

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6532666号
(P6532666)

(45) 発行日 令和1年6月19日(2019.6.19)

(24) 登録日 令和1年5月31日(2019.5.31)

(51) Int.Cl.		F I			
H04S	7/00	(2006.01)	H04S	7/00	300
H04R	3/00	(2006.01)	H04R	3/00	310

請求項の数 7 (全 13 頁)

(21) 出願番号	特願2014-227270 (P2014-227270)	(73) 特許権者	398058588
(22) 出願日	平成26年11月7日 (2014.11.7)		Dynabook株式会社
(65) 公開番号	特開2016-92683 (P2016-92683A)		東京都江東区豊洲五丁目6番15号
(43) 公開日	平成28年5月23日 (2016.5.23)	(74) 代理人	110002147
審査請求日	平成29年10月26日 (2017.10.26)		特許業務法人酒井国際特許事務所
		(72) 発明者	山口 隆一
			東京都港区芝浦一丁目1番1号 株式会社
			東芝内
		審査官	富澤 直樹

最終頁に続く

(54) 【発明の名称】 方法、電子機器、およびプログラム

(57) 【特許請求の範囲】

【請求項1】

複数の話者毎の複数の発話区間を含む音声信号を電子機器の複数のスピーカから再生出力するための方法であって、

前記複数の話者毎の前記複数の発話区間を含む前記音声信号を前記電子機器のメモリに記録し、

前記メモリから前記音声信号を再生する際に、前記複数の話者毎に発話区間を識別可能なように前記電子機器のディスプレイ画面に表示し、

前記ディスプレイ画面に表示された前記複数の話者毎の前記複数の発話区間のうち、第1話者による第1発話区間の第1音声タグ指定するための画面操作を受け取り、

前記複数のスピーカを用いて、前記タグ指定された前記第1発話区間の前記第1音声を前記電子機器の第1方向から聞こえるように再生し、

前記複数のスピーカを用いて、前記タグ指定がない第2話者による前記第1発話区間以外の第2発話区間の第2音声を前記電子機器の前記第1方向とは異なる第2方向から聞こえるように再生する、方法。

【請求項2】

複数の話者毎の複数の発話区間を含む音声信号を再生出力する複数のスピーカと、

前記複数の話者毎の複数の発話区間を含む前記音声信号を記録するメモリと、

前記音声信号を再生操作するための画像が表示されるディスプレイと、

前記音声信号の録音/再生プログラムを実行する処理手段と、

10

20

を具備する電子機器であって、

前記処理手段は、

前記メモリから前記音声信号を再生する際に、前記複数の話者毎に複数の発話区間を識別可能なように前記ディスプレイの画面に表示し、

前記ディスプレイの画面に表示された前記複数の話者毎の前記複数の発話区間のうち、第1話者による第1発話区間の第1音声をタグ指定するための画面操作を受け取り、

前記複数のスピーカを用いて、前記タグ指定された前記第1発話区間の前記第1音声を前記電子機器の第1方向から聞こえるように再生し、

前記複数のスピーカを用いて、前記タグ指定がない第2話者による前記第1発話区間以外の第2発話区間の第2音声を前記電子機器の前記第1方向とは異なる第2方向から聞こえるように再生する、

電子機器。

【請求項3】

前記処理手段は、

前記第1話者による前記第1発話区間の前記第1音声に基づいて前記複数のスピーカからそれぞれ再生出力される複数の音声が強め合う方向と、前記第2話者による前記第2発話区間の前記第2音声に基づいて前記複数のスピーカからそれぞれ再生出力される複数の音声が強め合う方向とを、前記第1音声および前記第2音声に対応した前記音声信号の記録時における前記第1話者と前記第2話者との位置関係、またはユーザの前記画面操作に基づいて設定し、

前記第1発話区間の前記第1音声に基づいて前記複数のスピーカからそれぞれ再生出力される複数の音声が、前記電子機器に対向する前記第1方向以外の前記第2方向で強め合うように、前記複数の音声間に位相差を設ける、請求項2に記載の電子機器。

【請求項4】

前記処理手段は、前記第2話者による前記第2発話区間の前記第2音声に基づいて前記複数のスピーカからそれぞれ再生出力される複数の音声が、前記第1音声と異なる方向に強め合うように、前記複数の音声間に位相差を設ける、請求項3に記載の電子機器。

【請求項5】

複数の話者毎の複数の発話区間を含む音声信号を電子機器の複数のスピーカから再生出力することをコンピュータに実行させるためのプログラムであって、

前記複数の話者毎の前記複数の発話区間を含む前記音声信号を前記電子機器のメモリに記録し、

前記メモリから前記音声信号を再生する際に、前記複数の話者毎に発話区間を識別可能なように前記電子機器のディスプレイ画面に表示し、

前記ディスプレイ画面に表示された前記複数の話者毎の前記複数の発話区間のうち、第1話者による第1発話区間の第1音声をタグ指定するための画面操作を受け取り、

前記複数のスピーカを用いて、前記タグ指定された前記第1発話区間の前記第1音声を前記電子機器の第1方向から聞こえるように再生し、

前記複数のスピーカを用いて、前記タグ指定がない第2話者による前記第1発話区間以外の第2発話区間の第2音声を前記電子機器の前記第1方向とは異なる第2方向から聞こえるように再生することを前記コンピュータに実行させる、プログラム。

【請求項6】

前記第1話者による前記第1発話区間の前記第1音声に基づいて前記複数のスピーカからそれぞれ再生出力される複数の音声が強め合う方向と、前記第2話者による前記第2発話区間の前記第2音声に基づいて前記複数のスピーカからそれぞれ再生出力される複数の音声が強め合う方向とを、前記第1音声および前記第2音声に対応した前記音声信号の記録時における前記第1話者と前記第2話者との位置関係、またはユーザの前記画面操作に基づいて設定し、

前記第1発話区間の前記第1音声に基づいて前記複数のスピーカからそれぞれ出力される複数の音声が、前記電子機器に対向する前記第1方向以外の前記第2方向で強め合うよ

10

20

30

40

50

うに、前記複数の音声間に位相差が設けられる、請求項 5 に記載のプログラム。

【請求項 7】

前記第 2 話者による前記第 2 発話区間の前記第 2 音声に基づいて前記複数のスピーカからそれぞれ再生出力される複数の音声が、前記第 1 音声と異なる方向に強め合うように、前記複数の音声間に位相差が設けられる、請求項 6 に記載のプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、方法、電子機器、およびプログラムに関する。

【背景技術】

10

【0002】

従来、複数の話者の複数の発話区間を含む音声を記録し、記録した音声を再生する技術が知られている。

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】特開 2003 - 006208 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

20

上記のような技術では、ユーザが指定した区間の音声と他の音声とを聴覚的に識別することができれば便利である。

【課題を解決するための手段】

【0005】

実施形態による方法は、複数の話者毎の複数の発話区間を含む音声信号を電子機器の複数のスピーカから再生出力するための方法である。この方法は、前記複数の話者毎の前記複数の発話区間を含む前記音声信号を前記電子機器のメモリに記録し、前記メモリから前記音声信号を再生する際に、前記複数の話者毎に発話区間を識別可能なように前記電子機器のディスプレイ画面に表示し、前記ディスプレイ画面に表示された前記複数の話者毎の前記複数の発話区間のうち、第 1 話者による第 1 発話区間の第 1 音声をタグ指定するための画面操作を受け取り、前記複数のスピーカを用いて、前記タグ指定された前記第 1 発話区間の前記第 1 音声を前記電子機器の第 1 方向から聞こえるように再生し、前記複数のスピーカを用いて、前記タグ指定がない第 2 話者による前記第 1 発話区間以外の第 2 発話区間の第 2 音声を前記電子機器の前記第 1 方向とは異なる第 2 方向から聞こえるように再生する。

30

【図面の簡単な説明】

【0006】

【図 1】図 1 は、実施形態による携帯端末の外観構成を示した例示図である。

【図 2】図 2 は、実施形態による携帯端末の内部構成を示した例示ブロック図である。

【図 3】図 3 は、実施形態による携帯端末が実行する録音 / 再生プログラムの機能的構成を示した例示ブロック図である。

40

【図 4】図 4 は、実施形態による携帯端末が記録音声を再生する際にディスプレイに表示される画像を示した例示図である。

【図 5】図 5 は、実施形態による携帯端末によって用いられる立体音響技術の概略を説明するための例示図である。

【図 6】図 6 は、実施形態による携帯端末を用いてユーザが話者毎の音声の到来方向を設定するための画像の一例を示した例示図である。

【図 7】図 7 は、実施形態による携帯端末を用いてユーザが話者毎の音声の到来方向を設定するための画像の他の一例を示した例示図である。

【図 8】図 8 は、実施形態による携帯端末が記録音声を再生する際に実行する処理を示し

50

た例示フローチャートである。

【図 9】図 9 は、実施形態において話者毎の音声の到来方向が設定される場合に携帯端末が実行する処理を示した例示フローチャートである。

【発明を実施するための形態】

【0007】

以下、実施形態を図面に基づいて説明する。

【0008】

まず、図 1 を参照して、実施形態による携帯端末 100 の外観構成について説明する。携帯端末 100 は、「電子機器」の一例である。図 1 は、タブレット型コンピュータとして実現された携帯端末 100 の外観を示している。なお、実施形態の技術は、スピーカを備えた電子機器であれば、スマートフォンなどの、タブレット型コンピュータ以外の携帯端末にも適用可能であるし、携帯型ではない一般的な情報処理装置にも適用可能である。

10

【0009】

図 1 に示すように、携帯端末 100 は、表示モジュール 101 と、カメラ 102 と、マイク 103 A および 103 B と、スピーカ 104 A および 104 B とを備える。

【0010】

表示モジュール 101 は、静止画や動画などの映像を表示（出力）する出力デバイスとしての機能と、ユーザの操作（タッチ操作）を受け付ける入力デバイスとしての機能とを有する。より具体的には、後述の図 2 に示すように、表示モジュール 101 は、静止画や動画などの映像を表示するためのディスプレイ 101 A と、携帯端末 100 に対する各種操作（タッチ操作）を行うための操作部として機能するタッチパネル 101 B とを備える。

20

【0011】

カメラ 102 は、カメラ 102 の正面側（Z 方向側）に位置する領域の画像を取得するための撮像デバイスである。マイク 103 A および 103 B は、表示モジュール 101 の周囲に居るユーザの音声を取得するための集音デバイスである。スピーカ 104 A および 104 B は、音声を出力するための出力デバイスである。なお、図 1 は、スピーカ 104 A および 104 B が 2 つ設けられた例を示しているが、実施形態では、スピーカ 104 A および 104 B の個数が 1 つであってもよいし、3 つ以上であってもよい。同様に、実施形態では、マイク 103 A および 103 B の個数が 1 つであってもよいし、3 つ以上であ

30

【0012】

次に、図 2 を参照して、携帯端末 100 の内部構成について説明する。

【0013】

図 2 に示すように、携帯端末 100 は、上記の表示モジュール 101、カメラ 102、マイク 103 A、103 B、スピーカ 104 A および 104 B に加えて、CPU 105 と、不揮発性メモリ 106 と、主メモリ 107 と、BIOS-ROM 108 と、システムコントローラ 109 と、グラフィクスコントローラ 110 と、サウンドコントローラ 111 と、通信コントローラ 112 と、オーディオキャプチャ 113 と、センサ群 114 とを備える。

40

【0014】

CPU 105 は、通常のコンピュータで用いられるプロセッサと同様のプロセッサであり、携帯端末 100 内の各種モジュールの動作を制御するように構成されている。この CPU 105 は、ストレージデバイスである不揮発性メモリ 106 から主メモリ 107 にロードされる各種ソフトウェアを実行するように構成されている。図 2 には、主メモリ 107 にロードされるソフトウェアの例として、OS（オペレーティングシステム）201 と、録音／再生プログラム 202 とが示されている。なお、録音／再生プログラム 202 の詳細については、後述する。

【0015】

また、CPU 105 は、BIOS-ROM 108 に格納された基本入出力システムプロ

50

グラム（ＢＩＯＳプログラム）も実行するように構成されている。なお、ＢＩＯＳプログラムとは、ハードウェアの制御を行うためのプログラムである。

【００１６】

システムコントローラ１０９は、ＣＰＵ１０５のローカルバスと、携帯端末１００に備えられた各種コンポーネントとの間を接続するためのデバイスである。

【００１７】

グラフィクスコントローラ１１０は、ディスプレイ１０１Ａを制御するデバイスである。ディスプレイ１０１Ａは、グラフィクスコントローラ１１０から入力される表示信号に基づいて画面イメージ（静止画や動画などの映像）を表示するように構成されている。

【００１８】

サウンドコントローラ１１１は、スピーカ１０４Ａおよび１０４Ｂを制御するデバイスである。スピーカ１０４Ａおよび１０４Ｂは、サウンドコントローラ１１１から入力される音声信号に基づいて音声を出力するように構成されている。

【００１９】

通信コントローラ１１２は、ＬＡＮなどを介した無線または有線の通信を実行するための通信デバイスである。オーディオキャプチャ１１３は、マイク１０３Ａおよび１０３Ｂにより取得された音声に対して各種信号処理を施す信号処理デバイスである。

【００２０】

センサ群１１４は、加速度センサや、方位センサや、ジャイロセンサなどを含む。加速度センサとは、携帯端末１００が移動する際における携帯端末１００の加速度の向きおよび大きさを検出する検出デバイスである。方位センサは、携帯端末１００の方位を検出する検出デバイスである。ジャイロセンサは、携帯端末１００が回転する際における携帯端末１００の角速度（回転角度）を検出する検出デバイスである。

【００２１】

次に、図３を参照して、ＣＰＵ１０５により実行される録音／再生プログラム２０２の機能的構成について説明する。この録音／再生プログラム２０２は、以下で説明するようなモジュール構成となっている。

【００２２】

図３に示すように、録音／再生プログラム２０２は、録音処理部２０３と、再生処理部２０４と、入力受付部２０５と、表示処理部２０６と、フィルタ係数算出部２０７と、到来方向設定部２０８とを備える。これらの各モジュールは、携帯端末１００のＣＰＵ１０５が不揮発性メモリ１０６から録音／再生プログラム２０２を読み出して実行した結果として主メモリ１０７上に生成される。

【００２３】

録音処理部２０３は、マイク１０３Ａおよび１０３Ｂを介して取得された音声信号を記録（録音）する処理を行うように構成されている。実施形態による録音処理部２０３は、複数の話者による複数の発話区間を含む音声を記録する際に、音声と同時に、各話者間の位置関係、すなわち各話者がどの方向からマイクに音声を入力したかを示す情報も記録することが可能なように構成されている。

【００２４】

再生処理部２０４は、録音処理部２０３により記録された音声（以下、記録音声という）を再生（出力）する処理を行うように構成されている。入力受付部２０５は、タッチパネル１０１Ｂなどを介したユーザの入力操作を受け付ける処理を行うように構成されている。表示処理部２０６は、ディスプレイ１０１Ａに出力する表示データを制御する処理を行うように構成されている。

【００２５】

フィルタ係数算出部２０７は、後述するフィルタ１１１Ｂおよび１１１Ｃ（図５参照）に設定するフィルタ係数を算出する処理を行うように構成されている。到来方向設定部２０８は、後述する到来方向を設定・変更する処理を行うように構成されている。

【００２６】

10

20

30

40

50

ここで、実施形態による表示処理部 206 は、再生処理部 204 が記録音声を再生する処理を行う際に、図 4 に示すような画像 IM1 をディスプレイ 101A に出力するように構成されている。この画像 IM1 は、記録音声に含まれる複数の話者の複数の発話区間を識別可能に表示するものである。

【0027】

画像 IM1 は、記録音声の大まかなステータスを表示する領域 R1 と、記録音声の詳細なステータスを表示する領域 R2 と、記録音声の再生の開始や停止などを行うための各種操作ボタンを表示する領域 R3 とを含む。

【0028】

領域 R1 には、記録音声の全体を示すバー B1 と、現在の再生位置を示すマーク M1 とが表示されている。また、領域 R1 には、記録音声の時間長（「03:00:00」という表示参照）も表示されている。

【0029】

領域 R2 には、現在の再生位置の前後の所定期間内における記録音声の詳細が表示されている。図 4 の例では、領域 R2 は、現在の再生位置の前後の所定期間内に、話者 [B] の発話区間 I1 と、話者 [A] の発話区間 I2 と、話者 [D] の発話区間 I3 と、話者 [B] の発話区間 I4 と、話者 [A] の発話区間 I5 とが含まれていることを示している。これらの発話区間 I1 ~ I5 は、話者毎に色分けされた状態で表示されていてもよい。

【0030】

領域 R2 の中央に表示されるバー B2 は、現在の再生位置を示している。図 4 の例では、バー B2 が話者 [D] の発話区間 I3 に重なるように表示されているため、現在再生されている音声の話者が [D] であることが分かる。なお、画像 IM1 は、記録音声に含まれる各発話区間の各話者を表示するための領域 R4 が含まれている。図 4 の例では、領域 R4 内の [D] という表示の近くに、現在再生されている音声の話者を示すマーク M2 が表示されているため、これによっても、現在再生されている音声の話者が [D] であることが分かる。

【0031】

また、領域 R2 には、発話区間 I1 ~ I5 に対応するように設けられる複数の星形のマーク M3 が表示されている。これらのマーク M3 は、たとえば、対応する発話区間のみを後で抽出して再生することを可能にするためのマーキング（いわゆるタグ付け）を行うためのものである。図 4 の例では、発話区間 I2 に対応するマーク M3 の周囲に細長い部分 P1 が表示されている。これにより、図 4 の例では、ユーザが発話区間 I2 に対応するマーク M3 をタッチすることによって発話区間 I2 に対してタグ付けを行ったことが分かる。

【0032】

なお、領域 R3 には、記録音声の再生の開始や停止などを行うための各種操作ボタンの他に、記録音声全体の中での現在の再生位置を示す時間（「00:49:59」という表示参照）が表示されている。

【0033】

ここで、実施形態による再生処理部 204 は、記録音声を再生する場合に、その記録音声に含まれる複数の発話区間のうちユーザが指定した第 1 発話区間の第 1 音声の出力形態を、第 1 発話区間以外の第 2 発話区間の第 2 音声と異ならせることが可能なように構成されている。

【0034】

たとえば、実施形態による再生処理部 204 は、ユーザが図 4 の画像 IM1 上でタグ付けを行った発話区間の音声の後ろ側から聴こえるとユーザに感じさせ、ユーザがタグ付けを行っていない発話区間の音声の正面側から聴こえるとユーザに感じさせるように、いわゆる立体音響技術を用いて記録音声を再生するように構成されている。

【0035】

ここで、図 5 を参照して、立体音響技術の概略について簡単に説明する。

【 0 0 3 6 】

図 5 に示すように、実施形態によるサウンドコントローラ 1 1 1 (図 2 参照) は、音声信号出力部 1 1 1 A と、2 つのフィルタ 1 1 1 B および 1 1 1 C と、信号増幅部 1 1 1 D とを備える。立体音響技術では、2 つのフィルタ 1 1 1 B および 1 1 1 C に設定するフィルタ係数を変更することにより、ユーザに感じさせる音声の到来方向を制御することができる。

【 0 0 3 7 】

フィルタ係数算出部 2 0 7 は、フィルタ係数を、スピーカ 1 0 4 A および 1 0 4 B とユーザとの位置関係に応じた頭部伝達関数と、設定したい到来方向に対応する仮想音源 V とユーザとの位置関係に応じた頭部伝達関数とに基づいて算出する。

10

【 0 0 3 8 】

たとえば、2 つのスピーカ 1 0 4 A および 1 0 4 B から出力される音声の後ろ側から聴こえるとユーザに感じさせたい場合、フィルタ係数算出部 2 0 7 は、図 5 に示す位置に仮想音源 V を設定し、一方のスピーカ 1 0 4 A の位置からユーザの両耳の位置までの 2 つの頭部伝達関数と、他方のスピーカ 1 0 4 B の位置からユーザの両耳の位置までの 2 つの頭部伝達関数と、仮想音源 V の位置からユーザの両耳の位置までの 2 つの頭部伝達関数とを用いて、2 つのフィルタ 1 1 1 B および 1 1 1 C の各々に設定するフィルタ係数を算出する。そして、再生処理部 2 0 4 は、算出されたフィルタ係数をフィルタ 1 1 1 B および 1 1 1 C に設定することにより、2 つのスピーカ 1 0 4 A および 1 0 4 B から出力される音声は仮想音源 V から聴こえるとユーザに感じさせるように、2 つのスピーカ 1 0 4 A および 1 0 4 B から出力される 2 つの音声間に位相差や音量差などを設ける。なお、実施形態では、状況に応じた複数の頭部伝達関数が携帯端末 1 0 0 に予め記憶されているものとする。

20

【 0 0 3 9 】

このように、実施形態による再生処理部 2 0 4 は、ユーザが指定した第 1 発話区間の第 1 音声に基づいて 2 つのスピーカ 1 0 4 A および 1 0 4 B からそれぞれ出力される 2 つの音声は、携帯端末 1 0 0 に対向する第 1 方向 (図 5 では方向 D 1) 以外の第 2 方向 (図 5 では方向 D 2) で強め合うように、2 つの音声間に少なくとも位相差を設けることが可能なように構成されている。

【 0 0 4 0 】

30

また、実施形態による再生処理部 2 0 4 は、上記の立体音響技術を用いて、発話区間の音声は話者毎に異なる到来方向から聴こえてくるとユーザに感じさせるように記録音声を再生することが可能なように構成されている。ここで、話者毎の音声の到来方向は、デフォルトでは、記録音声の記録時に録音処理部 2 0 3 により取得される各話者間の位置関係に基づいて設定される。また、デフォルトで設定された話者毎の音声の到来方向は、ユーザの操作によって変更することが可能である。このように到来方向を設定・変更する処理は、到来方向設定部 2 0 8 によって行われる。

【 0 0 4 1 】

たとえば、実施形態による表示処理部 2 0 6 は、話者毎の音声の到来方向をユーザに設定させるために、図 6 に示す画像 I M 2 や、図 7 に示す画像 I M 3 などをディスプレイ 1 0 1 A に表示することが可能なように構成されている。

40

【 0 0 4 2 】

図 6 の画像 I M 2 には、ユーザの位置を示すマーク M 1 0 と、マーク M 1 0 を囲む環状の点線 L 1 とが表示されている。そして、点線 L 1 上には、ユーザに対する話者 [A] ~ [D] の位置をそれぞれ示すマーク M 1 1 ~ M 1 4 が表示されている。ユーザは、各マーク M 1 1 ~ M 1 4 を点線 L 1 に沿って移動させるドラッグ操作を行うことにより、各話者 [A] ~ [D] の音声の到来方向を変更することができる。なお、図 6 の例では、話者 [A] の音声はユーザの正面側から聴こえ、話者 [B] の音声はユーザの左側から聴こえ、話者 [C] の音声はユーザの後ろ側から聴こえ、話者 [D] の音声はユーザの右側から聴こえるように、話者毎の音声の到来方向が設定されている。

50

【 0 0 4 3 】

同様に、図 7 の画像 I M 3 には、ユーザの位置を示すマーク M 2 0 と、ユーザに対するテーブル T を隔てた話者 [A] ~ [D] の位置をそれぞれ示すマーク M 2 1 ~ M 2 4 とが表示されている。ユーザは、各マーク M 2 1 ~ M 2 4 を移動させるドラッグ操作を行うことにより、各話者 [A] ~ [D] の音声の到来方向を変更することができる。なお、図 7 の例では、話者 [A] の音声はテーブル T を隔てて右側から聴こえ、話者 [B] の音声はテーブル T を隔てて正面側かつやや左寄りの位置から聴こえ、話者 [C] の音声はテーブル T を隔てて正面側かつやや右寄りの位置から聴こえ、話者 [D] の音声はテーブル T を隔てて右側から聴こえるように、話者毎の音声の到来方向が設定されている。

【 0 0 4 4 】

実施形態によるフィルタ係数算出部 2 0 7 は、話者毎に異なる到来方向から音声は聴こえるとユーザに感じさせるために、記録音声の記録時に取得された各話者の位置関係に応じた到来方向や、図 6 の画像 I M 2 または図 7 の画像 I M 3 を介して設定された到来方向などに基づいて、話者毎に異なるフィルタ係数を算出するように構成されている。そして、再生処理部 2 0 4 は、再生する音声の話者が切り替わる毎に、フィルタ 1 1 1 B および 1 1 1 C に設定するフィルタ係数を切り替えることにより、2 つのスピーカ 1 0 4 A および 1 0 4 B から出力される音声は話者毎に異なる到来方向から聴こえてくるとユーザに感じさせるように、2 つのスピーカ 1 0 4 A および 1 0 4 B から出力される 2 つの音声間に設ける位相差や音量差などを変化させる。

【 0 0 4 5 】

このように、実施形態による再生処理部 2 0 4 は、複数の話者のうち第 1 話者の発話区間に基づいて 2 つのスピーカ 1 0 4 A および 1 0 4 B から出力される 2 つの音声は強め合う方向と、第 1 話者とは異なる第 2 話者の発話区間に基づいて 2 つのスピーカ 1 0 4 A および 1 0 4 B から出力される 2 つの音声は強め合う方向とを異ならせるように、出力音声間に少なくとも位相差を設けることが可能なように構成されている。また、実施形態による到来方向設定部 2 0 8 は、これらの出力方向を、記録音声の記録時に取得される第 1 話者と第 2 話者との位置関係、またはユーザの操作に基づいて設定することが可能なように構成されている。

【 0 0 4 6 】

なお、上記では、ユーザが指定した第 1 発話区間の第 1 音声と、第 1 音声以外の第 2 音声とをユーザに聴覚的に識別させるために、立体音響技術を用いる例について説明した。しかしながら、実施形態では、第 1 音声と第 2 音声とで音量を異ならせることにより、立体音響技術を用いずに、第 1 音声と第 2 音声とをユーザに聴覚的に識別させてもよい。もちろん、第 1 音声と第 2 音声とで音量を異ならせることと、立体音響技術とを併用することにより、第 1 音声と第 2 音声とをユーザに聴覚的に識別させてもよい。

【 0 0 4 7 】

また、上記では、第 1 音声は後ろ側から聴こえ、第 2 音声は正面側から聴こえるとユーザに感じさせるように到来方向を設定することにより、第 1 音声と第 2 音声とをユーザに聴覚的に識別させる例について説明した。しかしながら、実施形態では、ユーザが第 1 音声と第 2 音声とを聴覚的に識別することが可能であれば、つまり第 1 音声と第 2 音声とで異なる到来方向から聴こえるとユーザに感じさせることが可能であれば、到来方向をどのように設定してもよい。なお、ユーザと携帯端末 1 0 0 とが互いに対向している場合、携帯端末 1 0 0 からの音声は正面側から聴こえるのが通常である。したがって、第 1 音声は後ろ側から聴こえるとユーザに感じさせるように到来方向を設定すれば、第 1 音声の再生時にユーザの注意を惹きやすい。

【 0 0 4 8 】

次に、図 8 を参照して、実施形態による携帯端末 1 0 0 の C P U 1 0 5 が記録音声を再生する際に実行する処理フローについて説明する。

【 0 0 4 9 】

この処理フローでは、図 8 に示すように、再生処理部 2 0 4 は、まず、ステップ S 1 に

10

20

30

40

50

において、次に再生する区間がユーザによりタグ付けされた区間であるか否かを判断する。

【0050】

ステップS1において、次に再生する区間がユーザによりタグ付けされた区間であると判断された場合には、ステップS2に処理が進む。そして、ステップS2において、フィルタ係数算出部207は、後ろ側から音声が増えたとユーザに感じさせるためのフィルタ係数を算出する。

【0051】

一方、ステップS1において、次に再生する区間がユーザによりタグ付けされた区間ではないと判断された場合には、ステップS3に処理が進む。そして、ステップS3において、再生処理部204は、次に再生する区間の話者を特定する。そして、ステップS4に処理が進む。

10

【0052】

ステップS4において、再生処理部204は、ステップS3において特定された話者に応じた到来方向を特定する。より具体的には、再生処理部204は、記録音声の記録時に取得された各話者の位置関係や、図6の画像IM2または図7の画像IM3上でのユーザの操作などに基づいて到来方向設定部208により設定された話者毎の音声の到来方向から、ステップS3において特定された話者に応じた到来方向を特定する。そして、ステップS5に処理が進む。

【0053】

ステップS5において、フィルタ係数算出部207は、ステップS4において特定された到来方向から音声が増えたとユーザに感じさせるためのフィルタ係数を算出する。

20

【0054】

ステップS2またはS5においてフィルタ係数が算出された場合、ステップS6に処理が進む。そして、ステップS6において、算出されたフィルタ係数をフィルタ111Bおよび111Cに設定する。そして、処理が戻る。

【0055】

次に、図9を参照して、実施形態において話者毎の音声の到来方向が設定される場合に携帯端末100のCPU105が実行する処理フローについて説明する。

【0056】

この処理フローでは、図9に示すように、到来方向設定部208は、まず、ステップS11において、デフォルトの設定として、記録音声の記録時に録音処理部203により取得された各話者間の位置関係に基づく到来方向を設定する。そして、ステップS12に処理が進む。

30

【0057】

ステップS12において、到来方向設定部208は、図6の画像IM2または図7の画像IM3上でのユーザの操作による到来方向の設定の変更が行われたか否かを判断する。このステップS12の処理は、ユーザの操作による設定の変更が行われたと判断されるまで繰り返される。ステップS12において、ユーザの操作による設定の変更が行われたと判断された場合、ステップS13に処理が進む。

【0058】

ステップS13において、到来方向設定部208は、ステップS12のユーザの操作に応じて、到来方向の設定を更新する。そして、ステップS12に処理が戻る。

40

【0059】

以上説明したように、実施形態によるCPU105は、録音/再生プログラム202を実行することにより、複数の話者の複数の発話区間を含む音声の信号を記録し、複数の話者の複数の発話区間を識別可能に表示し、複数の話者の複数の発話区間のうち第1話者の第1発話区間の第1音声を指定するための操作を受け取り、第1発話区間の第1音声を2つのスピーカ104Aおよび104Bを用いて第1出力形態により出力し、第1発話区間以外の第2発話区間の第2音声を2つのスピーカ104Aおよび104Bを用いて第2出力形態により出力する。ここで、第1音声の第1出力形態と、第2音声の第2出力形態と

50

【 0 0 6 0 】

【 0 0 6 1 】

10

【 0 0 6 2 】

20

【 0 0 6 3 】

【 0 0 6 4 】

30

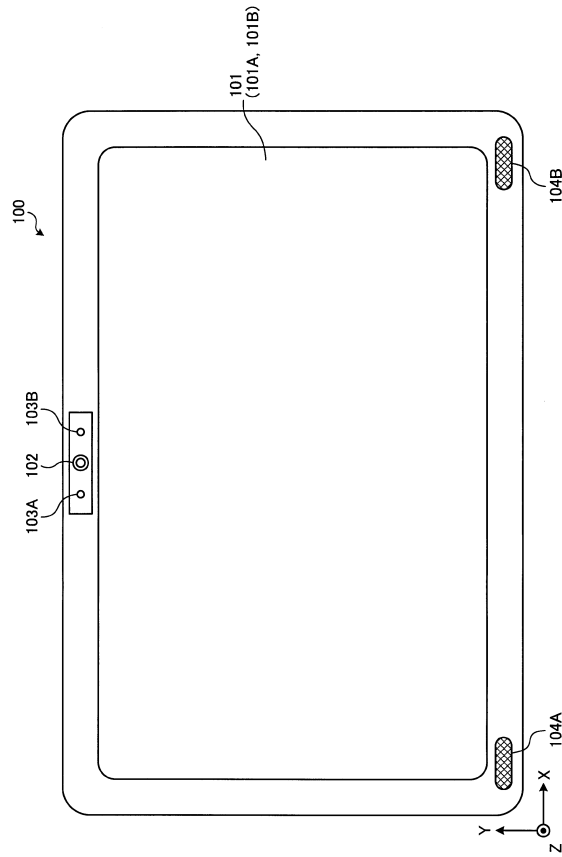
【 0 0 6 5 】

【符号の説明】

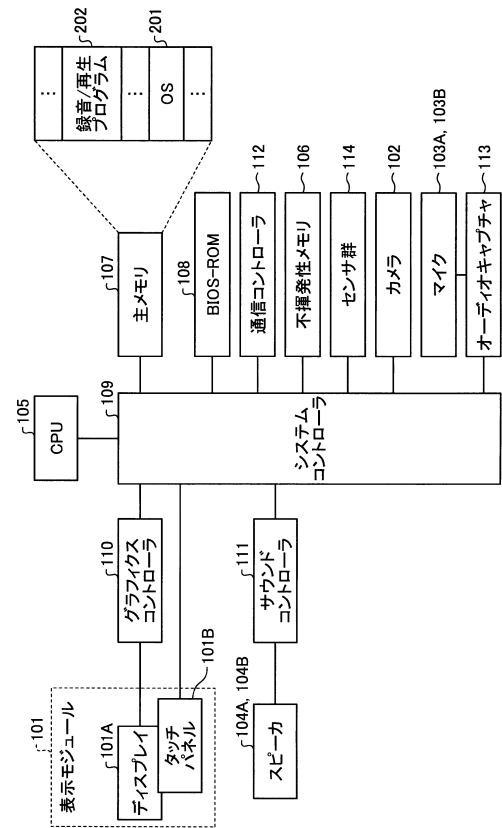
【 0 0 6 6 】

40

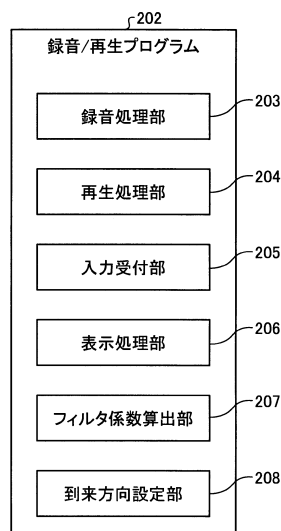
【図 1】



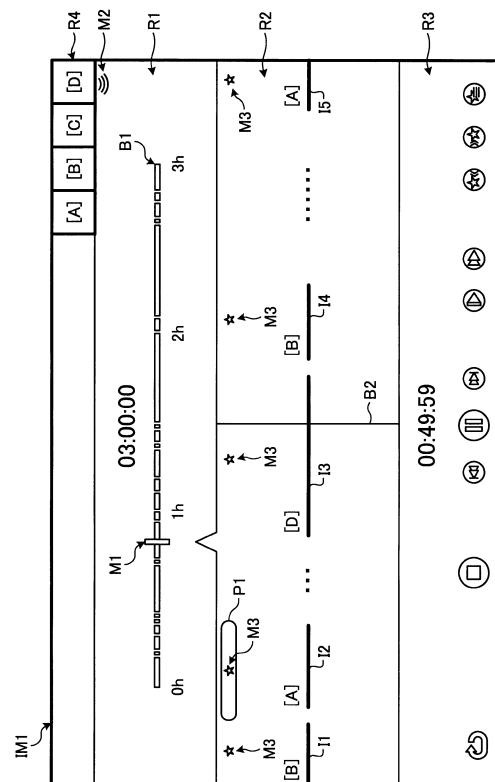
【図 2】



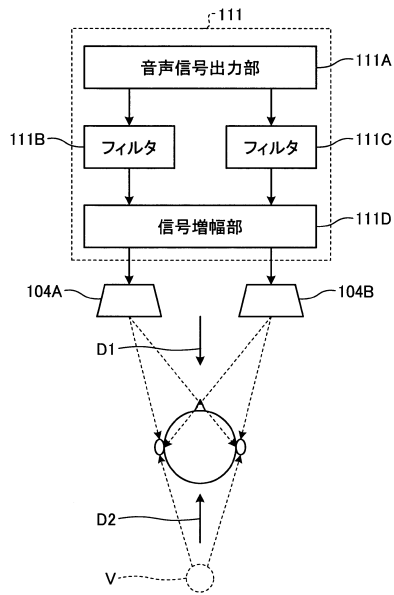
【図 3】



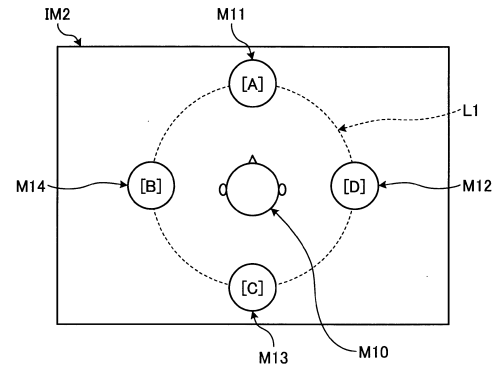
【図 4】



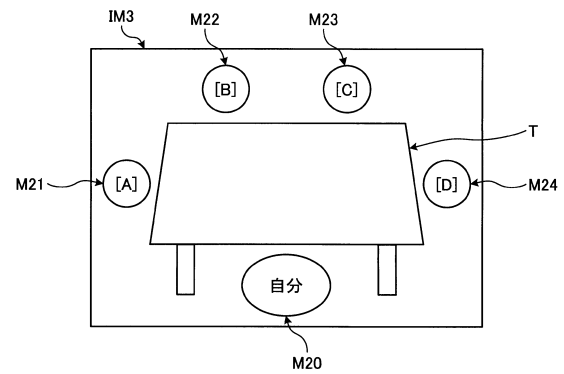
【図 5】



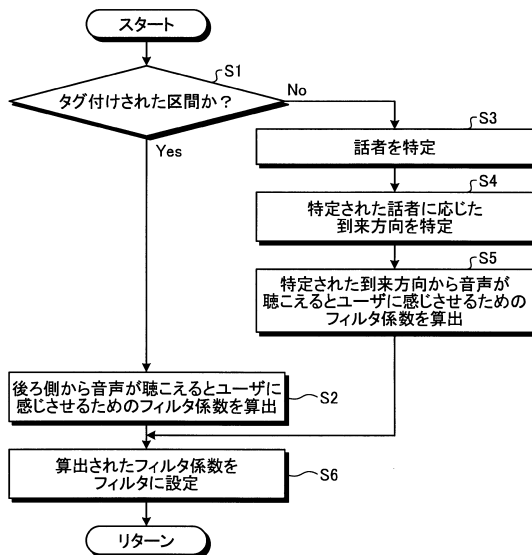
【図 6】



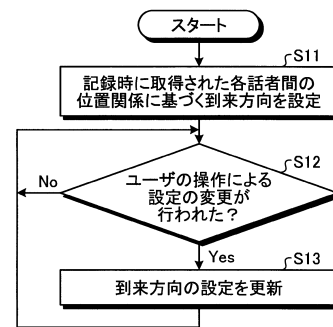
【図 7】



【図 8】



【図 9】



フロントページの続き

(56)参考文献 特開平 1 1 - 2 6 2 0 9 6 (J P , A)
特開 2 0 0 1 - 2 7 5 1 9 7 (J P , A)
特開平 0 3 - 2 5 2 2 5 8 (J P , A)
特開平 0 9 - 0 7 4 4 4 6 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

H 0 4 S 1 / 0 0 - 7 / 0 0
H 0 4 R 3 / 0 0
G 1 0 L 1 9 / 0 0 - 1 9 / 2 6