)。

【0089】

端末装置 10 は、ユーザによる操作部 15 の操作により、再生すべき楽曲の指定を受け取り (ステップ S102)、その楽曲を指定する楽曲指定情報をゲートサーバ 3 へ送信する (ステップ S103)。ゲートサーバ 3 は、受け取った楽曲指定情報に対応する楽曲の楽曲データ及び歌詞データをデータベースから取得し、端末装置 10 へ送信する (ステップ S104)。

40

【0090】

次に、端末装置 10 は、受信した楽曲データ及び歌詞データを利用して、ステップ S105 ~ S109 の処理を行う。ここで、ステップ S105 ~ S109 の処理は、図 2 におけるステップ S3 ~ S7 と同様であるので、説明を省略する。

【0091】

こうして、第 1 実施例の楽曲再生システムにおいては、車両 1 に搭載された端末装置 10 が主としてアシストボーカル処理を実行する。

【0092】

50

上記の例では、ステップ S 1 0 1 でゲートサーバ 3 はコンテンツプロバイダから楽曲データを取得しているが、楽曲データが端末装置 1 0 に保存されている場合には、ゲートサーバ 3 は端末装置 1 0 から楽曲データを取得してもよい。また、ゲートサーバ 3 内のデータベースに楽曲データが保存されている場合には、そこから楽曲データを取得してもよい。

【 0 0 9 3 】

なお、上記の構成において、ゲートサーバ 3 は本発明のサーバの一例であり、操作部 1 5 は本発明の入力手段の一例であり、通信部 1 1 は本発明の通信手段の一例である。また、制御部 1 2 は本発明の歌詞データ取得手段、歌詞音声データ生成手段、歌詞音声付楽曲データ生成手段の一例である。

【 0 0 9 4 】

[2 . 2] 第 2 実施例

第 2 実施例では、アシストボーカル処理の一部をゲートサーバ 3 側で実行する。第 2 実施例による楽曲再生システムの全体構成は、図 1 0 に示す第 1 実施例と同様であるので、説明を省略する。

【 0 0 9 5 】

次に、第 2 実施例の楽曲再生システムによるアシストボーカル処理について説明する。図 1 3 は、第 2 実施例に係るアシストボーカル処理のフローチャートである。この処理では、ゲートサーバ 3 がスピーチ情報を生成し、さらにスピーチ付楽曲データを生成して端末装置 1 0 へ送信する。端末装置 1 0 は、スピーチ付楽曲データを受信して再生する。以下、詳しく説明する。

【 0 0 9 6 】

まず、ゲートサーバ 3 は、ネットワーク 4 を介してコンテンツプロバイダ 2 に接続し、複数の楽曲について、楽曲データ及び歌詞データを取得し、内部のデータベースに保存する（ステップ S 2 0 1）。そして、ゲートサーバ 3 は、各楽曲について、取得した楽曲データと歌詞データとに基づいてスピーチ情報を生成する（ステップ S 2 0 2）。このスピーチ情報生成処理は、図 2 のステップ S 3 と同一であるので、説明を省略する。

【 0 0 9 7 】

スピーチ情報を生成すると、ゲートサーバ 3 は、楽曲データにスピーチを付加してスピーチ付楽曲データを生成する（ステップ S 2 0 3）。具体的に、ゲートサーバ 3 は、生成したスピーチ情報に基づいて、各スピーチに対応するスピーチ信号を、図 3 のステップ S 3 6 の処理により算出したタイミングで楽曲データに合成し、スピーチ付楽曲データを生成してデータベースに記憶する。つまり、スピーチ付楽曲データは、そのまま再生することにより、楽曲に加えてスピーチが再生されるデータである。

【 0 0 9 8 】

端末装置 1 0 は、ユーザによる操作部 1 5 の操作により、再生すべき楽曲の指定を受け取り（ステップ S 2 0 4）、その楽曲を指定する楽曲指定情報をゲートサーバ 3 へ送信する（ステップ S 2 0 5）。ゲートサーバ 3 は、受け取った楽曲指定情報に対応する楽曲のスピーチ付楽曲データを端末装置 1 0 へ送信する（ステップ S 2 0 6）。

【 0 0 9 9 】

次に、端末装置 1 0 は、受信したスピーチ付楽曲データを再生する（ステップ S 2 0 7）。これにより、楽曲の再生中の適切なタイミングで、スピーチが再生される。次に、端末装置 1 0 は、楽曲の再生を終了すべきか否かを判定する（ステップ S 2 0 8）。その楽曲が最後まで再生された場合、又は、ユーザが再生を中止した場合など、再生を終了すべき場合には（ステップ S 2 0 8 : Y e s）、端末装置 1 0 は再生を終了する。一方、楽曲の再生を終了すべきではない場合（ステップ S 2 0 8 : N o）、処理はステップ S 2 0 7 へ戻り、スピーチ付楽曲データの再生が継続される。

【 0 1 0 0 】

こうして、第 2 実施例の楽曲再生システムにおいては、ゲートサーバ 3 側でスピーチ付楽曲データが生成され、端末装置 1 0 へ提供される。端末装置 1 0 は、受信したスピーチ

10

20

30

40

50

付楽曲データを再生することにより、スピーチを含む楽曲を聞くことができる。

【0101】

上記の例では、ステップS201でゲートサーバ3はコンテンツプロバイダから楽曲データを取得しているが、楽曲データが端末装置10に保存されている場合には、ゲートサーバ3は端末装置10から楽曲データを取得してもよい。また、ゲートサーバ3内のデータベースに楽曲データが保存されている場合には、そこから楽曲データを取得してもよい。

【0102】

[3]スピーチのみを再生するアシストボーカル

上述のアシストボーカル処理では、端末装置10により再生している楽曲に対して、スピーチを付加して再生している。しかし、端末装置10以外のソース、例えば車内のラジオ、CDなど(以下、「外部ソース」と呼ぶ。)から再生されている楽曲に対してスピーチを付加することができれば便利である。この場合、端末装置10は、基本的に上述の方法でスピーチ情報を生成し、外部ソースから再生されている楽曲の再生位置に応じたタイミングでスピーチのみを再生すればよい。

【0103】

この場合のアシストボーカル処理のフローチャートを図14に示す。まず、端末装置10は、外部ソースから再生されている楽曲をマイク14により集音して再生楽曲データを取得し(ステップS151)、これをゲートサーバ3へ送信する(ステップS152)。

【0104】

ゲートサーバ153は、端末装置10から再生楽曲データを受信し、対応する楽曲及びその再生位置を特定する(ステップS153)。具体的には、ゲートサーバ3は、前述の音楽検索サーバの機能を有する音楽検索部を備え、再生楽曲データに基づいて、その楽曲を特定するとともに、その再生楽曲データの部分に対応する再生位置を特定する。そして、ゲートサーバ3は、特定した楽曲の楽曲名やアーティスト名とともに、歌詞データと、再生位置情報とを端末装置10へ送信する(ステップS154)。

【0105】

端末装置10は、受信した歌詞データを利用して、スピーチ情報を生成する(ステップS155)。なお、スピーチ情報の生成は、図3を参照して説明したのと同様の方法で行われる。なお、端末装置10は、マイク14で取得した再生楽曲データを解析することにより、楽曲解析データを取得することができる(図3のステップS32の処理)。

【0106】

次に、端末装置10は、ゲートサーバ3から取得した再生位置情報に基づいて、その楽曲における現在の再生位置を算出する(ステップS156)。この方法については後述する。次に、端末装置10は、スピーチ強調処理を行い(ステップS157)、外部ソースにより再生されている楽曲に合わせて適切なタイミングでスピーチを再生する(ステップS158)。これにより、外部ソースから再生されている楽曲に合わせて、スピーチが再生される。

【0107】

そして、端末装置10は、スピーチの再生を終了すべきか否かを判定し(ステップS159)、終了させるべきでない場合には、ステップS156へ戻って処理を継続する。一方、外部ソースからの楽曲の再生が終了した場合、再生されている楽曲が別の楽曲に変わった場合、再生すべきスピーチが無くなった場合など、スピーチの再生を終了すべき場合には(ステップS159:Yes)、処理を終了する。

【0108】

次に、図15を参照して、ステップS156において楽曲の現在の再生位置を特定する方法を説明する。端末装置10からゲートサーバ3へ送信される再生楽曲データは、実際には複数のオーディオフレームのデータとなる。即ち、端末装置10は、外部ソースにより再生されている楽曲をマイク14で集音し、複数のオーディオフレームとして順次ゲートサーバ3へ送信する。

10

20

30

40

50

【 0 1 0 9 】

図 1 5 の例では、端末装置 1 0 は、外部ソースにより再生されている楽曲のオーディオフレーム n 、 $(n+1)$ 、 $(n+2)$ 、... を、再生楽曲データとして順次ゲートサーバ 3 へ送信する。この際、端末装置 1 0 は、最初に再生楽曲データを送信した時刻、図 1 5 の例ではオーディオフレーム n を送信した時刻（以下、「基準時刻 t_0 」と呼ぶ。）を記憶しておく。

【 0 1 1 0 】

ゲートサーバ 3 の音楽検索部は、データベースに記憶された多数の楽曲の情報を参照し、受信した複数のオーディオフレームに基づいて楽曲を特定する。図 1 5 の例では、ゲートサーバ 3 の音楽検索部は、オーディオフレーム $n \sim (n+4)$ に基づいて楽曲を特定できたものとする。この場合、ゲートサーバ 3 は、楽曲判定結果として、楽曲名、アーティスト名などに加えて、端末装置 1 0 から最初に受信したオーディオフレーム n の曲先頭からの再生時間（ t_n ）を再生位置情報として端末装置 1 0 へ送信する。即ち、図 1 4 のステップ S 1 5 4 でゲートサーバ 3 から端末装置 1 0 へ送信される再生位置情報は、端末装置 1 0 がゲートサーバ 3 へ最初に送信したオーディオフレーム n の、その楽曲の先頭からの経過時間となっている。そこで、ステップ S 1 5 6 において、端末装置 1 0 は、予め記憶していた基準時刻 t_0 から現在までの経過時間の経過時間 t を算出し、これを再生時間 t_n に加算する。即ち、ゲートサーバ 3 から送信される再生時間 t_n は、その楽曲の先頭からオーディオフレーム n までの時間であり、経過時間 t はオーディオフレーム n から現在までの時間である。よって、現在の再生位置（再生時間） T_c は、以下の式で算出される。

【 0 1 1 1 】

$$T_c = t_n + t \quad (2)$$

以上のように、ゲートサーバ 3 に音楽検索機能を設け、再生楽曲データに基づいて楽曲及びその再生位置を特定することにより、外部ソースから再生されている楽曲に合わせてスピーチを再生することができる。また、ゲートサーバ 3 に音楽検索機能を設ける代わりに、外部の音楽検索サーバを利用しても良い。

【 0 1 1 2 】

なお、ステップ S 1 5 9 では、1 つの楽曲が終了したときに再生を終了してもよいが、1 つの楽曲が終了した後で別の楽曲が再生されているような場合には、処理を継続してもよい。即ち、端末装置 1 0 からゲートサーバ 3 への楽曲再生データの送信を継続している間は、スピーチの再生を継続することとしてもよい。これにより、外部ソースから再生される曲が変わっても、それに追従してスピーチの再生を継続することが可能となる。

【 符号の説明 】

【 0 1 1 3 】

- 1 車両
- 2 コンテンツプロバイダ
- 3 ゲートサーバ
- 4 ネットワーク
- 1 0、1 0 x 端末装置
- 1 2 制御部
- 1 3 記憶部
- 1 4 マイク
- 2 0 音楽再生装置
- 3 0 スピーカ

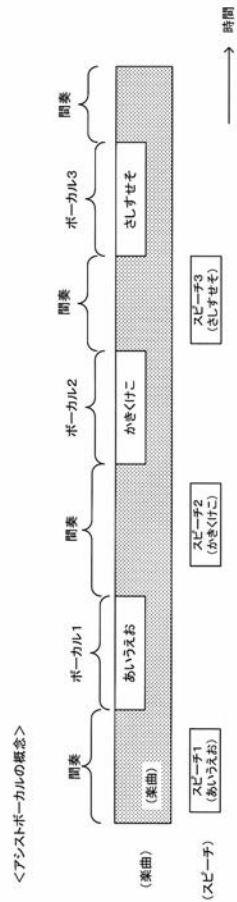
10

20

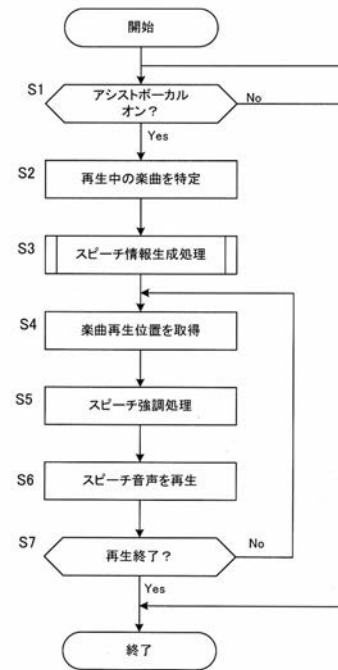
30

40

【図 1】



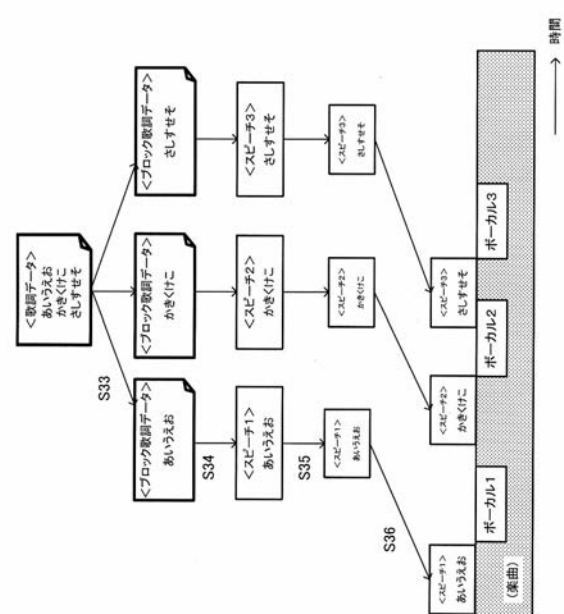
【図 2】



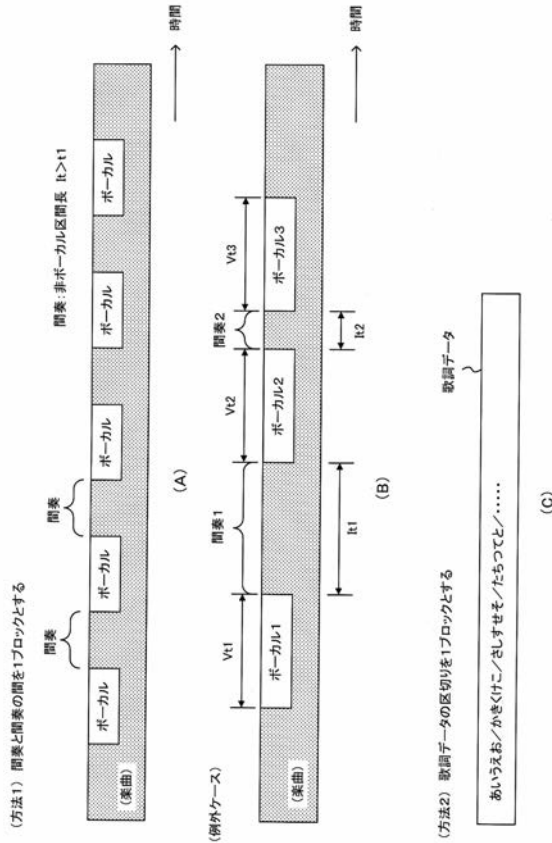
【図 3】



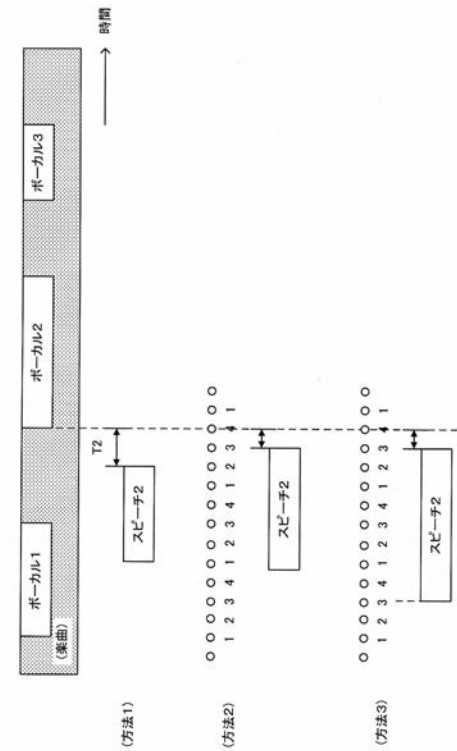
【図 4】



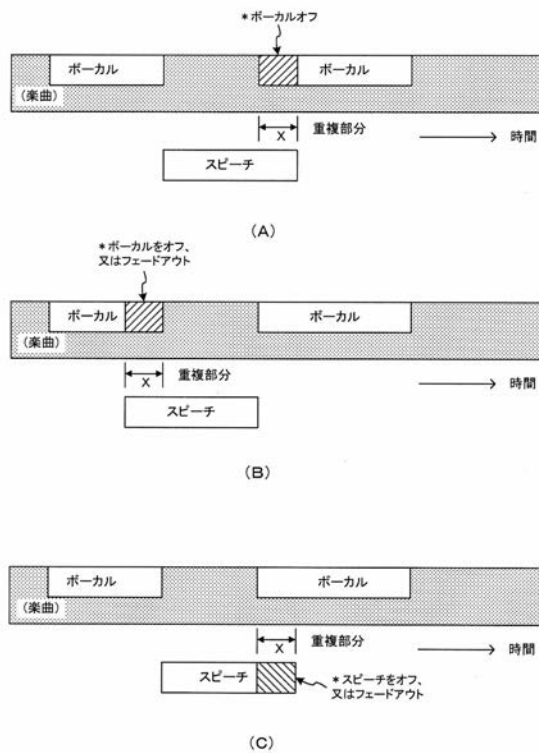
【図 5】



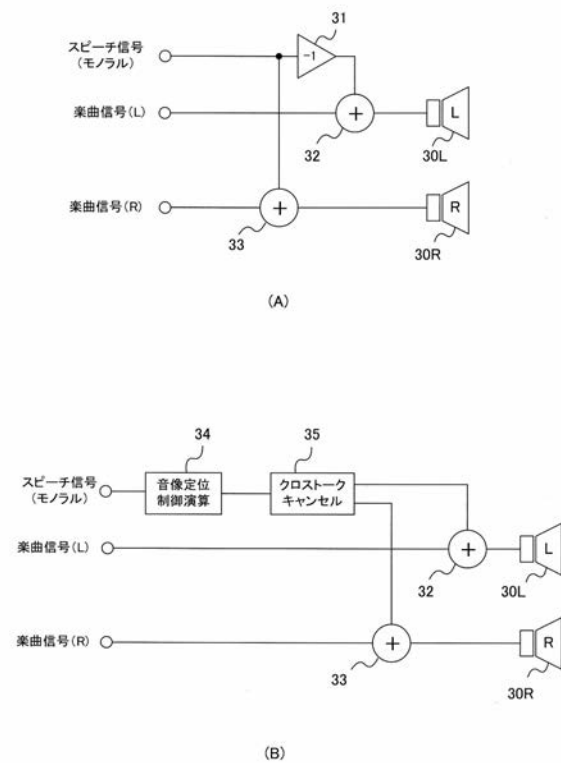
【図 6】



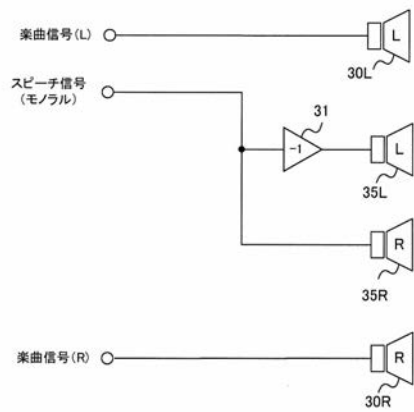
【図 7】



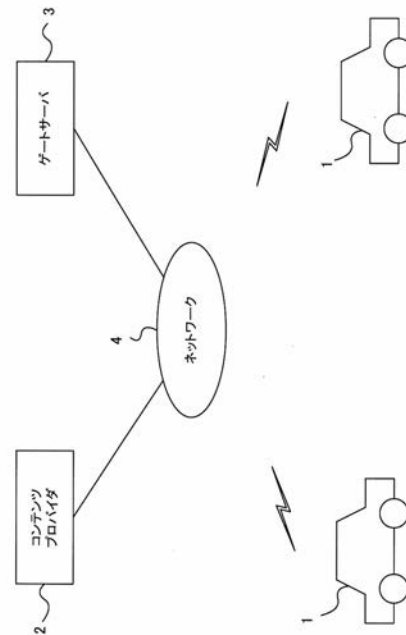
【図 8】



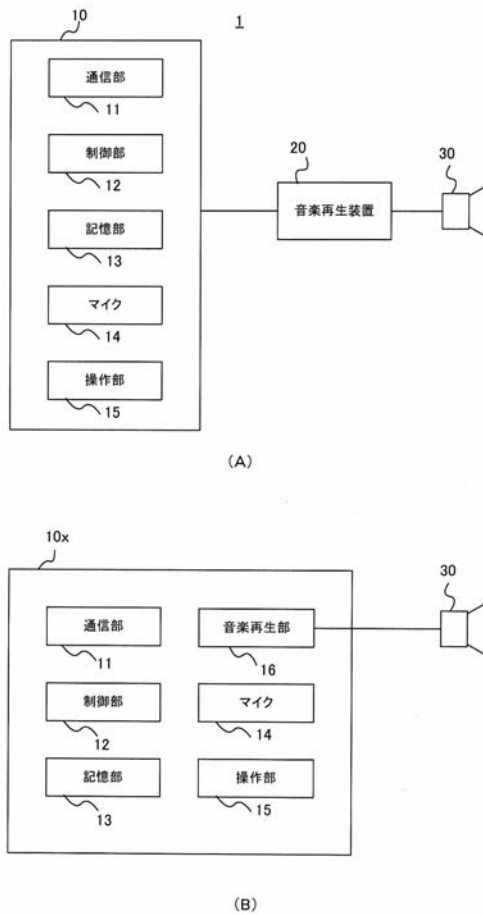
【図 9】



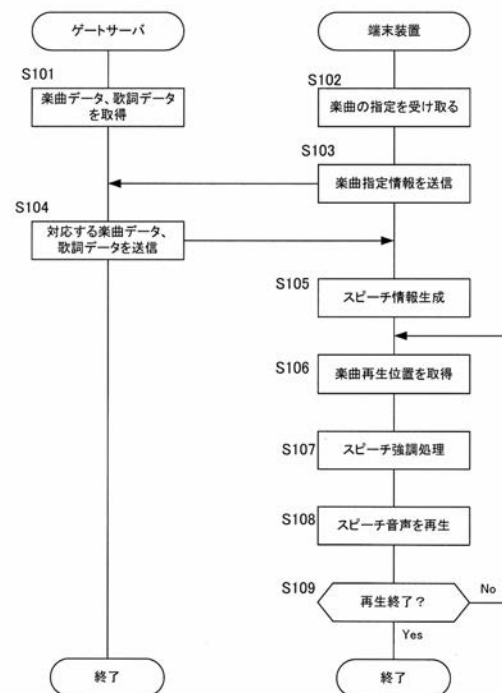
【図 10】



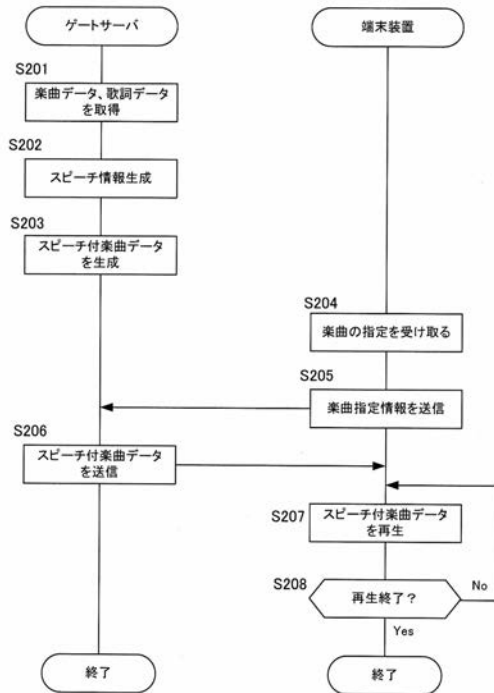
【図 11】



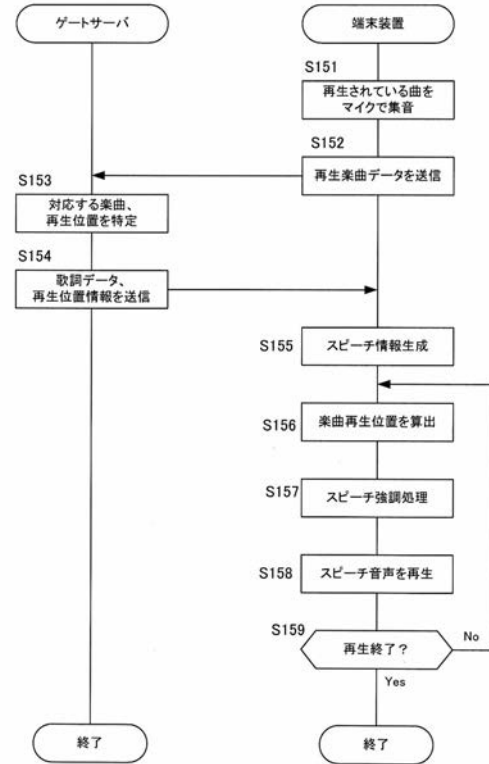
【図 12】



【図 13】



【図 14】



【図 15】

