(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5017441号
(P5017441)

(45)発行日　平成24年9月5日(2012.9.5)　(24)登録日　平成24年6月15日(2012.6.15)

(51)Int.Cl.		F I		
G10L 15/20	(2006.01)	G10L	15/20	370D
G10L 15/28	(2006.01)	G10L	15/28	400
G10L 15/00	(2006.01)	G10L	15/00	200C
G10L 21/02	(2006.01)	G10L	15/20	370E
		G10L	21/02	101Z

請求項の数 9　(全 17 頁)

(21)出願番号　特願2010-242474 (P2010-242474)
(22)出願日　平成22年10月28日 (2010.10.28)
(65)公開番号　特開2012-93641 (P2012-93641A)
(43)公開日　平成24年5月17日 (2012.5.17)
審査請求日　平成23年11月16日 (2011.11.16)

早期審査対象出願

(73)特許権者　000003078
株式会社東芝
東京都港区芝浦一丁目1番1号
(74)代理人　100108855
弁理士　蔵田　昌俊
(74)代理人　100091351
弁理士　河野　哲
(74)代理人　100088683
弁理士　中村　誠
(74)代理人　100109830
弁理士　福原　淑弘
(74)代理人　100075672
弁理士　峰　隆司
(74)代理人　100095441
弁理士　白根　俊郎

最終頁に続く

(54)【発明の名称】携帯型電子機器

(57)【特許請求の範囲】

【請求項１】

タッチスクリーンディスプレイを備えた本体を具備し、前記タッチスクリーンディスプレイ上のタップ位置に対応する表示オブジェクトに関連づけられた機能を実行するように構成された携帯型電子機器であって、

前記本体に取り付けられた少なくとも一つのマイクロホンと、

前記本体内に設けられ、前記少なくとも一つのマイクロホンからの入力音声信号を処理する音声処理手段と、

前記本体内に設けられ、前記音声処理手段によって処理された入力音声信号を認識および機械翻訳することによって得られる目的言語の翻訳結果を出力する翻訳結果出力手段とを具備し、

前記音声処理手段は、前記タッチスクリーンディスプレイ上をタップすることによって発生するタップ音信号の波形を示す予め用意された検出対象音波形と前記入力音声信号の波形との間の相関を算出することによって前記入力音声信号内に含まれる前記タップ音信号を検出し、前記検出されたタップ音信号に対応する信号部分を前記入力音声信号から削除する携帯型電子機器。

【請求項２】

前記翻訳結果出力手段は、前記目的言語の翻訳結果を示すテキストを音声信号に変換し、前記変換によって得られた音声信号に対応する音を出力する請求項１記載の携帯型電子機器。

【請求項３】
前記翻訳結果出力手段は、前記目的言語の翻訳結果を示すテキストを音声信号に変換し、前記変換によって得られた音声信号に対応する音を出力すると共に、前記目的言語の翻訳結果を示すテキストを前記タッチスクリーンディスプレイ上に表示する請求項１記載の携帯型電子機器。
【請求項４】
前記翻訳結果出力手段は、前記目的言語の翻訳結果を示すテキストを音声信号に変換し、少なくとも前記変換によって得られた音声信号に対応する音を含む音声信号を出力するように構成されており、
前記変換によって得られた音声信号に対応する音を含む音声信号の出力中における音声入力を可能にするために、前記入力音声信号から前記変換によって得られた音声信号を含む音声信号成分を軽減するエコーキャンセル手段をさらに具備する請求項１記載の携帯型電子機器。
【請求項５】
前記音声処理手段によって処理された入力音声信号を格納するバッファと、
前記バッファに格納された入力音声信号内の発話区間を検出し、前記バッファに格納された入力音声信号内に含まれ且つ前記検出された発話区間に属する音声信号を、認識対象の音声信号として出力する発話検出手段をさらに具備する請求項１記載の携帯型電子機器。
【請求項６】
前記本体には複数のマイクロホンが取り付けられており、
前記複数のマイクロホンそれぞれからの入力音声信号群を用いて、それら入力音声信号それぞれに対応する話者が位置する前記本体に対する方向を推定し、前記推定結果に基づいて、前記入力音声信号群から、前記本体に対して特定の方向からの入力音声信号を抽出する話者方向推定手段をさらに具備する請求項１記載の携帯型電子機器。
【請求項７】
前記本体には複数のマイクロホンが取り付けられており、
前記複数のマイクロホンそれぞれからの入力音声信号群を用いて、それら入力音声信号それぞれに対応する話者が位置する前記本体に対する方向を推定し、前記推定結果に基づいて、前記複数のマイクロホンそれぞれからの入力音声信号群を前記話者ごとに分類する話者分類手段をさらに具備する請求項１記載の携帯型電子機器。
【請求項８】
タッチスクリーンディスプレイを備えた本体を具備し、前記タッチスクリーンディスプレイ上に被案内者に対する案内画面を表示すると共に、前記タッチスクリーンディスプレイ上のタップ位置に対応する表示オブジェクトに関連づけられた機能を実行するように構成された携帯型電子機器であって、
前記本体に取り付けられた少なくとも一つのマイクロホンと、
前記本体内に設けられ、前記少なくとも一つのマイクロホンを用いて案内者および前記被案内者それぞれからの入力音声信号を処理する音声処理手段と、
前記本体内に設けられ、前記音声処理手段によって処理された前記案内者の入力音声信号を認識および機械翻訳することによって得られる、前記被案内者が使用する第２の言語の翻訳結果と、前記音声処理手段によって処理された前記被案内者の入力音声信号を認識および機械翻訳することによって得られる、前記案内者が使用する第１の言語の翻訳結果と出力する翻訳結果出力手段とを具備し、
前記音声処理手段は、前記タッチスクリーンディスプレイ上をタップすることによって発生するタップ音信号の波形を示す予め用意された検出対象音波形と前記案内者および前記被案内者それぞれからの入力音声信号の波形との間の相関を算出することによって前記各入力音声信号内に含まれる前記タップ音信号を検出し、前記検出されたタップ音信号に対応する信号部分を前記各入力音声信号から削除する携帯型電子機器。
【請求項９】

前記翻訳結果出力手段は、前記第２の言語の翻訳結果を示すテキストを第１の音声信号に変換し、前記第１の言語の翻訳結果を示すテキストを第２の音声信号に変換し、前記第１の音声信号に対応する音と前記第２の音声信号に対応する音とを出力する請求項８記載の携帯型電子機器。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明の実施形態は、音声信号を利用して各種サービスを実行するための携帯型電子機器に関する。

【背景技術】

【０００２】

近年、スマートフォン、ＰＤＡ、スレートＰＣといった様々な携帯型電子機器が開発されている。このような携帯型電子機器の多くはタッチスクリーンディスプレイ（タッチパネル式ディスプレイとも云う）を備えている。ユーザは、タッチスクリーンディスプレイ上を指でタップすることにより、そのタップ位置に関連付された機能の実行を携帯型電子機器に対して指示することができる。

【０００３】

また、最近では、音声認識機能および音声合成機能の性能が大幅に向上している。このため、携帯型電子機器においても、音声認識機能および音声合成機能等を用いたサービスを実行するための機能の搭載が要求され始めている。

【０００４】

音声認識機能を備えた機器の例としては、携帯型機械翻訳機器が知られている。この機械翻訳機器は、第１の言語の音声を認識し、その認識結果である文字データを第２の言語の文字データに翻訳する。この第２の言語の文字データは音声合成によって音声に変換され、そしてその音声がスピーカから出力される。

【先行技術文献】

【特許文献】

【０００５】

【特許文献１】特開２００３－１０８５５１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、音声認識の精度はノイズによって大きく影響される。一般に、音声認識技術の分野では、バックグラウンドノイズのような定常ノイズを除去するための様々な技術が利用されている。ここで、定常ノイズとは、時間的に連続して発生するノイズのことを意味する。定常ノイズの周波数特性は、例えば、無発話区間の音声信号を解析することによって算出することができる。周波数領域で入力音声信号から定常ノイズ成分を除去するための演算を行うことにより、定常ノイズによる影響を低減することができる。

【０００７】

しかし、携帯型電子機器において、定常ノイズのみならず、非定常ノイズが音声認識の精度に大きく影響を及ぼす可能性がある。非定常ノイズは、たとえば、いつ発生するかわからず、且つ瞬時的に発生するノイズである。この非定常ノイズとしては、音声入力中における、機器に対する接触音、周辺話者音声、機器のスピーカから再生される音、等があげられる。

【０００８】

音声認識機能を有する多くの携帯型電子機器においては、マイクロホンは、その携帯型電子機器の本体に取り付けられている。このため、もし音声入力中にユーザが機器の本体に触れると、機器の振動に対応する音がマイクロホンによって入力されてしまうことがある。特に、タッチスクリーンディスプレイを備えた機器においては、例えば、もし音声入力中にユーザがタッチスクリーンディスプレイをタップすると、そのタップ音によって入

力音声にノイズ（非定常ノイズ）が入り込む可能性がある。

【０００９】

音声入力中は他の操作を禁止するという方法を用いれば、入力音声にノイズ（非定常ノイズ）が入り込むことを軽減できる。しかし、もしこの方法を用いると、音声入力中は、ユーザは電子機器に対する他の操作を一切行うことができないので、携帯型電子機器の使い勝手が低下する。

【００１０】

本発明の目的は、非定常ノイズの影響を低減することによって音声入力中に他の操作を実行することができる携帯型電子機器を提供することである。

【課題を解決するための手段】

【００１１】

実施形態によれば、携帯型電子機器は、タッチスクリーンディスプレイを備えた本体を具備し、前記タッチスクリーンディスプレイ上のタップ位置に対応する表示オブジェクトに関連づけられた機能を実行するように構成されている。前記携帯型電子機器は、前記本体に取り付けられた少なくとも一つのマイクロホンと、前記本体内に設けられ、前記少なくとも一つのマイクロホンからの入力音声信号を処理する音声処理手段と、前記本体内に設けられ、前記音声処理手段によって処理された入力音声信号を認識および機械翻訳することによって得られる目的言語の翻訳結果を出力する翻訳結果出力手段とを具備する。前記音声処理手段は、前記タッチスクリーンディスプレイ上をタップすることによって発生するタップ音信号の波形を示す予め用意された検出対象音波形と前記入力音声信号の波形との間の相関を算出することによって前記入力音声信号内に含まれる前記タップ音信号を検出し、前記検出されたタップ音信号に対応する信号部分を前記入力音声信号から削除する。

【図面の簡単な説明】

【００１２】

【図１】実施形態に係る携帯型電子機器の外観を示す図。

【図２】同実施形態の携帯型電子機器のユースケースを示す図。

【図３】同実施形態の携帯型電子機器のシステム構成の例を示すブロック図。

【図４】同実施形態の携帯型電子機器によって検出されるタップ音信号の波形例を示す図。

【図５】同実施形態の携帯型電子機器によって検出されるサチレーション波形例を示す図。

【図６】同実施形態の携帯型電子機器に入力される、タップ音信号を含む入力音声信号の波形例を示す図。

【図７】同実施形態の携帯型電子機器によって実行される、タップ音信号を除去するための音声信号補正処理の例を説明するための図。

【図８】同実施形態の携帯型電子機器のシステム構成の別の例を示すブロック図。

【図９】同実施形態の携帯型電子機器のシステム構成のさらに別の例を示すブロック図。

【図１０】同実施形態の携帯型電子機器によって検出される発話区間の例を示す図。

【図１１】同実施形態の携帯型電子機器によって実行される発話区間検出処理の手順を示すフローチャート。

【図１２】同実施形態の携帯型電子機器のシステム構成のさらに別の例を示すブロック図。

【図１３】同実施形態の携帯型電子機器のシステム構成のさらに別の例を示すブロック図。

【発明を実施するための形態】

【００１３】

以下、図面を参照して、実施形態を説明する。

まず、図１を参照して、実施形態に係る携帯型電子機器の構成を説明する。この携帯型電子機器は、たとえば、スマートフォン、ＰＤＡ、またはスレートＰＣ等として実現する

ことができる。この携帯型電子機器は、タッチスクリーンディスプレイ１１を備えた本体１０を備えている。より詳しくは、本体１０は薄い箱状の筐体を有しており、その筐体の上面上にタッチスクリーンディスプレイ１１が設けられている。タッチスクリーンディスプレイ１１はその画面上のタップ位置（タッチ位置）を検出可能なディスプレイである。このタッチスクリーンディスプレイ１１は、たとえば、ＬＣＤのようなフラットパネルディスプレイとタッチパネルとから構成することができる。

【００１４】

この携帯型電子機器は、タッチスクリーンディスプレイ１１上のタップ位置に対応する表示オブジェクト（メニュー、ボタン、等）に関連づけられた機能を実行することができる。たとえば、この携帯型電子機器は、タッチスクリーンディスプレイ１１上に表示される画像（案内図等）と音声とを利用した様々なサービス、たとえば、旅行者に対して海外旅行における会話等をサポートするサービス、店員に対して外国人観光客に対する接客をサポートするサービス、等を実行することができる。これらサービスは、携帯型電子機器が有する音声入力機能、音声認識機能、機械翻訳機能、音声合成（テキスト・ツー・スピーチ）機能等を用いて実現することができる。これら機能の全てを携帯型電子機器によって実行してもよいが、これら機能の一部またはほとんど全てをネットワーク２０上のサーバ２１によって実行してもよい。たとえば、音声認識機能および機械翻訳機能をネットワーク２０上のサーバ２１によって実行し、音声入力機能および音声合成（テキスト・ツー・スピーチ）機能を携帯型電子機器によって実行してもよい。この場合、サーバ２１は、携帯型電子機器から受信した音声信号を認識する自動音声認識（ＡＳＲ）機能、ＡＳＲによって得られたテキストを目的言語に翻訳する機械翻訳（ＭＴ）機能等を有してればよい。携帯型電子機器は、機械翻訳（ＭＴ）によって得られる目的言語の翻訳結果をサーバ２１から受信することができる。携帯型電子機器は、受信した翻訳結果が示すテキストを音声信号に変換し、この音声信号に対応する音をスピーカから出力してもよい。また、携帯型電子機器は、受信した翻訳結果が示すテキストを、タッチスクリーンディスプレイ１１上に表示してもよい。

【００１５】

本体１０には１つ以上のマイクロホンが設けられている。これら１つ以上のマイクロホンは音声信号を入力するために用いられる。図１においては、本体１０の上端部の左端および右端にそれぞれマイクロホン１２Ａ，１２Ｂが設けられている構成例が例示されている。

【００１６】

ここで、ショッピングモールの店員（案内者）が外国人観光客（外国人）を接客するのをサポートするサービスを例示して、タッチスクリーンディスプレイ１１に表示される画面の例を説明する。図２に示すように、店員（案内者）３１と外国人（被案内者）３２の双方はタッチスクリーンディスプレイ１１の表示画面を見ながら会話する。店員３１は、たとえば左腕で携帯型電子機器を持ち、発話しながら、右手の指でタッチスクリーンディスプレイ１１の画面をタッチ操作（タップ操作、ドラッグ操作等、）する。

【００１７】

たとえば、ショッピングモールで外国人３２が「○○売り場はどこですか」と売り場を聞いてきたとき、店員３１は「○○売り場でございますね」などと発話しながら、タッチスクリーンディスプレイ１１を操作して「○○売り場」の売り場地図をタッチスクリーンディスプレイ１１上に表示する。その間、店員が発した音声「○○売り場でございますね」は目的言語（外国人３２が使用する言語）に翻訳され、その翻訳結果が携帯型電子機器から出力される。この場合、携帯型電子機器は、目的言語の翻訳結果を示すテキストを音声信号に変換し、この音声信号に対応する音を出力してもよい。また、携帯型電子機器は、目的言語の翻訳結果を示すテキストをタッチスクリーンディスプレイ１１上に表示してもよい。もちろん、携帯型電子機器は、目的言語の翻訳結果を示すテキストを音声信号に変換し、この音声信号に対応する音を出力すると共に、目的言語の翻訳結果を示すテキストをタッチスクリーンディスプレイ１１上に表示してもよい。

【００１８】
さらに、携帯型電子機器は、外国人３２の発話「○○売り場はどこですか」を認識および翻訳することによって得られる別の目的言語（店員３１が使用する言語）の翻訳結果を、音声またはテキストによって出力することもできる。
【００１９】
また、携帯型電子機器は、外国人３２の発話の認識結果を示す元言語のテキスト（外国人３２の使用する言語のテキスト）と外国人３２の発話を認識および翻訳することによって得られる翻訳結果を示すテキスト（店員３１が使用する言語のテキスト）とをタッチスクリーンディスプレイ１１上に表示してもよい。
【００２０】
以下では、説明をわかりやすくするために、店員３１が使用する言語が日本語であり、外国人３２の使用する言語が英語である場合を想定して説明するが、本実施形態は、これに限定されず、たとえば、店員３１が使用する言語が英語で外国人３２の使用する言語が中国語であるケース、店員３１が使用する言語が中国語で外国人３２の使用する言語が英語であるケース、等、他の様々なケースに対応できる。
【００２１】
図１に示されているように、タッチスクリーンディスプレイ１１上の表示画面には、たとえば、第１表示領域１３、第２表示領域１４と、第３表示領域１５、発話開始ボタン１８、言語表示領域切り替えボタン１９、等が表示される。第１表示領域１３は、たとえば、外国人３２の発話内容を示す英語のテキストを表示するために用いられる。第２表示領域１４は、たとえば、外国人３２の発話内容を翻訳することによって得られる日本語のテキストを表示するために用いられる。第３表示領域１５は、外国人３２に提示するための案内画面を表示するために用いられる。案内画面には、たとえば、案内図１６、メニュー１７等が表示される。メニュー１７には、案内図１６として表示すべき場所を指示するための様々な項目が表示されている。店員３１はメニュー１７上の複数の項目の一つをタップ操作することにより、案内図１６として表示すべき場所を指示することができる。図１においては、ショッピングモール内の７階のフロア内の売り場それぞれのレイアウトを示す売り場地図（フロア図）が表示される例が示されている。この売り場地図（フロア図）においては、各売り場の名称を示すたとえば日本語のテキストを表示してもよい。店員３１によって売り場マップ中の日本語テキスト（例えば「和食レストランコーナー」など）がタップされた時、そのタップされた日本語テキストを認識および翻訳し、「和食レストランコーナー」に対応する英語のテキストをタッチスクリーンディスプレイ１１上に表示してもよく、あるいはこの英語のテキストを音声信号に変換し、その変換によって得られた音声信号に対応する音を出力してもよい。
【００２２】
なお、売り場の名称を示す日本語文字列をイメージによって案内図１６上に表示してもよい。この場合、携帯型電子機器は、タップされた日本語文字列を文字認識することによって認識すればよい。
【００２３】
発話開始ボタン１８は、音声の入力および認識の開始を指示するためのボタンである。発話開始ボタン１８がタップされた時、携帯型電子機器は、音声の入力および認識を開始してもよい。言語表示領域切り替えボタン１９は、外国人３２の発話内容を示す英語のテキストを表示するため領域と外国人３２の発話内容を翻訳することによって得られる日本語のテキストを表示するための領域を、第１表示領域１３と第２表示領域１４との間で互いに切り替えるために用いられる。
【００２４】
なお、第１表示領域１３および第２表示領域１４それぞれの表示内容は上述の例のみではない。たとえば、店員３１の発話内容を示す日本語のテキストと外国人３２の発話内容を翻訳することによって得られる日本語のテキストの一方または双方を第２表示領域１４に表示し、店員３１の発話内容を翻訳することによって得られる英語のテキストと外国人

３２の発話内容を示す英語のテキストの一方または双方を第１表示領域１３に表示してもよい。

【００２５】

次に、図３を参照して、本実施形態の携帯型電子機器のシステム構成を説明する。

【００２６】

図３の例においては、携帯型電子機器は、入力音声処理部１１０、音声認識（ＡＳＲ）部１１７、機械翻訳（ＭＴ）部１１８、テキスト・ツー・スピーチ（ＴＴＳ）部１１９、メッセージ表示部１２０等を備えている。マイクロホン１２は上述のマイクロホン１２Ａ，１２Ｂを代表して示している。入力音声処理部１１０は、マイクロホン１２からの入力音声信号を処理する音声処理部である。

【００２７】

この入力音声処理部１１０は、店員３１が発話しながら携帯型電子機器を操作できるようにするために、入力音声信号内に含まれるタップ音信号を検出し、この検出されたタップ音信号による入力音声信号への影響を軽減するために、入力音声信号を補正するように構成されている。タップ音信号は、タッチスクリーンディスプレイ１１上をタップすることによって発生される音の信号である。上述のように、マイクロホン１２は本体１０に直接的に取り付けられているので、もし音声入力中に店員３１がタッチスクリーンディスプレイ１１をタップすると、そのタップ音によってマイクロホン１２からの入力音声信号にノイズが入る込む可能性がある。入力音声処理部１１０は、このタップ音を入力音声信号から自動的に除去し、タップ音が除去された入力音声信号を後段に出力する。これにより、たとえ店員３１または外国人３２の発話中に店員３１が携帯型電子機器を操作しても、入力音声信号の認識精度に与える影響を低減することができる。よって、店員３１は発話しながら携帯型電子機器を操作することができる。

【００２８】

タップ音は、たとえば、タップ音に対応する音声信号と入力音声信号との間の相関を算出することによって検出することができる。入力音声信号がタップ音に対応する音声信号の波形と類似する波形を含む場合、その類似する波形に対応する期間はタップ音発生期間として検出される。

【００２９】

またタップ音の発生時には、入力音声信号がサチュレーション状態になる可能性がある。このため、入力音声信号がサチュレーション状態である期間も、タップ音発生期間として検出してもよい。

【００３０】

入力音声処理部１１０は、以下の機能を有している。

【００３１】

（１）入力音声処理部１１０は、入力音声信号（入力波形）をフレーム単位で処理する。

【００３２】

（２）入力音声信号（入力波形）のサチレーション位置を検出する機能

（３）入力音声信号（入力波形）とタップ音に対応する音声信号の波形との間の相互相関を算出する機能

（４）入力音声信号（入力波形）を補正して、入力音声信号（入力波形）からタップ音の波形を除去する機能

以下、入力音声処理部１１０の構成例を説明する。

入力音声処理部１１０は、波形バッファ部１１１、波形補正部１１２、サチレーション位置検出部１１３、相互相関算出部１１４、検出対象音波形格納部１１５、タップ音判定部１１６等を含んでいる。

【００３３】

波形バッファ部１１１は、マイクロホン１２から受信した入力音声信号（入力波形）を一時的に格納するメモリである。波形補正部１１２は、入力音声信号（入力波形）からタップ音信号を除去するために、波形バッファ部１１１に格納された入力音声信号（入力波

形）を補正する。この補正では、入力音声信号からタップ音発生期間に対応する信号部分（タップ音発生期間に対応する波形部分）を削除してもよい。上述したようにタップ音は瞬時ノイズであるので、タップ音発生期間は非常に短い（たとえば、２０ｍｓから４０ｍｓ程度）。したがって、たとえ入力音声信号からタップ音発生期間に対応する信号部分を削除しても、入力音声信号に対する音声認識精度に悪影響を与えることはない。もし入力音声信号の周波数からタップ音の周波数を差し引くという周波数演算処理を行うと、この周波数演算処理によって入力音声信号に異音が入り込む可能性がある。よって、入力音声信号からタップ音発生期間に対応する信号部分を削除するとい方法は、周波数演算処理を用いるよりも、非定常ノイズの除去に好適である。

【００３４】

サチレーション位置検出部１１３は、マイクロホン１２から受信した入力音声信号（入力波形）内のサチレーション位置を検出する。入力音声信号の振幅レベルが最大振幅レベル付近または最小振幅レベル付近に達している状態がある期間中連続する場合、サチレーション位置検出部１１３は、その期間をサチレーション位置情報として検出してもよい。相互相関算出部１１４は、検出対象音波形（タップ波形）格納部１１５に格納された検出対象音波形（タップ音波形）と入力音声信号の波形との間の相互相関を算出する。検出対象音波形（タップ波形）格納部１１５には、タップ音信号の波形、つまりタッチパネルディスプレイをタップした時に発生する音声信号の波形が検出対象音波形として事前に格納されている。タップ音信号の波形の例を図４に示す。図４の横軸は時間を表し、また縦軸は振幅を表している。

【００３５】

タップ音判定部１１６は、入力音声信号に含まれるタップ音信号を検出するために、入力音声信号の現在のフレームがタップ音であるか否かを、サチレーション位置情報（サチレーション時間情報とも云う）と相互相関値とに基づいて判定する。この判定は、例えば、サチレーション位置情報と相互相関値との加重平均に基づいて行ってもよい。

【００３６】

もちろん、相互相関値とサチレーション位置情報とを個別に用いてもよい。入力音声信号がサチレーションを起こしている場合はその入力音声信号の波形が崩れるため、波形の相互相関では、タップ音を検出できない場合がある。しかし、サチレーション位置情報によってサチレーションを起こしている、入力音声信号内の期間を特定することにより、当該期間をタップ音発生期間として検出することができる。サチレーションは、たとえば、タップ操作によって指の爪がタッチスクリーンディスプレイ１１に接触したときに発生しやすい。サチレーションを起こしている音声信号の波形例を図５に示す。図５の横軸は時間を表し、縦軸は振幅を表している。サチレーションを起こしている音声信号の振幅のレベルは、最大振幅レベル付近または最小振幅レベル付近で一定期間継続する。

【００３７】

波形補正部１１２は、タップ音判定部１１６によってタップ音が検出された場合、つまりタップ音判定部１１６によって現在の入力音声信号がタップ音を含むと判定された場合、その入力音声信号からタップ音部分の波形を削除する。さらに、波形補正部１１２は、タップ音部分の前後の波形をオーバーラップ加算することによって、削除したタップ音部分の波形を、タップ音部分の前後の波形を用いて補間してもよい。

【００３８】

音声認識（ＡＳＲ）部１１７は、入力音声処理部１１０よって処理された音声信号を認識し、その音声認識結果を出力する。機械翻訳（ＭＴ）部１１８は、機械翻訳によって音声認識結果を示すテキスト（文字例）を目的言語のテキスト（文字例）に翻訳し、翻訳結果を出力する。

【００３９】

テキスト・ツー・スピーチ（ＴＴＳ）部１１９およびメッセージ表示部１２０は、入力音声処理部１１０によって処理された入力音声信号を認識および機械翻訳することによって得られる目的言語の翻訳結果を出力する翻訳結果出力部として機能する。より詳しくは

、テキスト・ツー・スピーチ（ＴＴＳ）部１１９は、音声合成処理によって、翻訳結果を示すテキストを音声信号に変換し、そして、スピーカ４０を用いて、その変換によって得られた音声信号に対応する音を出力するように構成されている。メッセージ表示部１２０は、翻訳結果を示すテキストをタッチパネルディスプレイ１１上に表示する。

【００４０】

なお、音声認識（ＡＳＲ）部１１７、機械翻訳（ＭＴ）部１１８、テキスト・ツー・スピーチ（ＴＴＳ）部１１９の内の少なくとも一つの機能はサーバ２１によって実行してもよい。たとえば、比較的負荷の小さいテキスト・ツー・スピーチ（ＴＴＳ）部１１９の機能を携帯型電子機器内で実行し、音声認識（ＡＳＲ）部１１７および機械翻訳（ＭＴ）部１１８それぞれの機能をサーバ２１によって実行してもよい。

【００４１】

携帯型電子機器はＣＰＵ（プロセッサ）、メモリ、無線通信部等をハードウェアコンポーネントとして備えている。テキスト・ツー・スピーチ（ＴＴＳ）部１１９の機能は、ＣＰＵによって実行されるプログラムによって実現してもよい。また、音声認識（ＡＳＲ）部１１７、機械翻訳（ＭＴ）部１１８それぞれの機能も、ＣＰＵによって実行されるプログラムによって実現してもよい。また、入力処理部１１０の一部または全ての機能も、ＣＰＵによって実行されるプログラムによって実現してもよい。もちろん、入力処理部１１０の一部または全ての機能を専用または汎用のハードウェアによって実行してもよい。

【００４２】

音声認識（ＡＳＲ）部１１７および機械翻訳（ＭＴ）部１１８それぞれの機能をサーバ２１によって実行する場合には、携帯型電子機器は、入力音声処理部１１０によって処理された音声信号をネットワーク２０を介してサーバ２１に送信し、翻訳結果をネットワーク２０を介してサーバ２１から受信すればよい。携帯型電子機器とネットワーク２０との間の通信は、無線通信部を用いて実行することができる。

【００４３】

次に、図６および図７を参照して、波形補正部１１２によって実行される処理の例を説明する。

【００４４】

図６はタップ音信号を含む入力音声信号の波形例を示している。図６の横軸は時間を表し、縦軸は入力音声信号の振幅を表している。入力音声信号の処理は所定時間のフレーム単位で実行される。ここでは、連続する２つのフレームが互いに半フレーム長だけオーバーラップする半フレームシフトを利用する場合を例示する。図６においては、ｎフレームにタップ音信号が含まれている。

【００４５】

図７は、タップ音信号を除去するための音声信号補正処理の例を示している。波形補正部１１２は、入力音声信号の波形から、タップ音信号を含むｎフレームを削除する。そして、波形補正部１１２は、ｎフレームの前後のフレーム、つまりｎ－１フレームとｎ＋１フレームとを用いて、削除したｎフレーム内の音声信号を補間する。この補間には、たとえば、ハニング窓のような窓関数を用いてもよい。この場合、波形補正部１１２は、ｎ－１フレーム内の信号に第１の窓関数を乗じることによって得られた信号とｎ＋１フレーム内の信号に第１の窓関数とは時間方向が逆の第２の窓関数を乗じることによって得られた信号とを加算し、その加算結果を、削除したｎフレーム内の音声信号の代わりに使用してもよい。

【００４６】

このように、本実施形態では、入力音声信号から非定常ノイズであるタップ音信号が自動的に削除されるので、音声認識精度の低下を招くことなく、音声入力中に他の操作を実行することができる。

【００４７】

図８は、携帯型電子機器のシステム構成の別の例を示している。図８のシステム構成は、テキスト・ツー・スピーチ（ＴＴＳ）部１１９によって得られた音声信号に対応する音

が発生している間も音声入力を行うことを可能にするために、エコーキャンセル部２０１を含んでいる。エコーキャンセル部２０１は、たとえば、音声入力部１１０の前段に設けてもよい。このエコーキャンセル部２０１は、入力音声信号から、テキスト・ツー・スピーチ（ＴＴＳ）部１１９から出力される音声信号がマイクに回り込んだ成分を除去する。これにより、入力音声信号に含まれる、スピーカ４０からの現在の出力音が除去される。よって、たとえば、店員３１は、自分の発話を認識、翻訳および音声合成することによって得られる音声出力の完了を待たずに、発話を行うことができる。

【００４８】

図９は、携帯型電子機器のシステム構成のさらに別の例を示している。図９のシステム構成は、任意のタイミングで音声入力を自動的に開始できるようにするために、発話区間検出部２０２を備えている。この発話区間検出部２０２は、たとえば、入力音声処理部１１０の後段に設けてもよい。

【００４９】

発話区間検出部２０２は、入力音声処理部１１０によって処理された入力音声信号を格納するバッファ（メモリ）２０２ａを備えている。発話区間検出部２０２は、バッファ２０２ａに格納された入力音声信号内の発話区間を検出する。発話区間は、話者が発話している期間である。そして、発話区間検出部２０２は、バッファ２０２ａに格納された入力音声信号内に含まれ且つ検出された発話区間に属する音声信号を、認識対象の音声信号として音声認識部１１７へ出力する。このように、発話区間検出部２０２によって発話区間を検出することにより、発話開始ボタン１９を押すことなく、音声認識および機械翻訳を適切なタイミングで開始することができる。

【００５０】

次に、図１０を参照して、発話区間の検出動作の例を説明する。図１０の横軸は時間を表し、縦軸は入力音声信号の信号強度レベル（パワー）を表している。入力音声信号の強度レベルはたとえばタイミングｔ１である基準値を超える。入力音声信号の強度レベルが基準値を超えている状態がタイミングｔ１からある期間Ｔ１だけ継続した場合、発話区間検出部２０２は、発話が開始されたことを検出する。この場合、発話区間検出部２０２は、たとえば、タイミングｔ１よりも少し前のタイミングｔ０から、入力音声信号の強度レベルが基準値よりも低下するタイミングｔ２までの期間、つまりＴ２で示される期間、を、発話区間として認識してもよい。発話区間検出部２０２は、発話区間に属する音声信号をバッファ２０２ａからリードし、リードした音声信号を後段に出力する。

【００５１】

図１１のフローチャートは、発話区間検出処理の手順を示している。入力音声処理部１１０はマイクロホン１２から音声信号を入力し、その入力音声信号を処理する（ステップＳ１１）。発話区間検出部２０２は、入力音声処理部１１０から出力される音声信号をバッファ２０２ａにバッファリングする（ステップＳ１２）。発話区間検出部２０２は、バッファリングされた音声信号の強度レベルに基づいて発話が開始されたか否かを判定する（ステップＳ１３）。発話が開始されたならば、発話区間検出部２０２は、発話区間を検出し（ステップＳ１４）、その発話区間に属する音声信号を音声認識（ＡＳＲ）部１１７へ出力する（ステップＳ１５）。

【００５２】

図１２は、携帯型電子機器のシステム構成のさらに別の例を示している。図１２のシステム構成は、複数人が同時に話している場合でも特定の人物の発話を入力および認識できるようにするために、複数のマイクロホン１２Ａ，１２Ｂと話者方向推定部２０３を備えている。話者方向推定部２０３は入力音声処理部１１０の前段に設けてもよい。

【００５３】

話者方向推定部２０３は、マイクロホン１２Ａ，１２Ｂと共同して、特定方向に位置する音源（話者）からの音を抽出可能なマイクロホンアレイとして機能する。話者方向推定部２０３は、マイクロホン１２Ａ，１２Ｂそれぞれからの入力音声信号群を用いて、それら入力音声信号それぞれに対応する音源（話者）が位置する、携帯型電子機器の本体１０

に対する方向（話者方向）を推定する。たとえば、携帯型電子機器の本体１０に対してたとえば左上方向に位置する話者の音声はマイクロホン１２Ａに先に到達し、少し遅れてマイクロホン１２Ｂに到達する。この遅延時間と、マイクロホン１２Ａとマイクロホン１２Ｂとの間の距離とから、入力音声信号に対応する音源方向（話者方向）を推定することができる。そして、この話者方向の推定結果に基づいて、話者方向推定部２０３は、マイクロホン１２Ａ，１２Ｂによって入力された入力音声信号群から、携帯型電子機器の本体１０に対して特定の方向からの入力音声信号を抽出（選択）する。たとえば、店員３１の音声を抽出する場合には、携帯型電子機器の本体１０に対してたとえば左上方向から入力される音声信号を抽出（選択）すればよい。また、外国人３２の音声を抽出する場合には、携帯型電子機器の本体１０に対してたとえば右上方向から入力される音声信号を抽出（選択）すればよい。入力音声処理部１１０は、抽出された特定の方向からの入力音声信号に対して上述の波形補正処理を施す。そして、波形補正処理された特定の方向からの入力音声信号に対して音声認識、機械翻訳、音声合成等の処理が施される。

【００５４】

よって、複数人が同時に話している場合でも、特定方向からの音声のみを処理することが可能となるので、店員３１または外国人３２以外の他の話者の音声に影響されることなく、特定の人物、たとえば、店員３１または外国人３２、の音声を正しく入力および認識することが可能となる。

【００５５】

また、代わりに、カメラを用いて携帯型電子機器の本体１０の周囲に存在する各人物の顔検出を行い、店員３１の顔に類似する顔が存在する方向を、店員３１が位置する携帯型電子機器の本体１０に対する方向として推定してもよい。また、店員３１の顔に類似する顔が存在する方向とは逆の方向を、外国人３２が位置する携帯型電子機器の本体１０に対する方向として推定してもよい。店員３１または外国人３２以外の他の話者の音声は非定常ノイズであるが、図１２のシステム構成によれば、店員３１または外国人３２のみを抽出できるので、この非定常ノイズによる影響を低減することができる。

【００５６】

また、携帯型電子機器においては、本体１０に対して第１の方向（たとえば左上方向）から入力される音声信号に対しては第１の言語（ここでは日本語）から第２の言語（ここでは英語）へ翻訳するための機械翻訳処理が施され、本体１０に対して第２の方向（たとえば右上方向）から入力される音声信号に対しては第２の言語（ここでは英語）から第１の言語（ここでは日本語）へ翻訳するための機械翻訳処理が施される。そして、左上方向から入力される音声信号に、第１の言語から第２の言語に翻訳するための機械翻訳を施すことによって得られる翻訳結果と、右上方向から入力される音声信号に、第２の言語から第１の言語に翻訳するための機械翻訳を施すことによって得られる翻訳結果とが、出力される。このように、音声信号に適用される機械翻訳の内容は、その音声信号の入力方向（話者方向）に応じて決定することができる。よって、店員３１の発話および外国人の発話を英語および日本語にそれぞれ容易に翻訳することができる。

【００５７】

図１３は、携帯型電子機器のシステム構成のさらに別の例を示している。図１３のシステム構成は、複数人が同時に話している場合に、発話者ごとに音声を入力および認識できるようにするために、複数のマイクロホン１２Ａ，１２Ｂと話者分類部２０４とを備えている。話者分類部２０４は入力音声処理部１１０の前段に設けてもよい。

【００５８】

話者分類部２０４もマイクロホンアレイとして機能する。この話者分類部２０４は話者方向推定部２０４ａと目的音声信号抽出部２０４ｂとを含む。話者方向推定部２０４ａは、複数のマイクロホン１２Ａ，１２Ｂそれぞれからの入力音声信号群を用いて、それら入力音声信号それぞれに対応する音源（話者）それぞれが位置する携帯型電子機器の本体１０に対する方向を推定する。目的音声信号抽出部２０４ｂは、複数の話者それぞれの方向の推定結果に基づいて、複数のマイクロホン１２Ａ，１２Ｂそれぞれからの入力音声信号

群を、話者ごと、つまり音源方向毎に、分類する。たとえば、携帯型電子機器の本体１０に対してたとえば左上方向からの音声信号は、店員３１の音声として決定され、話者＃１バッファ２０５に格納される。また、携帯型電子機器の本体１０に対してたとえば右上方向からの音声信号は、外国人３２の音声として決定され、話者＃２バッファ２０６に格納される。

【００５９】

スイッチ部２０７は話者＃１バッファ２０５と話者＃２バッファ２０６とを時分割形式で交互に選択する。これにより、入力音声処理部１１０は、店員３１の音声信号と外国人３２の音声信号とを時分割形式で交互に処理することができる。同様に、音声認識部１１０、機械翻訳部１１８、ＴＴＳ部１１９、メッセージ表示部１２０の各々も、店員３１の音声信号と外国人３２の音声信号とを時分割形式で交互に処理することができる。店員３１の音声の認識結果には日本語から英語へ翻訳するための機械翻訳が施され、その翻訳結果が音声出力またはテキスト表示によって出力される。また、外国人３２の音声の認識結果には英語から日本語へ翻訳するための機械翻訳が施され、その翻訳結果が音声出力またはテキスト表示によって出力される。

【００６０】

なお、入力音声処理部１１０、機械翻訳部１１８、ＴＴＳ部１１９、メッセージ表示部１２０をそれぞれ含む複数の音声処理ブロックを設け、複数の話者の音声信号を並列に処理してもよい。

【００６１】

以上説明したように、本実施形態によれば、タップ音信号のような非定常ノイズによる影響を低減することができるので、音声入力中にタップ操作を用いた他の各種操作を実行することができる。よって、たとえば店員は本実施形態の携帯型電子機器を用いて外国人との会話中においても、携帯型電子機器のタッチパネルディスプレイ１１をタップ操作して、売り場の紹介のような画像をタッチパネルディスプレイ１１上に表示させるといった操作を行うことができる。

【００６２】

なお、図８のエコーキャンセル部２０１、図９の発話区間検出部２０２、図１２の話者方向推定部２０３、図１３の話者分類部２０４の内の任意のいくつかまたは全てを併せ持つ構成を用いることもできる。

【００６３】

なお、本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

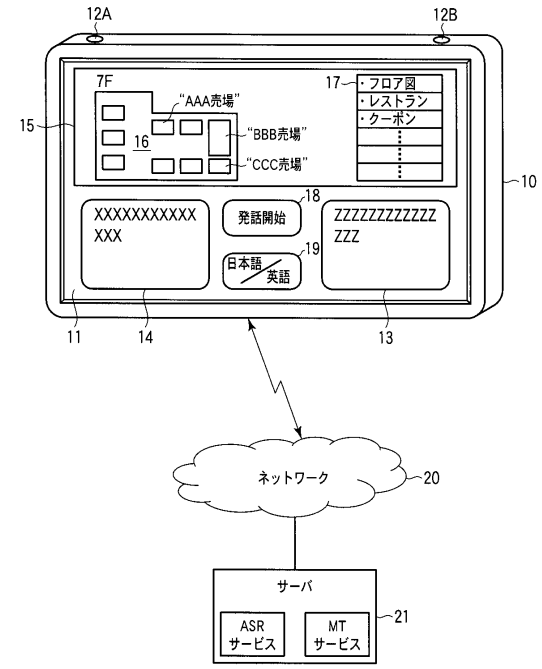
【符号の説明】

【００６４】

１０…携帯型電子機器の本体、１１…タッチスクリーンディスプレイ、１２Ａ，１２Ｂ…マイクロホン、１１０…入力音声処理部、２０１…エコーキャンセル部、２０２…発話区間検出部、２０３…話者方向推定部、２０４…話者分類部。

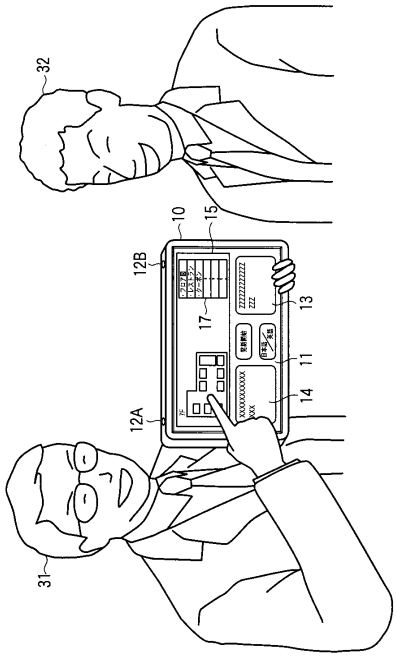
【図１】

図１



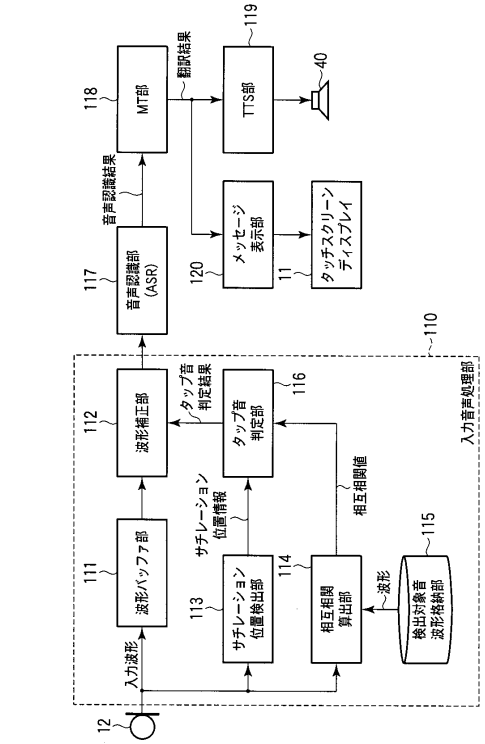
【図２】

図２



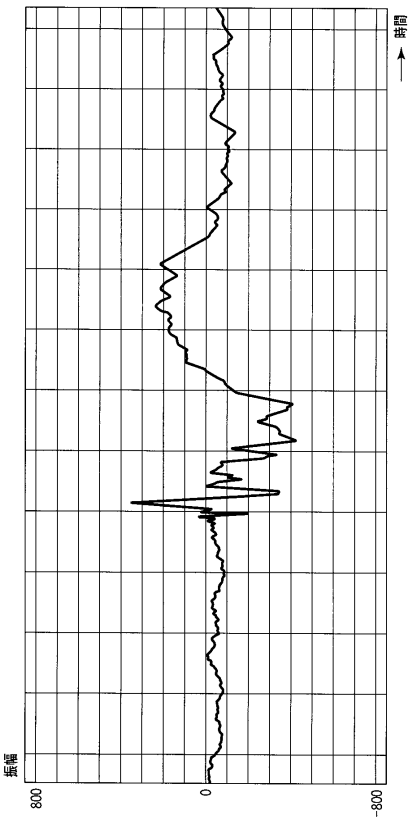
【図３】

図３



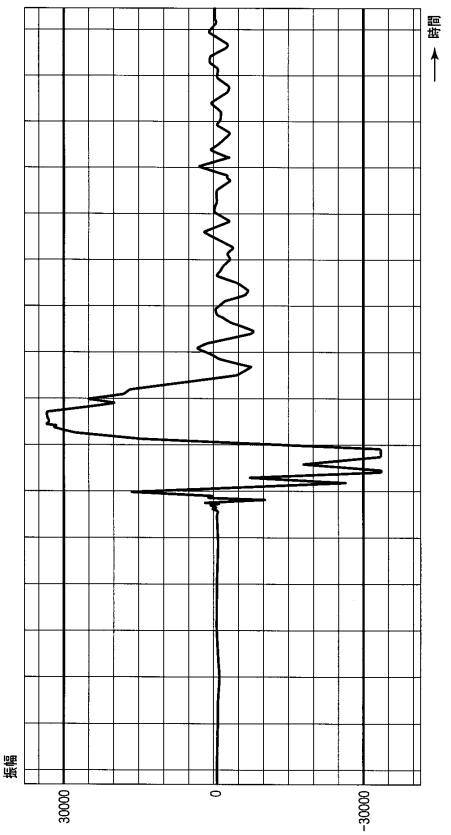
【図４】

図４



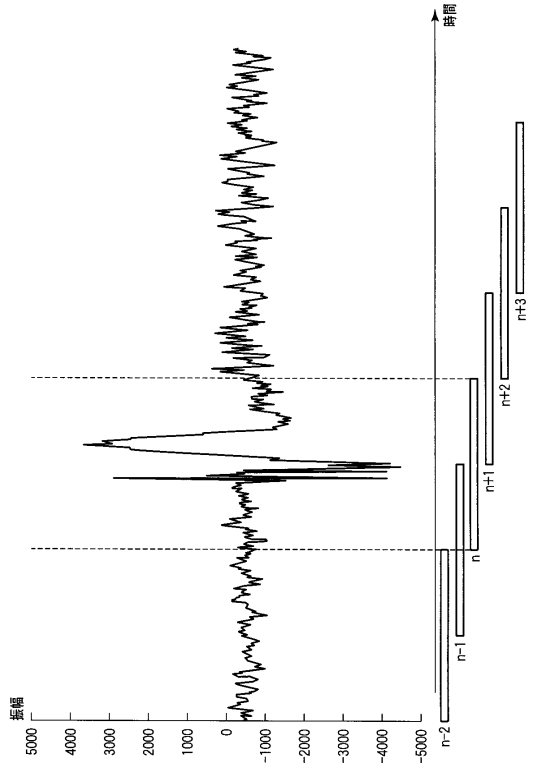
【図５】

図５



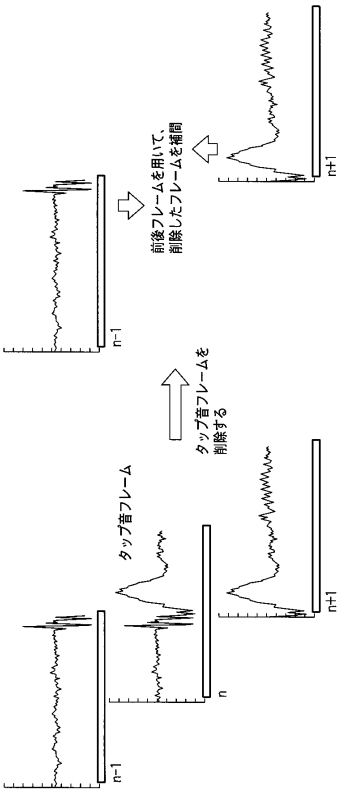
【図６】

図６



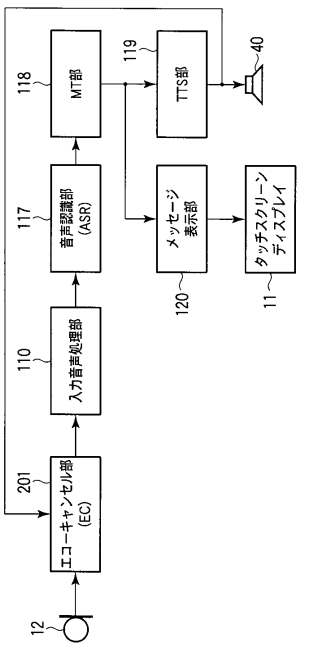
【図７】

図７



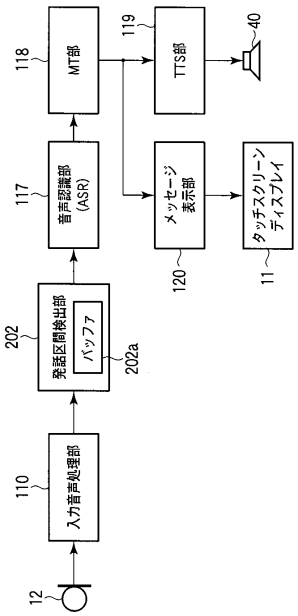
【図８】

図８



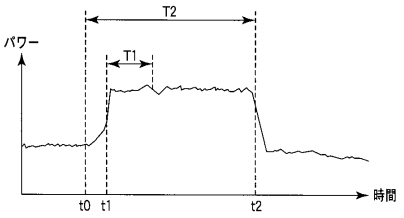
【図９】

図 9



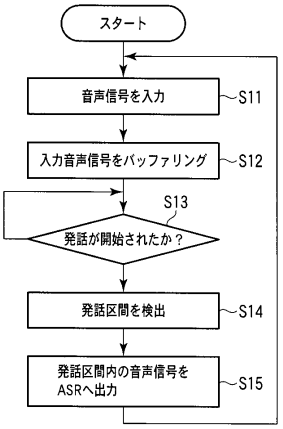
【図１０】

図 10



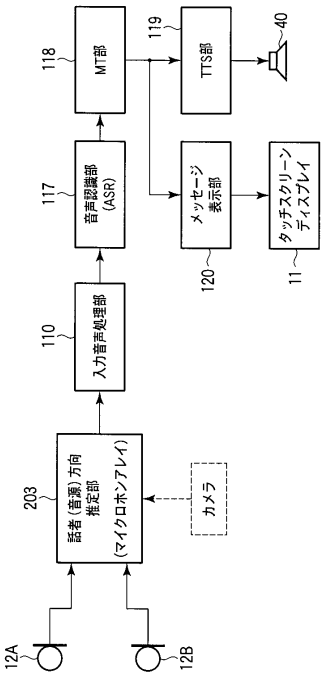
【図１１】

図 11



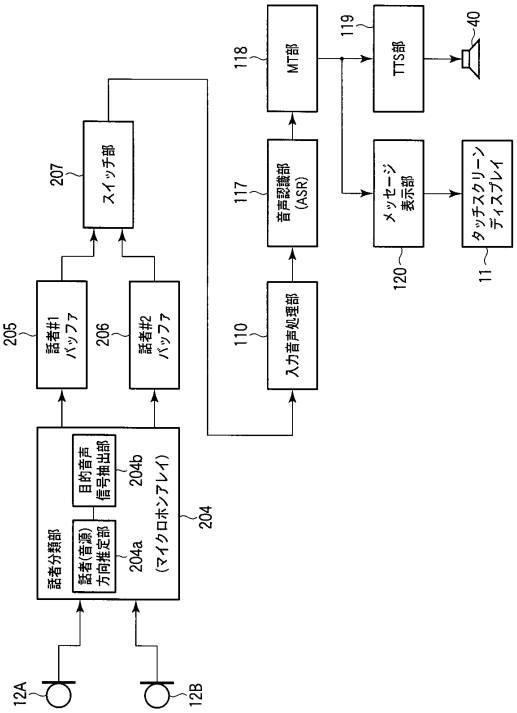
【図１２】

図 12



【図１３】

図 13



フロントページの続き

(74)代理人 100084618
弁理士 村松 貞男
(74)代理人 100103034
弁理士 野河 信久
(74)代理人 100119976
弁理士 幸長 保次郎
(74)代理人 100153051
弁理士 河野 直樹
(74)代理人 100140176
弁理士 砂川 克
(74)代理人 100101812
弁理士 勝村 紘
(74)代理人 100124394
弁理士 佐藤 立志
(74)代理人 100112807
弁理士 岡田 貴志
(74)代理人 100111073
弁理士 堀内 美保子
(74)代理人 100134290
弁理士 竹内 将訓
(74)代理人 100127144
弁理士 市原 卓三
(74)代理人 100141933
弁理士 山下 元
(72)発明者 杉浦 千加志
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 井阪 岳彦
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 須藤 隆
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 鈴木 真吾
東京都港区芝浦一丁目１番１号 株式会社東芝内

審査官 田部井 和彦

(56)参考文献 米国特許出願公開第２００９／０２１６５３１（ＵＳ，Ａ１）
特開２００３－２９５８９９（ＪＰ，Ａ）
米国特許出願公開第２０１０／０１０６４８３（ＵＳ，Ａ１）
特開２００３－１０８５５１（ＪＰ，Ａ）
特開２００７－２８８５６５（ＪＰ，Ａ）
国際公開第２０１１／００４５０３（WO，Ａ１）
特開２０１０－１０２１２９（ＪＰ，Ａ）
佐藤 幹，小型音声対話モジュールによる耐雑音音声認識，ＡＩチャレンジ研究会（第２６回）ＳＩＧ－Ｃｈａｌｌｅｎｇｅ－Ａ７０２，日本，社団法人人工知能学会ＡＩチャレンジ研究会，２００７年１１月

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １１／００－２１／０６

H04R 1/04
G06F 3/041