

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6342429号
(P6342429)

(45) 発行日 平成30年6月13日(2018.6.13)

(24) 登録日 平成30年5月25日(2018.5.25)

(51) Int.Cl.	F 1
HO4M 3/56 (2006.01)	HO4M 3/56 B
HO4N 7/15 (2006.01)	HO4N 7/15 120
HO4S 5/00 (2006.01)	HO4S 5/00
HO4S 1/00 (2006.01)	HO4S 1/00 700

請求項の数 15 (全 30 頁)

(21) 出願番号	特願2015-558018 (P2015-558018)
(86) (22) 出願日	平成26年1月23日 (2014.1.23)
(65) 公表番号	特表2016-514394 (P2016-514394A)
(43) 公表日	平成28年5月19日 (2016.5.19)
(86) 国際出願番号	PCT/US2014/012788
(87) 国際公開番号	W02014/130199
(87) 国際公開日	平成26年8月28日 (2014.8.28)
審査請求日	平成29年1月5日 (2017.1.5)
(31) 優先権主張番号	13/771,957
(32) 優先日	平成25年2月20日 (2013.2.20)
(33) 優先権主張国	米国(US)

(73) 特許権者	595020643 クアルコム・インコーポレイテッド QUALCOMM INCORPORATED アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドライブ 5775
(74) 代理人	100108855 弁理士 蔵田 昌俊
(74) 代理人	100109830 弁理士 福原 淑弘
(74) 代理人	100158805 弁理士 井関 守三
(74) 代理人	100194814 弁理士 奥村 元宏

最終頁に続く

(54) 【発明の名称】ステガノグラフィ方式で埋め込まれた音声データを用いた遠隔会議

(57) 【特許請求の範囲】

【請求項 1】

方法であって、

第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成することであって、前記第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、前記第2の音声データストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する音を表現すること、

前記第1の音声データストリームのより低質なモノラル表現である第1のサブストリームを生成することであって、ここにおいて、前記第1のサブストリームを生成することは、前記第1のサブストリームに対してデータ圧縮を行うためのボコーダを使用することを備えることと、

前記第2の音声データストリームのより低質なモノラル表現である第2のサブストリームを生成することであって、ここにおいて、前記第2のサブストリームを生成することは、前記第2のサブストリームに対してデータ圧縮を行うための前記ボコーダを使用することを備えることと、

少なくとも部分的に前記ミックスされた音声データストリーム内に前記第1のサブストリーム及び前記第2のサブストリームをステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データストリームを生成し、それによって、前記修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポ

10

20

ートするようにすることと、

前記修正されたミックスされた音声データストリームを前記遠隔会議に参加している第3の端末デバイスに出力することと、を備え、

前記第3の端末デバイスがマルチチャネル再生用に構成されている場合には、前記第3の端末において逆ステガノグラフィプロセスを実行することにより前記第1のサブストリーム及び前記第2のサブストリームが前記ミックスされた音声データストリームから抽出され、前記第3の端末デバイスがマルチチャネル再生用に構成されていない場合には、前記第1のサブストリーム及び前記第2のサブストリームを抽出することなしに、前記ミックスされた音声データストリームに基づいて音声データが生成される、方法。

【請求項2】

10

前記修正されたミックスされた音声データストリームに関連する音声歪みの量と前記第1及び第2のサブストリームを埋め込む前の前記ミックスされた音声データストリームに関連する音声歪みの量との間の差分は、歪みスレショルドよりも小さく、

前記修正されたミックスされた音声データストリームのビットレートは、前記1及び第2のサブストリームを埋め込む前の前記ミックスされた音声データストリームのビットレートよりも大きくない請求項1に記載の方法。

【請求項3】

前記修正されたミックスされた音声データストリームを生成することは、

データ圧縮形態を前記ミックスされた音声データストリームに適用することと、

前記第1及び第2のサブストリームを前記ミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことと、を備える請求項1に記載の方法。

20

【請求項4】

前記第1のサブストリームは、前記第1の音声データストリームよりも低いビットレートを有し、前記第2のサブストリームは、前記第2の音声データストリームよりも低いビットレートを有する請求項1に記載の方法。

【請求項5】

前記修正されたミックスされた音声データストリームを生成することは、前記第1及び第2のサブストリームを示すために前記ミックスされた音声データストリームの音声サンプルの最下位ビットを修正することを備える請求項1に記載の方法。

【請求項6】

30

方法であって、

ミックスされた音声データストリームから第1のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することであって、ここにおいて、前記第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現する第1の音声データストリームのより低質なモノラル表現であり、ここにおいて、前記第1のサブストリームは、前記第1のサブストリームに対してデータ圧縮を行うためのボコーダを使用して生成されることと、

前記ミックスされた音声データストリームから第2のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することであって、ここにおいて、前記第2のサブストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する音を表現する第2の音声データストリームのより低質なモノラル表現であり、ここにおいて、前記第2のサブストリームは、前記第2のサブストリームに対してデータ圧縮を行うための前記ボコーダを使用して生成されることと、

40

前記ミックスされた音声データストリーム、前記第1のサブストリーム及び前記第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成することと、を備える、方法。

【請求項7】

前記ミックスされた音声データストリームは、前記第1の端末デバイスに関連する前記音及び前記第2の端末デバイスに関連する前記音のミックスのモノラル表現を含み、

前記マルチチャネル音声データストリームを生成することは、空間化されたステレオ音

50

声データストリームを生成し、それによって、前記第1の端末デバイスに関連する前記音は、聴く人の左側の空間内のある地点から来ると知覚され、前記第2の端末デバイスに関連する音は、前記聴く人の右側の空間内のある地点から来ると知覚されることを備える請求項6に記載の方法。

【請求項8】

前記空間化されたステレオ音声データストリームを生成することは、

前記第1のサブストリームに少なくとも部分的に基づいて、前記第1の端末デバイスに関連する前記音を決定することと、

前記第2のサブストリームに少なくとも部分的に基づいて、前記第2の端末デバイスに関連する前記音を決定することと、

前記第1の端末デバイスに関連する前記音に少なくとも部分的に基づいて、前記ミックスされた音声データストリームの第1のコピーに頭部伝達関数(HRTF)を適用することであって、前記空間化されたステレオ音声データストリームの左チャネルは、前記ミックスされた音声データストリームの前記第1のコピーに少なくとも部分的に基づくことと、

前記第2の端末デバイスに関連する前記音に少なくとも部分的に基づいて、前記ミックスされた音声データストリームの第2のコピーに前記HRTFを適用することであって、前記空間化されたステレオ音声データストリームの右チャネルは、前記ミックスされた音声データストリームの前記第2のコピーに少なくとも部分的に基づくことと、を備える請求項7に記載の方法。

【請求項9】

前記第1の端末デバイスに関連する前記音は、前記第2の端末デバイスに関連する話声と同時並行である話声を含む請求項7に記載の方法。

【請求項10】

前記第1及び第2のサブストリームは、前記ミックスされた音声データストリームの音声サンプルの最下位ビットである請求項6に記載の方法。

【請求項11】

前記第1及び第2のサブストリームのビットレートは、前記ミックスされた音声データストリームのビットレートよりも低い請求項6に記載の方法。

【請求項12】

前記ミックスされた音声データストリームは、モノラル再生用に構成されてステレオ再生用に構成されていない端末デバイスと互換性がある請求項6に記載の方法。

【請求項13】

デバイスであって、

第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成するための手段であって、前記第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、前記第2の音声データストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する音を表現する手段と、

前記第1の音声データストリームのより低質なモノラル表現である第1のサブストリームを生成するための手段であって、ここにおいて、前記第1のサブストリームを生成することは、前記第1のサブストリームに対してデータ圧縮を行うためのボコーダを使用することを備える手段と、

前記第2の音声データストリームのより低質なモノラル表現である第2のサブストリームを生成するための手段であって、ここにおいて、前記第2のサブストリームを生成することは、前記第2のサブストリームに対してデータ圧縮を行うための前記ボコーダを使用することを備える手段と、

少なくとも部分的に前記ミックスされた音声データストリーム内に前記第1のサブストリーム及び前記第2のサブストリームをステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データストリームを生成し、それによって、前記修正された

10

20

30

40

50

ミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようにするための手段と、

前記修正されたミックスされた音声データストリームを前記遠隔会議に参加している第3の端末デバイスに出力するための手段と、を備え、

前記第3の端末デバイスがマルチチャネル再生用に構成されている場合には、前記第3の端末において逆ステガノグラフィプロセス実行することにより前記第1のサブストリーム及び前記第2のサブストリームが前記ミックスされた音声データストリームから抽出され、前記第3の端末デバイスがマルチチャネル再生用に構成されていない場合には、前記第1のサブストリーム及び前記第2のサブストリームを抽出することなしに、前記ミックスされた音声データストリームに基づいて音声データが生成される、デバイス。

10

【請求項14】

端末デバイスであって、

ミックスされた音声データストリームから第1のサブストリームを抽出するための逆ステガノグラフィプロセスを実行するための手段であって、ここにおいて、前記第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現する第1の音声データストリームのより低質なモノラル表現であり、ここにおいて、前記第1のサブストリームは、前記第1のサブストリームに対してデータ圧縮を行うためのボコーダを使用して生成される手段と、

前記ミックスされた音声データストリームから第2のサブストリームを抽出するための逆ステガノグラフィプロセスを実行するための手段であって、ここにおいて、前記第2のサブストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する音を表現する第2の音声データストリームのより低質なモノラル表現であり、ここにおいて、前記第2のサブストリームは、前記第2のサブストリームに対してデータ圧縮を行うための前記ボコーダを使用して生成される手段と、

20

前記ミックスされた音声データストリーム、前記第1のサブストリーム及び前記第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成するための手段と、を備える、端末デバイス。

【請求項15】

コンピュータによって読み取り可能な記憶媒体であって、デバイスの1つ以上のプロセッサによって実行されたときに、請求項1乃至12のうちのいずれか1項に記載の方法を実行することを前記デバイスに行わせる命令、を格納する、コンピュータによって読み取り可能な記憶媒体。

30

【発明の詳細な説明】

【技術分野】

【0001】

[0001] 本開示は、音声データの処理に関するものであり、より具体的には、遠隔会議のための音声データの処理に関するものである。

【背景技術】

【0002】

[0002] 遠隔会議は、概して、その遠隔会議の参加者によって使用される3つ以上の端末デバイスとの電気通信接続を確立することを含む。説明を簡単にする目的として、端末デバイスは、端末デバイスA、B及びCとして表すことができる。概して、参加者のうち1人、例えば、端末デバイスAのユーザが、他の参加者、例えば、端末デバイスのユーザB及びC、と会議を開くことによって遠隔会議を開始する。会議管理デバイスが会議を管理する。会議管理デバイスは、マルチパーティ制御ユニット(MCU)、又は代替として、ミキシングブリッジ、と呼ぶこともできる。MCUは、遠隔会議のホストとなるサービスプロバイダネットワーク内に配置することができる。

40

【0003】

[0003] MCUは、MCUにおいて端末デバイスから受信された音声データストリームを復号して3つの受信された音声データストリームのうちの2つによって表される音声

50

波形を合計し、ミックスされたモノラル（すなわち、モノ）の音声波形を生成することができる。MCUは、ミックスされたモノラルの音声波形を符号化してミックスされたモノデータストリームを生成し、それらのミックスされたモノデータストリームを端末デバイスの各々に送信することができる。

【0004】

[0004] 例えば、MCUは、端末デバイスA、B及びCから音声データストリームを受信し及び復号することができる。MCUは、受信された音声データストリームに基づいてミックスされた音声データを生成する。第1のミックスされた音声データストリームは、端末デバイスA及びBによって検出された音のモノラルミックス（すなわち、A+B）を表現する。第2のミックスされた音声データストリームは、端末デバイスA及びCによって検出された音のモノラルミックス（すなわち、A+C）を表現する。第3のミックスされた音声データストリームは、端末デバイスB及びCによって検出された音のモノラルミックス（すなわち、B+C）を表現する。MCUは、第1のミックスされた音声データストリームを端末デバイスCに送信し、第2のミックスされた音声データストリームを端末デバイスBに送信し、及び、第3のミックスされた音声データストリームを端末デバイスAに送信する。端末デバイスA、B、及びCは、ミックスされた音声データストリームを復号し、ミックスされた音声データストリームに基づいて音を生成する（すなわち、再生する）。

10

【0005】

[0005] 最近では、遠隔会議における三次元（3D）音声をサポートするMCU及び端末デバイスが開発されている。3D音声では、MCUは、ミックスされたステレオ音声データストリームを生成するために端末デバイスから受信された音声データストリームを処理する。ミックスされた音声ステレオデータストリームの各々は、2つ以上の（例えば、左右）チャネルを有する音を表現することができる。MCUは、これらのミックスされたステレオ音声データストリームを端末デバイスのうちの該当するそれに送信することができる。端末デバイスの各々は、2つ以上のスピーカーでミックスされたステレオ音声データストリームを再生することができる。MCUによってステレオ音声データストリームに適用される頭部伝達関数（HRTF）に起因して、端末デバイスのユーザは、他の端末デバイスのユーザの話声が空間内の様々な地点から来るのがわかることができる。例えば、端末デバイスAのユーザは、端末デバイスBの話声が端末デバイスAのユーザの左側の空間内のある地点から来るのがわかることができ及び端末デバイスCのユーザの話声が端末デバイスAのユーザの右側の空間内のある地点から来るのがわかることができる。このようにしてユーザの声を空間的に分離することは、ユーザが遠隔会議中に誰が話しているかを決定するのに役立つことができ、それによって3D音声遠隔会議の参加者間での通信を容易にすることができる。

20

30

【発明の概要】

【0006】

[0006] 概して、本開示は、遠隔会議のために音声データを処理するための技法について説明する。マルチパーティ制御ユニット（MCU）が、遠隔会議に参加している端末デバイスからやって来る音声データストリームを受信する。やって来る音声データの各々は、端末デバイスの各々によって検出された音のモノラル表現を備えることができる。端末デバイスの各々に関して、MCUは、端末デバイスのうちのその他のそれらによって検出された音のミックスのモノラル表現を備えるミックスされた音声データストリームを生成することができる。さらに、MCUは、端末デバイスによって検出された音の表現をミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことができる。説明を簡単にすることを目的として、端末デバイスによって検出された音の表現は、ここではサブストリームと呼ぶことができる。MCUは、ミックスされた音声データストリームを遠隔会議に参加している端末デバイスに出力する。

40

【0007】

[0007] 端末デバイスがステレオ再生用に構成されている場合は、端末デバイスは、

50

ミックスされた音声データストリームから、埋め込まれたサブストリームを抽出するための逆ステガノグラフィプロセスを実行する。端末デバイスは、抽出されたサブストリーム及びミックスされた音声データストリームに基づいてマルチチャネル音声データストリームを生成し及び出力する。端末デバイスは、マルチチャネル音声データストリームに基づいて音を出力することができる。端末デバイスがステレオ再生用に構成されていない場合は、端末デバイスは、埋め込まれているサブストリームを抽出せずにミックスされた音声データストリームに基づいて音を出力する。この意味で、サブストリームを埋め込むためにステガノグラフィを用いることによって、技法は、ミックスされた音声データストリームのマルチチャネル再生及びミックスされた音声データストリームのモノラル再生の両方を可能にすることができます。

10

【0008】

【0008】一例においては、本開示は、第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成することを備える方法であって、第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、第2の及びデータストリームは、遠隔会議に参加している第2の端末デバイスに関連する音を表現する方法、について説明する。その方法は、少なくとも部分的に第1のサブストリーム及び第2のサブストリームをミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データを生成して修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようにすることも備え、第1のサブストリームは、第1の端末デバイスに関連する音の表現を含み、第2のサブストリームは、第2の端末デバイスに関連する音の表現を含む。さらに、その方法は、修正されたモノ音声データストリームを遠隔会議に参加している第3の端末デバイスに出力することを備える。

20

【0009】

【0009】他の例においては、本開示は、第1のサブストリームをミックスされた音声データストリームから抽出するための逆ステガノグラフィプロセスを実行することを備える方法について説明し、第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する。その方法は、第2のサブストリームをミックスされた音声データから抽出するための逆ステガノグラフィプロセスを実行することも備え、第2のサブストリームは、遠隔会議に参加している第2の端末デバイスに関連する。さらに、その方法は、ミックスされた音声データストリーム、第1のサブストリーム及び第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成することを備える。

30

【0010】

【0010】他の例においては、本開示は、第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成するように構成された1つ以上のプロセッサを備えるデバイスであって、第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、第2の音声データストリームは、遠隔会議に参加している第2の端末デバイスに関連する音を表現するデバイス、について説明する。1つ以上のプロセッサは、少なくとも部分的に第1のサブストリーム及び第2のサブストリームをミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データを生成して修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするように構成され、第1のサブストリームは、第1の端末デバイスに関連する音の表現を含み、第2のサブストリームは、第2の端末デバイスに関連する音の表現を含む。さらに、1つ以上のプロセッサは、修正されたモノ音声データを遠隔会議に参加している第3の端末デバイスに出力するように構成される。

40

【0011】

【0011】他の例においては、本開示は、第1のサブストリームをミックスされた音声

50

データから抽出するための逆ステガノグラフィプロセスを実行するように構成された1つ以上のプロセッサを備える端末デバイスについて説明する。第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する。1つ以上のプロセッサは、第2のサブストリームをミックスされた音声データストリームから抽出するための逆ステガノグラフィプロセスを実行するようにも構成される。第2のサブストリームは、遠隔会議に参加している第2の端末デバイスに関連する。1つ以上のプロセッサは、ミックスされた音声データストリーム、第1のサブストリーム及び第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成するようにも構成される。

【0012】

[0012]他の例においては、本開示は、第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成するための手段であって、第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、第2の及びデータストリームは、遠隔会議に参加している第2の端末デバイスに関連する音を表現するデバイス、について説明する。そのデバイスは、少なくとも部分的に第1のサブストリーム及び第2のサブストリームをミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データを生成して修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようにするための手段も備え、第1のサブストリームは、第1の端末デバイスに関連する音の表現を含み、第2のサブストリームは、第2の端末デバイスに関連する音の表現を含む。さらに、そのデバイスは、修正されたモノ音声データストリームを遠隔会議に参加している第3の端末デバイスに出力するための手段を備える。
10

【0013】

[0013]他の例においては、本開示は、第1のサブストリームをミックスされた音声データストリームから抽出するための逆ステガノグラフィプロセスを実行するための手段を備える端末デバイスについて説明する。第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する。その端末デバイスは、第2のサブストリームをミックスされた音声データから抽出するための逆ステガノグラフィプロセスを実行するための手段も備え、第2のサブストリームは、遠隔会議に参加している第2の端末デバイスに関連する。さらに、その端末デバイスは、ミックスされた音声データストリーム、第1のサブストリーム及び第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成するための手段を備える。
20

【0014】

[0014]他の例においては、本開示は、デバイスの1つ以上のプロセッサによって実行されたときに、第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成することをデバイスに行わせる命令を格納するコンピュータによって読み取り可能な記憶媒体であって、第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、第2の音声データストリームは、遠隔会議に参加している第2の端末デバイスに関連する音を表現するコンピュータによって読み取り可能な記憶媒体、について説明する。それらの命令は、少なくとも部分的に第1のサブストリーム及び第2のサブストリームをミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データを生成して修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようにすることもデバイスに行わせ、第1のサブストリームは、第1の端末デバイスに関連する音の表現を含み、第2のサブストリームは、第2の端末デバイスに関連する音の表現を含む。さらに、それらの命令は、修正されたモノ音声データストリームを遠隔会議に参加している第3の端末デバイスに出力することをデバイスに行わせる。
40

【0015】

[0015]他の例においては、本開示は、デバイスの1つ以上のプロセッサによって実

10

20

30

40

50

行されたときに、第1のサブストリームをミックスされた音声データから抽出するための逆ステガノグラフィプロセスを実行することをデバイスに行わせる命令を格納するコンピュータによって読み取り可能な記憶媒体について説明する。第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する。それらの命令は、第2のサブストリームをミックスされた音声データストリームから抽出するための逆ステガノグラフィプロセスを実行することもデバイスに行わせる。第2のサブストリームは、遠隔会議に参加している第2の端末デバイスに関連する。さらに、それらの命令は、ミックスされた音声データストリーム、第1のサブストリーム及び第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成することをデバイスに行わせる。

10

【0016】

【0016】本開示の1つ以上の例の詳細が、添付された図面及び以下の説明において示される。これらの説明、図面、及び請求項からその他の特徴、目的、及び利点が明確なるであろう。

【図面の簡単な説明】

【0017】

【図1】【0017】本開示において説明される技法を利用することができる遠隔会議システム例を示したブロック図である。

【図2】【0018】本開示において説明される技法の様々な態様を実行することができるマルチパーティ制御ユニット（MCU）例を示したブロック図である。

20

【図3】【0019】本開示において説明される技法の様々な態様を実行する際の端末デバイス例を示したブロック図である。

【図4】【0020】本開示において説明される技法の様々な態様を実行する際の音声復号器例を示したブロック図である。

【図5】【0021】本開示において説明される技法の様々な態様を実行する際のMCUの動作例を示したフローチャート図である。

【図6】【0022】本開示において説明される技法の様々な態様を実行する際の音声復号器の動作例を示したフローチャートである。

【発明を実施するための形態】

【0018】

30

【0023】遠隔会議は、3人以上の当事者が音声データ（しばしば、話聲音声データ）をキャプチャし及びその他の端末デバイスによってキャプチャされた音声データを再生するための端末デバイスを用いて互いに会話するのを可能にする。遠隔会議に参加するユーザは、典型的には、別々の端末デバイス、例えば、電話又は専用の遠隔会議装置、と関連付けられる。さらに、遠隔会議システムは、マルチパーティ制御ユニット（MCU）を含むことができる。端末デバイスは、端末デバイスの1つ以上のマイクによって検出された音に基づいて音声データストリームを生成する。端末デバイスは、生成された音声データストリームをMCUに送信する。MCUは、音声データストリームによって表現される音をミックスすること及びその結果得られた音を符号化して複数のミックスされた音声データストリームを生成することができる。MCUが三次元（3D）音声遠隔会議用に構成されているときには、MCUは、ステレオのミックスされた音声データストリームを生成することができ、それらの各々は、左音声チャネルと右音声チャネルとを含むことができる。

40

【0019】

【0024】遠隔会議における3D音声は、参加しているユーザのために主観的音質を向上させることができ及び潜在的に遠隔会議経験を向上させることができる一方で、3D音声遠隔会議は、マルチチャネル（例えば、左右のチャネル）音声データストリームを提供するためにより高いデータレートを要求することがある。従って、MCUは、従来の3Dでない音声が遠隔会議において使用されるときよりも、3D音声が遠隔会議において使用されるときのほうが、有意な量だけより多くのデータを端末デバイスに送信する必要があ

50

る。さらに、ミックスされたステレオ音声データストリームは、後方互換可能（*b a c k w a r d c o m p a t i b l e*）でないことがある。すなわち、レガシーのモノ専用の端末デバイスは、ミックスされたステレオ音声データストリームを再生することができない。その結果、3D音声遠隔会議に参加するときには3D音声専用に設計された端末デバイスのみを使用することができる。3D音声遠隔会議をサポートする端末デバイスの費用及び入手可能性が低いことを考慮した場合、遠隔会議経験の向上と比較して3D音声遠隔会議をサポートするために端末デバイス及びMCUをアップグレードすることに関連するコストが高くなる可能性があるため、3D音声遠隔会議の採用は限定的になっている。

【0020】

[0025] 本開示の技法は、レガシーのモノ専用端末デバイスのサポートを通じて後方互換性も維持しながら3D音声遠隔会議を可能にするサブストリームを埋め込むことによって、有意に増大するアップグレードコストなしで、3D音声遠隔会議の採用を容易にすることができる。本開示の技法により、MCUは、遠隔会議に参加している端末デバイスによって検出された（例えば、関連する）音のモノラル表現を含む音声データストリームを受信することができる。MCUは、端末デバイスの各々のために高質のミックスされた音声データストリームを生成することができる。特定の端末デバイスのための高質のミックスされた音声データストリームは、ミックスされた音声データストリームが送信されるべき相手である特定の端末デバイス以外の端末デバイスによって検出された音のミックスのモノラル表現を含むことができる。さらに、MCUは、それらのその他の端末デバイスによって検出された音の低質の表現を高質のミックスされた音声データストリーム内に埋め込むための1つの形態のステガノグラフィ（共通して“ウォーターマーキング”と呼ぶことができる）を実行することができる。説明を簡単にする目的として、端末デバイスによって検出された音の低質の表現は、ここではサブストリームと呼ぶことができる。幾つかの例においては、MCUは、高質のミックスされた音声データストリームのビットレートを増大させずに高質のミックスされた音声データストリーム内にサブストリームを埋め込むことができる。ステガノグラフィを実行することは、典型的には、人間の視覚系又は聴覚系の限界に起因して、オリジナルのデータのサイズが増大しないようにさらには見る人、この場合は聴く人、によって知覚されるオリジナルのデータに影響を与えないように追加のデータを継ぎ目なく“隠蔽する”ために所定の量のデータにおける冗長なデータとみなすことができるデータを修正することを含む。

【0021】

[0026] MCUは、高質のミックスされた音声データストリームを遠隔会議に参加している端末デバイスに出力する。遠隔会議に参加している端末デバイスがマルチチャネル3D音声再生（例えば、ステレオ音声再生）用に構成されているときには、端末デバイスは、ステガノグラフィ方式で埋め込まれたサブストリームを高質のミックスされた音声データストリームから抽出するための逆ステガノグラフィプロセスを実行することができる。端末デバイスは、抽出されたサブストリーム及び高質のミックスされた音声データストリームに基づいて空間化された（“三次元”又は“3D”と呼ぶことができる）マルチチャネル（例えば、ステレオ）を生成することができる。端末デバイスは、空間化されたマルチチャネルデータに基づいてマルチチャネル（例えば、ステレオ）音を生成することができる。

【0022】

[0027] 遠隔会議に参加している端末デバイスがマルチチャネル再生用に構成されていないときには、端末デバイスは、ミックスされた音声データストリームからサブストリームを抽出せずにミックスされた音声データストリームに基づいてモノ音を生成することができる。サブストリームは、ステガノグラフィ方式でミックスされた音声データストリーム内に埋め込まれているため、マルチチャネル音声再生をサポートしていない端末デバイスのユーザは、サブストリームがミックスされた音声データストリーム内に埋め込まれているということを決定することができない。繰り返すと、サブストリームは、聴く人の観点からのミックスされた音声データストリームの完全性を損なわずにステガノグラフィ

10

20

30

40

50

を用いてマルチチャネル音声再生をサポートしない端末デバイスから“隠蔽”される。幾つかの例においては、3D音声は、タンデムフリーオペレーション(TFO) / トランスコーダフリーオペレーション(TrFO)ネットワークを要求することがあり、さもないと、品質がレガシーのモノに低下するおそれがある。

【0023】

[0028] このように、MCUは、マルチチャネル音声再生が可能であり及びマルチチャネル音声再生をサポートしていない端末デバイスと後方互換性がある音声データストリームを生成することができる。換言すると、ミックスされた音声データストリームは、モノラル再生用に構成され及びマルチチャネル(例えば、ステレオ)再生用に構成されていない端末デバイスと互換性がある。さらに、本開示の技法は、ミックスされた音声データストリームが複数のソースからの音を同時並行して含むときでも音に関連するソース(例えば、端末デバイス)を端末デバイスが決定するのを可能にすることができる。例えば、第1の端末デバイスに関連する音は、第2の端末デバイスに関連する話声と同時並行する話声を含むことができる。

10

【0024】

[0029] 添付される図面は、例を示す。添付される図面において参照数字によって示される要素は、以下の説明において同じ参照数字によって示される要素に対応する。本開示においては、序数語(例えば、“第1の”、“第2の”、“第3の”、等)で始まる名前を有する要素は、それらの要素が特定の順序を有することは必ずしも意味しない。むしろ、該序数語は、同じ又は同様のタイプの異なる要素を単に指し示すにすぎない。

20

【0025】

[0030] 図1は、本開示において説明される技法を実行することができる遠隔会議システム例10を示したブロック図である。図1において示されるように、遠隔会議システム10は、マルチパーティ制御ユニット(MCU)12を含む。さらに、会議システム10は、複数の端末デバイス14A乃至14C(総称して“端末デバイス14”と呼ばれる)を含む。図1の例は、例示するのを簡単にすることを目的として3つの端末デバイスを示しているにすぎない。しかしながら、その他の遠隔会議システム例は、4つ以上の端末デバイスを含むことができ、本開示の技法は、3つ以上の端末デバイスを含む遠隔会議システムによって実行することができる。

【0026】

30

[0031] MCU12は、様々なタイプのデバイスを備えることができる。例えば、MCU12は、専門化されたMCUデバイス、サーバコンピュータ、パソコン、又は他のタイプのデバイスを備えることができる。同様に、端末デバイスは、様々なタイプのデバイスを備えることができる。例えば、1つ以上の端末デバイス14は、携帯電話又はランドライン電話、デスクトップコンピュータ、モバイルコンピューティングデバイス、ノートブック(例えば、ラップトップ)コンピュータ、タブレットコンピュータ、テレビセットトップボックス、ビデオゲームコンソール、車内搭載コンピュータ、専門化された音声及び/又は映像会議装置、又は遠隔会議用に構成されたその他のタイプのデバイスを備えることができる。幾つかの例においては、MCU12は、遠隔会議のホストであるサービスプロバイダのネットワーク内に物理的に配置される。端末デバイス14は、様々な位置、例えば、ユーザの事務所、家庭、又は車両内、に存在することができる。

40

【0027】

[0032] MCU12は、端末デバイス14の各々と通信することができる。様々な例において、MCU12は、様々な方法で端末デバイス14とデータを通信する。例えば、MCU12は、通信ネットワーク、例えば、インターネット、ローカルエリアネットワーク(LAN)、プレーンオールドテレフォンシステム(Plain Old Telephone System)(POTS)ネットワーク、セララー又はモバイルネットワーク、セラーデータネットワーク、又はそれらの組み合わせ、を介して端末デバイス14のうちの1つ以上とデータを通信することができる。幾つかの例においては、MCU12は、有線及び/又は無線通信媒体を介して端末デバイス14と通信することができる。

50

【0028】

[0033] 1人以上のユーザが、遠隔会議に参加するために端末デバイス14の各々を使用することができる。遠隔会議中には、参加ユーザは、言葉で互いに会話することができる。例えば、3人のユーザA、B、及びCは、端末デバイス14A、14B、及び14Cをそれぞれ用いて遠隔会議に参加することができる。この例では、端末デバイス14Aは、音声データの形態でユーザAによって生成された音、例えば、話声、をキャプチャすることができ、及び、端末デバイス14B及び14Cによってキャプチャされた音、例えば、話声、を再生することができる。端末デバイス14Bは、ユーザBによって生成された音をキャプチャすることができ及び端末デバイス14A及び14Cによってキャプチャされた音を再生することができる。端末デバイス14Cは、ユーザCによって生成された音をキャプチャすることができ及び端末デバイス14A及び14Bによってキャプチャされた音を再生することができる。10

【0029】

[0034] 端末デバイス14は、端末デバイス14によってキャプチャされた音に基づいて音声データストリームを生成することができる。例えば、端末デバイス14は、端末デバイス14によって検出及びキャプチャされた音のモノラル表現を含む音声データストリームを生成することができる。端末デバイス14は、音声データストリームをMCU12に送信することができる。図1の例においては、矢印16A、16B、及び16Cは、端末デバイス14A、14B、及び14CからそれぞれMCU12に送信された音声データストリームを示す。音声データストリームは、経時で提供される音声データの流れである。例えば、端末デバイス14は、端末デバイス14が同じく一般的には話声の形態である音を検出、キャプチャ及び符号化するのとほぼリアルタイムでMCU12に音声データストリームを送信することができる。端末デバイス14は、各々、話声を符号化しそれによって音声データストリームを生成するための声又は話声符号器（共通して“ボコーダ”と呼ばれる）を含むことができる。ボコーダは、話声を符号化するように特化された符号化プロセスを特徴とする符号器を代表することができる。20

【0030】

[0035] MCU12は、端末デバイス14によって送信された音声データストリームを受信する。MCU12が音声データストリームを受信するのに応じて、MCU12は、複数のミックスされた音声データストリームを生成することができる。MCU12は、異なるミックスされた音声データストリームを異なる端末デバイス14に送信することができる。図1の例においては、MCU12は、第1のミックスされた音声データストリームを端末デバイス14Aに、第2のミックスされた音声データストリームを端末デバイス14Bに、及び第3のミックスされた音声データストリームを端末デバイス14Cに送信することができる。図1の例においては、矢印18A、18B、及び18Cは、端末デバイス14A、14B、及び14Cにそれぞれ送信されたミックスされた音声データストリームを示す。30

【0031】

[0036] 特定の端末デバイスのためのミックスされた音声データストリームは、遠隔会議に参加しているその他の端末デバイスによってキャプチャされた音のミックスを含むことができる。幾つかの例においては、MCU12は、その他の端末デバイスから受信された音声データストリーム内で示される対応する（例えば、同時並行する）音声サンプルを加えることによってその他の端末デバイスによってキャプチャされた音をミックスすることができる。例えば、MCU12は、第1の音声データストリームの音声サンプルを第2の音声データストリームの対応する音声サンプルに加えることによってその他の端末デバイスによって検出された音をミックスする。40

【0032】

[0037] ミックスされた音声データストリームを生成することに加えて、MCU12は、本開示の技法により、複数のサブストリームを生成することができる。それらのサブストリームの各々は、端末デバイス14のうちの異なる1つによってキャプチャされた音

1020304050

の表現を含むことができる。特定の端末デバイスによってキャプチャされた音の表現を含むサブストリームは、ここでは、その特定の端末デバイスに関連するサブストリームと呼ぶことができる。

【0033】

[0038] 幾つかの例においては、サブストリームは、MCU12によって受信される対応する音声データストリームよりも低いビットレートを有する。例えば、端末デバイス14Aに関連するサブストリームは、端末デバイス14AによってMCU12に送信される音声データストリームよりも低いビットレートを有することができる。同様に、端末デバイス14Bに関連するサブストリームは、端末デバイス14BによってMCU12に送信される音声データストリームよりも低いビットレートを有することができる。幾つかの例においては、サブストリームは、非常に低いビットレート、例えば、1 - 2 kb / 秒、を含むことができる。さらに、幾つかの例においては、サブストリームは、端末デバイス14によって送信される音声データストリーム内の対応する音声サンプルのビット深度よりも浅いビット深度を有するコーディングされた音声サンプルを含むことができる。10

【0034】

[0039] さらに、特定の端末デバイスへの送信のためにミックスされた音声データストリームを生成するために、MCU12は、本開示の技法により、その他の端末デバイスに関連するサブストリームをミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことによってミックスされた音声データストリームを修正することができる。MCU12は、その結果得られた修正されたミックスされた音声データストリームを端末デバイス14のうちの特定の1つに送信することができる。幾つかの例においては、修正されたミックスされた音声データに関連する音声歪みの量と修正されないミックスされた音声データに関連する音声歪みの量との間の差分は、検出スレショルドよりも低い。検出スレショルドは、典型的な人間の聴覚系によって気付くことができる歪みの量であることができる。さらに、幾つかの例においては、修正されたミックスされた音声データのビットレートは、サブストリームを埋め込む前のミックスされた音声データのビットレートよりも大きくない。MCU12は、標準的なコーディングされた話声ストリーム内の有意な量のデータを隠すことができる。例えば、MCU12は、相対的にほとんど劣化することなしに最大で2 kb / 秒を12.2 kb / 秒のエンハンストフルレート(EFR)ビットストリーム内に埋め込むことができる。20

【0035】

[0040] MCU12は、様々な方法でサブストリームをミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことができる。例えば、ミックスされた音声データストリームは、音を一連の符号化された音声サンプルとして表現することができる。それらの音声サンプルの各々は、固定数のビットから成ることができる。聴く人は、音声サンプルの最下位ビット(LSB)の変化が原因で生じた音の変化に気付くことができない。従って、MCU12は、ミックスされた音声データストリーム内の音声サンプルのLSBがサブストリームのデータを表現するようにそれらの音声サンプルのLSBのうちの1つ以上を変化させることによってミックスされた音声データストリーム内にステガノグラフィ方式でサブストリームを埋め込むことができる。すなわち、MCU12は、第1及び第2のサブストリームを示すためにミックスされたモノ音声データの音声サンプルの最下位ビットを修正することができる。30

【0036】

[0041] 他の例においては、ミックスされた音声データストリームのサンプルレートが十分に高い場合は、聴く人は、ミックスされた音声データストリーム内の音声サンプルをサブストリームからのサンプルと定期的に置き換えることが原因で生じる音の変化に気が付くことができない。従って、この例においては、MCU12は、ミックスされた音声データストリーム内の音声サンプルをサブストリームの音声サンプル又はその他のデータと定期的に置き換えることによってミックスされた音声データストリーム内にサブストリームをステガノグラフィ方式で埋め込むことができる。40

【0037】

[0042] 他の例においては、MCU12は、FCBトラック当たりの固定された数のビットを隠すことによって代数符号励振線形予測（ACELP）コーダ（例えば、適応型マルチレート狭帯域（“AMR-NB”）の固定されたコードブック（FCB）上のビットを隠すことによってミックスされたサブストリーム内にサブストリームをステガノグラフィ方式で埋め込むことができる。ビットは、許容されるパルス組み合わせの数を制限することによって隠される。AMR-NBの場合は、トラック当たり2つのパルスが存在し、1つのアプローチ法は、所定のトラック上の2つのパルス位置の排他的OR（XOR）が送信すべきウォーターマークと等しくなるようにパルス位置を制約することを含む。このようにしてトラック当たり1つ又は2つのビットを送信することができる。

10

【0038】

[0043] 他の例においては、ウォーターマークは、適応型である。前例において説明されるようにパルストラック当たり固定された数のビット（例えば、1つ又は2つ）を埋め込む代わりに、MCU12は、どのドラックが知覚的に最も重要であるかを決定する。幾つかの例においては、MCU12は、符号器及び復号器の両方において既に存在する情報を用いてこれを達成することができ、従って、どのドラックが知覚的に最も重要であるかを示す情報を追加で又は別個に送信する必要がない。一構成においては、最も重要なトラックをウォーターマークから保護するために長期予測（LTP）貢献（contribution）を使用することができる。例えば、LTP貢献は、通常は、主要なピッチパルスにおいてクリアなピークを呈し、符号器及び復号器の両方において既に利用可能である。この例では、AMR-NBを使用することができる。代替として、この例では、eAMRを使用することができる。eAMRは、狭帯域ビットストリーム内に隠されている広帯域情報の“薄”層をトランスポートする能力を有することができる。ここにおける引用によってその内容全体がここに組み入れられている米国特許出願13/275,997では、ウォーターマーキングにおけるeAMRの使用を記述している。

20

【0039】

[0044] このようにして、MCU12は、第1の音声データストリームを受信することができる。第1の音声データストリームは、遠隔会議に参加している第1の端末デバイス（例えば、端末デバイス14A）によってキャプチャされた（すなわち、第1の端末デバイスに関連する）音を表現することができる。さらに、MCU12は、第2の音声データストリームを受信することができる。第2の音声データストリームは、遠隔会議に参加している第2の端末デバイス（例えば、端末デバイス14B）によってキャプチャされた（すなわち、第2の端末デバイスに関連する）音を表現することができる。MCU12は、第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成することができる。さらに、MCU12は、少なくとも部分的に第1のサブストリーム及び第2のサブストリームをミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データストリームを生成して修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようにすることができます。第1のサブストリームは、第1の端末デバイスに関連する音の表現を含む。第2のサブストリームは、第2の端末デバイスに関連する音の表現を含む。MCU12は、修正されたミックスされた音声データストリームを遠隔会議に参加している第3の端末デバイス（例えば、端末デバイス14C）に出力することができる。

30

【0040】

[0045] 端末デバイス14Cは、ミックスされた音声データストリームをMCU12から受信することができる。端末デバイス14Cがマルチストリーム音声再生用に構成されていない場合は、端末デバイス14Cは、ミックスされた音声データストリームを復号し、モノラルチャネルを用いてミックスされた音声データストリームを再生することができる。

40

【0041】

50

[0046] しかしながら、端末デバイス14Cがマルチチャネル音声再生用に構成されている場合は、端末デバイス14Cは、ミックスされた音声データストリームからサブストリームを抽出するための逆ステガノグラフィプロセスを実行することができる。端末デバイス14Cは、ミックスされた音声データストリーム及び抽出されたサブストリームを復号することができる。さらに、端末デバイス14Cは、サブストリームに基づいて、端末デバイス14C以外の端末デバイスに関連する（例えば、端末デバイス14C以外の端末デバイスによって最初に検出された）ミックスされた音声データストリーム内の音を決定することができる。例えば、端末デバイス14Cは、第1のサブストリームに基づいて、第1の端末デバイス（例えば、端末デバイス14A）に関連するミックスされた音声データストリームの部分を決定することができ、及び、第2のサブストリームに基づいて、第2の端末デバイス（例えば、端末デバイス14B）に関連するミックスされた音声データストリームの部分を決定することができる。幾つかの例においては、第1の端末デバイス及び第2の端末デバイスに関連するミックスされた音声データストリームの部分は、例えば、端末デバイス14A及び14Bを使用する参加者が同時に（又は互いに同時並行して）話したときに、重なり合う。

【0042】

[0047] 端末デバイス14Cは、その他の端末デバイスに関連する音に基づいて、頭部伝達関数（HRTF）をミックスされた音声データストリームのコピーに適用することができる。HRTFは、耳が空間内のある地点からの音をどのように受け取るかの特徴を表す応答である。端末デバイス14Cは、2つの耳が空間内の特定の地点から来るようと思われるバイオーラル音を合成するために一対のHRTFを使用することができる。例えば、端末デバイス14Cは、左の音声チャネルが端末デバイス14Cのユーザの左側の空間内のある地点から来るように思われるバイオーラル音を合成するためにミックスされた音声データストリームのコピーにHRTFを適用することができる。同様に、端末デバイス14Cは、右の音声チャネルが端末デバイス14Cのユーザの右側の空間内のある地点から来るように思われるバイオーラル音を合成するためにミックスされた音声データストリームのコピーにHRTFを適用することができる。当業においては様々なHRTFが知られている。

【0043】

[0048] 端末デバイス14Cは、HRTFを適用することによって生成されたミックスされた音声データストリームのコピーに基づいてマルチチャネル音声データストリームを生成することができる。例えば、端末デバイス14Cは、空間化されたステレオ音声データストリームを生成することができ、従って、空間化されたステレオ音声データストリームの左チャネル及び右チャネルは、HRTFを適用することによって生成されたミックスされた音声データストリームのコピーに基づくことができる。このようにして、端末デバイス14Cは、空間化されたステレオ音声データストリームを生成することができ、従って、第1の端末デバイス（例えば、端末デバイス14A）に関連する音は、端末デバイス14Cに関連する聴く人の左側の空間内のある地点から来ると知覚され、第2の端末デバイス（例えば、端末デバイス14B）に関連する音は、端末デバイス14Cに関連する聴く人の右側の空間内のある地点から来ると知覚される。端末デバイス14C（又は他のデバイス）は、空間化されたミックスされたステレオ音声データストリームをアナログ信号に変換することができる。スピーカーは、アナログ信号を可聴音に変換することができる。

【0044】

[0049] このようにして、端末デバイス14Cは、ミックスされた音声データストリームを受信することができる。端末デバイス14Cは、ミックスされた音声データストリームから、第1のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することができる。第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する。さらに、端末デバイス14Cは、ミックスされた音声データストリームから、第2のサブストリームを抽出するための逆ステガノグラフィプロセスを実行するこ

10

20

30

40

50

できる。第2のサブストリームは、遠隔会議に参加している第2の端末デバイスに関連することができる。さらに、端末デバイス14Cは、ミックスされた音声データストリーム、第1のサブストリーム及び第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成することができる。端末デバイス14Cに関する上の説明は、該当する変更を行うことによって、端末デバイス14のうちの他の1つに対して適用することができる。

【0045】

[0050] さらに、このようにして、MCU12は、マルチチャネル音声再生が可能であり及びマルチチャネル音声再生をサポートしていない端末デバイスと後方互換性がある音声データストリームを生成することができる。換言すると、ミックスされた音声データストリームは、モノラル再生用に構成され、マルチチャネル（例えば、ステレオ）再生用に構成されていない端末デバイスと互換可能である。さらに、本開示の技法は、ミックスされた音声データストリームが複数のソースからの音を同時並行して含むときでも音に関連するソース（例えば、端末デバイス）を端末デバイス14が決定するのを可能にすることができる。例えば、第1の端末デバイスに関連する音は、第2の端末デバイスに関連する話声と同時並行する話声を含むことができる。従って、サブストリームを埋め込むためにステガノグラフィを使用することによって、それらの技法は、ミックスされた音声データストリームのマルチチャネル再生及びミックスされた音声データストリームのモノラル再生の両方を可能にすることができます。

【0046】

[0051] 本開示においては音声のみの遠隔会議（映像又はその他の媒体が伴わない遠隔会議を意味することができる）について説明される一方で、それらの技法は、映像会議（映像を伴う遠隔会議を意味することができる）、オンライン共同会議（オンラインでのプレゼンテーション又は共同での文書の編集及び閲覧を伴う遠隔会議を意味することができる）又は3人以上の参加者からの音声データを含むその他のタイプの会議に関して実行することができる。従って、本開示において説明される技法は、本開示において示される例に限定されないようにすべきである。

【0047】

[0052] 上述されるように、MCU12は、端末デバイス14によって検出及びキャプチャされた音のモノラル表現を受信することができる。その他の例においては、MCU12は、端末デバイス14によって検出及びキャプチャされた音のステレオ表現を受信することができる。該例においては、MCU12は、上述されるようにステレオ表現をモノラル表現に変換すること及びその結果得られたモノラル表現を処理することができる。

【0048】

[0053] 図2は、本開示において説明される技法の様々な態様を実行するMCU例12を示したブロック図である。図2の例においては、MCU12は、音声復号器50Aと、音声復号器50Bと、加算器52と、高質(HQ)符号器54と、低質(LQ)符号器56Aと、LQ符号器56Bと、ウォーターマークユニット58と、を備える。

【0049】

[0054] 音声復号器50Aは、端末デバイスから音声データストリームを受信し及び復号する。例えば、音声復号器50Aは、図1において矢印16Aによって示されるように、端末デバイス14Aから音声データストリームを受信し及び復号することができる。音声復号器50BBは、異なる端末デバイスから音声データストリームを受信し及び復号する。例えば、音声復号器50Bは、図1において矢印16Bによって示されるように、端末デバイス14Bから音声データストリームを受信し及び復号することができる。音声復号器50A及び50Bは、音声データストリームを復号するために様々な音声及び/又は話声コーディングプロセスを使用することができる。例えば、音声復号器50A及び/又は50Bは、符号励起線形予測(CELP)コーディング、コーデック2コーディング、適応型マルチレートワイドバンド(AMR-WB)コーディング、Speechコーディング、又はその他の形態の声又は音声データの圧縮を用いて音声データストリームを復号

10

20

30

40

50

することができる。

【0050】

[0055] 加算器52は、音声復号器50A及び50Bから音声データストリームを受信することができる。加算器52は、ミックスされた音声データストリーム60を生成することができる。ミックスされた音声データストリーム60は、音声データストリームによって表現される音のミックスのモノラル表現を含むことができる。幾つかの例においては、加算器52は、音声データストリームの音声サンプルを加えることによってミックスされた音声データストリーム60を生成することができる。

【0051】

[0056] HQ符号器54は、ミックスされた音声データストリーム60を符号化することができる。HQ符号器54は、ミックスされた音声データストリーム60を符号化することができ、従って、ミックスされた音声データストリームによって表現される音は、(比較的)高い忠実度(品質を意味することができる)を保持することができる。さらには、幾つかの例においては、HQ符号器54は、ミックスされた音声データストリーム60にデータ圧縮を適用する。例えば、HQ符号器54は、CELPコーディング、コーデック2コーディング、適応型マルチレートワイドバンド(AMR-WB)コーディング、Sapeexコーディング、又はその他の形態の声又は音声データの圧縮を用いてミックスされた音声データストリーム60を符号化することができる。

10

【0052】

[0057] LQ符号器56Aは、音声復号器50Aから音声データストリームを受信することができる。LQ符号器56Aは、サブストリームを生成するために音声データストリームを処理することができる。LQ符号器56Aは、様々な方法でサブストリームを生成することができる。例えば、LQ符号器56Aは、音声データストリームの(比較的)低質のバージョンであるサブストリームを生成することができる。例えば、この例においては、LQ符号器56Aは、部分的に、音声サンプルのビット深度を浅くするために音声データストリーム内の音声サンプルの値を量子化することによってサブストリームを生成することができる。他の例においては、LQ符号器56Aは、音声データストリームのサンプルレートを低下させることができる。幾つかの例においては、LQ符号器56Aは、音声データストリームの音声サンプルをデシメートすることによってサンプルレートを低下させることができます。さらに、幾つかの例においては、LQ符号器56Aは、音声データストリームにデータ圧縮、例えば、CELPコーディング、コーデック2コーディング、又はその他の形態の声又は音声データの圧縮、を適用することができる。他の例においては、サブストリームは、音声データストリームの利得のみを示すことができる。幾つかの例においては、LQ符号器56Aは、HQ符号器54によって生成される高質のミックスのサブストリームと異なる形でサブストリームのパラメータをコーディングすることができる。LQ符号器56Bは、音声復号器50Bから音声データストリームを受信することができ及びLQ符号器56Aと同様の方法で音声データストリームを処理することができる。

20

【0053】

[0058] 図2の例においては、ウォーターマークユニット58は、LQ符号器56A及び56Bからサブストリームを受信する。さらに、ウォーターマークユニット58は、高質のミックスされた音声データストリームをHQ符号器54から受信する。さらに、ウォーターマークユニット58は、高質のミックスされた音声データストリームをHQ符号器54から受信する。ウォーターマークユニット58は、サブストリームをミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むためにミックスされた音声データストリームを修正することができる。ウォーターマークユニット58は、サブストリームをミックスされた音声データストリーム内に埋め込むために上述されるステガノグラフィ技法のうちのいずれか、又はその他を使用することができる。ウォーターマークユニット58は、修正されたミックスされた音声データストリームを出力することができる。ウォーターマーキングは、サブストリームの個々の記述を送信する1つの方法であるに

30

40

50

すぎない。その他の例においては、サブストリームは、代替のメディアストリームで送信することができる。

【0054】

[0059] 図2の例では示されていないが、MCU12は、1つ以上の追加の端末デバイスから音声データストリームを受信することができる。例えば、MCU12は、3つ以上の端末デバイスから音声データストリームを受信することができる。この例においては、MCU12は、追加の端末デバイスに関連するサブストリームを生成するために追加の端末デバイスからのモノ音声データストリームを符号化することができる。さらに、MCU12は、追加の端末デバイスからのモノ音声データストリームをミックスされた音声データストリーム内に加え、その結果得られたミックスされた音声データストリームを符号化することができる。MCU12は、追加の端末デバイスに関連するサブストリームを符号化されたミックスされた音声データ内に埋め込むためにステガノグラフィ技法を使用することができる。10

【0055】

[0060] 図2は、単なる例であり、MCU12は、図2において示される方法以外の方法で実装することができる。例えば、MCU12は、その他の例においては、より多くの、より少ない、又は異なるコンポーネントを含むことができる。図2の例においては、MCU12は、LQ符号器56Aと、LQ符号器56Bと、を含む。その他の例においては、端末デバイス14は、LQ符号器56A及びLQ符号器56Bの機能を提供することができる。換言すると、端末デバイス14は、サブストリームを生成することができる。該例においては、端末デバイス14は、通常の品質の音声データストリームをMCU12に送信することに加えてサブストリームをMCU12に送信することができる。幾つかの例においては、端末デバイス14は、通常の品質の音声データストリーム内にステガノグラフィ方式でサブストリームを埋め込むことができる。20

【0056】

[0061] 図3は、本開示において説明される技法の様々な態様を実行する際の端末デバイス14のうちの1つの例を示したブロック図である。端末デバイス14は、図1の例において示される端末デバイス14A、14B、又は14Cのうちのいずれかを代表することができる。図3は単なる一例であり、端末デバイス14は、図3以外の方法で実装することができる。例えば、端末デバイス14は、その他の例においては、より多くの、より少ない、又は異なるコンポーネントを含むことができる。30

【0057】

[0062] 図3の例においては、端末デバイス14は、マイク100と、音声符号器102と、ネットワークインタフェース104と、音声復号器106と、を備える。端末デバイス14は、一組のスピーカー108A及び108B（総称して“スピーカー108”）も含む。マイク100は、端末デバイス14の近傍の音をキャプチャする。例えば、マイク100は、ユーザの声又は話声の音を検出及びキャプチャすることができる。マイク100は、検出された音に基づいて電気信号を生成する。音声符号器102は、マイク100によって生成された電気信号を音声データストリーム110に変換する。音声データストリーム110は、図1において矢印16によって示される音声データストリームのうちの1つであることができる。音声データストリーム110は、検出された音のモノラル表現を含むことができる。幾つかの例においては、音声符号器102は、音声データストリームのピットレートを低下させるために音声データストリーム110において様々なタイプのデータ圧縮を行うことができる。ネットワークインタフェース104は、音声データストリームをMCU、例えば、MCU12、に送信することができる（図1）。40

【0058】

[0063] 音声データストリーム110をMCU12に送信することに加えて、ネットワークインタフェース104は、ミックスされた音声データストリーム112をMCU12から受信することができる。ミックスされた音声データストリーム112は、図1において矢印18によって示されるミックスされた音声データストリームのうちの1つである50

ことができる。ネットワークインターフェース 104 は、ミックスされた音声データストリームを音声復号器 106 に提供することができる。音声復号器 106 は、音を再生するためにスピーカー 108 を駆動する信号を生成するために本開示において説明される技法によりミックスされた音声データストリームを処理することができる。

【0059】

[00764] 図 4 は、本開示において説明される技法の様々な態様を実行する際の音声復号器 106 の例を示すブロック図である。図 4 は、単なる例であり、音声復号器 106 は、図 4 において示される方法以外のそれで実装することができる。例えば、音声復号器 106 は、その他の例においては、より多い、より少ない、又は異なるコンポーネントを含むことができる。図 4 の例においては、音声復号器 106 は、HQ 復号器 150 と、LQ 復号器 152A と、LQ 復号器 152B と、ソース分離ユニット 154A と、ソース分離ユニット 154B と、HRTF ユニット 156A と、HRTF ユニット 156B と、ステレオ出力ユニット 158 と、を含む。
10

【0060】

[0065] HQ 復号器 150、LQ 復号器 152A、及び LQ 復号器 152B は、ミックスされた音声データストリームを受信する。HQ 復号器 150 は、ミックスされた音声データストリームに適用されるデータ圧縮を逆にするためにミックスされた音声データを復号することができる。ミックスされた音声データストリームを復号することによって、HQ 復号器 150 は、一連の音声サンプルを備える復号されたミックスされた音声データストリームを生成することができる。
20

【0061】

[0066] LQ 復号器 152A は、ミックスされた音声データストリームからサブストリームを抽出するための逆ステガノグラフィプロセスを実行することができる。例えば、より低い品質のサブストリームが音声サンプルの最下位ビットにおいてミックスされた音声データストリーム内にステガノグラフィ方式で埋め込まれる場合は、LQ 復号器 152A は、ミックスされた音声データストリームを復号することができる。この例においては、LQ 復号器 152A は、サブストリームを抽出するために音声サンプルの最下位ビットを分離することができる。幾つかの例においては、LQ 復号器 152A によって抽出されたサブストリームは、遠隔会議に参加している端末デバイスのうちの 1 つにおいて検出された音のモノラル表現を含む。幾つかの例においては、LQ 復号器 152A は、サブストリームに適用されたデータ圧縮を逆にするためにサブストリームを復号する。例えば、LQ 復号器 152A は、サブストリームを一連の音声サンプルに変換することができる。LQ 復号器 152B は、第 2 のサブストリームを抽出するために同様の動作を行うことができる。幾つかの例においては、LQ 復号器 152B によって抽出されたサブストリームは、遠隔会議に参加している端末デバイスのうちの他の 1 つにおいて検出された音のモノラル表現を含む。
30

【0062】

[0067] ミックスされた音声データストリームは、少なくとも第 1 の端末デバイス及び第 2 の端末デバイスに関連する音のミックスのモノラル表現を含む。例えば、ミックスされた音声データストリームによって表現される波形は、少なくとも第 1 の端末デバイス及び第 2 の端末デバイスによって検出された音によって表現される波形の和であることができる。ソース分離ユニット 154A は、復号されたミックスされた音声データストリームを HQ 復号器 150 から、サブストリームを LQ 復号器 152A から、及びサブストリームを LQ 復号器 152B から受信することができる。ソース分離ユニット 154A は、ミックスされた音声データストリーム内において、第 1 の端末デバイスに関連する音を決定するために復号されたミックスされた音声データストリーム及びサブストリームを解析することができる。
40

【0063】

[0068] 例えば、サブストリームは、第 1 の端末デバイスによって検出された音に関連する利得を示すことができる。この例においては、ソース分離ユニット 154B は、第
50

1 の端末デバイスによって検出された音に関連する利得に少なくとも部分的に基づいて、ミックスされた音声データストリームによって表現される音の幾つかの部分が第 1 の端末デバイスによって検出された音に起因すると決定することができる。他の例においては、サブストリームは、第 1 の端末デバイスによって検出された音の波形を示すことができる。この例においては、ソース分離ユニット 154A は、ミックスされた音声データストリームによって示される波形に対して、サブストリームによって示される波形に基づいてバンドパスフィルタを適用することができる。バンドパスフィルタは、ミックスされた音声データストリームの波形内で第 1 の端末デバイスによって検出された音の波形の周波数に対応しない周波数を抑止することができる。他の例においては、ソース分離ユニット 154A は、復号されたミックスされた音声データストリームを HQ 復号器 150 から受信し、サブストリームを LQ 復号器 152 から受信することに加えて、サブストリームを HQ 復号器 150 から受信することができる。この例においては、ソース分離ユニット 154A は、第 1 の端末デバイスに関連する音を決定するために 3 つの音声データストリームを使用することができる。
10

【0064】

[0069] 他の例においては、ソース分離ユニット 154A は、サブストリームに基づいて及び HQ 復号器 150 からのミックスされた音声データストリームに基づかずに、第 1 の端末デバイスに関連する音を決定することができる。他の例においては、ソース分離ユニット 154A は、フレームごとに及び LQ 復号器 1562A からのサブストリームと LQ 復号器 152B からのサブストリームの正規化されたエネルギー比に少なくとも部分的に基づいて、第 1 の端末デバイスに関連する音を決定することができる。従って、この例においては、ソース分離ユニット 154A は、現在のフレームに関する値 を以下のように計算することができる。
20

【0065】

$$= \text{NRG_LQA} / (\text{NRG_LQA} + \text{NRG_LQB})$$

上記の方程式において、NRG_LQA は、現在のフレームに関する、LQ 復号器 152A からのサブストリームのエネルギーであり、NRG_LQB は、現在のフレームに関する、LQ 復号器 152B からのサブストリームのエネルギーである。次に、ソース分離ユニット 154A は、現在のフレーム内において、少なくとも部分的には、ミックスされた音声データストリームの現在のフレームの音声サンプルに を乗じることによって第 1 の端末デバイスに関連する音を決定することができる。現在のフレーム内で第 2 の端末デバイスに関連する音は、ミックスされた音声データストリームの現在のフレームの音声サンプルに (1 -) を乗じることによって決定することができる。信号のうちの 2 つではなく 3 つのすべての信号を用いてソース分離アルゴリズムを実行することによってより良い性能を達成することができる。幾つかの例においては、サブストリームの低いビットレートに起因して、サブストリームは、サブストリームを個々に復号する上で十分なデータが入ることができないことがあり、ソース分離を補助する上で十分な補助情報のみを含むことができる。他の例においては、ソース分離ユニット 154A は、この段落の 2 つの前例の組み合わせ、例えば、線形の組み合わせ、を使用することができる。他の例においては、ソース分離ユニット 154A は、周波数帯域に基づいて（すなわち、周波数帯域当たり 1 つの 率を用いて）この段落の技法例を用いて第 1 の端末デバイスに関連する音を決定することができる。その他の例においては、ソース分離ユニット 154A は、第 1 の端末デバイスに関連する音を決定するためにコームフィルタリング（comb filtering）を実行するためにピッチ情報を使用することができる。
30
40

【0066】

[0070] ソース分離ユニット 154B は、HQ 復号器 150 からの復号されたミックスされた音声データストリーム、LQ 復号器 152A からのサブストリーム、及び LQ 復号器 152B からのサブストリームに少なくとも部分的に基づいて、ミックスされた音声
50

データストリームによって表現される幾つかの音の部分が第2の端末デバイスに起因すると決定することができる。ソース分離ユニット154Bは、ソース分離ユニット154Aと同様の方法で第2の端末デバイスに起因する音を決定することができる。

【0067】

[0071] HRTFユニット156Aは、第1の端末デバイスに関連する音に基づいて、ミックスされた音声データストリームの第1のコピーに1つ以上のHRTFを適用することができる。例えば、HRTFユニット156Aは、ミックスされた音声データストリームの第1のコピー内において、遠隔会議に参加しているその他の端末デバイスに関連する音に対して第1の端末デバイスに関連する音の量を増大することができる。さらに、HRTFユニット156Aは、ミックスされた音声データストリームの第1のコピー内において、遠隔会議に参加しているその他の端末デバイスに関連する音を時間的に遅延させることができる。該量の変更及び時間的遅延は、ユーザの左右の耳によって気付かれる音の量の違い及び受信時間を模倣することができる。幾つかの例においては、HRTFユニット156Aは、ソース分離ユニット154B（又は音声復号器106のその他のソース分離ユニット）から、第2の端末デバイス（又はその他の端末デバイス）に関連する音を示すデータを受信することができる。該例において、HRTFユニット156Aは、ミックスされた音声データストリームの第1のコピーにさらなるHRTFを適用するためにこのデータを使用することができる。HRTFユニット156Bは、第2の端末デバイスに関連する音に基づいて、ミックスされた音声データストリームの第2のコピーに1つ以上のHRTFを適用することができる。HRTFユニット156Bは、同様のHRTFを上述されるそれらに適用することができる。10

【0068】

[0072] ステレオ出力ユニット158は、ミックスされた音声データストリームの第1のコピーをHRTFユニット156Aから受信することができ及びミックスされた音声データストリームの第2のコピーをHRTFユニット156Bから受信することができる。ステレオ出力ユニット158は、マルチチャネル音声データストリームの左チャネルの一部としてミックスされた音声データストリームの第1のコピーを含むステレオ音声データストリームを生成することができ及びミックスされた音声データストリームの第2のコピーをマルチチャネル音声データストリームの右チャネルの一部として含むことができる。スピーカー108A（図3）は、左チャネルのデータによって表現される音を出力することができる。スピーカー108B（図3）は、右チャネルのデータによって表現される音を出力することができる。20

【0069】

[0073] 図4の例では示されていないが、音声復号器106は、3つ以上のサブストリームを抽出して処理することができる。例えば、音声復号器106は、幾つかの例においては、遠隔会議に参加している第3の端末デバイスに関連する音の表現を含む第3のサブストリームを抽出する追加のLQ復号器を含むことができる。この例においては、追加のソース分離ユニットは、第3のサブストリームに基づいて、ミックスされた音声データストリーム内において第3の端末デバイスに起因する音を決定することができる。追加のHRTFユニットは、第3の端末デバイスに関連する音に基づいて、ミックスされた音声データストリームの第3のコピーに1つ以上のHRTFを適用することができる。この例においては、ステレオ出力ユニット158は、マルチチャネル音声データストリームの左（又は右）チャネルがミックスされた音声データストリームの第3のコピー及びミックスされた音声データストリームの第1の（又は第2の）コピーの和に基づくような形でマルチチャネル音声データストリームを生成することができる。30

【0070】

[0074] 図5は、本開示において説明される技法の様々な態様を実行する際のMCU12の動作例200を示したフローチャートである。図5の例においては、音声復号器50A及び50Bは、第1及び第2の音声データストリームを受信及び復号する（202）。加算器52は、第1及び第2の音声データストリームに基づいて、ミックスされた音声40

データストリームを生成することができる(204)。HQ符号器54は、ミックスされた音声データストリームを符号化することができる(206)。さらに、LQ符号器54Aは、第1のサブストリームを生成する(208)。第1のサブストリームは、第1の音声データストリームにおいて表現される音のより低質なモノラル表現を含むことができる。幾つかの例においては、LQ符号器56Bは、第1のサブストリームに対してデータ圧縮を行うために声符号器(ボコーダ)を使用することができる。LQ符号器56Bは、第2のサブストリームを生成する(210)。第2のサブストリームは、第2の音声データストリームにおいて表現される音のより低質なモノラル表現を含むことができる。幾つかの例においては、LQ符号器56Bは、第2のサブストリームに対してデータ圧縮を行うためにボコーダを使用することができる。

10

【0071】

[0075]ウォーターマークユニット58は、第1及び第2のサブストリームをミックスされた音声データストリーム内に埋め込むためにステガノグラフィを実行することができる(212)。例えば、ウォーターマークユニット58は、第1及び第2の符号化されたサブストリームを符号化されたミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことができる。MCU12は、ミックスされた音声データストリームの結果的に得られたバージョンを出力することができる(214)。

【0072】

[0076]図6は、本開示において説明される技法の様々な態様を実行する際の音声復号器106の動作例250を示したフローチャートである。図5及び6のフローチャートは例である。その他の例においては、MCU12及び音声復号器106は、図5及び6の例において示される行動よりも多い、少ない、又は異なるそれらを含む動作を実行することができる。

20

【0073】

[0077]図6の例においては、音声復号器106は、ミックスされた音声データストリームを受信する(252)。さらに、音声復号器106は、ミックスされた音声データストリームから、第1のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することができる(254)。第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する。例えば、第1のサブストリームは、第1の端末デバイスによって検出された音のモノラル表現を含むことができる。音声復号器106は、ミックスされた音声データストリームから、第2のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することもできる(256)。第2のサブストリームは、遠隔会議に参加している第2の端末デバイスに関連する。例えば、第2のサブストリームは、第2の端末デバイスによって検出された音のモノラル表現を含むことができる。

30

【0074】

[0078]図6の例では、音声復号器106は、ミックスされた音声データストリーム、第1のサブストリーム及び第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成する(258)。音声復号器106は、2つ以上のスピーカーでの再生のためにマルチチャネル音声データストリームを出力することができる(260)。

40

【0075】

[0079]図6の例において、音声復号器106は、ミックスされた音声データストリームから2つのサブストリームを抽出する。その他の例においては、音声復号器106は、ミックスされた音声データストリームから3つ以上のサブストリームを抽出し、ミックスされた音声データストリーム及び抽出されたサブストリームの各々に基づいてマルチチャネル音声データストリームを生成することが評価されるべきである。

【0076】

[0080]1つ以上の例において、説明される機能は、ハードウェア、ソフトウェア、ファームウェア、又はそれらのあらゆる組み合わせにおいて実装することができる。ソフトウェアにおいて実装される場合は、それらの機能は、コンピュータによって読み取り可

50

能な媒体において1つ以上の命令又はコードとして格納又は送信すること及びハードウェアに基づく処理ユニットによって実行することができる。コンピュータによって読み取り可能な媒体は、コンピュータによって読み取り可能な記憶媒体を含むことができ、それは、有形な媒体、例えば、データ記憶媒体、又は、例えば、通信プロトコルにより、1つの場所から他へのコンピュータプログラムの転送を容易にするあらゆる媒体を含む通信媒体、に対応する。このように、コンピュータによって読み取り可能な媒体は、概して、(1)非一時的である有形なコンピュータによって読み取り可能な記憶媒体又は(2)通信媒体、例えば、信号又は搬送波、に対応することができる。データ記憶媒体は、本開示において説明される技法の実装のために命令、コード及び/又はデータ構造を取り出すために1つ以上のコンピュータ又は1つ以上のプロセッサによってアクセスすることができるあらゆる利用可能な媒体であることができる。コンピュータプログラム製品は、コンピュータによって読み取り可能な媒体を含むことができる。10

【0077】

[0081] 一例により、及び制限することなしに、該コンピュータによって読み取り可能な記憶媒体は、希望されるプログラムコードを命令又はデータ構造の形態で格納するために使用することができ及びコンピュータによってアクセス可能であるRAM、ROM、EEPROM、CD-ROM又はその他の光学ディスク記憶装置、磁気ディスク記憶装置、又はその他の磁気記憶デバイス、フラッシュメモリ、又はその他のいずれかの媒体を備えることができる。さらに、どのような接続も、コンピュータによって読み取り可能な媒体であると適切に呼ばれる。例えば、命令が、同軸ケーブル、光ファイバケーブル、より対線、デジタル加入者ライン(DSL)、又は無線技術、例えば、赤外線、無線、及びマイクロ波、を用いてウェブサイト、サーバ、又はその他の遠隔ソースから送信される場合は、該同軸ケーブル、光ファイバケーブル、より対線、DSL、又は無線技術、例えば赤外線、無線、及びマイクロ波、は、媒体の定義の中に含まれる。しかしながら、コンピュータによって読み取り可能な記憶媒体およびデータ記憶媒体は、コネクション、搬送波、信号、又はその他の遷移媒体は含まず、代わりに、非一時的な、有形の記憶媒体を対象とすることが理解されるべきである。ここにおいて用いられるときのディスク(disk及びdisc)は、コンパクトディスク(CD)(disc)と、レザーディスク(登録商標)(disc)と、光ディスク(disc)と、デジタルバーサタイルディスク(DVD)(disc)と、フロッピー(登録商標)ディスク(disk)と、Blu-ray(登録商標)ディスク(disc)と、を含み、ここで、diskは、通常は磁的にデータを複製し、discは、レーザを用いて光学的にデータを複製する。上記の組み合わせも、コンピュータによって読み取り可能な媒体の適用範囲内に含められるべきである。20

【0078】

[0082] 命令は、1つ以上のプロセッサ、例えば、1つ以上のデジタル信号プロセッサ(DSP)、汎用マイクロプロセッサ、特定用途向け集積回路(ASSIC)、フィールドプログラマブルロジックアレイ(FPGA)、又はその他の同等の集積又はディスクリート論理回路によって実行することができる。従って、ここにおいて用いられる場合の用語“プロセッサ”は、上記の構造又はここにおいて説明される技法の実装に適するあらゆるその他の構造のうちのいずれかを意味することができる。さらに、幾つかの態様では、ここにおいて説明される機能は、符号化および復号のために構成された専用のハードウェア及び/又はソフトウェアモジュール内において提供されること、又は組み合わされたコードック内に組み入れることができる。さらに、技法は、1つ以上の回路又は論理素子内に完全に実装することが可能である。30

【0079】

[0083] 本開示の技法は、無線ハンドセット、集積回路(IC)又は一組のIC(例えば、チップセット)を含む非常に様々なデバイス又は装置内に実装することができる。本開示では、開示される技法を実施するように構成されたデバイスの機能上の態様を強調するために様々なコンポーネント、モジュール、又はユニットが説明されるが、異なるハ40

ードウェアユニットによる実現は必ずしも要求しない。むしろ、上述されるように、様々なユニットは、適切なソフトウェア及び／又はファームウェアと関係させて、コーデックハードウェアユニット内において結合させること又は上述されるように1つ以上のプロセッサを含む相互運用的なハードウェアユニットの集合によって提供することができる。

【0080】

[0084] 様々な例が説明されている。これらの及びその他の例は、以下の請求項の範囲内である。

以下に本願出願当初の特許請求の範囲を付記する。

[C1] 方法であって、

第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成することであって、前記第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、前記第2の音声データストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する音を表現すること、

少なくとも部分的に前記ミックスされた音声データストリーム内に第1のサブストリーム及び第2のサブストリームをステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データストリームを生成して前記修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようすることであって、前記第1のサブストリームは、前記第1の端末デバイスに関連する前記音の表現を含み、前記第2のサブストリームは、前記第2の端末デバイスに関連する前記音の表現を含むこと、

前記修正されたモノ音声データストリームを前記遠隔会議に参加している第3の端末デバイスに出力することと、を備える、方法。

[C2] 前記修正されたミックスされた音声データストリームに関連する音声歪みの量と前記第1及び第2のサブストリームを埋め込む前の前記ミックスされた音声データストリームに関連する音声歪みの量との間の差分は、歪みスレショルドよりも小さく、

前記修正されたミックスされた音声データストリームビットレートは、前記1及び第2のサブストリームを埋め込む前の前記ミックスされたモノ音声データのビットレートよりも大きくないC1に記載の方法。

[C3] 前記ミックスされた音声データストリームを生成することは、前記第1の音声データストリームの音声サンプルを前記第2の音声データストリームの対応する音声サンプルに加えることを備えるC1に記載の方法。

[C4] 前記修正されたミックスされた音声データストリームを生成することは、
少なくとも部分的に前記第1の音声データストリームに関するデータ圧縮形態を適用することによって前記第1のサブストリームを生成することと、

少なくとも部分的に前記第2の音声データストリームに関する前記データ圧縮形態を適用することによって前記第2のサブストリームを生成すること、

同じ又は異なるデータ圧縮形態を前記ミックスされた音声データストリームに適用することと、

前記第1及び第2のサブストリームを前記ミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むことと、を備えるC1に記載の方法。

[C5] 前記第1のサブストリームは、前記第1の端末デバイスに関連する前記音のモノラル表現を含み、前記第2のサブストリームは、前記第2の端末デバイスに関連する前記音のモノラル表現を含むC4に記載の方法。

[C6] 前記第1のサブストリームは、前記第1の音声データストリームよりも低いビットレートを有し、前記第2のサブストリームは、前記第2の音声データストリームよりも低いビットレートを有するC5に記載の方法。

[C7] 前記第1のサブストリームは、前記第1の音声データストリームの電力利得を示し、

前記第2のサブストリームは、前記第2の音声データストリームの電力利得を示すC1

10

20

30

40

50

に記載の方法。

[C 8] 前記修正されたミックスされた音声データストリームを生成することは、前記第1及び第2のサブストリームを示すために前記ミックスされたモノ音声データの最下位ビットを修正することを備えるC 1に記載の方法。

[C 9] 方法であって、

ミックスされた音声データストリームから第1のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することであって、前記第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連することと、

前記ミックスされた音声データストリームから第2のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することであって、前記第2のサブストリームは、前記遠隔会議に参加している第2の端末デバイスに関連することと、

前記ミックスされた音声データストリーム、前記第1のサブストリーム及び前記第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成することと、を備える、方法。

[C 10] 前記ミックスされた音声データストリームは、前記第1の端末デバイスに関連する音及び前記第2の端末デバイスに関連する音のミックスのモノラル表現を含み、

前記マルチチャネル音声データストリームを生成することは、空間化されたステレオ音声データストリームを生成し、従って、前記第1の端末デバイスに関連する前記音は、聴く人の左側の空間内のある地点から来ると知覚され、第2の端末デバイスに関連する音は、前記聴く人の右側の空間内のある地点から来ると知覚されることを備えるC 9に記載の方法。

[C 11] 前記空間化されたステレオ音声データストリームを生成することは、

前記第1のサブストリームに少なくとも部分的に基づいて、前記第1の端末デバイスに関連する前記音を決定することと、

前記第2のサブストリームに少なくとも部分的に基づいて、前記第2の端末デバイスに関連する前記音を決定することと、

前記第1の端末デバイスに関連する前記音に少なくとも部分的に基づいて、前記ミックスされた音声データストリームの第1のコピーに頭部伝達関数(HRTF)を適用することであって、前記空間化されたステレオ音声データストリームの左チャネルは、前記ミックスされた音声データストリームの前記第1のコピーに少なくとも部分的に基づくことと、

前記第2の端末デバイスに関連する前記音に少なくとも部分的に基づいて、前記ミックスされた音声データストリームの第2のコピーに前記HRTFを適用することであって、前記空間化されたステレオ音声データストリームの右チャネルは、前記ミックスされた音声データストリームの前記第2のコピーに少なくとも部分的に基づくことと、を備えるC 10に記載の方法。

[C 12] 前記第1の端末デバイスに関連する前記音は、前記第2の端末デバイスに関連する話声と同時並行である話声を含むC 10に記載の方法。

[C 13] 前記第1及び第2のサブストリームを示すための前記ミックスされたモノ音声データの前記音声サンプルの最下位ビットであるC 9に記載の方法。

[C 14] 前記第1及び第2のサブストリームのビットレートは、前記ミックスされた音声データストリームのビットレートよりも低いC 9に記載の方法。

[C 15] 前記ミックスされた音声データストリームは、モノラル再生用に構成されてステレオ再生用に構成されていない端末デバイスと互換性があるC 9に記載の方法。

[C 16] 前記第1のサブストリームは、前記第1の音声データストリームの電力利得を示し、

前記第2のサブストリームは、前記第2の音声データストリームの電力利得を示すC 9に記載の方法。

[C 17] デバイスであって、

第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基

10

20

30

40

50

づいて、ミックスされた音声データストリームを生成するように構成され、前記第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、前記第2の音声データストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する音を表現し、

少なくとも部分的に前記ミックスされた音声データストリーム内に第1のサブストリーム及び第2のサブストリームをステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データストリームを生成して前記修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようになります、前記第1のサブストリームは、前記第1の端末デバイスに関連する前記音の表現を含み、前記第2のサブストリームは、前記第2の端末デバイスに関連する前記音の表現を含み、及び、

前記修正されたモノ音声データストリームを前記遠隔会議に参加している第3の端末デバイスに出力するように構成される1つ以上のプロセッサ、を備える、デバイス。

[C 18] 前記修正されたミックスされた音声データストリームに関連する音声歪みの量と前記第1及び第2のサブストリームを埋め込む前の前記ミックスされた音声データストリームに関連する音声歪みの量との間の差分は、歪みスレショルドよりも小さく、

前記修正されたミックスされた音声データストリームのビットレートは、前記1及び第2のサブストリームを埋め込む前の前記ミックスされたモノ音声データのビットレートよりも大きくなないC 17に記載のデバイス。

[C 19] 前記1つ以上のプロセッサは、少なくとも部分的に前記第1の音声データストリームの音声サンプルを前記第2の音声データストリームの対応する音声サンプルに加えることによって前記ミックスされた音声データストリームを生成するように構成されるC 17に記載のデバイス。

[C 20] 前記1つ以上のプロセッサは、

少なくとも部分的に前記第1の音声データストリームに関するデータ圧縮形態を適用することによって前記第1のサブストリームを生成し、

少なくとも部分的に前記第2の音声データストリームに関するデータ圧縮形態を適用することによって前記第2のサブストリームを生成し、

同じ又は異なるデータ圧縮形態を前記ミックスされた音声データストリームに適用し、及び、

前記第1及び第2のサブストリームを前記ミックスされた音声データストリーム内にステガノグラフィ方式で埋め込むように構成される、C 17に記載のデバイス。

[C 21] 前記第1のサブストリームは、前記第1の端末デバイスに関連する前記音のモノラル表現を含み、前記第2のサブストリームは、前記第2の端末デバイスに関連する前記音のモノラル表現を含むC 20に記載のデバイス。

[C 22] 前記第1のサブストリームは、前記第1の音声データストリームよりも低いビットレートを有し、前記第2のサブストリームは、前記第2の音声データストリームよりも低いビットレートを有するC 21に記載のデバイス。

[C 23] 前記第1のサブストリームは、前記第1の音声データストリームの電力利得を示し、

前記第2のサブストリームは、前記第2の音声データストリームの電力利得を示すC 17に記載のデバイス。

[C 24] 前記1つ以上のプロセッサは、第1及び第2のサブストリームを示すために少なくとも部分的に前記ミックスされたモノ音声データの音声サンプルの最下位ビットを修正することによって前記修正されたミックスされた音声データストリームを生成するように構成されるC 17に記載のデバイス。

[C 25] 前記デバイスは、マルチパーティ制御ユニット(MCU)]であるC 17に記載のデバイス。

[C 26] 端末デバイスであって、

ミックスされた音声データストリームから第1のサブストリームを抽出するための逆ス

10

20

30

40

50

テガノグラフィプロセスを実行するように構成され、前記第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連し、

前記ミックスされた音声データストリームから第2のサブストリームを抽出するための逆ステガノグラフィプロセスを実行するように構成され、前記第2のサブストリームは、前記遠隔会議に参加している第2の端末デバイスに関連し、及び

前記ミックスされた音声データストリーム、前記第1のサブストリーム及び前記第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成するように構成される、端末デバイス。

[C 27] 前記ミックスされた音声データストリームは、前記第1の端末デバイスに関連する音及び前記第2の端末デバイスに関連する音のミックスのモノラル表現を含み、 10

前記1つ以上のプロセッサは、少なくとも部分的に空間化されたステレオ音声データストリームを生成することによって前記マルチチャネル音声データストリームを生成し、従って、第1の端末デバイスに関連する前記音は、聴く人の左側の空間内のある地点から来ると知覚されるように構成され、前記第2の端末デバイスに関連する音は、前記聴く人の右側の空間内のある地点から来ると知覚されるように構成されるC 26に記載の端末デバイス。

[C 28] 前記少なくとも1つのプロセッサは、

前記第1のサブストリームに少なくとも部分的に基づいて、前記第1の端末デバイスに関連する前記音を決定するように構成され、

前記第2のサブストリームに少なくとも部分的に基づいて、前記第2の端末デバイスに関連する前記音を決定するように構成され、 20

前記第1の端末デバイスに関連する前記音に少なくとも部分的に基づいて、前記ミックスされた音声データストリームの第1のコピーに頭部伝達関数(HRTF)を適用するように構成され、前記空間化されたステレオ音声データストリームの左チャネルは、前記ミックスされた音声データストリームの前記第1のコピーに少なくとも部分的に基づき、及び

前記第2の端末デバイスに関連する前記音に少なくとも部分的に基づいて、前記ミックスされた音声データストリームの第2のコピーに前記HRTFを適用するように構成され、前記空間化されたステレオ音声データストリームの右チャネルは、前記ミックスされた音声データストリームの前記第2のコピーに少なくとも部分的に基づくC 27に記載の端末デバイス。 30

[C 29] 前記第1の端末デバイスに関連する前記音は、前記第2の端末デバイスに関連する話声と同時並行である話声を含むC 27に記載の端末デバイス。

[C 30] 前記第1及び第2のサブストリームを示すための前記ミックスされたモノ音声データの音声サンプルの最下位ビットであるC 26に記載の端末デバイス。

[C 31] 前記第1及び第2のサブストリームのビットレートは、前記ミックスされた音声データストリームのビットレートよりも低いC 26に記載の端末デバイス。

[C 32] 前記ミックスされた音声データストリームは、モノラル再生用に構成されてステレオ再生用に構成されていない端末デバイスと互換性があるC 26に記載の端末デバイス。 40

[C 33] 前記第1のサブストリームは、前記第1の音声データストリームの電力利得を示し、

前記第2のサブストリームは、前記第2の音声データストリームの電力利得を示すC 26に記載の端末デバイス。

[C 34] デバイスであって、

第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成するための手段であって、前記第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、前記第2の音声データストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する音を表現する手段と、 50

少なくとも部分的に前記ミックスされた音声データストリーム内に第1のサブストリーム及び第2のサブストリームをステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データストリームを生成して前記修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようにするための手段であって、前記第1のサブストリームは、前記第1の端末デバイスに関連する前記音の表現を含み、前記第2のサブストリームは、前記第2の端末デバイスに関連する前記音の表現を含む手段と、

前記修正されたモノ音声データストリームを前記遠隔会議に参加している第3の端末デバイスに出力するための手段と、を備える、デバイス。

[C 3 5] 端末デバイスであって、

10

ミックスされた音声データストリームから第1のサブストリームを抽出するための逆ステガノグラフィプロセスを実行するための手段であって、前記第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する手段と、

前記ミックスされた音声データストリームから第2のサブストリームを抽出するための逆ステガノグラフィプロセスを実行するための手段であって、前記第2のサブストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する手段と、

前記ミックスされた音声データストリーム、前記第1のサブストリーム及び前記第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成するための手段と、を備える、端末デバイス。

[C 3 6] コンピュータによって読み取り可能な記憶媒体であって、デバイスの1つ以上のプロセッサによって実行されたときに、

20

第1の音声データストリーム及び第2の音声データストリームに少なくとも部分的に基づいて、ミックスされた音声データストリームを生成することを前記デバイスに行わせる命令であって、前記第1の音声データストリームは、遠隔会議に参加している第1の端末デバイスに関連する音を表現し、前記第2の音声データストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する音を表現する命令、

少なくとも部分的に前記ミックスされた音声データストリーム内に第1のサブストリーム及び第2のサブストリームをステガノグラフィ方式で埋め込むことによって修正されたミックスされた音声データストリームを生成して前記修正されたミックスされた音声データストリームがモノ音声再生及びステレオ音声再生の両方をサポートするようすることを前記デバイスに行わせる命令であって、前記第1のサブストリームは、前記第1の端末デバイスに関連する前記音の表現を含み、前記第2のサブストリームは、前記第2の端末デバイスに関連する前記音の表現を含む命令、及び

30

前記修正されたモノ音声データストリームを前記遠隔会議に参加している第3の端末デバイスに出力をすることを前記デバイスに行わせるための命令、を格納する、コンピュータによって読み取り可能な記憶媒体。

[C 3 7] コンピュータによって読み取り可能な記憶媒体であって、デバイスの1つ以上のプロセッサによって実行されたときに、

ミックスされた音声データストリームから第1のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することを前記デバイスに行わせる命令であって、前記第1のサブストリームは、遠隔会議に参加している第1の端末デバイスに関連する命令、

40

前記ミックスされた音声データストリームから第2のサブストリームを抽出するための逆ステガノグラフィプロセスを実行することを前記デバイスに行わせる命令であって、前記第2のサブストリームは、前記遠隔会議に参加している第2の端末デバイスに関連する命令、及び

前記ミックスされた音声データストリーム、前記第1のサブストリーム及び前記第2のサブストリームに少なくとも部分的に基づいて、マルチチャネル音声データストリームを生成することを前記デバイスに行わせる命令、を格納する、コンピュータによって読み取り可能な記憶媒体。

【図1】

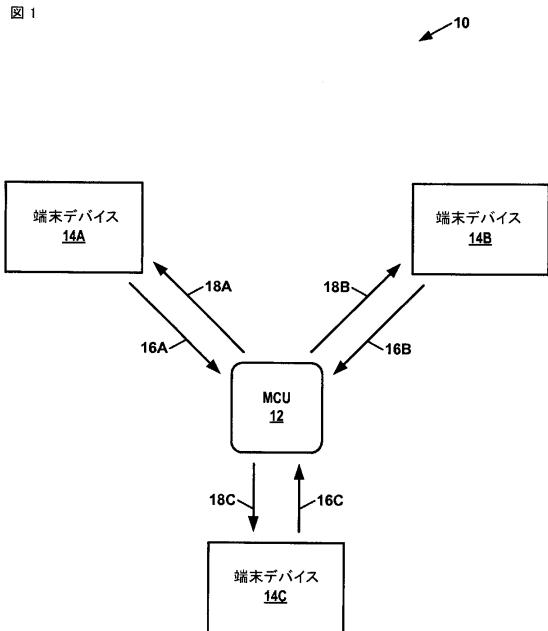


FIG. 1

【図2】

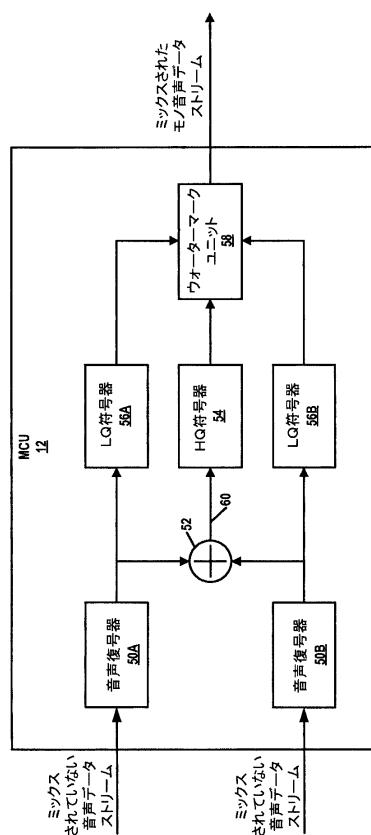


FIG. 2

【図3】

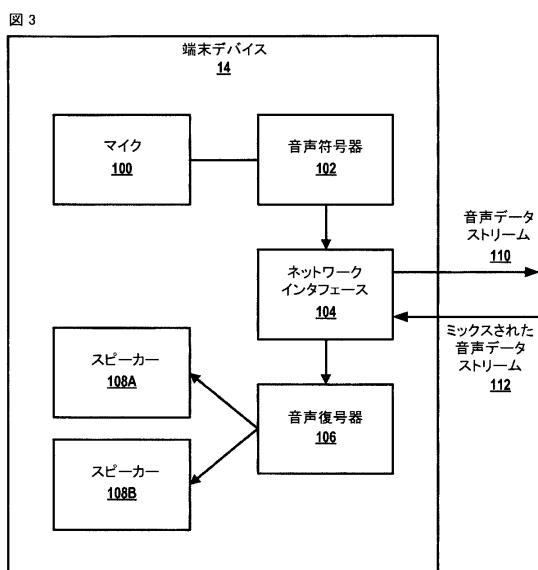


FIG. 3

【図4】

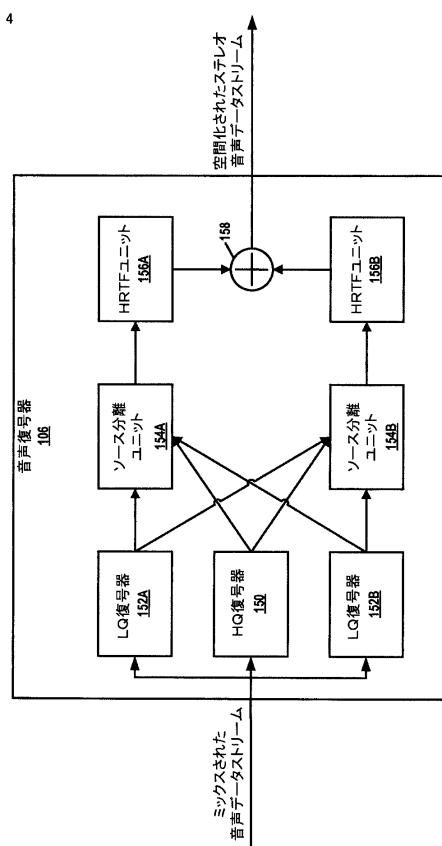
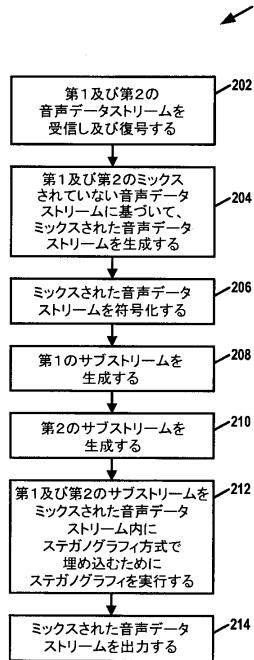


FIG. 4

【図5】

図5



【図6】

図6

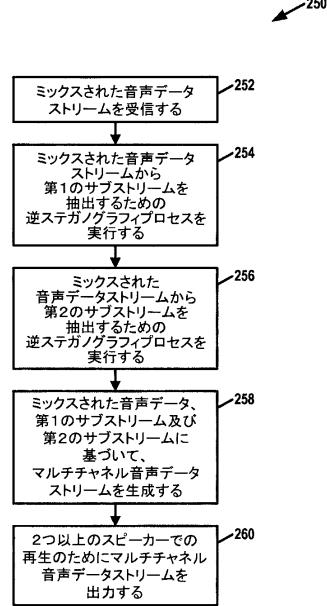


FIG. 6

FIG. 5

フロントページの続き

(72)発明者 ピレット、ステファン・ピエール

アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドライブ 5775

(72)発明者 シンダー、ダニエル・ジェイ.

アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドライブ 5775

審査官 石井 則之

(56)参考文献 特表2008-542816(JP,A)

特表2008-519301(JP,A)

特表2008-517334(JP,A)

特表2006-501514(JP,A)

特開2006-237841(JP,A)

特開2006-237839(JP,A)

特開2012-248986(JP,A)

特開2010-166425(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04M 3/00

3/16-3/20

3/38-3/58

7/00-7/16

11/00-11/10

H04N 7/10

7/14-7/173

7/20-7/56

21/00-21/858

H04S 1/00

1/00-7/00