

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6503559号
(P6503559)

(45) 発行日 平成31年4月24日 (2019. 4. 24)

(24) 登録日 平成31年4月5日 (2019. 4. 5)

(51) Int. Cl.	F I
HO 4 R 1/40 (2006. 01)	HO 4 R 1/40 3 2 O A
HO 4 R 1/00 (2006. 01)	HO 4 R 1/00 3 2 8 B
HO 4 R 3/00 (2006. 01)	HO 4 R 3/00 3 2 O

請求項の数 6 (全 17 頁)

(21) 出願番号	特願2016-555104 (P2016-555104)	(73) 特許権者	000002185
(86) (22) 出願日	平成27年7月13日 (2015. 7. 13)		ソニー株式会社
(86) 国際出願番号	PCT/JP2015/070040		東京都港区港南1丁目7番1号
(87) 国際公開番号	W02016/063587	(74) 代理人	100120499
(87) 国際公開日	平成28年4月28日 (2016. 4. 28)		弁理士 平山 淳
審査請求日	平成30年6月18日 (2018. 6. 18)	(74) 代理人	100095957
(31) 優先権主張番号	特願2014-213496 (P2014-213496)		弁理士 亀谷 美明
(32) 優先日	平成26年10月20日 (2014. 10. 20)	(74) 代理人	100096389
(33) 優先権主張国	日本国 (JP)		弁理士 金本 哲男
早期審査対象出願		(74) 代理人	100101557
			弁理士 萩原 康司
		(72) 発明者	関矢 俊之
			東京都港区港南1丁目7番1号 ソニー株式会社社内

最終頁に続く

(54) 【発明の名称】 音声処理システム

(57) 【特許請求の範囲】

【請求項 1】

ユーザに装着されたときに、前記ユーザの首の右側および左側から背中側にかけて半周回する形状を有するネックバンド型装着部と、

前記ネックバンド型装着部に設けられた第1の音声取得部と、

前記ネックバンド型装着部の、前記第1の音声取得部が設けられた位置と比較して前記半周回する形状の端側に、前記第1の音声取得部とは異なる向きに設けられた第2の音声取得部と

を有し、

前記第1の音声取得部は、前記ネックバンド型装着部が前記ユーザに装着されたときに、前記第2の音声取得部と比較して前記ユーザの口の方向に向くように設けられ、

前記第2の音声取得部は、前記ネックバンド型装着部が前記ユーザに装着されたときに、前記第1の音声取得部と比較して前記ユーザの足側に向くように設けられる情報処理装置。

【請求項 2】

前記第1の音声取得部及び前記第2の音声取得部によって取得された音声データに基づいて前記ユーザからの音声を取得する制御部を備える、請求項1に記載の情報処理装置。

【請求項 3】

無線によって他の装置と通信し、他の装置に音声を送信する通信部を備える、請求項1に記載の情報処理装置。

10

20

【請求項 4】

前記通信部は Bluetooth（登録商標）を用いて前記他の装置と通信する、請求項 3 に記載の情報処理装置。

【請求項 5】

イヤホンへの出力音声を出力する出力部を備える、請求項 1 に記載の情報処理装置。

【請求項 6】

前記第 2 の音声取得部は、前記ユーザの直立姿勢における前記第 1 の音声取得部より前記ユーザの足側に設けられる、請求項 1 に記載の情報処理装置。

【発明の詳細な説明】

10

【技術分野】

【0001】

本開示は、音声処理システムに関する。

【背景技術】

【0002】

近年、ユーザの体の任意の場所に装着して、ユーザの状態をセンシングしたり、周囲の様子を撮像又は録音等したり、多様な情報をユーザに出力したりするウェアラブルデバイスが普及しつつある。例えば、ウェアラブルデバイスは、ライフログの分野やスポーツ支援の分野等の多様な分野で利用されている。

【0003】

20

ウェアラブルデバイスが取得する情報は、装着場所やユーザの状態、周囲の環境から多大な影響を受け得る。例えば、音声に関しては、ユーザの口から発せられる音声（以下、ユーザ音声とも称する）は、ウェアラブルデバイスと衣服との摩擦音や振動に係る音、周囲の環境音等の雑音に埋もれてしまう場合がある。このため、ユーザ音声をより鮮明に取得するための技術が求められている。

【0004】

例えば、下記特許文献 1 では、ヘッドセットにマイクを 2 つ設け、各マイクから入力された音声信号をマイクロホンアレー処理することで、雑音を抑圧してユーザ音声を強調した音声信号を取得する技術が開示されている。

【先行技術文献】

30

【特許文献】

【0005】

【特許文献 1】特開 2005-303574 号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかし、本技術分野では、さらなる性能向上が望まれている。そこで、本開示では、より鮮明にユーザ音声を取得することが可能な、新規かつ改良された音声処理システムを提案する。

【課題を解決するための手段】

40

【0007】

本開示によれば、ユーザに装着される装着部を備え、前記装着部は、ビームフォーミングのための音声データを取得する音声取得部を少なくとも 3 つ有する、音声処理システムが提供される。

【発明の効果】

【0008】

以上説明したように本開示によれば、より鮮明にユーザ音声を取得することが可能である。

なお、上記の効果は必ずしも限定的なものではなく、上記の効果とともに、または上記の効果に代えて、本明細書に示されたいずれかの効果、または本明細書から把握され得る

50

他の効果が奏されてもよい。

【図面の簡単な説明】

【0009】

【図1】本実施形態に係る音声処理システムの外観構成の一例を示す図である。

【図2】本実施形態に係る音声処理システムの外観構成の一例を示す図である。

【図3】本実施形態に係る音声処理システムの外観構成の一例を示す図である。

【図4】本実施形態に係る音声処理システムの外観構成の他の一例を示す図である。

【図5】本実施形態に係る音声処理システムの外観構成の他の一例を示す図である。

【図6】比較例に係る音声処理システムの外観構成の一例を示す図である。

【図7】本実施形態に係る音声取得部の配置方針について説明するための図である。

10

【図8】本実施形態に係る音声取得部の配置方針について説明するための図である。

【図9】本実施形態に係る音声取得部の配置方針について説明するための図である。

【図10】本実施形態に係る音声取得部の配置方針について説明するための図である。

【図11】本実施形態に係る音声処理システムの内部構成の一例を示すブロック図である。

。

【図12】本実施形態に係る音声処理システムにおいて実行される音声信号処理の流れの一例を示すフローチャートである。

【発明を実施するための形態】

【0010】

以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

20

【0011】

また、本明細書及び図面において、実質的に同一の機能構成を有する要素を、同一の符号の後に異なるアルファベットを付して区別する場合もある。例えば、実質的に同一の機能構成を有する複数の要素を、必要に応じて音声取得部110A、110B及び110Cのように区別する。ただし、実質的に同一の機能構成を有する複数の要素の各々を特に区別する必要がない場合、同一符号のみを付する。例えば、音声取得部110A、110B及び110Cを特に区別する必要が無い場合には、単に音声取得部110と称する。

【0012】

30

なお、説明は以下の順序で行うものとする。

1. 外観構成

2. 音声取得部の配置

2-1. 配置方針

2-2. 実際の配置例

3. 内部構成

4. 動作処理

5. まとめ

【0013】

<1. 外観構成>

40

まず、図1～図6を参照して、本開示の一実施形態に係る音声処理システムの外観構成を説明する。

【0014】

図1～図3は、本実施形態に係る音声処理システムの外観構成の一例を示す図である。図1～図3に示すように、本実施形態に係る音声処理システム1は、首の両側から後ろ側（背中側）にかけて半周回するような形状の装着ユニット（装着部）を有する。そして、装着ユニットは、ユーザの首にかけられることでユーザに装着される。なお、図1～図3は、装着ユニットをユーザが装着した状態を各観点からみた図を示しており、具体的には、図1は、斜視図であり、図2はユーザの右側からみた側面図であり、図3はユーザの上側からみた平面図である。

50

【0015】

なお、本明細書では、上下左右前後といった方向を示す言葉を用いるが、これらの方向は後述する図8に示すようなユーザの直立姿勢における、ユーザの体の中心（例えば鳩尾の位置）からみた方向を示すものとする。例えば、「右」とはユーザの右半身側の方向を示し、「左」とはユーザの左半身側の方向を示し、「上」とはユーザの頭側の方向を示し、「下」とはユーザの足側の方向を示すものとする。また、「前」とはユーザの体が向く方向を示し、「後」とはユーザの背中側の方向を示すものとする。

【0016】

図1～図3に示すように、本実施形態に係る装着ユニットは、ユーザの首回りに装着される首かけ型であってもよい。装着ユニットは、ユーザの首に密着して装着されてもよいし、離間して装着されてもよい。首かけ型の装着ユニットの他の形状としては、例えば首下げ紐によりユーザに装着されるペンダント型や、頭にかけるヘッドバンドの代わりに首の後ろ側を通るネックバンドを有するヘッドセット型が考えられる。

10

【0017】

装着型ユニットの使用形態は、人体に直接的に装着されて使用される形態であってもよい。直接的に装着されて使用される形態とは、装着型ユニットと人体との間に何らの物体も存在しない状態で使用される形態を指す。例えば、図1～図3に示した装着ユニットが、ユーザの首の肌に接するように装着される場合は、本形態に該当する。他にも、頭部に直接的に装着されるヘッドセット型やメガネ型等の多様な形態が考えられる。

20

【0018】

装着型ユニットの使用形態は、人体に間接的に装着されて使用される形態であってもよい。間接的に装着されて使用される形態とは、装着型ユニットと人体との間に何らかの物体が存在する状態で使用される形態を指す。例えば、図1～図3に示した装着ユニットが、シャツの襟の下に隠れるように装着される等、服の上からユーザに接するように装着される場合は、本形態に該当する。他にも、首下げ紐によりユーザに装着されるペンダント型や、衣服に留め具等で留められるブローチ型等の多様な形態が考えられる。

【0019】

装着ユニットは、図1～図3に示すように、複数の音声取得部110（110A、110B、110C及び110D）を有している。音声取得部110は、ユーザ音声、ユーザの話し相手が発話した音声、又は周囲の環境音等の音声データを取得する。音声取得部110により取得された音声データは、ユーザ音声を鮮明にしたり、ユーザの話し相手が発話した音声を鮮明にしたり、他の雑音を抑圧したりするビームフォーミング処理の対象となる。図1～図3に示すように、音声取得部がユーザに直に接しないようユーザ側に面しない部分（例えば、ユーザに接する面と反対側の面）に設けられる場合、首と装着ユニットとの摩擦により発生する雑音の影響を低減することができる。なお、図1～図3では、装着ユニットに音声取得部110が4つ設けられる構成を示したが、本技術はかかる例に限定されない。例えば、装着ユニットは、音声取得部を少なくとも3つ有していてもよく、5つ以上有していてもよい。

30

【0020】

本実施形態に係る音声処理システム1は、装着ユニット単体として実現されてもよいし、複数の装置の組み合わせとして実現されてもよい。例えば、音声処理システム1は、図1～図3に示した首かけ型の装着ユニット及び腕に装着されるリストバンド型の装着ユニットの組み合わせとして実現されてもよい。そして、音声処理システム1は、複数の装置に設けられた複数の音声取得部により取得された音声データを用いてビームフォーミング処理を行ってもよい。なお、以下では、音声処理システム1は図1～図3に示した装着ユニット単体として実現されるものとして説明を行う。

40

【0021】

装着ユニットの他の例を、図4及び図5に示した。図4及び図5は、本実施形態に係る音声処理システムの外観構成の他の一例を示す図である。図4は、メガネ型の装着ユニット単体から成る音声処理システム1の外観構成を示している。図5は、ネックバンド型の

50

装着ユニット単体から成る音声処理システム1の外観構成を示している。図4及び図5に示した例においても、音声処理システム1は、図1～図3に示した例と同様に、複数の音声取得部110(110A、110B、110C及び110D)を有している。

【0022】

ここで、図6を参照して、比較例について説明する。図6は、比較例に係る音声処理システムの外観構成の一例を示す図である。図6の左図及び右図は、いわゆるBluetooth(登録商標)ヘッドセットの外観構成例を示している。図6の左図に示した例では、比較例に係る音声処理システムは、2つの音声取得部910(910A及び910B)を有し、ユーザの右耳にかけられることでユーザに装着される。図6の右図に示した例では、比較例に係る音声処理システムは、左右のイヤホンに連結されるケーブルに左右対称に設けられる2つの音声取得部910(910C及び910D)を有する。このように、図6の左図及び右図に示した例では、いずれも2つの音声取得部を有している。

10

【0023】

ここで、ウェアラブルデバイスでは、使用時にマイクとユーザの口との距離が離れてしまう場合があり、ユーザ音声が増雑音に埋もれてしまう場合があった。比較例のような2つの音声取得部により取得された音声データを用いたビームフォーミング処理が行われたとしても、このような問題を解決することは困難である。

【0024】

そこで、本実施形態では、ビームフォーミング処理により雑音抑圧する場合における、雑音抑圧性能を向上させるような音声取得部110の配置を提案する。

20

【0025】

<2. 音声取得部の配置>

[2-1. 配置方針]

まず、図7～図10を参照して、音声取得部110の配置方針について説明する。

【0026】

図7は、本実施形態に係る音声取得部110の配置方針について説明するための図である。図7の左図に示すように、目的音が到来する方向210に対して、直線的に音声取得部110を配置することを第1の配置方針とする。また、図7の左図に示すように、抑圧したい雑音が増雑音に到来する方向220に対して、直線的に音声取得部110を配置することを第2の配置方針とする。本実施形態では、目的音であるユーザ音声が増雑音に到来する方向である、ユーザの口の方向210に対して、直線的に音声取得部110A及び110Bが配置され得る。第1及び第2の配置方針によれば、反対方向220から到来する雑音成分を効果的に抑圧することが可能である。なぜならば、ユーザの口の方向210から到来するユーザ音声が増雑音取得部110A及び110Bに到達するまでの位相差(時間差)が大きい上に、反対方向220から到来する音声が増雑音取得部110B及び110Aに到達するまでの位相差も大きいためである。図7の右図に示したポーラパターンが示すように、後述する制御部160によるビームフォーミング処理により、ユーザの口の方向210から到来するユーザ音声が増雑音され、反対方向220A、220B及び220Cから到来する雑音成分が増雑音される。

30

【0027】

次いで、図8を参照して、抑圧したい雑音が増雑音に到来する方向に関する音声取得部110の配置方針について詳しく説明する。

40

【0028】

図8は、本実施形態に係る音声取得部110の配置方針について説明するための図である。図8に示すように、本実施形態では、下方向に対して直線的に音声取得部110A及び110Bを配置することを第3の配置方針とする。音声処理システム1が屋外で使用されることを想定すると、屋外で生じる雑音の多くは、図8に示すようにユーザの口を基準として地面の方向(下方向)又は水平方向から到来する。なお、地面の方向から到来する雑音は、ロードノイズとも称される。本配置方針によれば、ロードノイズが増雑音に到来する下方向に対して直線的に音声取得部110が配置されることになるので、ビームフォーミング

50

処理によりロードノイズが効率的に抑圧されることが可能となる。また、本配置方針によれば、下方向と水平方向との中間の斜め方向から到来する雑音についても、音声取得部110B及び110Aに到達するまでの位相差が存在するため、ビームフォーミング処理により雑音が抑圧され得る。このように、本配置方針によれば、屋外環境における雑音を効率的に抑圧することが可能である。なお、ヘリコプターのプロペラ音や頭上に設置された液晶ディスプレイからの音声等に関しては、抑圧が困難な場合もあり得る。もちろん、装着ユニットの形状に応じて、音声取得部110A及び110Bは、厳密に下方向（鉛直方向）へ直線的に配置されていなくてもよく、傾いて配置されてもよい。

【0029】

次いで、図9及び図10を参照して、4つ以上の音声取得部110が設けられる場合における配置方針について詳しく説明する。

【0030】

図9は、本実施形態に係る音声取得部110の配置方針について説明するための図である。図9の左図に示すように、本実施形態では、複数の音声取得部110を立体的に配置することを第4の配置方針とする。詳しくは、4つの音声取得部110が設けられる位置をそれぞれ結んで形成される形状は立体である。なお、立体であるとは、任意の3つの音声取得部110の位置を含む平面上に、残り1つの音声取得部110が存在しないとも捉えることができる。このように立体的に配置される場合には、どの方向から到来する音声であっても、いずれか2以上の音声取得部110により取得された音声データに位相差が生じるので、あらゆる方向から到来する雑音を抑圧することが可能になる。また、図9の左図に示すように、4つの音声取得部110が設けられる位置をそれぞれ結んで形成される形状は正四面体である。複数の音声取得部110が設けられる位置をそれぞれ結んで形成される形状は、各音声取得部110からユーザの口までの距離が等間隔となる、正四面体のような正多面体であることが望ましい。もちろん、図9の右図の示すように、装着ユニットの形状に応じて、4つの音声取得部110が設けられる位置を結んで形成される形状が正四面体ではない四面体であってもよい。

【0031】

図10は、本実施形態に係る音声取得部110の配置方針について説明するための図である。図10に示すように、本実施形態では、少なくともいずれかのひとつの音声取得部110をユーザの口に近づけることを第5の配置方針とする。本配置方針によれば、少なくともひとつの音声取得部110は、ユーザ音声を他の雑音と比較して大きな音量で取得することが可能となる。これにより、ビームフォーミング処理によるユーザ音声の強調効果がより増大し得る。例えば、図10に示すように、4面体を形成する4つの音声取得部110よりもユーザの口に近い位置に、5つ目の音声取得部110Eが設けられてもよい。他にも、例えば図9の右図に示したように、4面体の頂点に位置する音声取得部110のうちいずれか（図9の右図に示した例では音声取得部110A）が、他と比較して最もユーザの口に近い位置に設けられてもよい。

【0032】

以上、音声取得部110の配置方針について説明した。

【0033】

〔2-2. 実際の配置例〕

続いて、再度図1～図3を参照しながら、上述した配置方針に従った音声取得部110の実際の配置例を説明する。なお、音声取得部110の実際の配置は、装着ユニットの形状や各部品の重量等の制約条件により、上述した配置方針に必ずしも完全に従っていなくてもよい。

【0034】

まず、第1の配置方針に関して説明する。図2に示すように、装着ユニットがユーザに装着された状態で、音声取得部110A及び音声取得部110Bは、ユーザの口からみて同一方向上に配置されている。また、装着ユニットがユーザに装着された状態で、4つの音声取得部110に含まれる音声取得部110A（第1の音声取得部）とユーザの口との

10

20

30

40

50

距離と、4つの音声取得部110に含まれる音声取得部110B（第2の音声取得部）とユーザの口との距離と、を異ならせて設けられている。このように、図1～図3に示した例では、目的音が到来するユーザの口の方角に対して、直線的に音声取得部110A及び110Bが配置されているので、ビームフォーミング処理によりユーザ音声効率的に強調されることが可能となる。

【0035】

次いで、第2及び第3の配置方針に関して説明する。図2に示すように、装着ユニットがユーザに装着された状態で、音声取得部110A及び音声取得部110Bは、ユーザの口からみて同一方向上に配置されている。また、装着ユニットがユーザに装着された状態で、音声取得部110A（第1の音声取得部）及び音声取得部110B（第2の音声取得部）は、ユーザの直立姿勢におけるユーザの口より足側に設けられる。このように、図1～図3に示した例では、抑圧したい雑音が到来する地面の方角に対して、直線的に音声取得部110A及び110Bが配置されているので、ビームフォーミング処理により雑音が効率的に抑圧されることが可能となる。

10

【0036】

次に、第4の配置方針に関して説明する。図1～図3に示すように、音声取得部110A、110B、110C、及び110Dが設けられる位置をそれぞれ結んで形成される形状は立体である。このように、図1～図3に示した例では、複数の音声取得部110が立体的に配置されているので、ビームフォーミング処理によりあらゆる方向から到来する雑音を抑圧することが可能となる。

20

【0037】

次いで、第5の配置方針に関して説明する。図1～図3に示すように、装着ユニットがユーザに装着された状態で、音声取得部110A（第1の音声取得部）は、他の音声取得部と比較して最もユーザの口に近い位置に設けられている。このように、図1～図3に示した例では、音声取得部110Aがユーザの口に近い位置に設けられているので、ユーザ音声を他の雑音と比較して大きな音量で取得することが可能となる。また、第2及び第3の配置方針に関して、装着ユニットがユーザに装着された状態で、音声取得部110B（第2の音声取得部）は、ユーザの直立姿勢における、ユーザの口にも最も近い位置に設けられる音声取得部110A（第1の音声取得部）よりユーザの足側に設けられる。これにより、図1～図3に示した例では、ユーザ音声の強調効果と雑音の抑圧効果とを両立させることが可能となっている。なお、図1～図3に示した例では、音声取得部110Aもユーザの口より下側に設けられているが、音声取得部110Aは口より上側に設けられてもよい。

30

【0038】

以上、本実施形態に係る音声処理システム1における音声取得部110の配置について説明した。続いて、図11を参照して、本実施形態に係る音声処理システム1の内部構成について説明する。

【0039】

<3. 内部構成>

40

図11は、本実施形態に係る音声処理システム1の内部構成の一例を示すブロック図である。図11に示すように、音声処理システム1は、音声取得部110A～110D、撮像部120、操作部130、センサ部140、通信部150、及び制御部160を有する。

【0040】

(1) 音声取得部110

音声取得部110は、ビームフォーミングのための音声データを取得する機能を有する。例えば、音声取得部110は、音声処理システム1（装着ユニット）を装着したユーザが発するユーザ音声、または周囲の音声を取得する。例えば、音声取得部110は、マイクホンにより実現される。音声取得部110は、ひとつの装着ユニットに設けられても

50

よいし、装着ユニットとは別の装置に設けられていてもよいし、複数の装置に分散して設けられていてもよい。例えば、図1～図3に示した首かけ型の装着ユニットに加え、リストバンド型の装着ユニット、メガネ型の装着ユニット、及びスマートフォンに音声取得部110が設けられてもよい。

【0041】

音声取得部110は、有指向性のマイクロホンでなくてもよい。例えば、音声取得部110は、全方位に感度を有するマイクロホンであってもよい。全方位に感度を有するとは、ポーラパターンにおいて不感性の領域（方位）がないことを指す。このようなマイクロホンは、半指向性のマイクロホンとも称されてもよい。さらに、音声取得部110は、感度が全方位に一樣又は略一樣なマイクロホンであってもよい。感度が全方位に一樣又は略

10

【0042】

音声取得部110は、マイクロホンで得られた音声信号を増幅処理するマイクアンプ回路やA/D変換器を有していてもよい。音声取得部110は、取得した音声データを制御部160へ出力する。

【0043】

(2) 撮像部120

撮像部120は、撮像レンズ、絞り、ズームレンズ、及びフォーカスレンズ等により構成されるレンズ系、レンズ系に対してフォーカス動作やズーム動作を行わせる駆動系、レンズ系で得られる撮像光を光電変換して撮像信号を生成する固体撮像素子アレイ等を有する。固体撮像素子アレイは、例えばCCD (Charge Coupled Device) センサアレイや、CMOS (Complementary Metal Oxide Semiconductor) センサアレイにより実現されてもよい。例えば、撮像部120は、音声処理システム1（装着ユニット）がユーザに装着された状態で、ユーザの前方を撮像可能に設けられてもよい。この場合、撮像部120は、例えばユーザの話し相手を撮像することが可能となる。また、撮像部120は、音声処理システム1がユーザに装着された状態で、ユーザの顔を撮像可能に設けられてもよい。この場合、音声処理システム1は、撮像画像からユーザの口の位置を特定することが可能となる。撮像部120は、デジタル信号とされた撮像画像のデータを制御部160へ出力する。

20

30

【0044】

(3) 操作部130

操作部130は、ユーザにより操作され、ユーザからの入力を受け付ける機能を有する。例えば、操作部130は、撮像部120による静止画像の撮像を指示する入力、動画画像の撮像開始又は停止を指示する入力を受け付けるカメラボタンとして実現されてもよい。また、操作部130は、音声取得部110による音声入力の開始又は停止を指示する入力を受け付ける音声入力ボタンとして実現されてもよい。また、操作部130は、タッチ操作やスライド操作を受け付けるタッチスライダーとして実現されてもよい。また、操作部130は、音声処理システム1の電源ON又はOFFを指示する操作を受け付ける電源ボタンとして実現されてもよい。操作部130は、ユーザ入力を示す情報を制御部160へ

40

【0045】

(4) センサ部140

センサ部140は、音声処理システム1を装着したユーザの状態又は周囲の状態をセンシングする機能を有する。例えば、センサ部140は、加速度センサ、速度センサ、ジャイロセンサ、地磁気センサ、GPS (Global Positioning System) モジュール、又は振動センサの少なくともいずれかを有していてもよい。センサ部140は、装着ユニットとは別の装置に設けられていてもよいし、複数の装置に分散して設けられていてもよい。例えば、リストバンド型の装置に脈拍センサが設けられ、スマートフォンに振動センサが設けられてもよい。センサ部140は、センシング結果を示す情報を制御部160へ出力

50

する。

【0046】

(5) 通信部150

通信部150は、有線／無線により音声処理システム1と他の装置との間でデータの送受信を行うための通信モジュールである。通信部150は、例えば有線LAN (Local Area Network)、無線LAN、Wi-Fi (Wireless Fidelity、登録商標)、赤外線通信、Bluetooth、NFC (Near field communication) 等の方式で、外部機器と直接、またはネットワークアクセスポイントを介して無線通信する。

【0047】

例えば、後述する制御部160としての機能がスマートフォン又はクラウド上のサーバ等の他の装置に含まれる場合、通信部150は、音声取得部110、撮像部120、操作部130、センサ部140により取得されたデータを送信してもよい。この場合、他の装置により、ビームフォーミング処理や音声認識処理等が行われる。他にも、例えば音声取得部110、撮像部120、操作部130、又はセンサ部140が別箇の装置に設けられる場合には、通信部150は、それらにより取得されたデータを受信して制御部160に出力してもよい。また、通信部150は、制御部160によるビームフォーミング処理後の音声データを、当該音声データを記憶するための記憶装置へ送信してもよい。

【0048】

(6) 制御部160

制御部160は、演算処理装置および制御装置として機能し、各種プログラムに従って音声処理システム1内の動作全般を制御する。制御部160は、例えばCPU (Central Processing Unit)、マイクロプロセッサ等の電子回路によって実現される。なお、制御部160は、使用するプログラムや演算パラメータ等を記憶するROM (Read Only Memory)、及び適宜変化するパラメータ等を一時記憶するRAM (Random Access Memory) を含んでもよい。

【0049】

例えば、制御部160は、音声取得部110により取得された複数の音声データを用いて、ユーザの口の方向からの音声を取得するための指向性を形成するビームフォーミング処理を行う。ビームフォーミング処理とは、音の到来する領域ごとに強調の度合を変化させる処理である。具体的には、制御部160が行うビームフォーミング処理は、特定の領域から到来する音を抑圧する処理を含んでもよいし、所望する方位からの音声を強調する処理を含んでもよい。例えば、制御部160は、ユーザの口の方向以外の方向からの音声を雑音として抑圧してもよい。また、制御部160は、ユーザの口の方向からの音声を強調してもよい。上述したように、音声取得部110自体が指向性を有していなくてもよい。制御部160は、各音声取得部110により取得された音声データを対象としたビームフォーミング処理を行うことで、指向性を制御する。制御部160は、各音声取得部110により取得される音声データ間の位相差を用いて、ビームフォーミング処理を行い得る。

【0050】

制御部160は、多様な観点でビームフォーミング処理を制御することができる。例えば、制御部160は、一例として以下に説明する観点で、指向性を形成する方向及び／又は範囲を制御し得る。

【0051】

例えば、制御部160は、雑音発生源と音声取得部110との位置関係に基づいてビームフォーミング処理を制御してもよい。例えば上述したようにロードノイズの発生源は地面であるため、制御部160は、地面の方向からの音声を抑圧するようビームフォーミング処理を制御してもよい。また、例えば位置情報から特定の方向に交通量の多い道路や線路等があると判別可能な場合、制御部160は、当該方向からの音声を抑圧するようビームフォーミング処理を制御してもよい。他にも、例えば雑音発生源の位置を指定するユーザ指示がある場合、制御部160は、ユーザ指示が示す位置からの音声を抑圧するようビ

10

20

30

40

50

ームフォーミング処理を制御してもよい。

【0052】

例えば、制御部160は、ユーザ以外の話者の位置に基づいてビームフォーミング処理を制御してもよい。例えば、制御部160は、ユーザ以外の他の話者からの音声を強調するビームフォーミング処理を行ってもよい。また、制御部160は、ユーザ以外の他の話者からの音声を抑圧するビームフォーミング処理を行ってもよい。ユーザ以外の他の話者の存在又は位置（方向）を特定する方法は多様に考えられる。例えば、制御部160は、ユーザ以外の方向から話し声が取得された場合、他の話者が存在すると判定し、方向を特定してもよい。また、制御部160は、音声認識により他の話者の話し声が取得されたことが認識された場合、他の話者が存在すると判定してもよい。また、制御部160は、撮像部120により撮像された撮像画像の画像認識結果により、他の話者の存在及び位置を特定してもよい。また、制御部160は、センサ部140が有するGPSモジュールにより取得されたユーザの位置情報と、他の話者の位置情報とを比較することで、他の話者の存在及び位置を特定してもよい。また、制御部160は、他の話者が所持する装置から発せられる電波の電波強度（例えば、Wi-Fiの電波強度）を測定することで、他の話者の存在及び位置を特定してもよい。

10

【0053】

例えば、制御部160は、ユーザの状態を示す情報に基づいてビームフォーミング処理を制御してもよい。ユーザの状態とは、例えばユーザが走っている、歩いている、又は乗り物に乗っている等の運動状態を指していてもよい。例えば、制御部160は、センサ部140により取得されるセンシング結果に応じて、ユーザの運動状態を推定し得る。制御部160は、複数のセンシング結果を組み合わせることで、詳細な運動状態を推定してもよい。例えば、制御部160は、振動センサ及び速度センサによるセンシング結果を組み合わせることで、振動レベル及び速度が歩行時と比較して大きい場合は自転車に乗っていると推定してもよい。他にも、制御部160は、自転車に乗っている場合と比較して振動レベルが小さく速度が大きい場合は自動車に乗っていると推定してもよい。そして、制御部160は、推定したユーザの運動状態に応じて、形成する指向性の範囲を拡大又は縮小してもよい。例えば、制御部160は、運動状態が示す運動の激しさ（例えば、各センサから出力された数値）が相対的に大きい場合には、小さい場合と比較して指向性の範囲を拡大してもよい。なお、指向性の範囲を拡大又は縮小するとは、到来する音に対して所定の値以上の感度を示す領域の範囲を拡大又は縮小することであると捉えてもよい。他にも、ユーザの状態とは、ユーザの顔の向きや姿勢等のユーザの体勢を指していてもよい。例えば、制御部160は、撮像部120により撮像された撮像画像の画像認識結果等によりユーザの顔の向きを推定し、向きに応じて指向性の向きを制御してもよい。この場合、顔の向きが変わり、ユーザの口と音声取得部110との位置関係が変化した場合にも、制御部160は、ユーザが口から発する音声を鮮明に取得するよう指向性を制御することが可能である。

20

30

【0054】

また、制御部160は、ビームフォーミング処理を行った音声データに基づいて実行された音声認識の結果に応じた処理をしてもよい。音声認識処理は、制御部160により実行されてもよいし、クラウド上のサーバ等の他の装置により実行されてもよい。例えば、制御部160は、音声認識の結果に基づいて音声処理システム1の動作を制御してもよい。具体的には、制御部160は、音声認識の結果に基づいてビームフォーミング処理に係る指向性を制御してもよい。これにより、ユーザは、例えば記録したい音声の方向に指向性を向けるよう、音声で指示することが可能となる。他にも、制御部160は、音声認識の結果に基づいて、カメラでの撮像を開始又は停止させたり、特定のセンシング結果を記録したりしてもよい。これにより、ユーザは、例えば記録したい風景や運動状態を記録するよう、音声で指示することが可能となる。

40

【0055】

なお、制御部160は、例えばモバイルプロセッサとして実現され得る。上述したよう

50

に、制御部 160 は、装着ユニットが有していてもよく、スマートフォン又はクラウド上のサーバ等の他の任意の装置が有していてもよい。

【0056】

(7) その他

他にも、音声処理システム 1 は、多様な構成要素を有し得る。例えば、音声処理システム 1 は、バッテリーを有していてもよい。図 1～図 3 に示すように、装着ユニットが湾曲した形状を有し得るため、バッテリーは曲面状の曲面バッテリーであることが望ましい。また、音声処理システム 1 は、バッテリーに充電するためのケーブルを接続可能な充電コネクタを有していてもよい。充電コネクタは、通信ケーブルを接続可能な通信コネクタとしての機能を兼ね備える、充電通信コネクタであってもよい。また、音声処理システム 1 は、ユーザへの出力装置として機能するバイブレータを有していてもよい。また、音声処理システム 1 は、ユーザへの出力装置として機能するスピーカを有していてもよい。また、音声処理システム 1 は、ユーザへの出力装置として機能するイヤホン接続可能なイヤホンコネクタを有していてもよい。イヤホンコネクタは、磁力を有していてもよく、磁力によりイヤホンコネクタとイヤホンとが着脱可能であってもよい。また、音声処理システム 1 は、制御部 160 によるビームフォーミング処理後の音声データを記憶するための記憶部を有していてもよい。

10

【0057】

以上、本実施形態に係る音声処理システム 1 の内部構成について説明した。続いて、図 12 を参照して、本実施形態に係る音声処理システム 1 の動作処理について説明する。

20

【0058】

< 4. 動作処理 >

図 12 は、本実施形態に係る音声処理システム 1 において実行される音声信号処理の流れの一例を示すフローチャートである。

【0059】

図 12 に示すように、まず、ステップ S102 で、音声処理システム 1 は、音声データを取得する。例えば、音声取得部 110A、110B、110C 及び 110D は、それぞれ音声データを取得して制御部 160 へ出力する。

【0060】

次いで、ステップ S104 で、音声処理システム 1 は、音源と音声取得部 110 との位置関係を示す情報を取得する。音源とは、雑音発生源であってもよいし、ユーザ音声の発生源であるユーザの口であってもよいし、ユーザ以外の話者であってもよい。制御部 160 は、これらの音源と音声取得部 110 との位置関係、詳しくは音声取得部 110 からみた方向を示す情報を取得する。このような情報としては、例えば、音声取得部 110 により取得された音声の音声認識結果、撮像部 120 により撮像された撮像画像の画像認識結果、操作部 130 により取得されたユーザ入力を示す情報、センサ部 140 によるセンシング結果、通信部 150 により他の装置から取得された情報等が挙げられる。

30

【0061】

次に、ステップ S106 で、音声処理システム 1 は、ユーザの状態を示す情報を取得する。例えば、制御部 160 は、ユーザの運動状態又はユーザの体勢を示す情報を取得する。このような情報としては、例えば、音声取得部 110 により取得された音声の音声認識結果、撮像部 120 により撮像された撮像画像の画像認識結果、操作部 130 により取得されたユーザ入力を示す情報、センサ部 140 によるセンシング結果、通信部 150 により他の装置から取得された情報等が挙げられる。

40

【0062】

そして、ステップ S108 で、音声処理システム 1 は、ビームフォーミング処理を行う。例えば、制御部 160 は、上記ステップ S102 において取得された複数の音声データを用いて、ユーザの口の方向からの音声を取得するための指向性を形成するビームフォーミング処理を行う。このとき、制御部 160 は、雑音発生源と音声取得部 110 との位置関係に基づいて、雑音を抑圧するようビームフォーミング処理を制御してもよい。また、

50

制御部160は、ユーザ以外の話者の位置に基づいて、ユーザ以外の他の話者からの音声を強調する又は抑圧するビームフォーミング処理を行ってもよい。また、制御部160は、ユーザの状態に応じて、指向性を形成する方向及び／又は範囲を制御してもよい。

【0063】

その後、ステップS110で、音声処理システム1は、音声認識処理を行う。例えば、制御部160は、ビームフォーミング処理を行った音声データに基づいて音声認識処理を実行する。そして、制御部160は、音声認識結果に応じて音声処理システム1の動作を制御してもよい。

【0064】

以上、音声処理システム1において実行される音声信号処理の流れの一例を説明した。

10

【0065】

<5. まとめ>

以上、図1～図12を参照して、本開示の一実施形態について詳細に説明した。上記説明したように、本実施形態に係る音声処理システム1は、装着ユニットに少なくとも3つの音声取得部を有する。これにより、音声処理システム1は、ユーザ音声をより鮮明にするビームフォーミング処理を行うために適した音声データを取得することが可能となる。

【0066】

以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

20

【0067】

なお、本明細書において説明した各装置による一連の処理は、ソフトウェア、ハードウェア、及びソフトウェアとハードウェアとの組合せのいずれを用いて実現されてもよい。ソフトウェアを構成するプログラムは、例えば、各装置の内部又は外部に設けられる記憶媒体（非一時的な媒体：non-transitory media）に予め格納される。そして、各プログラムは、例えば、コンピュータによる実行時にRAMに読み込まれ、CPUなどのプロセッサにより実行される。

【0068】

また、本明細書においてフローチャート及びシーケンス図を用いて説明した処理は、必ずしも図示された順序で実行されなくてもよい。いくつかの処理ステップは、並列的に実行されてもよい。また、追加的な処理ステップが採用されてもよく、一部の処理ステップが省略されてもよい。

30

【0069】

また、本明細書に記載された効果は、あくまで説明的または例示的なものであって限定的ではない。つまり、本開示に係る技術は、上記の効果とともに、または上記の効果に代えて、本明細書の記載から当業者には明らかな他の効果を奏しうる。

【0070】

なお、以下のような構成も本開示の技術的範囲に属する。

40

(1)

ユーザに装着される装着部を備え、

前記装着部は、ビームフォーミングのための音声データを取得する音声取得部を少なくとも3つ有する、音声処理システム。

(2)

前記装着部は、少なくとも4つの前記音声取得部を有し、

4つの前記音声取得部が設けられる位置をそれぞれ結んで形成される形状は立体である、前記(1)に記載の音声処理システム。

(3)

前記装着部が前記ユーザに装着された状態で、前記4つの音声取得部に含まれる第1の

50

音声取得部と前記ユーザの口との距離と、前記４つの音声取得部に含まれる第２の音声取得部と前記ユーザの口との距離と、を異ならせて設けられる、前記（１）又は（２）に記載の音声処理システム。

（４）

前記装着部が前記ユーザに装着された状態で、

前記第１の音声取得部は、他の前記音声取得部と比較して最も前記ユーザの口に近い位置に設けられ、

前記第２の音声取得部は、前記ユーザの直立姿勢における前記第１の音声取得部より前記ユーザの足側に設けられる、前記（３）に記載の音声処理システム。

（５）

前記第１の音声取得部及び前記第２の音声取得部は、前記ユーザの直立姿勢における前記ユーザの口より足側に設けられる、前記（３）又は（４）に記載の音声処理システム。

（６）

前記音声取得部は、全方位に感度を有するマイクロホンである、前記（２）～（５）のいずれか一項に記載の音声処理システム。

（７）

前記音声取得部は、感度が全方位に一樣又は略一樣なマイクロホンである、前記（６）に記載の音声処理システム。

（８）

前記音声処理システムは、前記音声取得部により取得された複数の音声データを用いて、前記ユーザの口の方向からの音声を取得するための指向性を形成するビームフォーミング処理を行う制御部をさらに備える、前記（２）～（７）のいずれか一項に記載の音声処理システム。

（９）

前記ビームフォーミング処理は、音の到来する領域ごとに強調の度合を変化させる処理である、前記（８）に記載の音声処理システム。

（１０）

前記ビームフォーミング処理は、特定の領域から到来する音を抑圧する処理を含む、前記（９）に記載の音声処理システム。

（１１）

前記制御部は、雑音発生源と前記音声取得部との位置関係に基づいて前記ビームフォーミング処理を制御する、前記（８）～（１０）のいずれか一項に記載の音声処理システム。

（１２）

前記制御部は、前記ユーザ以外の話者の位置に基づいて前記ビームフォーミング処理を制御する、前記（８）～（１１）のいずれか一項に記載の音声処理システム。

（１３）

前記制御部は、前記ユーザの状態を示す情報に基づいて前記ビームフォーミング処理を制御する、前記（８）～（１２）のいずれか一項に記載の音声処理システム。

（１４）

前記制御部は、前記ビームフォーミング処理を行った音声データに基づいて実行された音声認識の結果に応じた処理をする、前記（８）～（１３）のいずれか一項に記載の音声処理システム。

（１５）

前記制御部は、前記音声認識の結果に基づいて前記音声処理システムの動作を制御する、前記（１４）に記載の音声処理システム。

（１６）

前記制御部は、前記音声認識の結果に基づいて前記指向性を制御する、前記（１５）に記載の音声処理システム。

（１７）

前記装着部は、前記制御部を有する、前記（８）～（１６）のいずれか一項に記載の音声処理システム。

（１８）

前記装着部は、前記ユーザの首回りに装着される、前記（２）～（１７）のいずれか一項に記載の音声処理システム。

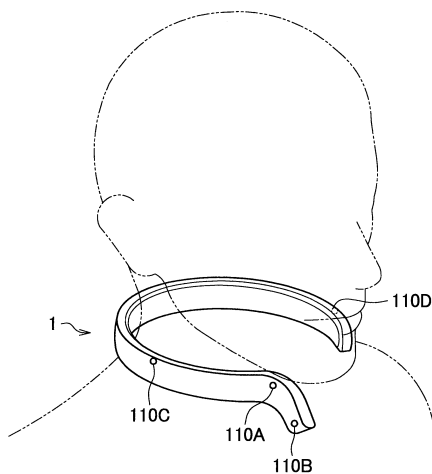
【符号の説明】

【００７１】

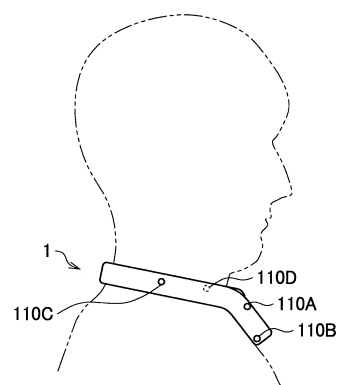
- １ 音声処理システム
- １１０ 音声取得部
- １２０ 撮像部
- １３０ 操作部
- １４０ センサ部
- １５０ 通信部
- １６０ 制御部

10

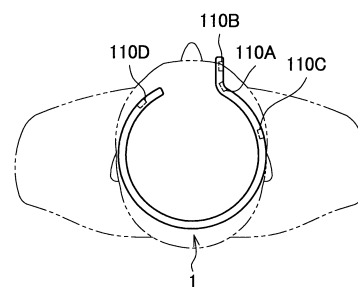
【図１】



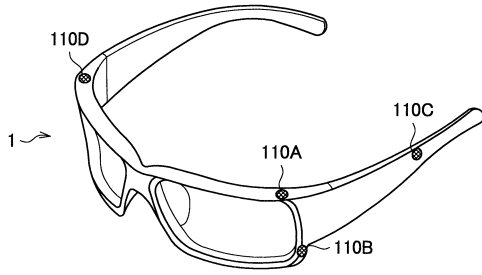
【図２】



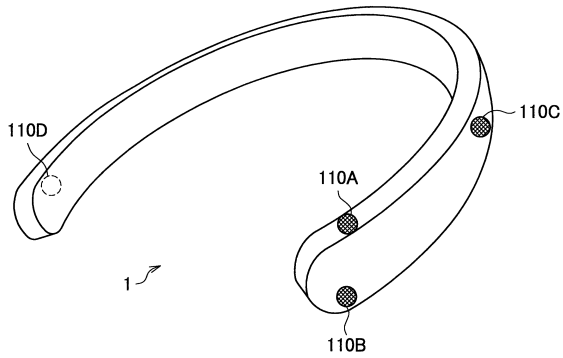
【図３】



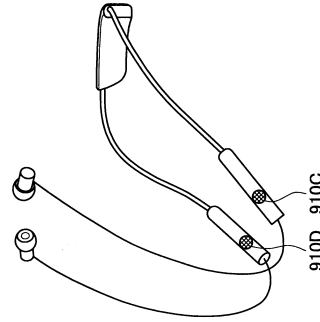
【図 4】



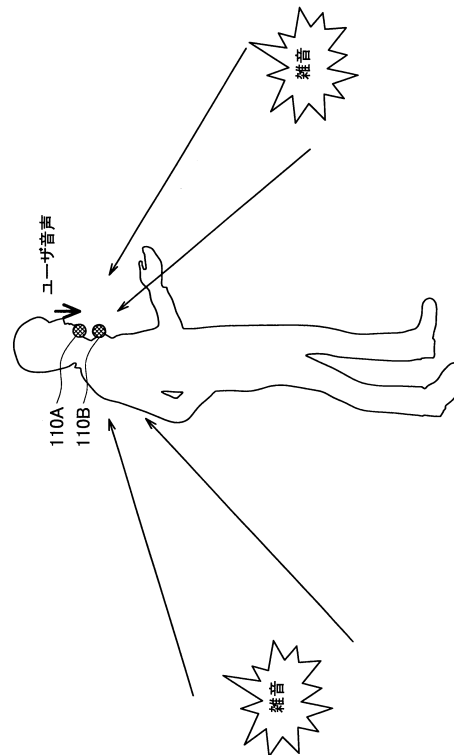
【図 5】



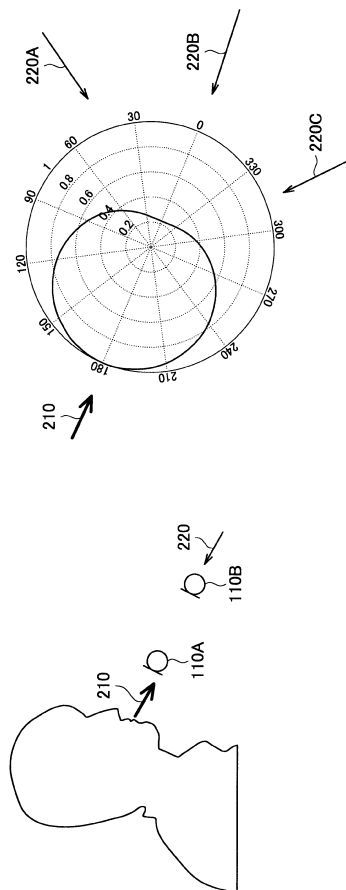
【図 6】



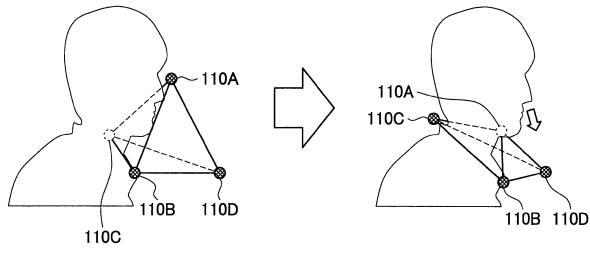
【図 8】



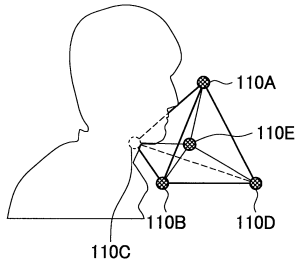
【図 7】



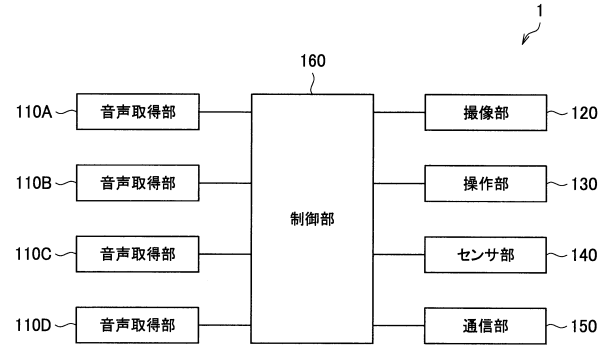
【図 9】



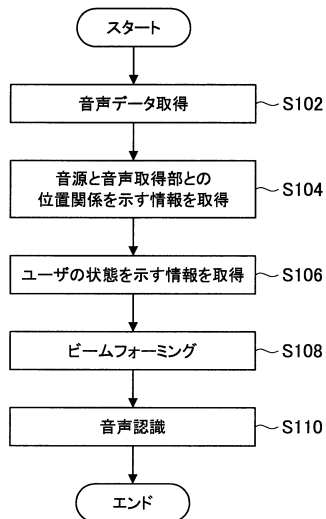
【図 10】



【図 11】



【図 12】



フロントページの続き

- (72)発明者 森 秀人
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 羽生田 誠
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 東ヶ崎 優
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 平野 雄哉
東京都港区港南1丁目7番1号 ソニー株式会社内

審査官 渡邊 正宏

- (56)参考文献 特開2012-133250 (JP, A)
特表2012-524505 (JP, A)
特表2009-514312 (JP, A)
特開2009-017083 (JP, A)
特開2001-095098 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L	13/00-13/10
G10L	19/00-19/26
G10L	21/00-21/18
G10L	25/00-25/93
G10L	99/00
H04R	1/00- 1/08
H04R	1/10
H04R	1/12- 1/14
H04R	1/20- 1/40
H04R	1/42- 1/46
H04R	3/00- 3/14
H04R	5/00- 5/04
H04R	25/00-25/04