

(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2014-207568
(P2014-207568A)

(43) 公開日 平成26年10月30日(2014. 10. 30)

(51) Int. Cl.		F I				テーマコード (参考)
HO4S	1/00	(2006.01)	HO4S	1/00	L	5D062
HO4R	3/00	(2006.01)	HO4R	3/00	310	5D220
HO4S	5/02	(2006.01)	HO4S	5/02	D	

審査請求 未請求 請求項の数 7 O L (全 30 頁)

(21) 出願番号 特願2013-84162 (P2013-84162)
(22) 出願日 平成25年4月12日 (2013. 4. 12)

(71) 出願人 00005223
富士通株式会社
神奈川県川崎市中原区上小田中4丁目1番1号
(74) 代理人 100107766
弁理士 伊東 忠重
(74) 代理人 100070150
弁理士 伊東 忠彦
(74) 代理人 100146776
弁理士 山口 昭則
(72) 発明者 ▲角▼岡 幹篤
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】 情報処理装置、音声処理方法、及び音声処理プログラム

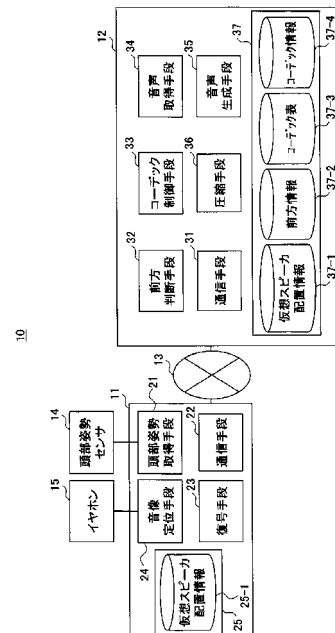
(57) 【要約】

【課題】 適切な音声出力を実現する。

【解決手段】 情報処理装置において、ユーザの姿勢情報から前記ユーザの前方を判断する前方判断手段と、予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成する音声生成手段と、前記音声生成手段により生成された前記音声データに対し、前記前方判断手段により得られる前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとで異なる圧縮を行う圧縮手段と、前記圧縮手段により圧縮された前記音声データを送信する通信手段とを有する。

【選択図】 図1

第1実施形態における音声処理システムの構成例を示す図



【特許請求の範囲】**【請求項 1】**

ユーザの姿勢情報から前記ユーザの前方を判断する前方判断手段と、
予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成する音声生成手段と、

前記音声生成手段により生成された前記音声データに対し、前記前方判断手段により得られる前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとで異なる圧縮を行う圧縮手段と、

前記圧縮手段により圧縮された前記音声データを送信する通信手段とを有することを特徴とする情報処理装置。

10

【請求項 2】

前記圧縮手段は、

前記ユーザの前方に対応する音声データに対して、高周波数成分が復元可能な圧縮を行い、前記ユーザの前方以外の方向に対応する音声データに対して低周波数成分が復元可能な圧縮を行うことを特徴とする請求項 1 に記載の情報処理装置。

【請求項 3】

前記通信手段は、

前記圧縮手段により得られる前記ユーザの前方に対応する音声データと、前記前方以外の方向に対応する音声データとを、それぞれ異なる通信路を用いて送信することを特徴とする請求項 1 又は 2 に記載の情報処理装置。

20

【請求項 4】

前記前方判断手段により得られる前方情報に対応させて、前記音声生成手段により得られる前記音声データを振り分ける振り分け手段を有し、

前記圧縮手段は、前記振り分け手段により振り分けられた音声データ毎に、前記異なる圧縮を行うことを特徴とする 1 乃至 3 の何れか 1 項に記載の情報処理装置。

【請求項 5】

前記圧縮手段は、

前記音声生成手段により生成された全ての仮想音源に対応する音声データを低周波数成分と高周波成分とに分離して圧縮し、

前記圧縮手段により得られる前記高周波成分の音声データから、前記前方判断手段により得られる前記ユーザの前方に対応する前記高周波成分の音声データを抽出する抽出手段を有し、

30

前記通信手段は、前記圧縮手段により圧縮された前記低周波成分の音声データの全てと、前記抽出手段により抽出された前記ユーザの前方に対応する前記高周波成分の音声データとを送信することを特徴とする請求項 1 乃至 4 の何れか 1 項に記載の情報処理装置。

【請求項 6】

情報処理装置が、

ユーザの姿勢情報から前記ユーザの前方を判断し、

予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成し、

40

生成された前記音声データに対し、前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとで異なる圧縮を行い、

前記異なる圧縮により圧縮された前記音声データを送信することを特徴とする音声処理方法。

【請求項 7】

ユーザの姿勢情報から前記ユーザの前方を判断し、

予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成し、

生成された前記音声データに対し、前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとで異なる圧縮を行い、

50

前記異なる圧縮により圧縮された前記音声データを送信する、処理をコンピュータに実行させるための音声処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、音声処理方法、及び音声処理プログラムに関する。

【背景技術】

【0002】

ある地点を基準にした周囲の音声環境を、限られた数の仮想スピーカ（仮想音源）で集約し、別の地点で再現する音声 Augmented Reality（AR、拡張現実）技術が検討されている。音声AR技術では、周囲の多数の方向（例えば、8方向）からの音を他の空間上で再現するため、それぞれの方向で捉えた多数の音声ストリームを再生装置側に伝送する通信帯域が必要になる。

10

【0003】

例えば、サーバからユーザ端末にコンテンツを配信する場合に、ユーザの注目が向けられている部分にはネットワークで大きな通信帯域を割り当て、注目が向けられていない部分には小さな通信帯域を割り当てる手法がある（例えば、特許文献1参照）。

【先行技術文献】

【特許文献】

【0004】

20

【特許文献1】特開2011-172250号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

上述したように、多数の音を伝送するには、多くの通信帯域が必要になる。そのため、例えば Wireless Local Area Network（WLAN）やキャリア網等の帯域が制限される環境では、音声AR技術を利用することが難しい。

【0006】

なお、通信するデータ量を削減するために、伝送前の音声に対して、可逆圧縮や不可逆圧縮等を行うことが考えられ、圧縮率等を考慮すると高圧縮が可能な不可逆圧縮が好ましい。しかしながら、不可逆圧縮は、音質が劣化し、例えば音源の上下方向を判定するキーとなる高周波成分が脱落することにより、ユーザ（聴取者）の前方の音像定位感が悪化する。そのため、ユーザに対する前方の音が仮想音源として割り当てた位置より上方に聞こえてしまう等の現象が生じ、前方の音像定位感が適切に定位されない。

30

【0007】

1つの側面では、本発明は、適切な音声出力を実現することを目的とする。

【課題を解決するための手段】

【0008】

一態様における情報処理装置は、ユーザの姿勢情報から前記ユーザの前方を判断する前方判断手段と、予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成する音声生成手段と、前記音声生成手段により生成された前記音声データに対し、前記前方判断手段により得られる前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとで異なる圧縮を行う圧縮手段と、前記圧縮手段により圧縮された前記音声データを送信する通信手段とを有する。

40

【発明の効果】

【0009】

適切な音声出力を実現することができる。

【図面の簡単な説明】

【0010】

【図1】第1実施形態における音声処理システムの構成例を示す図である。

50

- 【図 2】再生装置のハードウェア構成例を示す図である。
- 【図 3】提供サーバのハードウェア構成例を示す図である。
- 【図 4】音声処理システムの処理の一例を示すシーケンス図である。
- 【図 5】音声処理システムで用いられる各種データ例を説明するための図である、
- 【図 6】仮想スピーカの配置例を説明するための図である。
- 【図 7】第 2 実施形態における音声処理システムの構成例を示す図である。
- 【図 8】第 2 実施形態における音声処理システムの動作を説明するための図である。
- 【図 9】第 2 実施形態における圧縮手段の処理の一例を示すフローチャートである。
- 【図 10】第 2 実施形態における提供サーバの通信手段の処理の一例を示すフローチャートである。
- 【図 11】第 2 実施形態における再生装置の通信手段の処理の一例を示すフローチャートである。
- 【図 12】第 3 実施形態における音声処理システムの構成例を示す図である。
- 【図 13】第 3 実施形態における音声処理システムの動作を説明するための図である。
- 【図 14】第 3 実施形態における圧縮手段及び抽出手段の処理の一例を示すフローチャートである。
- 【図 15】第 3 実施形態における提供サーバの通信手段の処理の一例を示すフローチャートである。
- 【図 16】第 3 実施形態における再生装置の通信手段の処理の一例を示すフローチャートである。
- 【図 17】第 3 実施形態における再生装置の復号手段の処理の一例を示すフローチャートである。

10

20

【発明を実施するための形態】

【0011】

以下、添付図面を参照しながら実施例について詳細に説明する。

【0012】

< 第 1 実施形態における音声処理システムの概略構成例 >

図 1 は、第 1 実施形態における音声処理システムの構成例を示す図である。第 1 実施形態では、サンプリングレート（サンプリング周波数）を変えて音声通信を行う例を示している。例えば、第 1 実施形態では、データ圧縮機能としてダウンサンプリング（サンプリング周波数を下げる変換）を用いる。

30

【0013】

図 1 に示す音声処理システム 10 は、通信端末の一例としての再生装置 11 と、情報処理装置の一例としての提供サーバ 12 とを有する。再生装置 11 と、提供サーバ 12 とは、例えばインターネットや W L A N、L A N 等に代表される通信ネットワーク 13 により、データの送受信が可能な状態で接続されている。

【0014】

再生装置 11 は、提供サーバ 12 から送信された音声データを受信し、受信した音声データを再生する。音声データとは、例えば音声 A R 用の音データ、音楽データであるが、これに限定されるものではなく、その他の音響データでもよい。

40

【0015】

再生装置 11 は、ユーザの頭部の姿勢を検出する姿勢検出手段の一例としての頭部姿勢センサ 14、及び音声を出力する音声出力手段の一例としてのイヤホン 15 と接続されている。再生装置 11 は、例えば頭部姿勢センサ 14 からリアルタイムにユーザの正面方向等の姿勢情報を取得し、取得した姿勢情報を、通信ネットワーク 13 を介して提供サーバ 12 に送信する。更に、再生装置 11 は、提供サーバ 12 により姿勢情報に基づいて生成された音声 A R を実現する複数の仮想スピーカ（仮想音源）に対応する複数チャンネル（複数 c h）の音声データを受信し、受信した各音声データを復号する。再生装置 11 は、復号した各音声データを右耳用、左耳用に集約してイヤホン 15 から音の出力を行う。

【0016】

50

提供サーバ12は、通信ネットワーク13を介して再生装置11から得られるユーザの姿勢情報等に基づいて、ユーザの前方の向きを判断する。更に、提供サーバ12は、判断されたユーザの前方に配置される仮想スピーカに対応する音声データには高周波成分の情報を有する音声データを再生装置11に送信する。また、提供サーバ12は、ユーザの後方（前方以外）に相当する音声データには高周波成分の情報を削減した高圧縮（低周波成分）の音声データを再生装置11に送信する。

【0017】

ここで、ユーザの前方とは、ユーザの頭部を軸として回転させた360°の範囲において、ユーザの頭部の両耳を結んだ直線を基準としたときの前側180°の範囲とすることができるが、これに限定されるものではない。例えば、ユーザの前方とは、ユーザの正面方向を基準として左右に所定の角度（±45°）を基準とした範囲としてもよい。また、ユーザの後方とは、上述した前方以外の範囲であるが、これに限定されるものではない。例えば、ユーザの周囲360°のうち、ユーザの視界の範囲を前方とし、視界の範囲外を後方としてもよい。

10

【0018】

高周波成分とは、例えば約11～12kHz以上の周波数成分である。また、低周波成分とは、高周波成分よりも低い、例えば約11～12kHz未満の周波数成分であるが、各成分については、これに限定されるものではない。

【0019】

頭部姿勢センサ14は、例えばリアルタイム、所定時間間隔毎、又は頭部の移動を検知する毎に、ユーザの頭部の姿勢を取得する。頭部姿勢センサ14は、例えば加速度センサや方位センサ等をユーザの頭部に取り付けることで頭部姿勢（方位）を取得してもよく、例えばカメラ等の撮像手段により撮影した映像に映っている被写体（例えば、構造物等）等からユーザの頭部姿勢を取得してもよいが、これに限定されるものではない。

20

【0020】

イヤホン15は、ユーザ（聴取者）の耳等に装着することで、左右の耳からユーザに仮想スピーカによる音声ARの音を出力する。なお、音声出力手段としては、イヤホン15に限定されるものではなく、例えばヘッドホンやサラウンドスピーカ等を用いることができるが、これに限定されるものではない。姿勢検出手段と、音声出力手段とは、例えば、イヤホン15やヘッドホンとして一体に形成されていてもよい。

30

【0021】

音声処理システム10において、再生装置11及び提供サーバ12の数は、図1の例に限定されるものではなく、例えば1つの提供サーバ12に対して複数の再生装置11が通信ネットワーク13を介して接続されていてもよい。また、提供サーバ12は、1以上の情報処理装置を有するクラウドコンピューティングにより構成されてもよい。

【0022】

上述したように、第1実施形態では、例えば人間の特性と、圧縮の特性とを鑑みて、音像定位の維持とデータ圧縮を両立することで、適切な音声出力を実現する。なお、人間の特性とは、例えば音像の定位感には方向毎に異なった周波数特性があり、前方の定位感には高周波数成分が必要であること等をいう。また、圧縮の特性とは、例えば音声圧縮では高周波成分の情報量の削減が音質を維持しつつ圧縮率を高めるのに効果的であること等をいうが、これらの特性については、これに限定されるものではない。

40

【0023】

次に、上述した音声処理システム10における再生装置11及び提供サーバ12の機能構成例について説明する。

【0024】

<再生装置11の機能構成例>

図1に示す再生装置11は、頭部姿勢取得手段21と、通信手段22と、復号手段23と、音像定位手段24と、記憶手段25とを有する。記憶手段25は、仮想スピーカ配置情報25-1を有している。

50

【 0 0 2 5 】

頭部姿勢取得手段 2 1 は、頭部姿勢センサ 1 4 からユーザの頭部の姿勢情報（方位）を取得する。頭部姿勢センサ 1 4 の出力値は、例えばある方向（例えば「北」）を基準（ $= 0^\circ$ ）として、左右何れかの方向に回転させたときの角度に対応させることができる。例えば、北を基準として右回りに回転させた角度の場合、ユーザが「東」を向いているときの頭部姿勢センサ 1 4 の出力値は、 90° となる。

【 0 0 2 6 】

頭部姿勢取得手段 2 1 は、例えば頭部姿勢センサ 1 4 から約 1 0 0 m s 毎等の周期的なタイミングで姿勢情報を取得してもよく、またユーザからの取得要求があった場合や頭部の変位量が所定数以上の場合に姿勢情報を取得してもよい。

10

【 0 0 2 7 】

通信手段 2 2 は、頭部姿勢取得手段 2 1 から得られた姿勢情報を、通信ネットワーク 1 3 を介して提供サーバ 1 2 に送信する。通信手段 2 2 は、通信ネットワーク 1 3 を介して提供サーバ 1 2 から音声 A R を実現する複数の仮想スピーカに対応して所定の形式で圧縮（符号化）された各音声データ（例えば、圧縮デジタル音声（8 c h ステレオ）等）を受信する。

【 0 0 2 8 】

通信手段 2 2 は、提供サーバ 1 2 から、音声データの他にも例えば各種パラメータ等を受信してもよい。例えば、通信手段 2 2 は、提供サーバ 1 2 から音声データ、音声データを識別するシーケンス番号、音声データに対するコーデック情報等をパケットから読み取る。コーデック情報とは、例えば音声 A R を実現する複数の仮想スピーカに対応する各音声データに対する圧縮の有無、又はどのような形式（例えば、符号化方式等）で圧縮したかを示す情報等であるが、これに限定されるものではない。

20

【 0 0 2 9 】

復号手段 2 3 は、通信手段 2 2 で受信したデータに対して、コーデック（符号化方式）に対応するデコーデック（復号化方式）や各種パラメータ等を用いて復号する。例えば、復号手段 2 3 は、予め設定された複数の仮想スピーカ（仮想音源）# 1 ~ # 8 のそれぞれについて、コーデック情報から仮想スピーカの識別情報（例えば、I D 等）に合ったコーデックとパラメータを取得し、取得した内容に合わせて音声データを復号する。復号手段 2 3 により、低圧縮又は無圧縮の音声データに対しては、高周波成分を有する音声データが復元され、高圧縮の音声データに対しては、低周波成分（高周波成分を含まない）の音声データが復元される。

30

【 0 0 3 0 】

音像定位手段 2 4 は、頭部姿勢取得手段 2 1 から取得したユーザの姿勢情報と、予め記憶手段 2 5 に記憶された仮想スピーカ配置情報 2 5 - 1 とに基づいて、復号手段 2 3 から得られる各音声データを集約して音声 A R 再生用の音像定位を行う。更に、音像定位手段 2 4 は、音像が定位された音声データをアナログ音声（例えば、2 c h ステレオ）等により、イヤホン 1 5 に出力する。

【 0 0 3 1 】

ここで、音像定位手段 2 4 は、例えば Head Related Transfer Function（H R T F、頭部伝達関数）等を用いて、任意の方角に対応する H R T F を音声データ（音源信号）に畳み込む処理を行う。これにより、あたかも音声は任意の方角から聞こえたような効果を得ることができる。

40

【 0 0 3 2 】

音像定位手段 2 4 は、複数の仮想スピーカのそれぞれに対し、ユーザの前方に対する方向に応じて伝達関数を畳み込むことにより、イヤホン 1 5 に出力可能な左右の音（例えば、2 c h ステレオ）を生成する。この場合、音像定位手段 2 4 は、例えばユーザの前方に対応する予め設定された仮想スピーカに対応する音声データに高周波成分を出力させるが、これに限定されるものではない。

【 0 0 3 3 】

50

記憶手段 25 の仮想スピーカ配置情報 25 - 1 は、音声 AR を実現するために予め設定された多方向に配置される仮想スピーカの配置情報である。この仮想スピーカ配置情報 25 - 1 は、例えば提供サーバ 12 でも管理されており、再生装置 11 と提供サーバ 12 とでデータの同期が取られている。

【0034】

また、記憶手段 25 は、再生装置 11 が第 1 実施形態における各処理を実行するための各種情報（例えば、設定情報等）を記憶するが、記憶される情報としては、これに限定されるものではない。例えば、記憶手段 25 には、頭部姿勢センサ 14 により取得した頭部姿勢情報、提供サーバ 12 より得られる音声データ、コーデック情報を記憶することができる。

10

【0035】

上述した再生装置 11 における各処理は、例えば再生装置 11 にインストールされた専用のアプリケーション（プログラム）を実行することにより実現することができる。

【0036】

< 提供サーバ 12 の機能構成例 >

図 1 に示す提供サーバ 12 は、通信手段 31 と、前方判断手段 32 と、コーデック制御手段 33 と、音声取得手段 34 と、音声生成手段 35 と、圧縮手段 36 と、記憶手段 37 とを有する。記憶手段 37 は、仮想スピーカ配置情報 37 - 1 と、前方情報 37 - 2 と、コーデック表 37 - 3 と、コーデック情報 37 - 4 とを有する。

【0037】

20

通信手段 31 は、再生装置 11 から通信ネットワーク 13 を介してユーザ（聴取者）の頭部の姿勢情報を受信する。また、通信手段 31 は、圧縮手段 36 等により所定の符号化方式に圧縮された仮想スピーカに対応した各音声データ（例えば、圧縮デジタル音声（8ch ステレオ）等）を再生装置 11 に送信する。

【0038】

通信手段 31 が再生装置 11 に送信する情報としては、例えばシーケンス番号、コーデック情報、音声データ（バイナリ列）等であるが、これに限定されるものではなく、またそれぞれの情報の組を送信してもよい。例えば、通信手段 31 は「シーケンス番号、コーデック情報、音声データ（バイナリ列）」 = 「1, { (#1, 圧縮なし, 44kHz ···), ···, (#8, サンプリング, 22kHz ···) }, { (3R1T0005 ···), ···, (4F1191 ···) }」等の情報を送信する。

30

【0039】

前方判断手段 32 は、通信手段 31 が受信した姿勢情報からユーザの前方の方向を判断する。前方判断手段 32 は、ユーザの姿勢情報と仮想スピーカ配置情報 37 - 1 とを比較し、ユーザの前方（正面方向）に最も近い仮想スピーカを所定数（例えば、2つ）選択する。前方判断手段 32 は、選択した前方の仮想スピーカを識別するための識別情報（仮想スピーカ ID）等をコーデック制御手段 33 に出力したり、前方情報 37 - 2 として記憶手段 37 に記憶する。

【0040】

コーデック制御手段 33 は、記憶手段 37 に記憶された前方情報 37 - 2 及びコーデック表 37 - 3 等を参照し、全ての仮想スピーカ（例えば #1 ~ #8 の 8 チャンネル）に対するコーデック（符号化情報等）とパラメータ（符号化パラメータ等）とを取得する。例えば、コーデック制御手段 33 は、前方の仮想スピーカと、それ以外の仮想スピーカにそれぞれ対応する音声データに対し、コーデックやパラメータ等を用いた符号化等による圧縮手法（符号化手法）を、圧縮手段 36 に出力する。

40

【0041】

例えば、コーデック制御手段 33 は、処理対象の仮想スピーカがユーザの前方であるか否かを判断し、前方である場合にはコーデック表 37 - 3 から、前方用のコーデックとパラメータとを取得し、圧縮手段 36 に出力する。また、コーデック制御手段 33 は、処理対象の仮想スピーカが前方でない場合にはコーデック表 37 - 3 から前方以外の他スピー

50

力用のコーデックとパラメータとを取得し、圧縮手段 3 6 に出力する。

【 0 0 4 2 】

コーデック制御手段 3 3 は、ユーザの正面方向の変化に対して音声が届かないようなタイミングで仮想スピーカ # 1 ~ # 8 に対する圧縮手法を切り替える。また、コーデック制御手段 3 3 は、各仮想スピーカ（各方位）のコーデック（符号化情報）とパラメータとを記憶手段 3 7 のコーデック情報 3 7 - 4 に記憶することもできる。

【 0 0 4 3 】

音声取得手段 3 4 は、再生装置 1 1 側で音声 A R を実現するための音声データを取得する。例えば、音声取得手段 3 4 は、実際の空間上に多方向に配置した複数のマイクロホン（以下、「マイク」と略称する）から同時に音を取得してもよい。また、音声取得手段 3 4 は、例えばアプリケーションを用いて、仮想空間で出力された音声をその空間上の所定の位置に配置された複数の仮想マイクから得られる音声データを取得してもよい。

【 0 0 4 4 】

音声生成手段 3 5 は、音声取得手段 3 4 で取得された各方向からの音声データに対応させて、予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成する。例えば、音声生成手段 3 5 は、音声取得手段 3 4 で取得された各方向からの音声データに対応させた仮想スピーカ（仮想音源）の配置位置から音声データを出力させるための音声データを生成する。

【 0 0 4 5 】

圧縮手段 3 6 は、音声生成手段 3 5 から得られる仮想スピーカ毎の音声データに対して、コーデック制御手段 3 3 で制御されたコーデック及びパラメータの組み合わせに基づいて圧縮（この場合は、リサンプリング）する。例えば、圧縮手段 3 6 は、前方判断手段 3 2 により得られるユーザの前方に対応する音声データと、ユーザの前方以外の音声データとで、異なる圧縮を行う。

【 0 0 4 6 】

例えば、圧縮手段 3 6 は、音声生成手段 3 5 から複数の仮想スピーカ（例えば、# 1 ~ # 8）に対応する音声データを取得すると、各音声データについて、コーデック情報 3 7 - 4 から仮想スピーカの I D に合ったコーデックとパラメータとを参照する。圧縮手段 3 6 は、参照したパラメータ等に基づいて各音声データを圧縮する。

【 0 0 4 7 】

例えば、圧縮手段 3 6 は、ユーザの前方に対応する音声データに対して、再生装置 1 1 側で高周波数成分が復元可能な圧縮（低圧縮）を行い、前方以外の音声データに対して、再生装置 1 1 側で低周波数成分のみが復元可能な圧縮（高圧縮）を行う。なお、圧縮手段 3 6 は、ユーザの前方に対応する仮想スピーカの音声データに対し、高周波成分を残すために圧縮を行わなくてもよい（無圧縮）。

【 0 0 4 8 】

圧縮手段 3 6 は、例えば元の音声データに対する圧縮手法として P u l s e C o d e M o d u l a t i o n (P C M) 等を用いることができる。また、圧縮手段 3 6 は、可逆圧縮として F r e e L o s s l e s s A u d i o C o d e c (F L A C) 等を用いることができる。また、圧縮手段 3 6 は、例えば不可逆（音声用）として G . 7 1 1、G . 7 2 2 . 1、G . 7 1 9 等を用いたり、不可逆（音楽用）として M P 3、A d v a n c e d A u d i o C o d i n g (A A C) 等を用いることができる。圧縮手段 3 6 は、コーデック制御手段 3 3 による制御により上述した圧縮手法のうち、少なくとも 1 つを用いて圧縮を行うが、圧縮手法はこれらに限定されるものではない。

【 0 0 4 9 】

通信手段 3 1 は、圧縮手段 3 6 により圧縮された仮想スピーカの音声データと、コーデック情報 3 7 - 4 等とを関連付けて、再生装置 1 1 に送信する。例えば、通信手段 3 1 は、圧縮手段 3 6 から所定の符号化方式により圧縮された、又は無圧縮の音声データを取得し、シーケンス番号やコーデック情報等をパケットに含めて、音声データの各チャンネル（c h）に対し、コーデックに合わせた音声データ領域を設定する。通信手段 3 1 は、設

10

20

30

40

50

定した各領域を用いて各チャンネルの音声データを、通信ネットワーク 13 を介して再生装置 11 に送信する。

【0050】

記憶手段 37 は、上述した仮想スピーカ配置情報 37 - 1、前方情報 37 - 2、コーデック表 37 - 3、及びコーデック情報 37 - 4 等のうち、少なくとも 1 つの情報を記憶する。記憶手段 37 は、提供サーバ 12 が第 1 実施形態における各処理を実行するための各種情報（例えば、設定情報等）を記憶するが、記憶される情報としては、これに限定されるものではない。例えば、記憶手段 37 は、再生装置 11 を使用するユーザの識別情報や、再生装置 11 から得られる姿勢情報等を記憶してもよい。

【0051】

第 1 実施形態では、上述した提供サーバ 12 の処理により、定位感を維持したままの音声データを圧縮して通信することができる。上述した提供サーバ 12 における各処理は、例えば提供サーバ 12 にインストールされた専用のアプリケーション（プログラム）を実行することにより実現することができる。

【0052】

上述した再生装置 11 は、例えば Personal Computer（PC）であるが、これに限定されるものではなく、例えばタブレット端末、スマートフォン等の通信端末でもよく、音楽再生装置、ゲーム機器等でもよい。また、提供サーバ 12 は、例えば PC やサーバ等であるが、これに限定されるものではない。

【0053】

<再生装置 11 のハードウェア構成例>

図 2 は、再生装置のハードウェア構成の一例を示す図である。図 2 に示す再生装置 11 は、入力装置 41 と、出力装置 42 と、通信インタフェース 43 と、オーディオインタフェース 44 と、主記憶装置 45 と、補助記憶装置 46 と、Central Processing Unit（CPU）47 と、ネットワーク接続装置 48 とを有し、これらはシステムバス B で相互に接続されている。

【0054】

入力装置 41 は、再生装置 11 のユーザからのプログラムの実行指示、各種操作情報、ソフトウェア等を起動するための情報等の入力を受け付ける。入力装置 41 は、例えばタッチパネルや所定の操作キー等である。入力装置 41 に対する操作に応じた信号が CPU 47 に送信される。

【0055】

出力手段 42 は、本実施形態における再生装置 11 を操作するのに必要な各種ウィンドウやデータ等を表示するディスプレイを有し、CPU 47 が有する制御プログラムによりプログラムの実行経過や結果等を表示することができる。

【0056】

通信インタフェース 43 は、上述した頭部姿勢センサ 14 によるユーザの頭部の姿勢情報を取得する。オーディオインタフェース 44 は、CPU 47 から送信されたデジタル音声をアナログ音声に変換したり、変換したアナログ音声を増幅して、上述したイヤホン 15 等に出力する。

【0057】

主記憶装置 45 は、CPU 47 に実行させる Operating System（OS）プログラムやアプリケーションプログラムの少なくとも一部を一時的に記憶する。また、主記憶装置 45 は、CPU 47 による処理に必要な各種データを記憶する。主記憶装置 45 は、例えば Read Only Memory（ROM）や Random Access Memory（RAM）等である。

【0058】

補助記憶装置 46 は、内蔵した磁気ディスクに対して、磁気的にデータの書き込み及び読み出し等を行う。補助記憶装置 46 は、OS プログラム、アプリケーションプログラム、及び各種データをお記憶する。補助記憶装置 46 は、例えばフラッシュメモリや、Ha

10

20

30

40

50

rd Disk Drive (HDD)、Solid State Drive (SSD)等のストレージ手段等である。主記憶装置45及び補助記憶装置46は、例えば上述した記憶手段25に対応している。

【0059】

CPU47は、OS等の制御プログラム、及び主記憶装置45に格納されている実行プログラムに基づいて、各種演算や各ハードウェア構成部とのデータの入出力等、再生装置11等のコンピュータ全体の処理を制御して各処理を実現することができる。プログラムの実行中に必要な各種情報等は、例えば補助記憶装置46から取得することができ、また実行結果等を格納することもできる。

【0060】

例えば、CPU47は、例えば入力装置41から得られるプログラムの実行指示等に基づき、補助記憶装置46にインストールされたプログラム(例えば、音声処理プログラム)を実行させることにより、主記憶装置45上でプログラムに対応する処理を行う。

【0061】

例えば、CPU47は、音声処理プログラムを実行させることで、上述した頭部姿勢取得手段21による頭部姿勢の取得、通信手段22における各種データの送受信、復号手段23による復号、音像定位手段24による音像定位等の処理を行う。なお、CPU47における処理内容は、これに限定されるものではない。CPU47により実行された内容は、必要に応じて補助記憶装置46に記憶される。

【0062】

ネットワーク接続装置48は、CPU47からの制御信号に基づき、通信ネットワーク13等と接続することにより、実行プログラムやソフトウェア、設定情報等を、通信ネットワーク13に接続されている外部装置(例えば、提供サーバ12等)等から取得する。ネットワーク接続装置48は、プログラムを実行することで得られた実行結果又は本実施形態における実行プログラム自体を外部装置等に提供することができる。また、ネットワーク接続装置48は、例えばWi-Fi(登録商標)やBluetooth(登録商標)等による通信を可能にする通信手段を有していてもよい。また、ネットワーク接続装置48は、電話端末との通話を可能にする通話手段を有していてもよい。

【0063】

上述したようなハードウェア構成により、本実施形態における音声処理を実行することができる。本実施形態は、各機能をコンピュータに実行させることができる実行プログラム(音声処理プログラム)を例えば通信端末等にインストールすることで、本実施形態における音声処理を容易に実現することができる。

【0064】

更に、ネットワーク接続装置47は、例えばWi-Fi(登録商標)やBluetooth(登録商標)等による通信を可能にする通信手段を有していてもよい。また、ネットワーク接続装置47は、電話端末との通話を可能にする通話手段を有していてもよい。

【0065】

<提供サーバ12のハードウェア構成例>

図3に示す提供サーバ12は、入力装置51と、出力装置52と、ドライブ装置53と、主記憶装置54と、補助記憶装置55と、CPU56と、ネットワーク接続装置57とを有し、これらはシステムバスBで相互に接続されている。

【0066】

入力装置51は、提供サーバ12の管理者等のユーザからのプログラムの実行指示、各種操作情報、ソフトウェア等を起動するための情報等の入力を受け付ける。入力装置51は、提供サーバ12のユーザ等が操作するキーボード及びマウス等のポインティングデバイスや、マイク等の音声入力デバイスを有する。

【0067】

出力装置52は、本実施形態における提供サーバ12を操作するのに必要な各種ウィンドウやデータ等を表示するディスプレイを有し、CPU56が有する制御プログラムによ

10

20

30

40

50

りプログラムの実行経過や結果等を表示することができる。

【0068】

ここで、提供サーバ12等のコンピュータ本体にインストールされる実行プログラムは、例えばUniversal Serial Bus (USB)メモリやCD-ROM、DVD等の可搬型の記録媒体58等により提供される。プログラムを記録した記録媒体58は、ドライブ装置53にセット可能であり、CPU56からの制御信号に基づき、記録媒体58に含まれる実行プログラムが、記録媒体58からドライブ装置53を介して補助記憶装置55にインストールされる。

【0069】

主記憶装置54は、CPU56に実行させるOSプログラムやアプリケーションプログラムの少なくとも一部を一時的に記憶する。また、主記憶装置54は、CPU56による処理に必要な各種データを記憶する。主記憶装置54は、ROMやRAM等である。

【0070】

補助記憶装置55は、CPU56からの制御信号に基づき、本実施形態における実行プログラムや、コンピュータに設けられた制御プログラム等を記憶し、必要に応じて入出力を行う。補助記憶装置55は、CPU56からの制御信号等に基づいて、記憶された各情報から必要な情報を読み出したり、書き込んだりすることができる。補助記憶装置55は、例えばHDDやSSD等のストレージ手段等である。主記憶装置54及び補助記憶装置55は、例えば上述した記憶手段37に対応している。

【0071】

CPU56は、OS等の制御プログラム、及び主記憶装置54に格納されている実行プログラムに基づいて、各種演算や各ハードウェア構成部とのデータの入出力等、提供サーバ12等のコンピュータ全体の処理を制御して各処理を実現することができる。プログラムの実行中に必要な各種情報等は、例えば補助記憶装置55から取得することができる。また実行結果等を格納することもできる。

【0072】

例えば、CPU56は、例えば入力装置51から得られるプログラムの実行指示等に基づき、補助記憶装置55にインストールされたプログラム(例えば、音声処理プログラム)を実行させることにより、主記憶装置54上でプログラムに対応する処理を行う。

【0073】

例えば、CPU56は、音声処理プログラムを実行させることで、上述した前方判断手段32による前方判断、コーデック制御手段33によるコーデック制御、音声取得手段34による音声データの取得等の処理を行う。更に、CPU56は、音声生成手段35による仮想スピーカ音声生成、圧縮手段36による圧縮等の処理を行う。なお、CPU56における処理内容は、これに限定されるものではない。CPU56により実行された内容は、必要に応じて補助記憶装置55に記憶される。

【0074】

ネットワーク接続装置57は、CPU56からの制御信号に基づき、通信ネットワーク13等と接続することにより、実行プログラムやソフトウェア、設定情報等を、通信ネットワーク13に接続されている外部装置等から取得する。また、ネットワーク接続装置57は、プログラムを実行することで得られた実行結果又は本実施形態における実行プログラム自体を外部装置等に提供することができる。

【0075】

上述したようなハードウェア構成により、本実施形態における音声処理を実行することができる。本実施形態は、各機能をコンピュータに実行させることができる実行プログラム(音声処理プログラム)を例えば汎用のPC等にインストールすることで、本実施形態における音声処理を容易に実現することができる。

【0076】

<音声処理システム10における処理の一例>

次に、上述した音声処理システム10における処理(音声通信処理)の一例についてシ

10

20

30

40

50

ーケンス図を用いて説明する。図4は、音声処理システムの処理の一例を示すシーケンス図である。図4の例では、上述した再生装置11と提供サーバ12とを有している。

【0077】

図4の例において、再生装置11の頭部姿勢取得手段21は、頭部姿勢センサ14等からユーザの頭部姿勢情報を取得する(S01)。再生装置11の通信手段22は、S01の処理により取得した頭部姿勢情報を提供サーバ12に送信する(S02)。

【0078】

提供サーバ12の前方判断手段32は、S02の処理により取得した再生装置11からの頭部姿勢情報や予め記憶手段37に記憶されている仮想スピーカ配置情報37-1に基づいて、ユーザの前方判断を行い、前方に対応する仮想スピーカを選択する(S03)。

10

【0079】

次に、提供サーバ12のコーデック制御手段33は、前方判断結果に基づいて各仮想スピーカに対応する音声データの圧縮時のコーデック制御を行う(S04)。次に、提供サーバ12の音声取得手段34は、再生装置11で実現される音声ARに対応する複数の仮想スピーカから出力させる元となる音声データを取得する(S05)。次に、提供サーバ12の音声生成手段35は、S05の処理により取得した音声データから仮想スピーカ用の音声データを生成する(S06)。

【0080】

次に、提供サーバ12の圧縮手段36は、記憶手段37に記憶されているコーデック表37-3に基づいて、各仮想スピーカに対応する圧縮手法を用いて各音声データを圧縮(符号化)する(S07)。S07の処理では、例えば上述したS03の処理で得られた前方に対応するチャンネルに対して、例えば高周波成分を有する音声データの圧縮(低圧縮又は無圧縮)を行い、前方以外のチャンネルに対して、例えば高周波成分が復元されない程度の高圧縮を行う。

20

【0081】

また、提供サーバ12の通信手段31は、S07の処理により圧縮された音声データやコーデック情報等をパケットデータ等により通信ネットワーク13を介して再生装置11に送信する(S08)。

【0082】

再生装置11の通信手段22は、S08の処理により提供サーバ12から送信された情報を受信する。再生装置11の復号手段23は、受信した情報からS07の処理で圧縮された音声データを取得し、取得した音声データをコーデック情報に対応させた復号手法で復号する(S09)。なお、S09の処理は、S08の処理において、音声データと共に送信されたチャンネル毎のコーデック情報等を用いることで、適切な復号を実現できる。

30

【0083】

また、再生装置11の音像定位手段24は、S09の処理で復号された各チャンネルの音声データを左右の耳用に集約してイヤホン15から音声ARによる出力ができるように音像の定位処理を行い(S10)、処理された音声データをイヤホン15等に出力する(S11)。

【0084】

なお、上述の処理は、再生装置11から再生される音声終了するまで、又は、ユーザの指示により第1実施形態における音声通信処理が終了されるまで繰り返し行われる。したがって、ユーザの頭部姿勢のリアルタイムな動きに対応させて音像定位された音声データをユーザに提供することができる。

40

【0085】

<各種データ例等>

次に、上述した音声処理システム10における各種データ例等について、図を用いて説明する。図5は、音声処理システムで用いられる各種データ例を説明するための図である。図5(A)は、頭部姿勢情報の一例を示す。図5(B)は、仮想スピーカ配置情報25-1, 37-1の一例を示す。図5(C)は、前方情報37-2の一例を示す。図5(D)

50

)は、コーデック表37-3の一例を示す。図5(E)は、コーデック情報の一例を示す。

【0086】

図5(A)に示す頭部姿勢情報の項目としては、例えば「識別情報」、「時間」、「姿勢情報」等であるが、これに限定されるものではない。図5(A)に示す「識別情報」は、提供サーバ12が再生装置11を識別するための識別情報である。図5(A)に示す「時間」は、頭部姿勢センサ14からユーザの頭部の姿勢情報を取得した時間である。図5(A)に示す「姿勢情報」は、頭部姿勢センサ14により取得したユーザの頭部の姿勢情報が示されている。なお、図5(A)の例では、姿勢情報として、ユーザの前方(真正面)の角度が示されているが、これに限定されるものではない。

10

【0087】

図5(B)に示す仮想スピーカ配置情報25-1, 37-1の項目としては、例えば「仮想スピーカID」、「配置位置x」、「配置位置y」等があるが、これに限定されるものではなく、角度情報であってもよい。図5(B)の例では、8つの仮想スピーカ(ID:#1~#8)に対する配置情報を座標で設定しているが、これに限定されるものではなく、各仮想スピーカに対応する設置角度を設定してもよい。

【0088】

ここで、図6は、仮想スピーカの配置例を説明するための図である。図6の例では、8つの仮想スピーカがユーザ(聴取者)の頭部の位置を中心として、半径1の円形状に45°間隔で配置された例を示している。図5(B)に示す仮想スピーカ配置情報25-1, 37-1では、図6に示す配置例に対応する仮想スピーカのx y座標が記憶されている。

20

【0089】

第1実施形態では、前方判断手段32が、図5(A)に示す頭部姿勢情報と、図5(B)に示す仮想スピーカ配置情報とを比較し、ユーザの前方を基準にして最も近い仮想スピーカを判断し、更に近い順に所定数の仮想スピーカを選択する。

【0090】

例えば、前方判断手段32は、姿勢情報と同一の角度に仮想スピーカが割り当てられている場合には、その仮想スピーカ1つを選択し、姿勢情報と同一の角度に仮想スピーカが割り当てられていない場合には、その角度に近い方から2つの仮想スピーカを選択する。

【0091】

例えば、図6に示す配置例を基準に前方にある仮想スピーカを判断すると、 $\theta = 15^\circ$ の場合、前方判断手段32は、その前方(正面)に仮想スピーカが存在していないと判断し、例えば正面に近い方から2つの仮想スピーカ#1、#2を選択する。また、 $\theta = 90^\circ$ の場合、前方判断手段32は、その前方(正面)に仮想スピーカ#3が存在していると判断し、例えば仮想スピーカ#3を選択する。

30

【0092】

なお、仮想スピーカの選択については、上述した例に限定されるものではない。例えば、前方判断手段32は、姿勢正面に仮想スピーカが割り当てられていない場合には、前方を基準に左右のスピーカを2個ずつ(計4個)を選択してもよい。また、前方判断手段32は、姿勢正面に仮想スピーカが割り当てられている場合には、その仮想スピーカと、その両側にある仮想スピーカ(計3個)を選択してもよい。

40

【0093】

図5(C)に示す前方情報37-2の項目としては、例えば「前方の仮想スピーカ」等があるが、これに限定されるものではなく、例えば「後方の仮想スピーカ」の情報を有していてもよい。また、前方情報37-2として、例えば前方と後方の両方の仮想スピーカの情報を有していてもよいが、この場合には、例えば前方と後方のどちらの仮想スピーカであるかを識別する識別情報を有する。図5(C)の例では、前方判断手段32により判断された前方の仮想スピーカIDとして#1、#2が記憶されている。

【0094】

図5(D)に示すコーデック表37-3の項目としては、例えば「仮想スピーカ種別」

50

、「コーデック」、「パラメータ」等であるが、これに限定されるものではない。コーデック表 37-3 は、コーデック制御手段 33 により制御される情報である。図 5 (D) に示す「仮想スピーカ種別」は、コーデック及びパラメータ等を設定する対象の仮想スピーカを識別する情報である。図 5 (D) の例では、「前方」と「その他」とで識別されているが、これに限定されるものではなく、例えば仮想スピーカ毎に識別してもよい。コーデック表 37-3 を用いることで、仮想スピーカ種別毎にコーデックやパラメータを任意に設定することができる。

【0095】

図 5 (D) に示す「コーデック」は、例えば仮想スピーカ種別毎に設定されるコーデック手法である。「コーデック」において、「圧縮なし」とは無圧縮 (Null Codec) を示し、「サンプリング」とは例えばパラメータ等で設定された条件で圧縮 (ダウンサンプリング) することを意味するが、これに限定されるものではない。

10

【0096】

図 5 (D) に示す「パラメータ」は、「コーデック」で設定された条件で圧縮する時の各種パラメータである。例えば、図 5 (D) の例では、パラメータとして周波数 (例えば、44 kHz 等)、データ量 (例えば、16 bit)、及びフレーム量 (例えば、1024 frame) 等が設定される。なお、パラメータは、これに限定されるものではなく、例えば上述した周波数、データ量、及びフレーム量のうち、少なくとも 1 つでもよく、その他の情報が含まれていてもよい。

【0097】

図 5 (E) に示すコーデック情報の項目としては、例えば「コーデック情報」等であるが、これに限定されるものではない。図 5 (E) に示す「コーデック情報」は、上述した図 5 (D) に示すコーデック表 37-3 に基づいて、仮想スピーカ種別毎に圧縮手段 36 で各音声データを圧縮したときの内容等であるが、これに限定されるものではない。

20

【0098】

図 5 (E) に示すコーデック情報では、例えば ID が #1, #2 の仮想スピーカに対しては、圧縮なしの高周波成分 (44 kHz) の音声データであることを示している。また、図 5 (E) に示すコーデック情報では、例えば ID が #3 ~ #8 の仮想スピーカに対しては、サンプリングレート (周波数) を 22 kHz に圧縮 (ダウンサンプリング) した音声データであることを示している。

30

【0099】

上述したように、第 1 実施形態では、適切な音声出力を実現することができる。また、第 1 実施形態では、提供サーバ 12 から送信される全ての音声データ (チャンネル) において高周波成分を含む場合と比較して通信帯域を削減することができる。また、第 1 実施形態では、再生装置 11 において、前方の音像定位感が適切に定位された音声出力を実現することができる。

【0100】

< 第 2 実施形態における音声処理システムの概略構成例 >

次に、音声処理システムの第 2 実施形態について説明する。図 7 は、第 2 実施形態における音声処理システムの構成例を示す図である。上述した第 1 実施形態では、ダウンサンプリングによる圧縮例を示したが、第 2 実施形態では、音声ストリームの切り替え例を示す。

40

【0101】

なお、図 7 に示す音声処理システム 60 において、上述した音声処理システム 10 と同様の構成等については、同一の符号を付するものとし、ここでの具体的な説明は省略する。また、音声処理システム 60 における再生装置や提供サーバのハードウェア構成も上述した第 1 実施形態におけるハードウェア構成を適用することができるため、ここでの具体的な説明は省略する。

【0102】

図 7 に示す音声処理システム 60 は、再生装置 61 と、提供サーバ 62 とを有する。再

50

生装置 6 1 と、提供サーバ 6 2 とは、例えばインターネットや W L A N、L A N 等に代表される通信ネットワーク 1 3 により、データの送受信が可能な状態で接続されている。第 2 実施形態における通信ネットワーク 1 3 は、コネクション接続により常時接続されているネットワーク形態を示している。

【 0 1 0 3 】

再生装置 6 1 は、頭部姿勢取得手段 2 1 と、通信手段 7 1 と、復号手段 7 2 と、音像定位手段 2 4 と、記憶手段 7 3 とを有する。記憶手段 7 3 は、仮想スピーカ配置情報 2 5 - 1 と、コーデック表 7 3 - 1 とを有する。第 2 実施形態における再生装置 6 1 は、上述した第 1 実施形態における再生装置 1 1 と同一の構成であるが、通信手段 7 1、復号手段 7 2 による処理が異なる。また、記憶手段 7 3 は、再生装置 6 1 が、提供サーバ 6 2 とのセッション開始後に提供サーバ 6 2 から取得されるコーデック表 7 3 - 1 が記憶される。

10

【 0 1 0 4 】

提供サーバ 6 2 は、通信手段 8 1 と、前方判断手段 3 2 と、コーデック制御手段 3 3 と、音声取得手段 3 4 と、音声生成手段 3 5 と、振り分け手段 8 2 と、圧縮手段 8 3 と、記憶手段 3 7 とを有する。第 2 実施形態における提供サーバ 6 2 は、上述した第 1 実施形態における提供サーバ 1 2 と比較すると、振り分け手段 8 2 を有しており、通信手段 8 1、圧縮手段 8 3 の処理も異なる。

【 0 1 0 5 】

第 2 実施形態において、提供サーバ 6 2 の通信手段 8 1 は、圧縮手段 8 2 により得られるユーザの前方に対応する音声データと、前方以外の方向に対応する音声データを、それぞれ異なる通信路を用いて送信する。例えば、通信手段 8 1 は、通信ネットワーク 1 3 を介して再生装置 6 1 と通信する際、予め圧縮率の高い（高圧縮）通信路と、圧縮率の低い（低圧縮）通信路（無圧縮でもよい）とによるコネクションを確立する。

20

【 0 1 0 6 】

更に、通信手段 8 1 は、再生装置 6 1 に対してコーデック表 3 7 - 3 を送信する。第 2 実施形態におけるコーデック表 3 7 - 3 には、どの通信路でどのようなコーデック及びパラメータを用いるかの情報等を有するが、コーデック表 3 7 - 3 の情報としては、これに限定されるものではなく、例えば仮想スピーカ種別等が含まれていてもよい。

【 0 1 0 7 】

提供サーバ 6 2 の振り分け手段 8 2 は、コーデック制御手段 3 3 により生成されたコーデック表 3 7 - 3 に基づいて、音声生成手段 3 5 から得られる各仮想スピーカ（各チャンネル）に対応する音声データを 2 種類の圧縮条件のうちの何れかに振り分ける。圧縮手段 8 3 は、振り分け手段 8 2 により振り分けた各仮想スピーカに対応する圧縮条件で圧縮を行う。

30

【 0 1 0 8 】

例えば、振り分け手段 8 2 は、再生装置 6 1 から得られるユーザの姿勢情報からユーザの前方にある所定数の仮想スピーカに対しては、低圧縮の圧縮条件とし、前方以外の仮想スピーカに対しては、高圧縮の圧縮条件となるように振り分けを行う。なお、前方の仮想スピーカの判断手法については、上述した第 1 実施形態と同様であるため、ここでの説明は省略する。

40

【 0 1 0 9 】

ここで、図 8 は、第 2 実施形態における音声処理システムの動作を説明するための図である。なお、図 8 の例では、第 2 実施形態における音声処理システム 6 0 の概略的な部分のみを記載している。

【 0 1 1 0 】

第 2 実施形態では、図 8 の例に示すように、再生装置 6 1 と提供サーバ 6 2 との間のデータ通信において、所定数の高圧縮データ用の通信路と、所定数の低圧縮データ用の通信路とを用いたコネクションを確立する。例えば、第 2 実施形態では、再生装置 6 1 側の通信手段 7 1 と、提供サーバ 6 2 側の通信手段 8 1 とにおいて、例えば 8 チャンネルの仮想スピーカに対応する音声データを通信するためのコネクションを確立する。例えば、通信

50

手段 7 1 , 8 1 は、高圧縮の音声データを送信するための 6 つの狭帯域の通信路 a ~ f と、低圧縮の音声データを送信するための 2 つの広帯域の通信路 A , B とを用いたコネクションを確立する。なお、第 2 実施形態におけるコネクションの数については、これに限定されるものではない。

【 0 1 1 1 】

振り分け手段 8 2 では、例えば多方向 (8 チャンネル) の仮想スピーカに対する音声データを生成し、生成した各音声データに対して、前方の音声データであるか否かに基づいて振り分け処理を行う。

【 0 1 1 2 】

圧縮手段 8 3 は、2 つの通信路 A , B で通信させる前方の音声データに対して低圧縮を行うか、又は圧縮しない (無圧縮) 。したがって、復元時に高周波成分が残ったままの音声データとなる。また、圧縮手段 8 3 は、6 つの通信路 a ~ f で通信させる前方以外の音声データに対して高圧縮を行う。したがって、復元時に高周波成分を含まない音声データとなる。

10

【 0 1 1 3 】

例えば、図 8 の例において、頭部姿勢情報が北を 0 ° にした方位を基準にして、頭部姿勢センサ 1 4 の値が最初 = 1 5 ° であり、所定時間経過後に = 6 0 ° に変化したとする。この場合、前方判断手段 3 2 は、上述した図 5 (B) や図 6 を参照すると、最初 = 1 5 ° に対応して 2 つの仮想スピーカ # 1 及び # 2 を選択する。したがって、2 つの通信路 A , B には、# 1 及び # 2 に対する音声データが送信される。また、6 つの通信路 a ~ f は、他の仮想スピーカ # 3 ~ # 8 に対する高圧縮された音声データが送信される。

20

【 0 1 1 4 】

また、その後の姿勢情報 = 6 0 ° となった場合に、前方判断手段 3 2 は、前方の仮想スピーカとして # 2 及び # 3 を選択する。つまり、選択される 2 つの仮想スピーカは、「 # 1、# 2 」から「 # 2、# 3 」に変化する。このような場合に、振り分け手段 8 2 は、姿勢情報が変化するタイミングに対応させて、通信路 A , B と、通信路 a ~ f とに対する音声データの振り分けを変えることで、シームレスに情報を送信することができる。

【 0 1 1 5 】

例えば、通信手段 8 1 は、2 つの通信路 A , B を用いて、仮想スピーカ # 2 及び # 3 に対する音声データを送信する。また、通信手段 8 1 は、6 つの通信路 a ~ f を用いて、他の仮想スピーカ # 1、# 4 ~ # 8 に対する高圧縮された音声データを送信する。

30

【 0 1 1 6 】

なお、第 2 実施形態では、通信ネットワーク 1 3 の回線がコネクション状態のままであるため、コーデック情報の送受信を 1 回で済ませることができる。また、第 2 実施形態では、使用する通信路が固定となるため、そのためのメモリの確保を固定にすることができる。

【 0 1 1 7 】

第 2 実施形態における再生装置 6 1 では、通信手段 7 1 が、上述した 2 種類の通信路で送信される音声データを受信する。復号手段 7 2 は、それぞれの通信路から送られたデータに対して予め受信したコーデック表 7 3 - 1 を用いて、通信路毎の復号化方式により復号し、その結果を集約して、音像が定位された音声データをイヤホン 1 5 から出力する。

40

【 0 1 1 8 】

< 第 2 実施形態における圧縮手段 8 3 の処理の一例 >

図 9 は、第 2 実施形態における圧縮手段の処理の一例を示すフローチャートである。図 9 の例において、圧縮手段 8 3 は、コーデック制御手段 3 3 から再生装置 6 1 とのセッション開始が通知される (S 2 1) 。次に、圧縮手段 8 3 は、記憶手段 3 7 に記憶されたコーデック表 3 7 - 3 のコーデックを準備する (S 2 2) 。

【 0 1 1 9 】

次に、圧縮手段 8 3 は、音声生成手段 3 5 から仮想スピーカ用の音声データを取得すると (S 2 3) 、前方情報 3 7 - 2 を参照し、前方以外の仮想スピーカの音声データを圧縮

50

する (S 2 4)。この場合、前方の仮想スピーカの音声データは無圧縮とする。

【 0 1 2 0 】

次に、圧縮手段 8 3 は、通信手段 3 1 に仮想スピーカの識別情報 (仮想スピーカ I D) と、I D に対応する音声データと、I D に対して前方か否かを示す情報とを通信手段 8 1 に出力する (S 2 5)。

【 0 1 2 1 】

< 第 2 実施形態における提供サーバ 6 2 の通信手段 8 1 の処理の一例 >

図 1 0 は、第 2 実施形態における提供サーバの通信手段の処理の一例を示すフローチャートである。なお、以下の処理では、上述したように 8 チャンネルの音声データのうち、低圧縮 (無圧縮) の音声データを 2 つのコネクション (通信路) A , B で伝送し、高圧縮の音声データを 6 つのコネクション a ~ f で伝送する例について説明するが、これに限定されるものではない。

10

【 0 1 2 2 】

図 1 0 の例において、通信手段 8 1 は、再生装置 6 1 とセッションを開始し (S 3 1)、再生装置 6 1 にコーデック表 3 7 - 3 を送信する (S 3 2)。次に、通信手段 8 1 は、例えば高圧縮の音声データ用のコネクション a ~ f と、無圧縮の音声データ用のコネクション A , B を確立する (S 3 2)。

【 0 1 2 3 】

次に、通信手段 8 1 は、圧縮手段 8 3 から仮想スピーカ毎に圧縮又は無圧縮の音声データを取得し (S 3 4)、コネクション A , B、コネクション a ~ f にそれぞれ未使用フラグを付与する (S 3 5)。次に、通信手段 8 1 は、所定の仮想スピーカに対応する音声データを取得し (S 3 6)、その音声データは、前方か否かを判断する (S 3 7)。所定の仮想のスピーカとは、例えば全ての仮想スピーカ (# 1 ~ # 8) のうち、まだ再生装置 6 1 に送信していない音声データに対応する仮想スピーカである。

20

【 0 1 2 4 】

S 3 7 の処理において、通信手段 8 1 は、音声データが前方の場合 (S 3 7 において、Y E S)、コネクション A , B のうち、未使用フラグのついたコネクションを 1 つ割り当て、そのコネクションの未使用フラグを消す (S 3 8)。未使用フラグを消すとは、そのコネクションを使用したことを示す。

【 0 1 2 5 】

また、通信手段 8 1 は、音声データが前方でない場合 (S 3 7 において、N O)、コネクション a ~ f のうち、未使用フラグのついたコネクションを 1 つ割り当て、そのコネクションの未使用フラグを消す (S 3 9)。

30

【 0 1 2 6 】

次に、通信手段 8 1 は、割り当てられたコネクションに { 仮想スピーカ I D , 音声データ } の組を有する通信データを設定し (S 4 0)、その通信データを割り当てたコネクションを用いて再生装置 6 1 に送信する (S 4 1)。

【 0 1 2 7 】

ここで、通信手段 8 1 は、全ての音声データに対して処理を実行したか否かを判断し (S 4 2)、全ての音声データに対して処理を実行していない場合 (S 4 2 において、N O)、S 3 6 に戻り、未処理の音声データに対して処理を行う。また、通信手段 8 1 は、全ての音声データに対して処理を実行した場合 (S 4 2 において、Y E S)、処理を終了する。

40

【 0 1 2 8 】

< 第 2 実施形態における再生装置 6 1 の通信手段 7 1 の処理の一例 >

次に、第 2 実施形態における再生装置 6 1 の通信手段 7 1 の処理の一例について、フローチャートを用いて説明する。図 1 1 は、第 2 実施形態における再生装置の通信手段の処理の一例を示すフローチャートである。なお、図 1 1 の例では、上述した図 1 0 に示す処理により提供サーバ 6 2 から送信された通信データに対応する処理について説明するが、これに限定されるものではない。

50

【 0 1 2 9 】

図 1 1 の例において、通信手段 7 1 は、提供サーバ 6 2 とのセッションを開始し (S 5 1)、提供サーバ 6 2 からコーデック表 3 7 - 3 を受信する (S 5 2)。また、通信手段 7 1 は、高圧縮の音声データ用のコネクション a ~ f と、無圧縮の音声データ用のコネクション A , B を確立する (S 5 3)。次に、通信手段 7 1 は、復号手段 7 2 にコーデック表 3 7 - 3 の情報を出力する (S 5 4)。なお、コーデック表 3 7 - 3 は、コーデック表 7 3 - 1 として記憶手段 7 3 に記憶しておき、復号手段 7 2 による復号時に記憶手段 7 3 からコーデック表 7 3 - 1 を参照してもよい。

【 0 1 3 0 】

次に、通信手段 7 1 は、提供サーバ 6 2 からの通信データを受信すると (S 5 5)、その通信データをコネクション A , B から受信したか否かを判断する (S 5 6)。通信手段 7 1 は、通信データをコネクション A , B から受信した場合 (S 5 6 において、 Y E S)、前方用のフラグを付けて復号手段 7 2 に出力する (S 5 7)。また、通信手段 7 1 は、通信データをコネクション A , B から受信していない場合 (S 5 6 において、 N O)、前方用でない (前方以外である) ことを示すフラグを付けて復号手段 7 2 に出力する (S 5 8)。なお、 S 5 7 の処理において、前方用のフラグを付けているため、そのフラグがついていない通信データは、前方用ではないと判断ができる。したがって、上述した S 5 8 の処理は、省略してもよい。

【 0 1 3 1 】

これにより、復号手段 7 2 は、例えば前方用のフラグがある通信データは、無圧縮であるため復号を行わず、前方以外の通信データはコーデック表 7 3 - 1 等のコーデックに対応する復号化方式 (デコーデック) で復号を行う。また、復号手段 7 2 は、復号された音声データ等を音像定位手段 2 4 に出力する。これにより、音像定位手段 2 4 は、復号手段 7 2 から得られる音声データを集約して前方に高周波数成分を有し、音像が定位された適切な音声データをイヤホン 1 5 から出力することができる。

【 0 1 3 2 】

上述したように、第 2 実施形態では、適切な音声出力を実現することができる。また、第 2 実施形態では、高圧縮の通信路 (低域) と、低圧縮の通信路 (高域) を固定で用意しておくことで、コーデック情報の送受信を 1 回で済ませることができる。また、第 2 実施形態では、メモリの確保を固定にすることができる。

【 0 1 3 3 】

< 第 3 実施形態における音声処理システムの概略構成例 >

次に、第 3 実施形態について説明する。図 1 2 は、第 3 実施形態における音声処理システムの構成例を示す図である。第 3 実施形態では、上述した第 2 実施形態とは異なる音声ストリームの切り替え例を示している。

【 0 1 3 4 】

図 1 2 に示す音声処理システム 9 0 において、上述した音声処理システム 1 0 , 8 0 と同様の構成等については、同一の符号を付するものとし、ここでの具体的な説明は省略する。また、音声処理システム 9 0 における再生装置や提供サーバのハードウェア構成も上述した第 1 実施形態におけるハードウェア構成を適用することができるため、ここでの具体的な説明は省略する。

【 0 1 3 5 】

図 1 2 に示す音声処理システム 9 0 は、再生装置 9 1 と、提供サーバ 9 2 とを有する。再生装置 9 1 と、提供サーバ 9 2 とは、例えばインターネットや W L A N 等に代表される通信ネットワーク 1 3 により、データの送受信が可能な状態で接続されている。なお、第 3 実施形態における通信ネットワーク 1 3 は、コネクション接続により常時接続されているネットワーク形態を示している。

【 0 1 3 6 】

再生装置 9 1 は、頭部姿勢取得手段 2 1 と、前方判断手段 1 0 1 と、通信手段 1 0 2 と、復号手段 1 0 3 と、音像定位手段 2 4 と、記憶手段 1 0 4 とを有する。記憶手段 1 0 4

10

20

30

40

50

は、仮想スピーカ配置情報 25 - 1 と、コーデック表 73 - 1 と、前方情報 104 - 1 とを有する。

【0137】

また、提供サーバ 92 は、通信手段 111 と、前方判断手段 32 と、コーデック制御手段 33 と、音声取得手段 34 と、音声生成手段 35 と、圧縮手段 112、抽出手段 113 と、記憶手段 37 とを有する。

【0138】

第 3 実施形態では、図 12 に示すように、再生装置 91 及び提供サーバ 92 の両方に前方判断手段 32、101 を有し、両方でユーザの前方を判断し、前方に対応する仮想スピーカを選択する。これにより、第 3 実施形態は、再生装置 91 と提供サーバ 92 との間で前方に対応する音声はどれであるかという情報の送受信を省略することができるため、通信量を削減して通信効率を向上させることができる。

10

【0139】

また、第 3 実施形態では、音声生成手段 35 で生成された各仮想スピーカに対応する音声データを圧縮する際、低周波成分と高周波成分とに分離して圧縮を行う。更に、第 3 実施形態では、全ての仮想スピーカに対応する低周波成分の音声データを再生装置 91 に送信すると共に、ユーザの前方に対応する仮想スピーカに対して高周波成分の音声データを送信する。

【0140】

ここで、図 13 は、第 3 実施形態における音声処理システムの動作を説明するための図である。なお、図 13 の例では、第 3 実施形態における音声処理システム 90 の概略的な部分のみを記載している。

20

【0141】

第 3 実施形態では、再生装置 91 における通信手段 102 と、提供サーバ 92 における通信手段 111 とにおけるセッション開始時に、例えば低周波成分用のコネクション（通信路）8 つ（a ~ h）と、高周波成分用のコネクション 2 つ（A, B）を確立する。なお、第 3 実施形態におけるコネクションの数については、これに限定されるものではない。

【0142】

提供サーバ 92 の圧縮手段 112 は、音声生成手段 35 により生成される仮想スピーカ毎の音声データ（例えば、8 チャンネル）の全てに対して高周波成分と低周波成分とに分離して圧縮を行う。圧縮手段 112 による圧縮手法は、例えば M P E G 2 - A A C の S c a l a b l e S a m p l e R a t e (S S R) 等のスケーラブルな音声符号化を用いることができるが、これに限定されるものではない。

30

【0143】

抽出手段 113 は、前方判断手段 32 による判断結果に応じて、圧縮手段 112 により得られる各仮想スピーカに対応する高周波成分の圧縮音声データから、ユーザの前方に対応するデータを抽出する。第 3 実施形態では、図 13 に示すように、8 つのコネクション a ~ h では、8 チャンネル全ての低周波成分の音声データを再生装置 91 に送信し、その他に 2 つのコネクション A, B に対して前方のチャンネル用の高周波成分の音声データを再生装置 91 に送信する。

40

【0144】

再生装置 91 では、頭部姿勢取得手段 21 により得られる頭部姿勢センサ 14 からの取得情報に基づいて、前方判断手段 101 により前方を判断し、仮想スピーカ配置情報 25 - 1 を参照して、前方に対応する仮想スピーカを選択する。なお、選択された前方情報 104 - 1 は、記憶手段 104 に記憶される。

【0145】

復号手段 103 は、前方情報 104 - 1 を用いて、上述したコネクション A, B の 2 つの高周波成分の音声データを、コネクション a ~ h の 8 つの低周波成分の音声データのうち、前方に対応する音声データに付加して復号する。また、復号手段 103 は、これらの復号結果を音像定位手段 24 に出力する。音像定位手段 24 は、得られた音声データを

50

集約して音像が定位された音声データをイヤホン15から出力する。

【0146】

例えば、図13の例では、頭部姿勢情報が、北を0°にした方位を基準にして、頭部姿勢センサ14の値が最初 = 15°であり、所定時間経過後に = 60°に変化したとする。この場合、上述した第2実施形態と同様に、図6や図5(B)の例を参照すると、前方の仮想スピーカは、最初「#1、#2」であり、その後「#2、#3」に変化する。

【0147】

このような場合、抽出手段113は、圧縮手段112によりそれぞれの周波数成分（高周波、低周波）で圧縮した音声データのうちの高周波成分について、最初は、前方と判断された仮想スピーカ#1、#2に対応する高周波成分の音声データを抽出する。また、抽出手段113は、上述した頭部姿勢情報の変化（例えば、 = 15° 60°）により、仮想スピーカ#2、#3に対応する高周波成分の音声データを抽出する。

10

【0148】

通信手段111は、全ての仮想スピーカ#1～#8に対応する低周波成分の音声データを送信すると共に、抽出手段113により抽出された高周波成分の音声データを切り替えながら送信する。

【0149】

これにより、第3実施形態では、低周波成分の音声データが継続的に送信されるため、音声データをシームレスに出力することができる。また、第3実施形態では、通信回線がコネクション状態のままであるため、コーデック表37-3の送受信を1回で済ませることができる。また、第3実施形態では、前方判断を再生装置91と、提供サーバ92の両方で行うため、例えば前方情報に対応する情報等の送受信が不要となり、通信効率を向上させることができる。

20

【0150】

上述したよう、第3実施形態では、高周波成分用のコネクションA、Bに、コネクションa～hで送信される低周波成分の音声データと元の音声データとの差分情報（高周波成分）を送ることで、再生装置91において適切な音声出力を実現することができる。

【0151】

<第3実施形態における圧縮手段112及び抽出手段113の処理の一例>

図14は、第3実施形態における圧縮手段及び抽出手段の処理の一例を示すフローチャートである。図14の例において、圧縮手段112は、コーデック制御手段33から再生装置91とのセッション開始が通知されると（S61）、コーデック表37-3のコーデックを準備する（S62）。

30

【0152】

次に、圧縮手段112は、音声生成手段35から仮想スピーカ用の音声データを取得し（S63）、低周波数成分と高周波数成分とに分離して圧縮する（S64）。なお、S64の処理では、予め設定された仮想スピーカの各チャンネルに対応する全ての音声データに対して低周波数成分と、高周波数成分とに分離して圧縮する。なお、圧縮形式は、低周波成分と高周波成分とで同一でもよく異なってもよい。圧縮形式は、低周波成分及び高周波成分の成分毎に選択することができる。次に、圧縮手段112は、圧縮された低周波数成分の音声データを通信手段111等へ出力する（S65）。

40

【0153】

次に、抽出手段113は、前方情報判断手段32により判断された前方情報37-2を参照し（S66）、圧縮された高周波成分の音声データのうち、前方に対応する音声データを抽出し、抽出した音声データに高周波成分フラグを付与して通信手段111等へ出力する（S67）。なお、S67の処理では、再生装置91側においてどのコネクションから受信したかを検出することにより高周波成分の音声データか否かを判断することが可能である。したがって、その場合には、S67の処理において高周波成分フラグを付与しなくてもよい。

【0154】

50

< 第3実施形態における提供サーバ92の通信手段111の処理の一例 >

図15は、第3実施形態における提供サーバの通信手段の処理の一例を示すフローチャートである。図15の例において、通信手段111は、再生装置91とセッションを開始し(S71)、再生装置91にコーデック表37-3を送信する(S72)。また、通信手段111は、低周波成分の音声データ用のコネクションa~hと、高周波成分の音声データ用のコネクションA,Bを確立する(S73)。

【0155】

次に、通信手段111は、圧縮手段112から圧縮された音声データを取得し(S74)、低周波成分の音声データ8つをコネクションa~hに割り当て、前方の高周波成分の音声データ2つをコネクションA,Bに割り当てる(S75)。次に、通信手段111は、コネクションを通じてデータを再生装置91に送信する(S76)。

10

【0156】

< 第3実施形態における再生装置91の通信手段102の処理の一例 >

図16は、第3実施形態における再生装置の通信手段の処理の一例を示すフローチャートである。上述した提供サーバ92により送信された通信データに対応する処理について説明するが、これに限定されるものではない。

【0157】

図16の例において、通信手段81は、提供サーバ92とのセッションを開始し(S81)、提供サーバ92からコーデック表を受信する(S82)。また、通信手段81は、低周波成分の音声データ用のコネクションa~fと、高周波成分の音声データ用のコネクションA,Bを確立する(S83)。

20

【0158】

次に、通信手段81は、復号手段103にコーデック表37-3の情報を出力する(S84)。なお、コーデック表37-3は、コーデック表73-1として記憶手段104に記憶しておき、復号手段103による復号時に記憶手段104からコーデック表73-1を参照してもよい。

【0159】

次に、通信手段81は、提供サーバ92から通信データを受信し(S85)、通信データをコネクションA,Bから受信したか否かを判断する(S86)。なお、S86の処理では、受信した通信データに対して、上述した高周波成分フラグが付与されているか否かで判断してもよい。

30

【0160】

通信手段81は、通信データをコネクションA,Bから受信した場合(S86において、YES)、再生装置91の前方情報104-1から前方の仮想スピーカIDを取得する(S87)。なお、S87の処理では、予め頭部姿勢取得手段21により頭部姿勢センサ14から頭部姿勢情報を取得し、取得した頭部姿勢情報から前方判断手段101により前方がどこであるかが判断され、その結果が前方情報104-1に記憶されている。

【0161】

次に、通信手段81は、仮想スピーカIDに一致する復号手段103の高周波用の入力に、コネクションA,Bからの音声データを割り当てて復号手段103に出力する(S88)。また、S86の処理において、通信手段81は、通信データをコネクションA,Bから受信していない場合(S86において、NO)、低周波成分用のコネクションa~hから受信したものと判断し、コネクションa~hからの音声データを復号手段103の低周波成分用の入力1~8に割り当てて復号手段103に出力する(S89)。

40

【0162】

< 第3実施形態における再生装置91の復号手段103の処理の一例 >

図17は、第3実施形態における再生装置の復号手段の処理の一例を示すフローチャートである。図17の例において、復号手段103は、コーデック表73-1を取得すると(S91)、復号用のコーデックを準備し、低周波成分用の入力口1~8と、高周波成分用の入力口1'~8'を設定する(S92)。

50

【 0 1 6 3 】

次に、復号手段 1 0 3 は、通信手段 1 0 2 から音声データを取得し (S 9 3)、低周波成分の音声データのみが通知された場合、低周波成分のみで復号し、低周波成分と高周波成分の情報が両方通知された場合は、両方を用いて復号する (S 9 4)。

【 0 1 6 4 】

次に、復号手段 1 0 3 は、復号した音声データを音像定位手段 2 4 に出力する (S 9 5)。これにより、音像定位手段 2 4 は、取得した音声データを集約してユーザの前方に高周波数成分を有する音像が定位した音声データをイヤホン 1 5 から出力することができる。

【 0 1 6 5 】

上述したように第 3 実施形態では、再生装置 9 1 と提供サーバ 9 2 との両側で前方を判断することで、前方がどれであるかという情報を送信する必要がなくなる。このため、通信量を削減し、通信効率を向上させることができる。

【 0 1 6 6 】

なお、上述した第 1 ~ 第 3 実施形態は、複数の実施形態の一部又は全部を組み合わせることができる。また、上述した実施形態に限定されるものではなく、例えば音源に高周波数成分を含めて圧縮や伸長 (復号) するのではなく、例えば提供サーバ側から低周波数成分の音声と音源の位置だけを送信する。そして、再生装置側で、ユーザの前方に対応する低周波数の音声を用いて高周波数の音声を生成し、それらを集約することで音像に定位感を与えることができる。

【 0 1 6 7 】

上述したように本実施形態によれば、適切な音声出力を実現することができる。例えば、本実施形態では、人間の特性と、圧縮の特性を鑑みて、音像定位の維持と圧縮を両立する。例えば、本実施形態では、ユーザの姿勢情報に対応させて高周波数成分の音声データを処理する。また、本実施形態では、第 2 実施形態や第 3 実施形態に示すように、同じ帯域幅を用いて、帯域幅を変更する仮想スピーカを切り替える。このとき、例えば、ユーザの前方に存在する音源は高周波成分を含めて通信し、それ以外 (後方) は圧縮した低周波の音源を伝送することで、圧縮と音質を両立させた適切な音声通信を実現することができる。

【 0 1 6 8 】

また、本実施形態では、通信量を削減しつつ、ある地点の周囲の音声を、方向感を含めて別の地点で適切に再現することができる。したがって、本実施形態は、例えば博物館や美術館、展示会、テーマパーク等において、イヤホンやヘッドホン等の耳装着型の再生装置を用いた聴取者が、展示物等の方向から、その展示物に係る展示案内の音声や音楽を聴取可能にするシステム等に適用可能である。

【 0 1 6 9 】

以上、各実施例について詳述したが、特定の実施例に限定されるものではなく、特許請求の範囲に記載された範囲内において、上記変形例以外にも種々の変形及び変更が可能である。

【 0 1 7 0 】

なお、以上の実施例に関し、更に以下の付記を開示する。

(付記 1)

ユーザの姿勢情報から前記ユーザの前方を判断する前方判断手段と、
予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成する音声生成手段と、

前記音声生成手段により生成された前記音声データに対し、前記前方判断手段により得られる前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとで異なる圧縮を行う圧縮手段と、

前記圧縮手段により圧縮された前記音声データを送信する通信手段とを有することを特徴とする情報処理装置。

10

20

30

40

50

(付記 2)

前記圧縮手段は、

前記ユーザの前方に対応する音声データに対して、高周波数成分が復元可能な圧縮を行い、前記ユーザの前方以外の方向に対応する音声データに対して低周波数成分が復元可能な圧縮を行うことを特徴とする付記 1 に記載の情報処理装置。

(付記 3)

前記通信手段は、

前記圧縮手段により得られる前記ユーザの前方に対応する音声データと、前記前方以外の方向に対応する音声データとを、それぞれ異なる通信路を用いて送信することを特徴とする付記 1 又は 2 に記載の情報処理装置。

10

(付記 4)

前記前方判断手段により得られる前方情報に対応させて、前記音声生成手段により得られる前記音声データを振り分ける振り分け手段を有し、

前記圧縮手段は、前記振り分け手段により振り分けられた音声データ毎に、前記異なる圧縮を行うことを特徴とする 1 乃至 3 の何れか 1 項に記載の情報処理装置。

(付記 5)

前記圧縮手段は、

前記音声生成手段により生成された全ての仮想音源に対応する音声データを低周波数成分と高周波成分とに分離して圧縮し、

前記圧縮手段により得られる前記高周波成分の音声データから、前記前方判断手段により得られる前記ユーザの前方に対応する前記高周波成分の音声データを抽出する抽出手段を有し、

20

前記通信手段は、前記圧縮手段により圧縮された前記低周波成分の音声データの全てと、前記抽出手段により抽出された前記ユーザの前方に対応する前記高周波成分の音声データとを送信することを特徴とする付記 1 乃至 4 の何れか 1 項に記載の情報処理装置。

(付記 6)

前記前方判断手段は、

前記ユーザの姿勢情報と、予め前記仮想音源の配置位置が設定された配置情報とを用いて、前記ユーザの前方に最も近い少なくとも 1 つの仮想音源を選択することを特徴とする付記 1 乃至 5 の何れか 1 項に記載の情報処理装置。

30

(付記 7)

前記前方判断手段により得られる前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとに対する圧縮時の符号化情報及び符号化パラメータを制御する制御手段を有することを特徴とする付記 1 乃至 6 の何れか 1 項に記載の情報処理装置。

(付記 8)

情報処理装置が、

ユーザの姿勢情報から前記ユーザの前方を判断し、

予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成し、

40

生成された前記音声データに対し、前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとで異なる圧縮を行い、

前記異なる圧縮により圧縮された前記音声データを送信することを特徴とする音声処理方法。

(付記 9)

ユーザの姿勢情報から前記ユーザの前方を判断し、

予め設定した複数の方向に配置される仮想音源のそれぞれに割り当てた音声データを生成し、

生成された前記音声データに対し、前記ユーザの前方に対応する音声データと、前記ユーザの前方以外の方向に対応する音声データとで異なる圧縮を行い、

50

前記異なる圧縮により圧縮された前記音声データを送信する、処理をコンピュータに実行させるための音声処理プログラム。

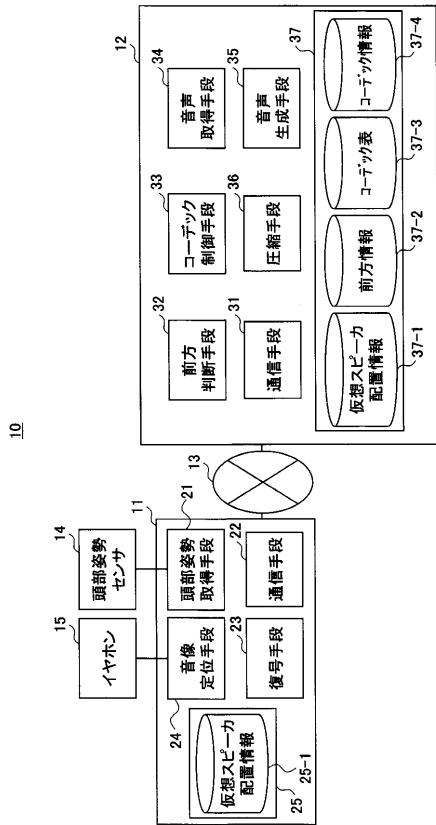
【符号の説明】

【0171】

- | | | |
|--------------------------|-----------------|----|
| 10, 60, 90 | 音声処理システム | |
| 11, 61, 91 | 再生装置(通信端末) | |
| 12, 62, 92 | 提供サーバ(情報処理装置) | |
| 13 | 通信ネットワーク | |
| 14 | 頭部姿勢センサ(姿勢検出手段) | |
| 15 | イヤホン(音声出力手段) | 10 |
| 21 | 頭部姿勢取得手段 | |
| 22, 31, 71, 81, 102, 111 | 通信手段 | |
| 23, 72 | 復号手段 | |
| 24 | 音像定位手段 | |
| 25, 37, 73, 94 | 記憶手段 | |
| 32, 101 | 前方判断手段 | |
| 33 | コーデック制御手段 | |
| 34 | 音声取得手段 | |
| 35 | 音声生成手段 | |
| 36, 83, 112 | 圧縮手段 | 20 |
| 41, 51 | 入力装置 | |
| 42, 52 | 出力装置 | |
| 43 | 通信インタフェース | |
| 44 | オーディオインタフェース | |
| 45, 54 | 主記憶装置 | |
| 46, 55 | 補助記憶装置 | |
| 47, 56 | CPU | |
| 48, 57 | ネットワーク接続装置 | |
| 53 | ドライブ装置 | |
| 58 | 記録媒体 | 30 |
| 82 | 振り分け手段 | |
| 113 | 抽出手段 | |

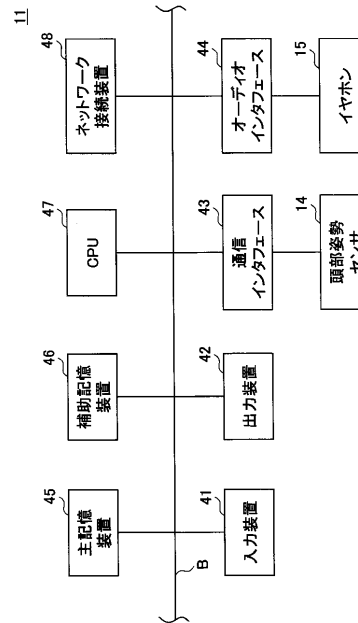
【 図 1 】

第1実施形態における音声処理システムの構成例を示す図



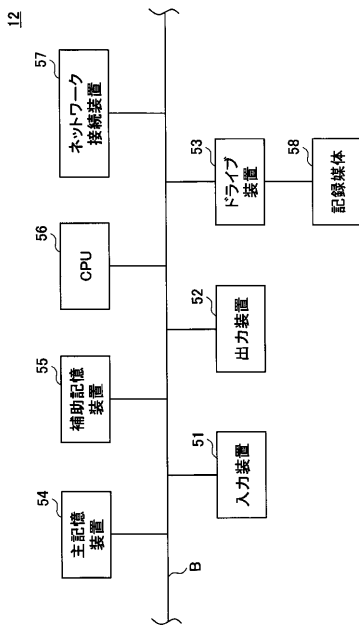
【 図 2 】

再生装置のハードウェア構成例を示す図



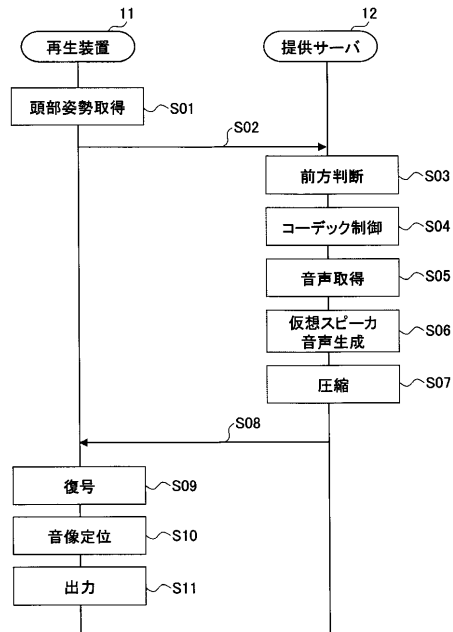
【 図 3 】

提供サーバのハードウェア構成例を示す図



【 図 4 】

音声処理システムの処理の一例を示すシーケンス図



【 図 5 】

音声処理システムで用いられる各種データ例を説明するための図

(A)

種別情報	時間	姿勢情報(θ)	...
A001	16:56:10	15°	...
A001	16:56:40	60°	...
...

(B)

仮想スピーカID	配置位置x	配置位置y
#1	0	1
#2	0.71	0.71
#3	1	0
#4	0.71	-0.71
#5	0	-1
#6	-0.71	-0.71
#7	-1	0
#8	-0.71	0.71

(C)

前方の仮想スピーカ
#1
#2

(D)

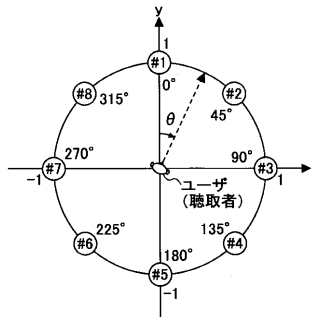
仮想スピーカ種別	コーデック	パラメータ
前方	圧縮なし	44kHz,16bit,1024frame
その他	サンプリング	22kHz,16bit,1024frame

(E)

コーデック情報
{(#1,圧縮なし,44kHz,16bit,1024frame),(#2,圧縮なし,44kHz,16bit,1024frame), (#3,サンプリング,22kHz,16bit,1024frame),(#4,サンプリング,22kHz,16bit,1024frame), (#5,サンプリング,22kHz,16bit,1024frame),(#6,サンプリング,22kHz,16bit,1024frame), (#7,サンプリング,22kHz,16bit,1024frame),(#8,サンプリング,22kHz,16bit,1024frame)}

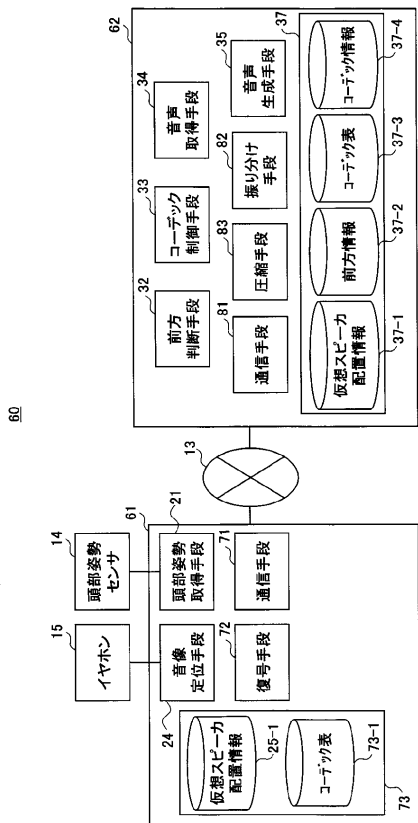
【 図 6 】

仮想スピーカの配置例を説明するための図



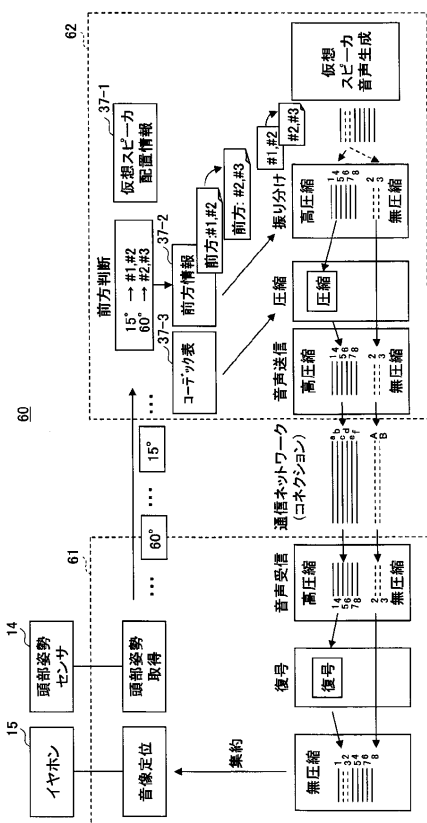
【 図 7 】

第2実施形態における音声処理システムの構成例を示す図



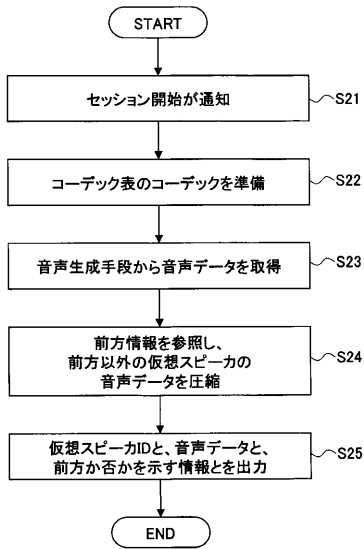
【 図 8 】

第2実施形態における音声処理システムの動作を説明するための図



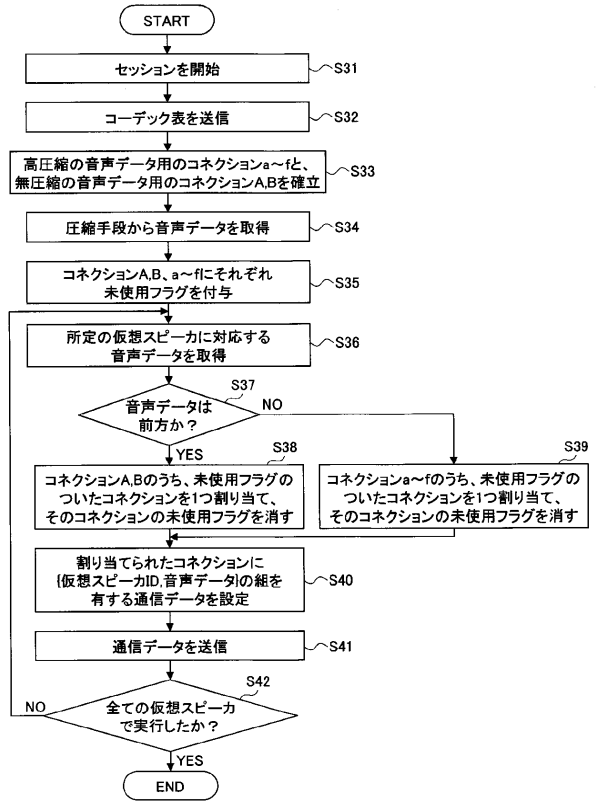
【図9】

第2実施形態における圧縮手段の処理の一例を示すフローチャート



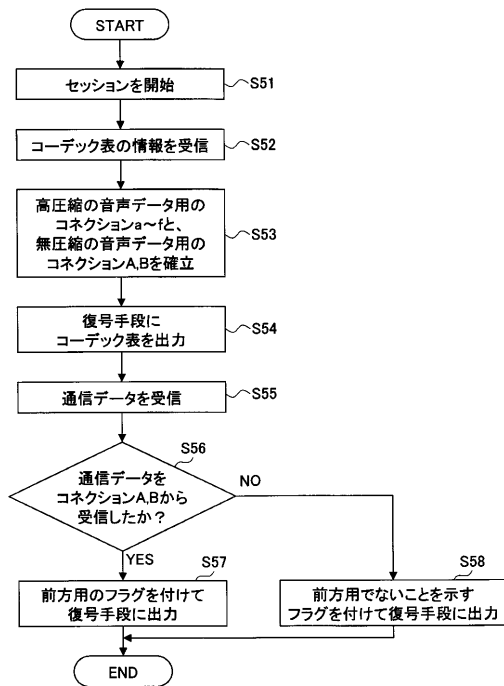
【図10】

第2実施形態における提供サーバの通信手段の処理の一例を示すフローチャート



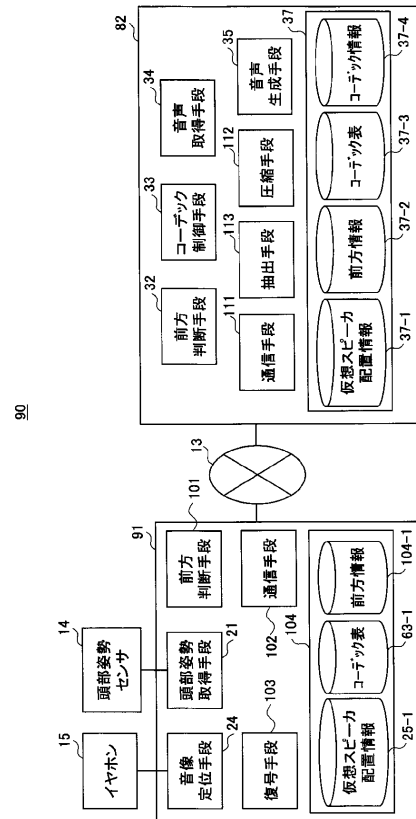
【図11】

第2実施形態における再生装置の通信手段の処理の一例を示すフローチャート



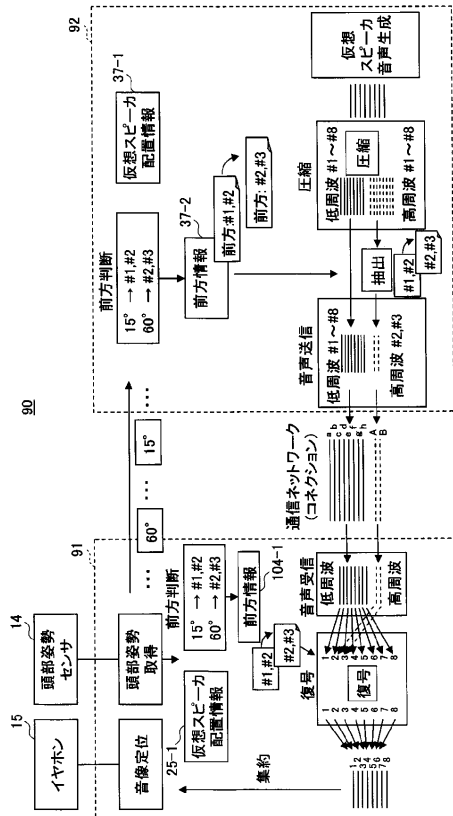
【図12】

第3実施形態における音声処理システムの構成例を示す図



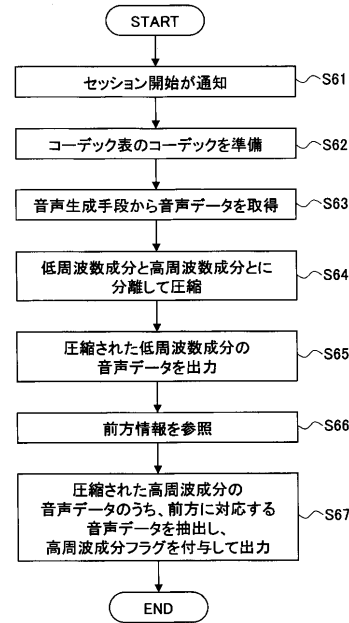
【図 1 3】

第3実施形態における音声処理システムの動作を説明するための図



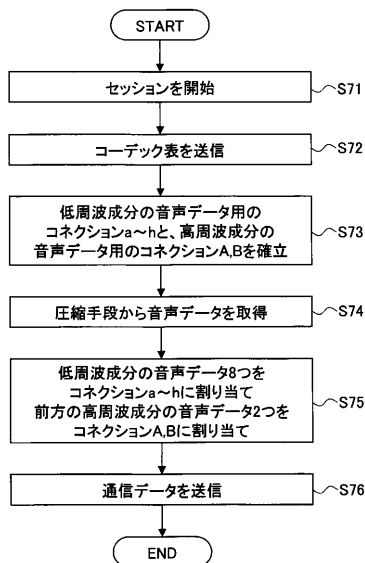
【図 1 4】

第3実施形態における圧縮手段及び抽出手段の処理の一例を示すフローチャート



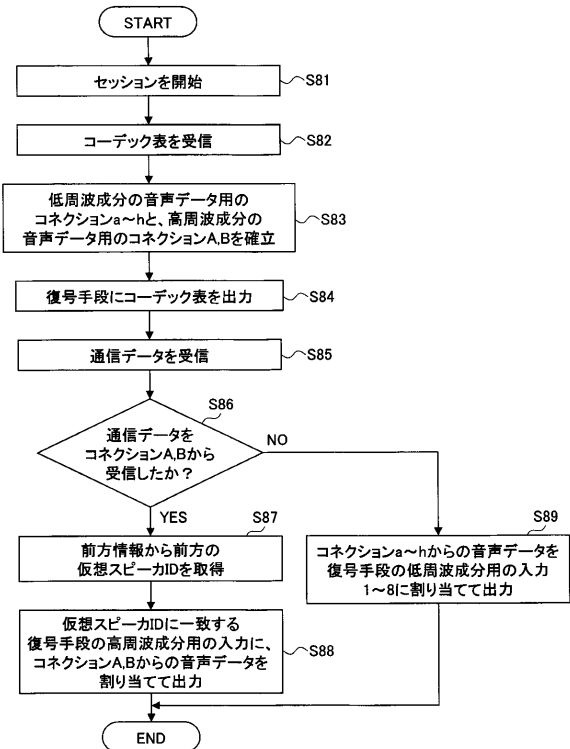
【図 1 5】

第3実施形態における提供サーバの通信手段の処理の一例を示すフローチャート



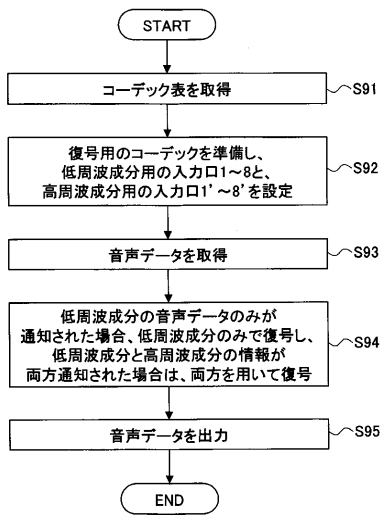
【図 1 6】

第3実施形態における再生装置の通信手段の処理の一例を示すフローチャート



【 図 1 7 】

第3実施形態における再生装置の復号手段の処理の一例を示すフローチャート



フロントページの続き

- (72)発明者 佐々木 和雄
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
- (72)発明者 野田 政秀
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
- (72)発明者 大谷 武
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
- Fターム(参考) 5D062 AA70 BB10
5D220 AA05 AA11 AB03 AB04 AB08