

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5929786号  
(P5929786)

(45) 発行日 平成28年6月8日(2016.6.8)

(24) 登録日 平成28年5月13日(2016.5.13)

(51) Int.Cl.		F I			
<b>G 1 O K</b>	<b>11/178</b>	<b>(2006.01)</b>	<b>G 1 O K</b>	<b>11/16</b>	<b>H</b>
<b>G 1 O L</b>	<b>15/10</b>	<b>(2006.01)</b>	<b>G 1 O L</b>	<b>15/10</b>	<b>5 0 0 Z</b>
<b>G O 6 F</b>	<b>3/16</b>	<b>(2006.01)</b>	<b>G O 6 F</b>	<b>3/16</b>	

請求項の数 15 (全 22 頁)

(21) 出願番号	特願2013-45230 (P2013-45230)	(73) 特許権者	000002185
(22) 出願日	平成25年3月7日(2013.3.7)		ソニー株式会社
(65) 公開番号	特開2014-174255 (P2014-174255A)		東京都港区港南1丁目7番1号
(43) 公開日	平成26年9月22日(2014.9.22)	(74) 代理人	100095957
審査請求日	平成27年2月3日(2015.2.3)		弁理士 亀谷 美明
		(74) 代理人	100096389
			弁理士 金本 哲男
		(74) 代理人	100101557
			弁理士 萩原 康司
		(74) 代理人	100128587
			弁理士 松本 一騎
		(72) 発明者	浅田 宏平
			東京都港区港南1丁目7番1号 ソニー株式会社内

最終頁に続く

(54) 【発明の名称】 信号処理装置、信号処理方法及び記憶媒体

(57) 【特許請求の範囲】

【請求項1】

ユーザ音声を收音し、オーディオ信号を生成する收音部と、  
前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成する信号処理部と、  
前記マスキング音声信号を再生する第1のスピーカと、  
前記マスキング音声信号の逆相信号を再生する第2のスピーカと、  
を備え、  
前記逆相信号は、前記第1のスピーカより再生された前記マスキング音声信号と前記第2のスピーカより再生された前記逆相信号とが前記收音部が收音する空間において打ち消し合うよう再生される、信号処理装置。

【請求項2】

前記信号処理部は、前記オーディオ信号のうち前記ユーザ音声が含まれる時間区間にのみ前記マスキング音声信号を生成する、請求項1に記載の信号処理装置。

【請求項3】

前記信号処理装置は、除去部をさらに備え、  
前記除去部は、前記收音部が前記第1のスピーカから再生された前記マスキング音声信号を前記ユーザ音声と共に收音して前記オーディオ信号を生成した場合、特定の伝達関数および前記信号処理部が生成した前記マスキング音声信号に基づいて、前記收音部により生成された前記オーディオ信号から前記マスキング音声信号を除去する、請求項1に記載

の信号処理装置。

【請求項 4】

前記信号処理装置は、前記逆相信号を遅延させる遅延部をさらに備え、

前記第 2 のスピーカは、前記遅延部により遅延された前記逆相信号を再生する、請求項 1 に記載の信号処理装置。

【請求項 5】

前記信号処理部は、前記ユーザ音声を組成する周波数成分ごとのデータ量に応じて前記マスキング音声信号を生成する、請求項 1 に記載の信号処理装置。

【請求項 6】

前記マスキング音声信号は、音声帯域の帯域ノイズである、請求項 1 に記載の信号処理装置。

10

【請求項 7】

前記マスキング音声信号は、母音を主な成分とする音声信号である、請求項 1 に記載の信号処理装置。

【請求項 8】

前記信号処理装置は、前記收音部により收音された前記ユーザ音声を記録する記録部をさらに備え、

信号処理部は、前記記録部により記録された前記ユーザ音声により前記マスキング音声信号を生成する、請求項 1 に記載の信号処理装置。

【請求項 9】

20

前記信号処理装置は、前記收音部により收音される前記ユーザ音声の言語を認識する言語認識部をさらに備え、

前記信号処理部は、前記言語認識部により認識された前記言語に応じて前記マスキング音声信号を生成する、請求項 1 に記載の信号処理装置。

【請求項 10】

前記信号処理部は、前記言語認識部により認識された前記言語と同じ言語により前記マスキング音声信号を生成する、請求項 9 に記載の信号処理装置。

【請求項 11】

前記信号処理部は、前記言語認識部により認識された前記言語と異なる言語により前記マスキング音声信号を生成する、請求項 9 に記載の信号処理装置。

30

【請求項 12】

前記信号処理装置は、前記オーディオ信号を外部に送信し、外部からのオーディオ信号を受信する通信部をさらに備える、請求項 1 に記載の信号処理装置。

【請求項 13】

前記信号処理装置は、

前記オーディオ信号から制御情報を認識する制御情報認識部と、

前記制御情報認識部により認識された前記制御情報に基づいて前記信号処理装置を制御する制御部と、

をさらに備える、請求項 1 に記載の信号処理装置。

【請求項 14】

40

ユーザ音声を收音し、オーディオ信号を生成するステップと、

前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成するステップと、

前記マスキング音声信号を再生するステップと、

前記マスキング音声信号の逆相信号を再生するステップと、

を備え、

前記逆相信号は、再生された前記マスキング音声信号と再生された前記逆相信号とが前記ユーザ音声<sub>1</sub>が收音される空間において打ち消し合うよう再生される、信号処理方法。

【請求項 15】

コンピュータに、

50

ユーザ音声を収録し、オーディオ信号を生成するステップと、  
 前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成するステップと、  
 前記マスキング音声信号を再生するステップと、  
前記マスキング音声信号の逆相信号を再生するステップと、  
を実行させるためのプログラムが記憶された記憶媒体であり、  
前記逆相信号は、再生された前記マスキング音声信号と再生された前記逆相信号とが前記ユーザ音声が収録される空間において打ち消し合うよう再生される、記憶媒体。

【発明の詳細な説明】

【技術分野】

10

【0001】

本開示は、信号処理装置、信号処理方法及び記憶媒体に関する。

【背景技術】

【0002】

近年、スマートフォンやタブレット端末等の携帯端末の普及に伴い、通話のためにユーザが発話する機会が増えている。また、携帯端末をユーザの発話内容に基づいて制御する、音声認識機能の普及に伴い、ユーザが発話する機会はさらに増加している。このようなユーザが発話する機会の増加、および携帯端末が騒音環境下で使用されることが多いことに鑑み、収録したユーザの音声から外部騒音を抑制するノイズリダクション技術が多く提案されている。

20

【0003】

一方で、携帯端末は、周囲に他人がいる状況で使用されることも多く、その際ユーザが発話した声が周囲の他人に聞かれる可能性が高い。ユーザは、発話内容が他人に聞かれることについて、恥ずかしく思う場合や、セキュリティの観点から回避したいと思う場合がある。よって、発話内容を周囲の他人に聞き取れないよう妨害するマスキング技術が求められている。

【0004】

例えば、下記特許文献1では、携帯端末においてマスキング技術を利用するために、マスキング音声信号をサーバからダウンロードして再生することで、周囲の他人がユーザの発話内容を聞き取ることを妨害する技術が開示されている。

30

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2012-119785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、上記特許文献1では、マスキング音声信号を生成するためには専用の装置を要するため、携帯端末のみでマスキング技術を利用することはできなかった。

【0007】

40

そこで、本開示では、ユーザ音声に応じたマスキング音声信号を生成および再生することが可能な、新規かつ改良された信号処理装置、信号処理方法および記憶媒体を提案する。

【課題を解決するための手段】

【0008】

本開示によれば、ユーザ音声を収録し、オーディオ信号を生成する収録部と、前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成する信号処理部と、前記マスキング音声信号を再生する第1のスピーカと、を備える信号処理装置が提供される。

【0009】

50

また、本開示によれば、ユーザ音声を收音し、オーディオ信号を生成するステップと、前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成するステップと、前記マスキング音声信号を再生するステップと、を備える信号処理方法が提供される。

【0010】

また、本開示によれば、コンピュータに、ユーザ音声を收音し、オーディオ信号を生成するステップと、前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成するステップと、前記マスキング音声信号を再生するステップと、を実行させるためのプログラムが記憶された記憶媒体が提供される。

【発明の効果】

10

【0011】

以上説明したように本開示によれば、ユーザ音声に応じたマスキング音声信号を生成および再生することが可能である。

【図面の簡単な説明】

【0012】

【図1】本開示の一実施形態に係る信号処理装置の概要を示す説明図である。

【図2】比較例に係るスマートフォンの構成を示すブロック図である。

【図3】第1の実施形態に係るスマートフォンの構成を示すブロック図である。

【図4A】第1の実施形態に係る信号処理部が生成するマスキング音声信号の一例を示す説明図である。

20

【図4B】第1の実施形態に係る信号処理部が生成するマスキング音声信号の一例を示す説明図である。

【図5】第1の実施形態に係る信号処理部の構成例を示す説明図である。

【図6】第1の実施形態に係る信号処理部の構成例を示す説明図である。

【図7】第1の実施形態に係るスマートフォンの動作を示すフローチャートである。

【図8】変形例1に係るスマートフォンの構成を示すブロック図である。

【図9】第2の実施形態に係るスマートフォンの構成を示すブロック図である。

【図10】第3の実施形態に係るスマートフォンの構成を示すブロック図である。

【図11】第3の実施形態に係るスマートフォンのキャンセル領域を示す説明図である。

【図12】変形例3に係るヘッドセットを示す説明図である。

30

【発明を実施するための形態】

【0013】

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。

【0014】

なお、説明は以下の順序で行うものとする。

1. 本開示の一実施形態に係る信号処理装置の概要

2. 実施形態

2-1. 第1の実施形態

(2-1-1. スマートフォンの構成)

(2-1-2. 動作処理)

(2-1-3. 変形例1)

2-2. 第2の実施形態

2-3. 第3の実施形態

(2-3-1. 基本形態)

(2-3-2. 変形例2)

(2-3-3. 変形例3)

3. まとめ

【0015】

40

50

<< 1. 本開示の一実施形態に係る信号処理装置の概要 >>

図1を参照して、本開示の一実施形態に係る信号処理装置の概要を説明する。図1は、本開示の一実施形態に係る信号処理装置の概要を示す説明図である。図1に示すように、本実施形態に係る信号処理装置は、一例としてスマートフォン1により実現される。

【0016】

スマートフォン1は、通話用スピーカ2、マイクロフォン3（以下、マイク3と称する）、およびマスキング用スピーカ4を有する。ユーザ8は、通話用スピーカ2およびマイク3により通話相手と通話したり、マイク3に対して制御情報を発話することで、音声認識によりスマートフォン1を制御したりする。

【0017】

ここで、比較例に係るスマートフォンとして一般的な構成を、図2を参照して説明する。図2は、比較例に係るスマートフォン100の構成を示すブロック図である。図2に示した各ブロックは、スマートフォン100が内部に有している。図2に示したように、スマートフォン100は、通話用スピーカ2、マイク3、制御部11、マイクアンプ21、パワーアンプ23、送話部31、および受話部32を有する。ユーザ8がスマートフォン100により通話する際、受話部32が受信した通話相手の音声は、パワーアンプ23により増幅され、通話用スピーカ2により再生される。ユーザ8が発話した音声はマイク3により收音され、マイクアンプ21により増幅されて、送話部31により通話相手の端末に送信される。また、制御部11は、ユーザ8が発話した音声を音声認識することで、スマートフォン100を制御する。

【0018】

ユーザ8がスマートフォン100に対して発話した音声は、周囲にいる他人に聞かれ得るが、ユーザ8は、発話内容が他人に聞かれることが恥ずかしい、またはセキュリティの観点から回避したいと思う場合がある。しかしながら、比較例に係るスマートフォン100は、ユーザ8の発話音声を他人に聞かれないための構成を何ら有さないため、これを防ぐことはできない。

【0019】

そこで、上記事情を一着眼点にして本開示の一実施形態に係る信号処理装置を創作するに至った。本開示の一実施形態に係る信号処理装置は、マスキング音声信号を再生することで、周囲にいる他人にユーザ8の発話音声が聞き取られることを防止することが可能である。本実施形態に係るスマートフォン1は、図1に示したように、マスキング用スピーカ4を有し、マスキング用スピーカ4からマスキング音声信号を再生することで、周囲の他人9がユーザ8の発話内容を聞き取ることを妨害する。

【0020】

ただし、マスキング用スピーカ4がマスキング音声信号として白色雑音等の単なるノイズを再生した場合、マスキング音声信号とユーザ8の発話音声とを他人9に容易に識別されて、ユーザ8の発話内容が聞き取られてしまう可能性がある。そこで、本実施形態に係るスマートフォン1は、ユーザ8が発話した音声をマイク3により收音して、收音したユーザ音声に応じたマスキング音声信号を生成および再生することで、発話内容の聞き取りを妨害する。

【0021】

以上、本開示の一実施形態に係る信号処理装置の概要について説明した。続いて、本開示の一実施形態に係る信号処理装置の詳細な内容について説明する。

【0022】

なお、図1に示した例では、信号処理装置の一例としてスマートフォン1を用いたが、本開示に係る情報処理装置はこれに限定されない。例えば、信号処理装置は、HMD（Head Mounted Display）、ヘッドセット、デジタルカメラ、デジタルビデオカメラ、PDA（Personal Digital Assistants）、PC（Personal Computer）、ノートPC、タブレット端末、携帯電話端末、携帯用音楽再生装置、携帯用映像処理装置または携帯用ゲーム機器等であってもよ

10

20

30

40

50

い。

【 0 0 2 3 】

< < 2 . 実施形態 > >

< 2 - 1 . 第 1 の実施形態 >

[ 2 - 1 - 1 . スマートフォンの構成 ]

まず、図 3 を参照して、本実施形態に係るスマートフォン 1 - 1 の構成について説明する。図 3 は、第 1 の実施形態に係るスマートフォン 1 - 1 の構成を示すブロック図である。図 3 に示した各ブロックは、スマートフォン 1 - 1 が内部に有している。図 3 に示すように、スマートフォン 1 - 1 は、通話用スピーカ 2、マイク 3、マスキング用スピーカ 4、制御部 1 1、信号処理部 1 2、マイクアンプ 2 1、パワーアンプ 2 2、パワーアンプ 2 3、送話部 3 1、受話部 3 2、およびマスキング用音源 4 1 を有する。以下、スマートフォン 1 - 1 の各構成要素について詳細に説明する。

10

【 0 0 2 4 】

( 受話部 3 2 )

受話部 3 2 は、外部からのオーディオ信号を受信する通信部としての機能を有する。具体的には、受話部 3 2 は、通話相手の端末から通話相手の音声を示すオーディオ信号を受信する。受話部 3 2 は、受信したオーディオ信号をパワーアンプ 2 3 に出力する。

【 0 0 2 5 】

( パワーアンプ 2 3 )

パワーアンプ 2 3 は、受話部 3 2 から出力されたオーディオ信号を増幅する機能を有する。パワーアンプ 2 3 は、増幅したオーディオ信号を通話用スピーカ 2 に出力する。

20

【 0 0 2 6 】

( 通話用スピーカ 2 )

通話用スピーカ 2 は、パワーアンプ 2 3 から出力されたオーディオ信号を再生する出力装置である。なお、本実施形態では、ユーザ 8 が通話用スピーカ 2 に耳を当てながらスマートフォン 1 - 1 を使用することを想定している。

【 0 0 2 7 】

( マイク 3 )

マイク 3 は、ユーザ音声を收音し、オーディオ信号を生成する收音部としての機能を有する。より詳しくは、マイク 3 は、ユーザ 8 が発話した音声を收音して、オーディオ信号を生成する。このとき、マイク 3 は、後述するマスキング用スピーカ 4 により再生されたマスキング音声信号もユーザ 8 の音声と共に收音して、オーディオ信号を生成し得る。つまり、マイク 3 が生成するオーディオ信号には、ユーザ音声およびマスキング音声信号が含まれ得る。以下では、マイク 3 が生成するオーディオ信号を、收音信号とも称する。マイク 3 は、生成した收音信号をマイクアンプ 2 1 に出力する。

30

【 0 0 2 8 】

( マイクアンプ 2 1 )

マイクアンプ 2 1 は、マイク 3 から出力された收音信号を増幅する機能を有する。マイクアンプ 2 1 は、増幅した收音信号を制御部 1 1、送話部 3 1、および信号処理部 1 2 に出力する。

40

【 0 0 2 9 】

( 制御部 1 1 )

制御部 1 1 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スマートフォン 1 - 1 内の動作全般を制御する。制御部 1 1 は、例えば CPU ( Central Processing Unit )、マイクロプロセッサによって実現される。なお、制御部 1 1 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 )、および適宜変化するパラメータ等を一時記憶する RAM ( Random Access Memory ) を含んでいてもよい。

【 0 0 3 0 】

制御部 1 1 は、收音信号に含まれるユーザ音声から、制御情報を認識する制御情報認識

50

部としての機能を有する。より詳しくは、制御部 1 1 は、マイクアンプ 2 1 から出力された收音信号から、ユーザ音声に含まれる制御情報を認識する。例えば、制御部 1 1 は、ユーザの発話内容に基づいて、電話をかける、メッセージを送信する、検索する等の制御情報を認識する。また、制御部 1 1 は、認識された制御情報に基づいてスマートフォン 1 - 1 を制御する機能を有する。例えば、制御部 1 1 は、電話をかける、メッセージを送信する、検索する等の制御情報に基づいて、スマートフォン 1 - 1 を制御して実際に電話発信、メッセージ送信、または検索等を行う。また、制御部 1 1 は、マイク 3 により收音されるユーザ音声の言語を認識する言語認識部としての機能を有する。例えば、制御部 1 1 は、ユーザ 8 が話す言語が日本語、英語、または中国語等のどの言語であるかを認識する。また、制御部 1 1 は、ユーザ 8 の発音やイントネーション等に応じて、ユーザ 8 の母国語や出身地方を認識してもよい。

10

## 【 0 0 3 1 】

( 送話部 3 1 )

送話部 3 1 は、收音信号を外部に送信する通信部としての機能を有する。より詳しくは、送話部 3 1 は、マイクアンプ 2 1 から出力された收音信号を、通話相手の端末に送信する。

## 【 0 0 3 2 】

( パワーアンプ 2 2 )

パワーアンプ 2 2 は、後述の信号処理部 1 2 から出力されたマスキング音声信号を増幅する機能を有する。パワーアンプ 2 2 は、増幅した收音信号をマスキング用スピーカ 4 に出力する。なお、パワーアンプ 2 2 は、マスキング用スピーカ 4 により再生されるマスキング音声信号が周囲の他人 9 に聞こえ、且つ、周囲の他人 9 がユーザ 8 の発話内容を聞き取れない程度の音量となるよう増幅する。

20

## 【 0 0 3 3 】

( マスキング用スピーカ 4 )

マスキング用スピーカ 4 は、マスキング音声信号を再生する出力装置 ( 第 1 のスピーカ ) である。より詳しくは、マスキング用スピーカ 4 は、パワーアンプ 2 2 から出力されたマスキング音声信号を再生する。

## 【 0 0 3 4 】

( マスキング用音源 4 1 )

マスキング用音源 4 1 は、マスキング音声信号を生成するための元となる音源を記録する記録部としての機能を有する。例えば、マスキング用音源 4 1 は、音源として、300 Hz ~ 3 kHz とされる音声帯域の帯域ノイズ、無意味列の音声信号、男女含む複数名による人声、白色雑音、有色雑音などの多様なノイズを記録する。他にも、マスキング用音源 4 1 は、音源として、マイク 3 により收音されたユーザ音声を記録してもよい。後述する信号処理部 1 2 は、マスキング用音源 4 1 に記録された音源に基づいて、マスキング音声信号を生成する。

30

## 【 0 0 3 5 】

( 信号処理部 1 2 )

信号処理部 1 2 は、收音信号に応じて、ユーザ音声をマスキングするためのマスキング音声信号を生成する。より詳しくは、信号処理部 1 2 は、マイクアンプ 2 1 から出力された收音信号に基づいて、マスキング用音源 4 1 に記録された音源を用いたマスキング音声信号を生成する。ここで、ユーザ音声をマスキングするとは、ユーザ 8 の発話をマスキング用スピーカ 4 により再生されるマスキング音声信号に埋没させて、他人 9 に聞き取れないよう秘匿することを指す。このような、ユーザ音声をマスキングするためのマスキング音声信号には、多様な種類が考えられる。

40

## 【 0 0 3 6 】

例えば、信号処理部 1 2 は、一般的に 300 Hz ~ 3 kHz とされる音声帯域の帯域ノイズ、または無意味列の音声信号、男女含む複数名による人声によりマスキング音声信号を生成する。この場合、マスキング音声信号は、ユーザ 8 の音声と同等の帯域のノイズや

50

人声となるため、他人9に対してユーザ8の発話とマスキング音声信号との混同を生じさせて、ユーザ8の発話をマスキングすることができる。また、信号処理部12は、マスキング用音源41により記録されたユーザ8自身の音声によりマスキング音声信号を生成してもよい。ユーザ8自身の過去の音声によるマスキング音声信号は、ユーザ8が現在発話した音声とより強く混同するため、ユーザ8の発話をより強くマスキングすることができる。

【0037】

さらに、信号処理部12は、他人9にとって意味を持つ内容のマスキング音声信号を生成してもよい。マスキング音声信号が他人9にとって意味を持つ内容である場合、マスキング音声信号は、他人9の注意をユーザ8の発話内容から逸らすことで、ユーザ8の発話をマスキングすることができる。

10

【0038】

例えば、信号処理部12は、制御部11により認識されたユーザ8の言語に応じてマスキング音声信号を生成してもよい。具体的には、信号処理部12は、ユーザ8の使用言語と同じ言語、または異なる言語により、マスキング音声信号を生成してもよい。このとき、マスキング音声信号が他人9の使用言語と同じ言語である場合、他人9はマスキング音声信号が示す内容が理解可能であるためマスキング音声信号に注意を引かれる。一方で、マスキング音声信号が他人9の使用言語と異なる言語である場合、他人9は珍しい外国語や方言が聞こえることに関心を持ち、同様にマスキング音声信号に注意を引かれる。このようなマスキング音声信号は、他人9の注意をユーザ8の発話内容から逸らすことで、ユーザ8の発話が聞き取られることを妨害することができる。なお、信号処理部12は、制御部11により認識されたユーザ8の母国語や出身地方等に基づいて、ユーザ8が母国または出身地方にいるものとして周囲の他人9の使用言語を推定し、周囲の他人9の言語に応じたマスキング音声信号を生成してもよい。なお、マスキング音声信号は、ユーザ8の使用言語と同じ場合、ユーザ8の発話と同じ周波数帯域となるため、他人9に対してユーザ8の発話との混同を生じさせることもできる。他に、他人9にとって意味を持つ、注意を引き得るマスキング音声信号としては、有名人・著名人の話し声によるものが考えられる。

20

【0039】

また、スマートフォン1-1は、マスキング音声信号の再生音量をユーザ8の発話よりも大きくすることで、ユーザ8の発話をマスキングしてもよい。

30

【0040】

さらに、信号処理部12は、收音信号のうちユーザ音声が含まれる時間区間にのみマスキング音声信号を生成してもよい。この場合、マスキング音声信号が一様に再生されないため、他人9がマスキング音声信号に耳慣れすることを防止することができる。また、ユーザ8の発話と同時にマスキング音声信号が再生されるため、ユーザ8の発話とマスキング音声信号とを他人9が識別し辛くすることができる。以下、図4A、図4Bを参照して、連続してマスキング音声信号を生成する例と、收音信号のうちユーザ音声が含まれる時間区間にのみマスキング音声信号を生成する例とを、対比させながら説明する。

【0041】

40

図4A、図4Bは、第1の実施形態に係る信号処理部12が生成するマスキング音声信号の一例を示す説明図である。図4A、図4Bでは、スマートフォン1-1が通話や音声認識を行う動作モードに切り替えられた時から、その動作モード終了までの間における、收音信号およびマスキング音声信号を示す音声信号例120-1、120-2を示している。

【0042】

音声信号例120-1は、信号処理部12が、收音信号に何ら依拠することなく、連続するマスキング音声信号を生成した場合の波形である。音声信号例120-1に示したように、マスキング音声信号は一定の音量および帯域で再生されるため、他人9はマスキング音声信号に耳慣れし得る。

50

## 【 0 0 4 3 】

音声信号例 1 2 0 - 2 は、信号処理部 1 2 が、ユーザ 8 が発話中、即ち收音信号のうちユーザ音声が含まれる時間区間にのみマスキング音声信号を生成した場合の波形である。音声信号例 1 2 0 - 2 に示したように、マスキング音声信号はユーザ 8 が発話していない時間区間に再生が停止されるため、他人 9 の耳慣れを防ぐことができる。そこで、図 5 および図 6 を参照して、收音信号のうちユーザ音声が含まれる時間区間にのみマスキング音声信号を生成するための、具体的な信号処理部 1 2 の構成例を説明する。

## 【 0 0 4 4 】

図 5 は、第 1 の実施形態に係る信号処理部 1 2 の構成例を示す説明図である。図 5 に示したように、信号処理部 1 2 - 1 は、解析用 B P F ( B a n d P a s s F i l t e r ) 群 1 2 1、可変ゲインブロック群 1 2 2、合成用 B P F 群 1 2 3、および加算器 1 2 4 を有する。信号処理部 1 2 - 1 は、B P F バンクにより発話音声解析して、ユーザ音声を組成する周波数成分ごとのデータ量に応じてマスキング音声信号を生成する機能を有する。以下、信号処理部 1 2 - 1 の各構成要素について詳細に説明する。

## 【 0 0 4 5 】

## ・ 解析用 B P F 群 1 2 1

解析用 B P F 群 1 2 1 は、複数の B P F のアレイから成るフィルタバンクである。解析用 B P F 群 1 2 1 は、ユーザ音声を組成する周波数帯域成分ごとに、振幅等のデータ量に基づいて対応係数を算出する。例えば、解析用 B P F 群 1 2 1 を構成する解析用 B P F は、それぞれ所定の周波数帯域を通過させて、所定時間幅でのデータ二乗和により対応係数を算出する。ここで、対応係数は、ユーザ音声を組成する各周波数帯域成分の構成比率を示し、信号処理部 1 2 - 1 が生成するマスキング音声信号の、各周波数帯域成分の配分比となる。解析用 B P F 群 1 2 1 を構成する解析用 B P F は、それぞれ対応する可変ゲインブロック群 1 2 2 を構成する可変ゲインブロックに、算出した対応係数を出力する。

## 【 0 0 4 6 】

## ・ 可変ゲインブロック群 1 2 2

可変ゲインブロック群 1 2 2 は、マスキング用音源 4 1 から取得した音声信号を増幅する機能を有する。可変ゲインブロック群 1 2 2 を構成する可変ゲインブロックは、対応する解析用 B P F から出力された対応係数によりマスキング用音源 4 1 から取得した音声信号を増幅して、それぞれ対応する合成用 B P F 群 1 2 3 を構成する合成用 B P F に出力する。

## 【 0 0 4 7 】

## ・ 合成用 B P F 群 1 2 3

合成用 B P F 群 1 2 3 は、複数の B P F のアレイから成るフィルタバンクである。合成用 B P F 群 1 2 3 を構成する合成用 B P F は、対応する可変ゲインブロックから出力された音声信号から、対応する解析用 B P F と同じ周波数帯域成分を通過させて、合成用音声信号を生成する。合成用 B P F 群 1 2 3 は、生成した音声信号を加算器 1 2 4 に出力する。

## 【 0 0 4 8 】

## ・ 加算器 1 2 4

加算器 1 2 4 は、合成用 B P F 群 1 2 3 から出力された音声信号を合成することで、マスキング音声信号を生成する。

## 【 0 0 4 9 】

このように、解析用 B P F 群 1 2 1 を構成する各 B P F の応答量と、可変ゲインブロック群 1 2 2 を構成する各可変ゲインブロックの可変ゲイン量との対応関係が、対応係数により規定される。よって、信号処理部 1 2 - 1 は、收音信号の周波数帯域成分ごとのデータ量に応じたマスキング音声信号を生成することができる。つまり、信号処理部 1 2 - 1 は、收音信号のうちユーザ音声が含まれる時間区間にのみマスキング音声信号を生成することができる。さらに、信号処理部 1 2 - 1 は、ユーザ音声と同様の周波数帯域成分の配分比を有する、即ちユーザ 8 の発話音声と似たマスキング音声信号を生成することができ

10

20

30

40

50

る。このため、信号処理部 12 - 1 により生成されたマスキング音声信号は、他人 9 に対してユーザ 8 の発話との混同を生じさせて、ユーザ 8 の発話をより強くマスキングすることができる。

【 0 0 5 0 】

以上、BPFバンク解析を用いてマスキング音声信号を生成する信号処理部 12 の構成例を説明した。続いて、図 6 を参照して、信号処理部 12 の他の構成例を説明する。

【 0 0 5 1 】

図 6 は、第 1 の実施形態に係る信号処理部 12 の構成例を示す説明図である。図 6 に示したように、信号処理部 12 - 2 は、VAD (Voice Activity Detection) 125、およびスイッチ 126 を有する。以下、信号処理部 12 - 2 の各構成要素について詳細に説明する。

10

【 0 0 5 2 】

・ VAD 125

VAD 125 は、入力された收音信号から、音声が発話された音声区間とそれ以外のノイズ区間とを検出する機能を有する。VAD 125 は、音声区間かノイズ区間かに応じて、スイッチ 126 を制御する。

【 0 0 5 3 】

・ スイッチ 126

スイッチ 126 は、VAD 125 による制御に基づいて、マスキング用音源 41 から取得した音声信号を通過または非通過させて、マスキング音声信号として出力する。より詳しくは、スイッチ 126 は、收音信号の音声区間に相当する時間区間ではマスキング用音源 41 から取得した音声信号を通過させ、ノイズ区間に相当する時間区間では非通過とする。

20

【 0 0 5 4 】

このように、信号処理部 12 - 2 は、音声区間かノイズ区間かに応じてマスキング用音源 41 から取得した音声信号の通過 / 非通過を制御することで、收音信号のうちユーザ音声が含まれる時間区間にのみマスキング音声信号を生成することができる。

【 0 0 5 5 】

以上、VAD の手法を用いてマスキング音声信号を生成する信号処理部 12 の構成例を説明した。

30

【 0 0 5 6 】

( 補足 )

なお、スマートフォン 1 - 1 は、ADC (Analog - to - Digital Converter) および DAC (Digital - to - Analog Converter) を有していてもよい。ADC とは、アナログ信号をデジタル信号に変換する電子回路であり、DAC とは、デジタル信号をアナログ信号に変換する電子回路である。例えば、マイクアンプ 21 の後段に ADC が設けられていてもよい。また、パワーアンプ 22、およびパワーアンプ 23 の前段に DAC が設けられていてもよい。

【 0 0 5 7 】

以上、スマートフォン 1 - 1 の構成について説明した。

40

【 0 0 5 8 】

[ 2 - 1 - 2 . 動作処理 ]

続いて、図 7 を参照して、スマートフォン 1 - 1 の動作処理について説明する。図 7 は、第 1 の実施形態に係るスマートフォン 1 - 1 の動作を示すフローチャートである。なお、他の実施形態における動作は、スマートフォン 1 - 1 の動作と同様である。図 7 に示すように、まず、ステップ S 11 で、マイク 3 は、ユーザ音声を收音し、收音信号を生成する。

【 0 0 5 9 】

次いで、ステップ S 12 で、信号処理部 12 は、マイク 3 により生成された收音信号に応じてマスキング音声信号を生成する。より詳しくは、信号処理部 12 は、図 5 および図

50

6を参照して上記説明したように、BPFバンク解析やVADの手法により、ユーザ音声をマスキングするマスキング音声信号を生成する。

【0060】

そして、ステップS13で、マスキング用スピーカ4は、信号処理部12により生成されたマスキング音声信号を再生する。なお、スマートフォン1-1は、マスキング音声信号を再生しながら、送話部31および受話部32による通話や、制御部11により音声認識された制御情報に基づく動作を行う。

【0061】

以上、第1の実施形態について説明した。続いて、第1の実施形態に係る変形例について説明する。

【0062】

[2-1-3. 変形例1]

本変形例は、通話用スピーカ2が、通話相手の音声と共にマスキング音声信号を再生する形態である。以下、図8を参照して、本変形例に係るスマートフォン1-2について説明する。

【0063】

図8は、変形例1に係るスマートフォン1-2の構成を示すブロック図である。図8に示した各ブロックは、スマートフォン1-2が内部に有している。図8に示すように、本変形例に係るスマートフォン1-2は、図3を参照して上記説明した第1の実施形態に係るスマートフォン1-1から、マスキング用スピーカ4およびパワーアンプ22を省き、加算器13を加えた構成を有する。

【0064】

信号処理部12により生成されたマスキング音声信号は、加算器13に出力される。加算器13は、入力された信号を合成する機能を有し、信号処理部12から出力されたマスキング音声信号と、受話部32から出力された通話相手のオーディオ信号を合成する。加算器13により合成されたマスキング音声信号および通話相手のオーディオ信号は、パワーアンプ23により増幅され、通話用スピーカ2により出力される。つまり、通話用スピーカ2は、通話相手の音声とマスキング音声信号とを再生する。

【0065】

本変形例に係るスマートフォン1-2は、通話用スピーカ2のマスキング用スピーカ4としても用いることで、複数のスピーカを要することなく、マスキング音声信号を再生してユーザ音声をマスキングすることができる。なお、本変形例は、ユーザ8が通話用スピーカ2に耳を当てずにスマートフォン1-2を使用する、ハンズフリー通話や音声認識入力において用いられることを想定している。ユーザ8は、通話用スピーカ2に耳を当てながら、即ち口元をマイク3に近づけて使用する第1の実施形態と比較して大きな声で話し得る。そこで、パワーアンプ23は、第1の実施形態と比較して強くマスキング音声信号を増幅する。

【0066】

以上、変形例1について説明した。

【0067】

<2-2. 第2の実施形態>

本実施形態は、マスキング用スピーカ4から再生されたマスキング音声信号がマイク3により收音された場合に、電氣的に收音信号からマスキング音声信号成分を除去する形態である。マスキング用スピーカ4から再生されたマスキング音声信号は、マイク3とマスキング用スピーカ4との位置関係や向き、再生音量、收音感度等によってはマイク3に收音されてしまい、通話や音声認識の妨げになり得る。この点、本実施形態によれば、收音信号からマスキング音声信号成分を除去することで、雑音を低減した高品質な通話や音声認識を実現することができる。以下、図9を参照して、本実施形態に係るスマートフォン1-3について説明する。

【0068】

10

20

30

40

50

図9は、第2の実施形態に係るスマートフォン1-3の構成を示すブロック図である。図9に示した各ブロックは、スマートフォン1-3が内部に有している。図9に示すように、本実施形態に係るスマートフォン1-3は、図3を参照して上記説明した第1の実施形態に係るスマートフォン1-1に、エコーキャンセラ14および加算器15を加えた構成を有する。以下、エコーキャンセラ14および加算器15が有する機能について説明する。

【0069】

(エコーキャンセラ14)

エコーキャンセラ14は、マスキング用スピーカ4から再生されたマスキング音声信号がマイク3により收音された場合に、收音信号からマスキング音声信号を除去する、除去部としての機能を有する。なお、エコーキャンセラ14および後述の加算器15により、除去部として機能すると捉えてもよい。

10

【0070】

エコーキャンセラ14は、特定の伝達関数および信号処理部12が生成したマスキング音声信号に基づいて、收音信号に含まれるマスキング音声信号を生成する。エコーキャンセラ14は、信号処理部12が生成したマスキング音声信号、マイク3およびマスキング用スピーカ4の特性に基づいて、マイク3とマスキング用スピーカ4との間の空間の伝達関数を推定する。エコーキャンセラ14は、スマートフォン1-3とユーザ8との位置関係に応じて、随時伝達関数を更新してもよい。また、エコーキャンセラ14は、デジタルフィルタとして実現されてもよい。なお、伝達関数は、信号処理部12が生成したマスキング音声信号と、マイク3により收音されたマスキング音声信号との対応関係とも捉えることができる。

20

【0071】

エコーキャンセラ14は、生成した收音信号に含まれるマスキング音声信号を、加算器15に出力する。

【0072】

(加算器15)

加算器15は、收音信号から、エコーキャンセラ14により生成されたマスキング音声信号を減算する機能を有する。このため、收音信号から、マスキング用スピーカ4から再生されマイク3により收音されたマスキング音声信号が除去される。加算器15は、マスキング音声信号を除去した收音信号を、制御部11、送話部31、および信号処理部12に出力する。

30

【0073】

このように、本実施形態は、エコーキャンセラ14および加算器15により收音信号からマスキング音声信号成分を除去することができ、雑音を低減した高品質な通話や音声認識を実現することができる。また、信号処理部12に入力される收音信号も雑音が低減されているため、信号処理部12は、よりユーザ8の音声に即したマスキング音声信号を生成することができる。

【0074】

以上、第2の実施形態について説明した。

40

【0075】

< 2-3. 第3の実施形態 >

[ 2-3-1. 基本形態 ]

本実施形態は、マスキング音声信号を再生するスピーカを複数設け、互いに打ち消し合わせることで、空間音響的に收音信号からマスキング音声信号成分を除去する形態である。以下、図10を参照して、本実施形態に係るスマートフォン1-4について説明する。なお、以下ではマスキング音声信号を再生するスピーカを2つ設ける例を説明するが、3つ以上であってもよい。

【0076】

図10は、第3の実施形態に係るスマートフォン1-4の構成を示すブロック図である

50

。図10に示した各ブロックは、スマートフォン1-4が内部に有している。図10に示すように、本実施形態に係るスマートフォン1-4は、図9を参照して上記説明した第2の実施形態に係るスマートフォン1-3に、逆相信号生成部16、パワーアンプ24、マスキング用スピーカ4-2を加えた構成を有する。なお、第2の実施形態におけるマスキング用スピーカ4を、本実施形態ではマスキング用スピーカ4-1と称する。以下、逆相信号生成部16、パワーアンプ24、マスキング用スピーカ4-2が有する機能について説明する。

【0077】

(逆相信号生成部16)

逆相信号生成部16は、信号処理部12から出力されたマスキング音声信号の逆相信号を生成する機能を有する。逆相信号生成部16は、生成した逆相信号をパワーアンプ24に出力する。

10

【0078】

(パワーアンプ24)

パワーアンプ24は、逆相信号生成部16から出力された逆相信号を増幅する機能を有する。パワーアンプ24は、パワーアンプ22と同程度に増幅してもよい。パワーアンプ24は、増幅した逆相信号をマスキング用スピーカ4-2に出力する。

【0079】

(マスキング用スピーカ4-2)

マスキング用スピーカ4-2は、マスキング音声信号の逆相信号を再生する出力装置(第2のスピーカ)である。具体的には、マスキング用スピーカ4-2は、パワーアンプ24から出力された逆相信号を、マスキング用スピーカ4-1によるマスキング音声信号の再生と同時に再生する。マスキング用スピーカ4-2は、マスキング用スピーカ4-1より再生されたマスキング音声信号と、マスキング用スピーカ4-2より再生された逆相信号とが、マイク3が収音する空間において打ち消し合うよう設置される。マスキング用スピーカ4-2は、マスキング用スピーカ4-1と同一のスピーカ特性を有する。また、図10に示したように、マスキング用スピーカ4-2は、マイク3の位置を中心として、マスキング用スピーカ4-1と幾何学的に対称な位置に設置される。

20

【0080】

マスキング用スピーカ4-1から再生されたマスキング音声信号とマスキング用スピーカ4-2が再生する逆相信号とは、ぶつかり合う領域で互いに打ち消し合う。このような領域を、以下ではキャンセル領域とも称する。スマートフォン1-4におけるキャンセル領域について、図11を参照して説明する。

30

【0081】

図11は、第3の実施形態に係るキャンセル領域を示す説明図である。図11(A)に示した各ブロックは、スマートフォン1-4が内部に有している。図11(A)に示すように、スマートフォン1-4のキャンセル領域5-1は、マスキング音声信号および逆相信号が同時に再生されるため、マスキング用スピーカ4-1およびマスキング用スピーカ4-2のほぼ中間地点に形成される。キャンセル領域5-1がマイク3を覆うため、マイク3が収音する空間でマスキング音声信号が打ち消される。このようにして、スマートフォン1-4は、空間音響的に収音信号からマスキング音声信号成分を除去することができる。さらに、キャンセル領域5-1は、マイク3が収音する空間、即ちユーザ8の口元に位置するため、ユーザ8は、マスキング音声信号による邪魔を受けずに、発話することができる。

40

【0082】

また、一般的に、逆相信号による打消しの効果は、低域周波数であるほど高い。このため、マスキング音声信号が低域であるほど、逆相信号と強く打ち消し合うようになり、マイク3はユーザ8の音声より明瞭に収音することができる。このような低域のマスキング音声信号としては、例えば母音を主な成分とする音声信号がある。また、低域のマスキング音声信号がマスキング用スピーカ4-2により空間音響的に除去されるため、エコー

50

キャンセラ 14 は、特に中高域において電氣的にマスキング音声信号を除去してもよい。スマートフォン 1-4 は、マスキング用スピーカ 4-2 とエコーキャンセラ 14 との併用により、全音域でマスキング音声信号を除去することができる。

【0083】

以上、第 3 の実施形態について説明した。続いて、第 3 の実施形態に係る変形例について説明する。

【0084】

[ 2-3-2 . 変形例 2 ]

本変形例は、マスキング用スピーカ 4-2 は、遅延させた逆相信号を再生することで、マスキング用スピーカ 4-1 およびマスキング用スピーカ 4-2 の中間地点以外の領域にキャンセル領域を形成する形態である。以下、図 11 (B) を参照し、本変形例に係るスマートフォン 1-5 について説明する。

10

【0085】

図 11 (B) に示すように、本変形例に係るスマートフォン 1-5 において、マスキング用スピーカ 4-1 およびマスキング用スピーカ 4-2 は、マイク 3 の位置を中心とした幾何学的に対称な位置に設置されていない。また、スマートフォン 1-5 は、図 10 を参照して上記説明したスマートフォン 1-4 と同様の内部構成を有する。ただし、スマートフォン 1-5 は、図 11 (B) に示したように、ディレイ 17 をさらに有する。以下、ディレイ 17 が有する機能について説明する。

【0086】

( ディレイ 17 )

ディレイ 17 は、入力された音声信号を遅延させて出力する機能を有する。本変形例では、ディレイ 17 は、逆相信号生成部 16 により生成された逆相信号を遅延させる遅延部として機能する。より詳しくは、ディレイ 17 は、マスキング用スピーカ 4-1 より再生されたマスキング音声信号とマスキング用スピーカ 4-2 より再生された逆相信号とが、マイク 3 が收音する空間において打ち消し合うよう、逆相信号を遅延させる。ディレイ 17 は、遅延させた逆相信号を、パワーアンプ 24 に出力する。なお、ディレイ 17 は、特定のフィルタ形式であってもよい。

20

【0087】

ディレイ 17 により遅延された逆相信号は、パワーアンプ 24 により増幅され、マスキング用スピーカ 4-2 により再生される。そして、マスキング用スピーカ 4-2 から再生された逆相信号は、ディレイ 17 により遅延した分だけマスキング用スピーカ 4-2 に近い位置で、マスキング用スピーカ 4-1 から出力されたマスキング音声信号と打ち消し合う。つまり、図 11 (B) に示したように、キャンセル領域 5-2 は、マスキング用スピーカ 4-2 に近い位置に形成され、マスキング用スピーカ 4-1 と比較してマスキング用スピーカ 4-2 に近い位置に設置されたマイク 3 を覆う。

30

【0088】

このため、スマートフォン 1-5 は、マスキング用スピーカ 4-1 およびマスキング用スピーカ 4-2 を、マイク 3 の位置を中心とした幾何学的に対称な位置に設置することなく、收音信号からマスキング音声信号成分を除去することができる。さらに、マスキング用スピーカ 4-2 は、マスキング用スピーカ 4-1 と異なるスピーカ特性を有していてもよい。このように、スマートフォン 1-5 は、マスキング用スピーカ 4-2 を設置する位置およびスピーカ特性に係る制約を、ディレイ 17 による遅延の効果により緩和することができる。このため、スマートフォン 1-5 は、マスキング用スピーカ 4-2、マスキング用スピーカ 4-1 の大きさ、位置関係、全体のデザイン等について自由に設計され得る。

40

【0089】

以上、変形例 2 について説明した。続いて、第 3 の実施形態に係る他の変形例について説明する。

【0090】

50

## [ 2 - 3 - 3 . 変形例 3 ]

本変形例は、ヘッドセット 6 により、本開示の一実施形態に係る信号処理装置を実現する形態である。以下、図 1 2 を参照し、本変形例に係るヘッドセット 6 について説明する。

## 【 0 0 9 1 】

図 1 2 は、変形例 3 に係るヘッドセット 6 を示す説明図である。図 1 2 に示すように、ヘッドセット 6 は、マスキング用スピーカ 4 - 1、マスキング用スピーカ 4 - 2、およびマイク 3 を有し、ユーザ 8 の頭部に装着される。ヘッドセット 6 は、図 1 1 ( B ) を参照して上記説明したスマートフォン 1 - 5 と同様の構成を有する。図 1 2 に示すように、マイク 3 は、マスキング用スピーカ 4 - 2 に近い位置に設置されているので、ヘッドセット 6 は、ディレイ 1 7 により遅延された逆相信号をマスキング用スピーカ 4 - 2 から再生することで、マイク 3 をキャンセル領域で覆うことができる。このように、ヘッドセット 6 においても、空間音響的に收音信号からマスキング音声信号成分を除去することができる。

10

## 【 0 0 9 2 】

以上、変形例 3 について説明した。

## 【 0 0 9 3 】

## &lt; &lt; 3 . まとめ &gt; &gt;

以上説明したように、本開示の一実施形態に係るスマートフォン 1 は、ユーザ音声に応じたマスキング音声信号を生成および再生することで、ユーザ 8 の発話内容が聞き取られることを防ぐことができる。より詳しくは、スマートフォン 1 は、他人 9 に対して混同を生じさせる、または注意を逸らせるマスキング音声信号を生成および再生することで、ユーザ 8 の発話をマスキング音声信号に埋没させ、発話内容の聞き取りを妨害することができる。また、スマートフォン 1 は、收音信号のうちユーザ音声が含まれる時間区間にのみマスキング音声信号を再生することで、他人 9 がマスキング音声信号に耳慣れすることを防止することができる。

20

## 【 0 0 9 4 】

さらに、スマートフォン 1 は、收音信号からマスキング音声信号成分を電氣的に除去することで、雑音を低減した高品質な通話や音声認識を実現することができる。また、スマートフォン 1 は、マスキング音声信号を再生するスピーカを複数設け、互いに打ち消し合わせることで、空間音響的に收音信号からマスキング音声信号成分を除去することもできる。

30

## 【 0 0 9 5 】

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、特許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。

## 【 0 0 9 6 】

例えば、上記実施形態では、ユーザ 8 が通話や音声認識入力を行う際にマスキング音声信号を生成および再生するものとして説明したが、本技術はかかる例に限定されない。例えば、ユーザ 8 の寝言や独り言、愚痴を他人には聞かれないための消音装置として、本技術が用いられてもよい。

40

## 【 0 0 9 7 】

また、情報処理装置に内蔵される CPU、ROM および RAM などのハードウェアを、上述したスマートフォン 1 の各構成と同等の機能を発揮させるためのコンピュータプログラムも作成可能である。また、該コンピュータプログラムを記憶した記憶媒体も提供される。

## 【 0 0 9 8 】

なお、以下のような構成も本開示の技術的範囲に属する。

50

( 1 )

ユーザ音声を收音し、オーディオ信号を生成する收音部と、  
前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成する信号処理部と、  
前記マスキング音声信号を再生する第 1 のスピーカと、  
を備える信号処理装置。

( 2 )

前記信号処理部は、前記オーディオ信号のうち前記ユーザ音声が含まれる時間区間にのみ前記マスキング音声信号を生成する、前記 ( 1 ) に記載の信号処理装置。

( 3 )

前記信号処理装置は、除去部をさらに備え、  
前記除去部は、前記收音部が前記第 1 のスピーカから再生された前記マスキング音声信号を前記ユーザ音声と共に收音して前記オーディオ信号を生成した場合、特定の伝達関数および前記信号処理部が生成した前記マスキング音声信号に基づいて、前記收音部により生成された前記オーディオ信号から前記マスキング音声信号を除去する、前記 ( 1 ) または ( 2 ) に記載の信号処理装置。

( 4 )

前記信号処理装置は、前記マスキング音声信号の逆相信号を再生する第 2 のスピーカをさらに備え、

前記第 2 のスピーカは、前記第 1 のスピーカより再生された前記マスキング音声信号と前記第 2 のスピーカより再生された前記逆相信号とが前記收音部が收音する空間において打ち消し合うよう設置される、前記 ( 1 ) ~ ( 3 ) のいずれか一項に記載の信号処理装置。

( 5 )

前記信号処理装置は、前記逆相信号を遅延させる遅延部をさらに備え、  
前記第 2 のスピーカは、前記遅延部により遅延された前記逆相信号を再生する、前記 ( 4 ) に記載の信号処理装置。

( 6 )

前記信号処理部は、前記ユーザ音声を組成する周波数成分ごとのデータ量に応じて前記マスキング音声信号を生成する、前記 ( 1 ) ~ ( 5 ) のいずれか一項に記載の信号処理装置。

( 7 )

前記マスキング音声信号は、音声帯域の帯域ノイズである、前記 ( 1 ) ~ ( 6 ) のいずれか一項に記載の信号処理装置。

( 8 )

前記マスキング音声信号は、母音を主な成分とする音声信号である、前記 ( 1 ) ~ ( 6 ) のいずれか一項に記載の信号処理装置。

( 9 )

前記信号処理装置は、前記收音部により收音された前記ユーザ音声を記録する記録部をさらに備え、

信号処理部は、前記記録部により記録された前記ユーザ音声により前記マスキング音声信号を生成する、前記 ( 1 ) ~ ( 8 ) のいずれか一項に記載の信号処理装置。

( 10 )

前記信号処理装置は、前記收音部により收音される前記ユーザ音声の言語を認識する言語認識部をさらに備え、

前記信号処理部は、前記言語認識部により認識された前記言語に応じて前記マスキング音声信号を生成する、前記 ( 1 ) ~ ( 9 ) のいずれか一項に記載の信号処理装置。

( 11 )

前記信号処理部は、前記言語認識部により認識された前記言語と同じ言語により前記マスキング音声信号を生成する、前記 ( 10 ) に記載の信号処理装置。

10

20

30

40

50

( 1 2 )

前記信号処理部は、前記言語認識部により認識された前記言語と異なる言語により前記マスキング音声信号を生成する、前記( 1 0 )に記載の信号処理装置。

( 1 3 )

前記信号処理装置は、前記オーディオ信号を外部に送信し、外部からのオーディオ信号を受信する通信部をさらに備える、前記( 1 ) ~ ( 1 2 )のいずれか一項に記載の信号処理装置。

( 1 4 )

前記信号処理装置は、  
前記オーディオ信号から制御情報を認識する制御情報認識部と、  
前記制御情報認識部により認識された前記制御情報に基づいて前記信号処理装置を制御する制御部と、  
をさらに備える、前記( 1 ) ~ ( 1 3 )のいずれか一項に記載の信号処理装置。

10

( 1 5 )

ユーザ音声を収録し、オーディオ信号を生成するステップと、  
前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成するステップと、  
前記マスキング音声信号を再生するステップと、  
を備える信号処理方法。

20

( 1 6 )

コンピュータに、  
ユーザ音声を収録し、オーディオ信号を生成するステップと、  
前記オーディオ信号に応じて、前記ユーザ音声をマスキングするためのマスキング音声信号を生成するステップと、  
前記マスキング音声信号を再生するステップと、  
を実行させるためのプログラムが記憶された記憶媒体。

20

【符号の説明】

【 0 0 9 9 】

- 1、 1 - 1、 1 - 2、 1 - 3、 1 - 4、 1 - 5 スマートフォン
- 2 通話用スピーカ
- 3 マイク
- 4、 4 - 1、 4 - 2 マスキング用スピーカ
- 5 - 1、 5 - 2 キャンセル領域
- 6 ヘッドセット
- 8 ユーザ
- 9 他人
- 1 1 制御部
- 1 2、 1 2 - 1、 1 2 - 2 信号処理部
- 1 3 加算器
- 1 4 エコーキャンセラ
- 1 5 加算器
- 1 6 逆相信号生成部
- 1 7 ディレイ
- 2 1 マイクアンプ
- 2 2、 2 3、 2 4 パワーアンプ
- 3 1 送話部
- 3 2 受話部
- 4 1 マスキング用音源
- 1 0 0 スマートフォン
- 1 2 0 - 1、 1 2 0 - 2 音声信号例

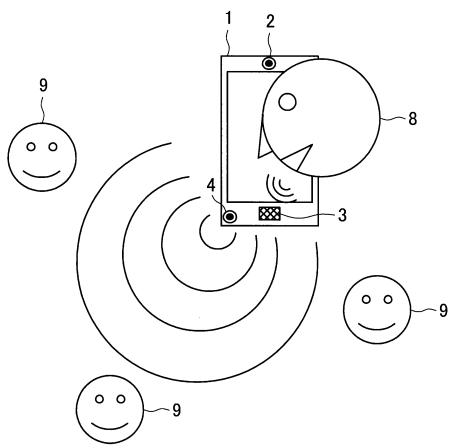
30

40

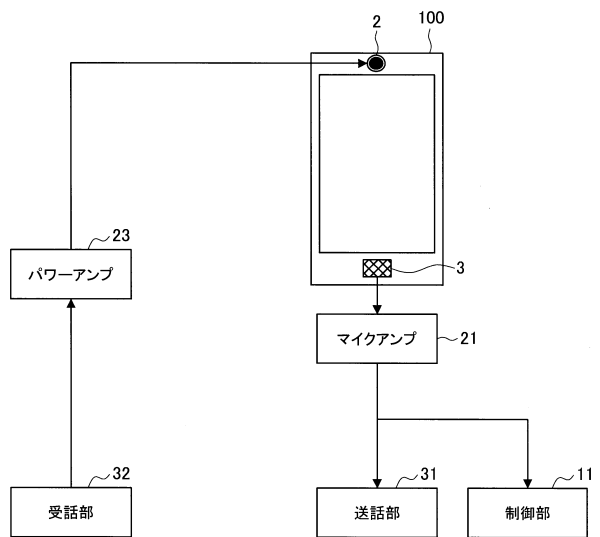
50

- 1 2 1 解析用 B P F 群
- 1 2 2 可変ゲインブロック群
- 1 2 3 合成用 B P F 群
- 1 2 4 加算器
- 1 2 5 V A D
- 1 2 6 スイッチ

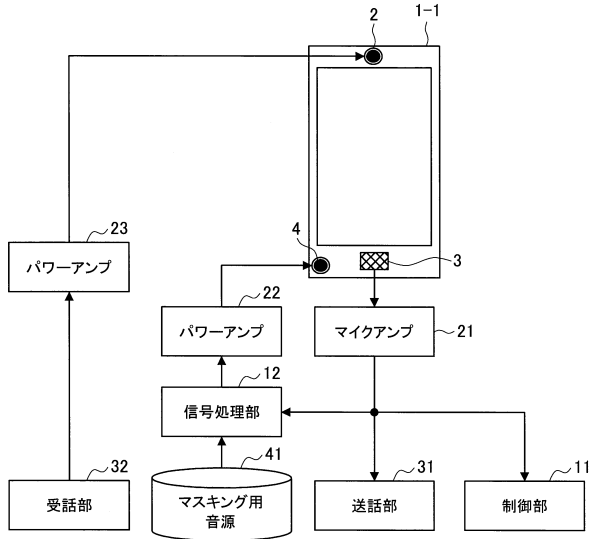
【図 1】



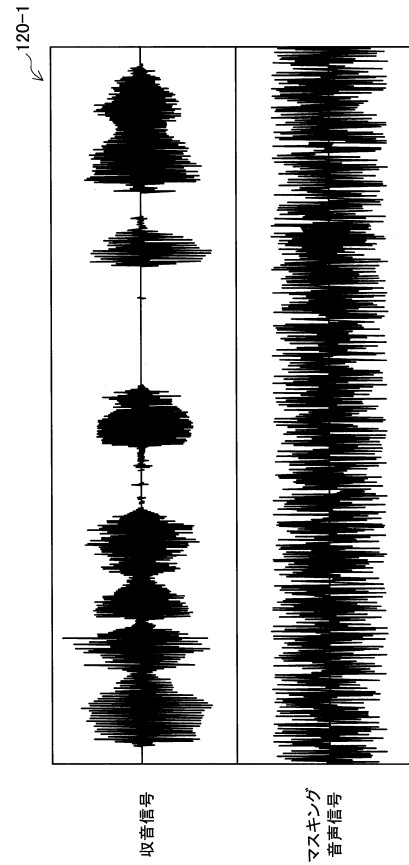
【図 2】



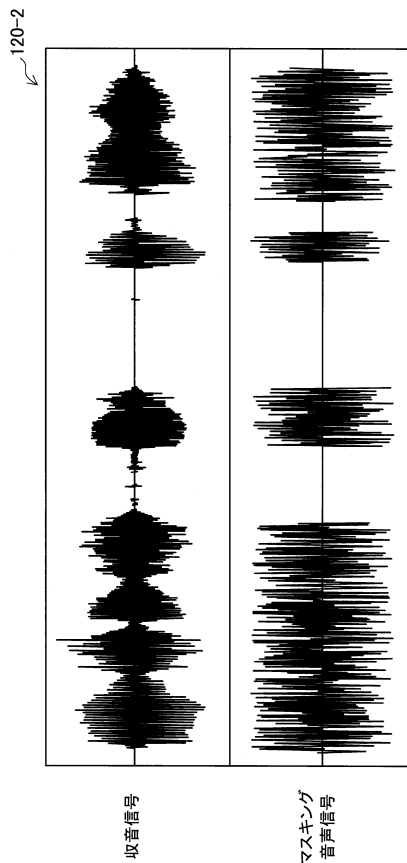
【図3】



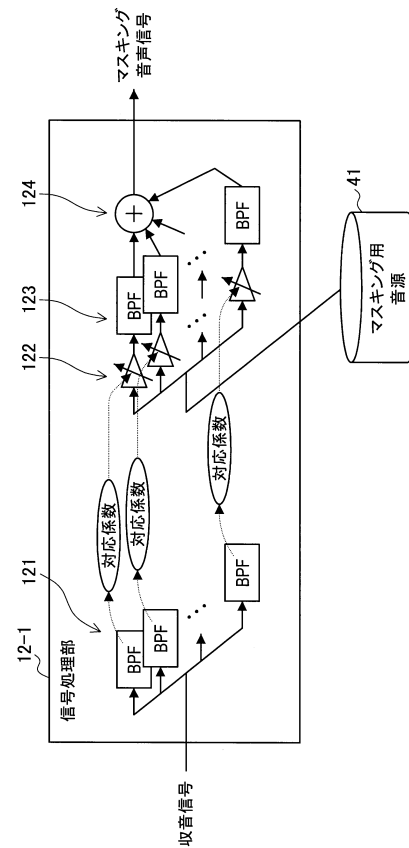
【図4A】



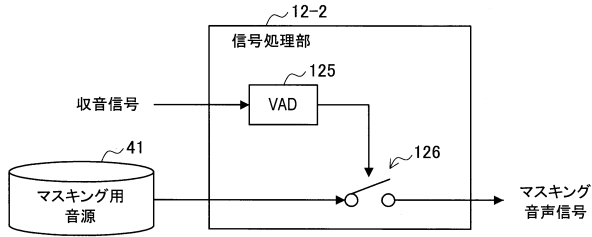
【図4B】



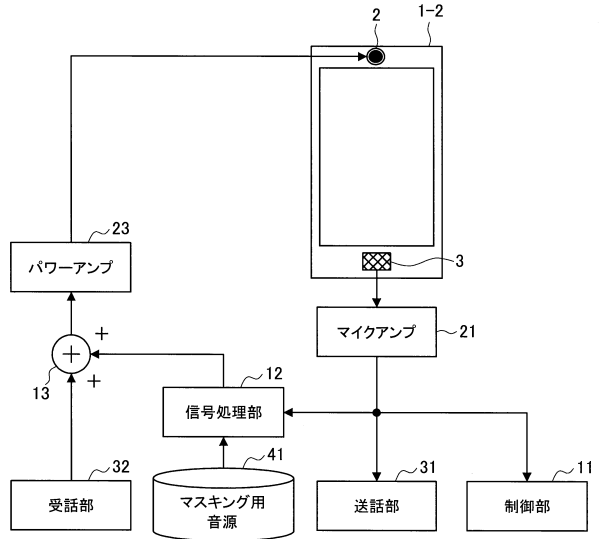
【図5】



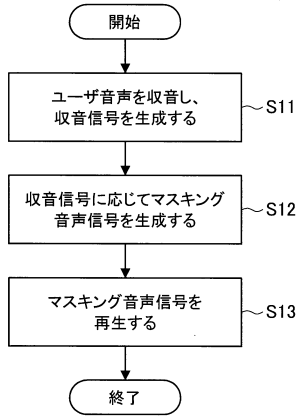
【図6】



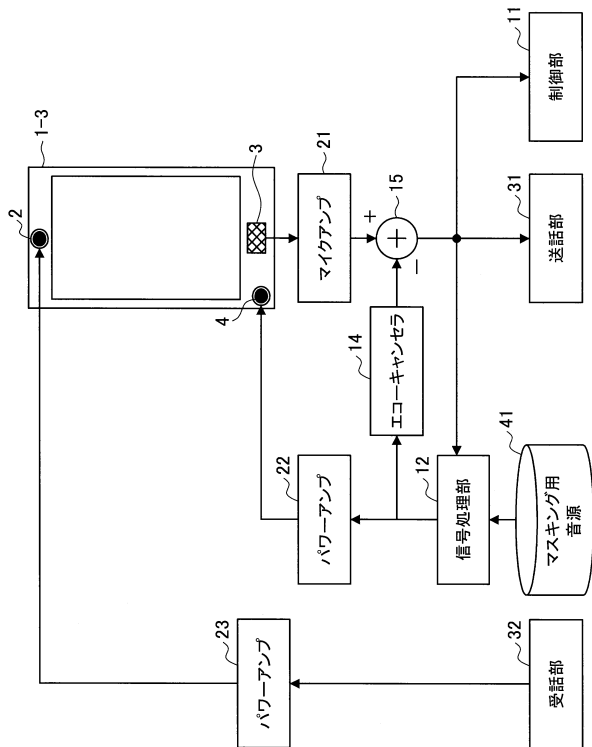
【図8】



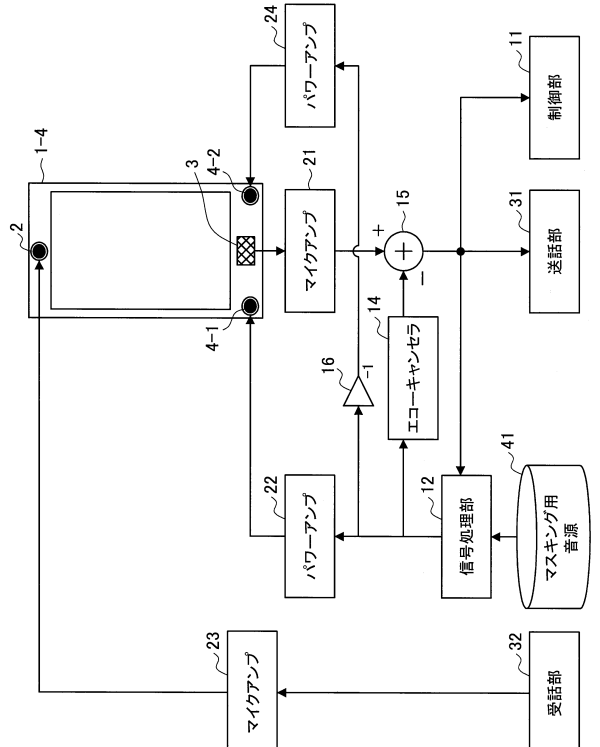
【図7】



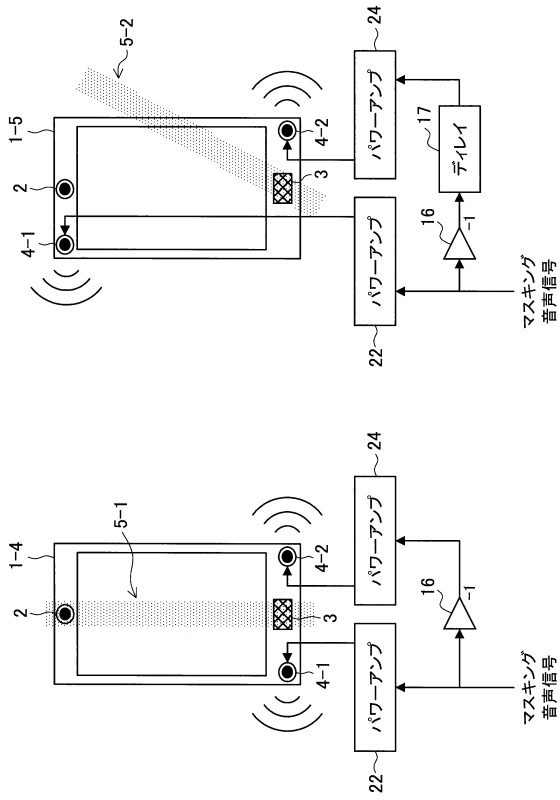
【図9】



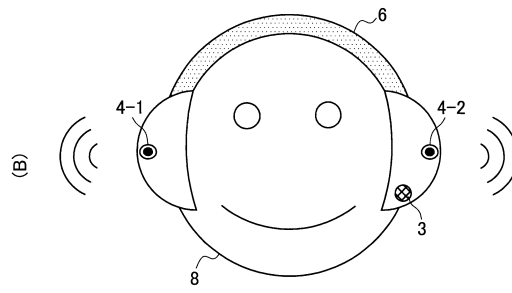
【図10】



【図11】



【図12】



(A)

(B)

## フロントページの続き

- (72)発明者 佐古 曜一郎  
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 迫田 和之  
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 竹原 充  
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 中村 隆俊  
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 丹下 明  
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 花谷 博幸  
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 甲賀 有希  
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 大沼 智也  
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2006-243178(JP,A)  
特開2007-006363(JP,A)  
特開平08-296335(JP,A)  
特開2011-123141(JP,A)  
特開2008-233672(JP,A)  
特開2005-084645(JP,A)

## (58)調査した分野(Int.Cl., DB名)

G06F 3/16  
G10L 15/00 - 17/26  
G10K 11/00 - 13/00  
H04M 1/00  
H04M 1/02 - 1/23  
H04M 1/24 - 1/82  
H04M 99/00