

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6897677号  
(P6897677)

(45) 発行日 令和3年7月7日(2021.7.7)

(24) 登録日 令和3年6月14日(2021.6.14)

(51) Int.Cl. F I  
**G 1 0 L 15/22 (2006.01)** G 1 0 L 15/22 4 7 0 F  
 G 1 0 L 15/22 4 6 0 Z

請求項の数 18 (全 38 頁)

(21) 出願番号	特願2018-523319 (P2018-523319)	(73) 特許権者	000002185
(86) (22) 出願日	平成29年3月15日 (2017.3.15)		ソニーグループ株式会社
(86) 国際出願番号	PCT/JP2017/010406		東京都港区港南1丁目7番1号
(87) 国際公開番号	W02017/217046	(74) 代理人	110002147
(87) 国際公開日	平成29年12月21日 (2017.12.21)		特許業務法人酒井国際特許事務所
審査請求日	令和2年2月6日 (2020.2.6)	(72) 発明者	加藤 亜由美
(31) 優先権主張番号	特願2016-118622 (P2016-118622)		東京都港区港南1丁目7番1号 ソニー株式会社内
(32) 優先日	平成28年6月15日 (2016.6.15)	(72) 発明者	河野 真一
(33) 優先権主張国・地域又は機関	日本国 (JP)		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	滝 祐平
			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】 情報処理装置及び情報処理方法

(57) 【特許請求の範囲】

【請求項 1】

音声の集音に関する期間における情報であって、前記音声の発話の速度に関する情報を含むコンテキスト情報を取得する取得部と、

前記音声を変換された文字情報として、前記音声の発音を示す文字情報の候補を、前記発話の速度に応じて、所定の出力部に提示させる制御部と、

を備える、情報処理装置。

【請求項 2】

前記コンテキスト情報は、前記音声が集音される環境の状態に関する情報を含む、請求項 1 に記載の情報処理装置。

【請求項 3】

前記コンテキスト情報は、入力情報の状態または当該入力情報の入力状況に関する情報を含む、請求項 1 に記載の情報処理装置。

【請求項 4】

前記制御部は、前記発話の速度に応じて、前記出力部に提示させる前記候補の数を制限する、

請求項 1 に記載の情報処理装置。

【請求項 5】

前記コンテキスト情報は、前記音声が発話したユーザの状態に関する情報を含む、請求項 1 に記載の情報処理装置。

## 【請求項 6】

前記制御部は、前記音声の集音に関する期間における前記コンテキスト情報に応じて、当該音声に変換された前記文字情報の1以上の前記候補のうち少なくとも一部の候補の優先度を制御する、請求項1～5のいずれか一項に記載の情報処理装置。

## 【請求項 7】

前記コンテキスト情報は、雑音の影響に関する情報を含み、

前記制御部は、前記雑音の影響が閾値よりも小さい場合には、前記音声の発音に基づく類似度に応じて前記優先度を制御する、請求項6に記載の情報処理装置。

## 【請求項 8】

前記コンテキスト情報は、雑音の影響に関する情報を含み、

前記制御部は、前記雑音の影響が閾値よりも大きい場合には、前記音声に変換された複数の文字情報間の共起関係に応じて前記優先度を制御する、請求項6に記載の情報処理装置。

10

## 【請求項 9】

前記制御部は、

集音された前記音声である音響の解析結果に応じて、当該音声の発音を示す文字情報の1以上の候補を前記出力部に提示させ、

提示された当該発音を示す1以上の候補からの少なくとも一部の当該候補に対する選択の結果に基づき前記優先度を制御する、

請求項6に記載の情報処理装置。

20

## 【請求項 10】

前記制御部は、前記コンテキスト情報に応じて、前記優先度を制御するために参照するデータを切り替える、請求項6に記載の情報処理装置。

## 【請求項 11】

前記制御部は、

前記コンテキスト情報に応じて、互いに異なる複数のアルゴリズムに基づき、当該アルゴリズムごとに前記優先度を制御し、

前記アルゴリズムごとに、当該アルゴリズムに基づき制御された前記優先度に応じて、前記文字情報の前記候補を前記出力部に提示させる、

請求項6に記載の情報処理装置。

30

## 【請求項 12】

前記制御部は、前記アルゴリズムごとに異なる態様で、前記文字情報の前記候補を提示する、請求項11に記載の情報処理装置。

## 【請求項 13】

前記制御部は、前記音声に変換された文字情報に関する条件の指定を受けて、当該条件に基づき、前記出力部に提示させる当該文字情報の前記候補を制限する、請求項1～12のいずれか一項に記載の情報処理装置。

## 【請求項 14】

前記制御部は、前記音声に変換された文字情報の前記候補の数を示す情報を、当該文字情報に関連付けて提示する、請求項1～13のいずれか一項に記載の情報処理装置。

40

## 【請求項 15】

前記取得部は、逐次集音される前記音声に基づき実行される音声認識処理の結果を逐次取得し、

前記制御部は、逐次取得される前記音声認識処理の結果に基づき、集音された当該音声のうち少なくとも一部の音声に変換された文字情報の1以上の候補を、前記出力部に提示させる、

請求項1～14のいずれか一項に記載の情報処理装置。

## 【請求項 16】

所定の取得部により取得された、音声の集音に関する期間における情報であって、前記音声の発話の速度に関する情報を含むコンテキスト情報を外部装置に送信する送信部と、

50

外部装置から送信される前記音声を変換された文字情報として、前記音声の発音を示す文字情報の候補を、前記発話の速度に応じて提示する出力部と、  
を備える、情報処理装置。

【請求項 17】

コンピュータシステムが、  
音声の集音に関する期間における情報であって、前記音声の発話の速度に関する情報を含むコンテキスト情報を取得することと、

前記音声を変換された文字情報として、前記音声の発音を示す文字情報の候補を、前記発話の速度に応じて、所定の出力部に提示させることと、  
を含む、情報処理方法。

10

【請求項 18】

コンピュータシステムが、  
所定の取得部により取得された音声の集音に関する期間における情報であって、前記音声の発話の速度に関する情報を含むコンテキスト情報を外部装置に送信することと、

外部装置から送信される前記音声を変換された文字情報として、前記音声の発音を示す文字情報の候補を、前記発話の速度に応じて提示することと、  
を含む、情報処理方法。

【発明の詳細な説明】

【技術分野】

20

【0001】

本開示は、情報処理装置及び情報処理方法に関する。

【背景技術】

【0002】

PC (Personal Computer) やゲーム機器等のような所謂情報処理装置の動作を制御するために、ユーザからの入力を取得するための入力インタフェースとして多様な方式のものが提案されている。特に近年では、所謂音声認識技術や自然言語処理技術の発展に伴い、各種情報処理装置に対して、ユーザが音声により各種情報を入力可能としたユーザインタフェース (UI: User Interface) が普及してきている。例えば、特許文献 1 には、所謂音声認識技術を利用した仕組みの一例が開示されている。

30

【0003】

上述したような音声認識技術や自然言語処理技術を応用することで、例えば、ユーザが入力した音声を所謂文字情報に変換して提示することが可能となる。また、このように音声を変換された文字情報は、例えば、当該文字情報の意味内容を認識することで、各種制御への入力として応用することも可能となる。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開 2003 - 140691 号公報

【発明の概要】

40

【発明が解決しようとする課題】

【0005】

ところで、音声を入力として使用することが可能なユーザインタフェースは、ウェアブル端末等のような所謂入力インタフェースが制限された情報処理装置への応用も可能であり、その利用用途は多岐にわたる。このような状況から、特に近年では、音声を入力としたユーザインタフェースの操作性の向上が望まれている。

【0006】

そこで、本開示では、音声を入力としたユーザインタフェースの操作性をより向上させることが可能な、情報処理装置及び情報処理方法を提案する。

【課題を解決するための手段】

50

## 【 0 0 0 7 】

本開示によれば、音声の集音に関する期間におけるコンテキスト情報を取得する取得部と、前記音声に変換された文字情報の候補を、前記コンテキスト情報に応じた態様で、所定の出力部に提示させる制御部と、を備える、情報処理装置が提供される。

## 【 0 0 0 8 】

また、本開示によれば、所定の取得部により取得された音声の集音に関する期間におけるコンテキスト情報を外部装置に送信する送信部と、外部装置から送信される前記音声に変換された文字情報の候補を、前記コンテキスト情報に応じた態様で提示する出力部と、を備える、情報処理装置が提供される。

## 【 0 0 0 9 】

また、本開示によれば、コンピュータシステムが、音声の集音に関する期間におけるコンテキスト情報を取得することと、前記音声に変換された文字情報の候補を、取得された前記コンテキスト情報に応じた態様で、所定の出力部に提示させることと、を含む、情報処理方法が提供される。

## 【 0 0 1 0 】

また、本開示によれば、コンピュータシステムが、所定の取得部により取得された音声の集音に関する期間におけるコンテキスト情報を外部装置に送信することと、外部装置から送信される前記音声に変換された文字情報の候補を、前記コンテキスト情報に応じた態様で提示することと、を含む、情報処理方法が提供される。

## 【 発明の効果 】

## 【 0 0 1 1 】

以上説明したように本開示によれば、音声を入力としたユーザインタフェースの操作性をより向上させることが可能な、情報処理装置及び情報処理方法が提供される。

## 【 0 0 1 2 】

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他の効果が奏されてもよい。

## 【 図面の簡単な説明 】

## 【 0 0 1 3 】

【 図 1 】 本実施形態に係る情報処理システムの概略的なシステム構成の一例について説明するための説明図である。

【 図 2 】 本開示の第 1 の実施形態に係る情報処理システムが提供する UI の一例について説明するための説明図である。

【 図 3 】 同実施形態に係る情報処理システムが提供する UI の他の一例について説明するための説明図である。

【 図 4 】 同実施形態に係る情報処理システムの機能構成の一例を示したブロック図である。

【 図 5 】 同実施形態に係る情報処理システムが提供する UI の他の一例について説明するための説明図である。

【 図 6 】 同実施形態に係る情報処理システムが提供する UI の他の一例について説明するための説明図である。

【 図 7 】 同実施形態に係る情報処理システムの一連の処理の流れの一例について示したフローチャートである。

【 図 8 】 同実施形態に係る情報処理システムの一連の処理の流れの一例について示したフローチャートである。

【 図 9 】 同実施形態の変形例 1 - 1 に係る情報処理システムが提供する UI の一例について説明するための説明図である。

【 図 1 0 】 同実施形態の変形例 1 - 2 に係る情報処理システムの概要について説明するための説明図である。

【 図 1 1 】 同実施形態の変形例 1 - 2 に係る情報処理システムが提供する UI の一例につ

10

20

30

40

50

いて説明するための説明図である。

【図 1 2】同実施形態の変形例 1 - 2 に係る情報処理システムが提供する U I の一例について説明するための説明図である。

【図 1 3】同実施形態の変形例 1 - 3 に係る情報処理システムが提供する U I の一例について説明するための説明図である。

【図 1 4】同実施形態の変形例 1 - 3 に係る情報処理システムが提供する U I の一例について説明するための説明図である。

【図 1 5】同実施形態の変形例 1 - 3 に係る情報処理システムが提供する U I の一例について説明するための説明図である。

【図 1 6】同実施形態の変形例 1 - 3 に係る情報処理システムが提供する U I の一例について説明するための説明図である。 10

【図 1 7】本開示の第 2 の実施形態に係る情報処理システムの概要について説明するための説明図である。

【図 1 8】同実施形態に係る情報処理システムが提供する U I の概要について説明するための説明図である。

【図 1 9】同実施形態に係る情報処理システムの機能構成の一例を示したブロック図である。

【図 2 0】同実施形態に係る情報処理システムの一連の処理の流れの一例について示したフローチャートである。

【図 2 1】実施形態に係る情報処理システムによる制御の一例について説明するための説明図である。 20

【図 2 2】変形例 2 - 1 に係る情報処理システムが提供する U I の一例について説明するための説明図である。

【図 2 3】変形例 2 - 2 に係る情報処理システムが提供する U I の一例について説明するための説明図である。

【図 2 4】本開示の一実施形態に係る情報処理システムを構成する情報処理装置のハードウェア構成の一例を示す機能ブロック図である。

【発明を実施するための形態】

【 0 0 1 4 】

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。 30  
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。

【 0 0 1 5 】

なお、説明は以下の順序で行うものとする。

1 . システム構成

2 . 第 1 の実施形態

2 . 1 . 概要

2 . 2 . 構成

2 . 3 . 処理

2 . 4 . 変形例 40

2 . 4 . 1 . 変形例 1 - 1 : 修正候補の提示するための U I の一例

2 . 4 . 2 . 変形例 1 - 2 : 集音結果に基づく入力情報の状態に応じた制御

2 . 4 . 3 . 変形例 1 - 3 : 情報入力の態様に応じた制御

2 . 5 . 実施例

2 . 6 . 評価

3 . 第 2 の実施形態

3 . 1 . 概要

3 . 2 . 構成

3 . 3 . 処理

3 . 4 . 変形例 50

3.4.1. 変形例 2 - 1 : 認識結果を確定するためのインタフェースの一例

3.4.2. 変形例 2 - 2 : 状態や状況に応じた修正候補の制御

3.5. 評価

4. ハードウェア構成

5. 適用例

6. むすび

【0016】

<< 1. システム構成 >>

まず、図1を参照して、本開示の一実施形態に係る情報処理システムの概略的な構成について説明する。図1は、本実施形態に係る情報処理システムの概略的なシステム構成の一例について説明するための説明図である。

10

【0017】

図1に示すように、本実施形態に係る情報処理システムは、情報処理装置10と、入出力装置30とを含む。情報処理装置10と入出力装置30とは、所定のネットワークN1を介して互いに情報を送受信可能に構成されている。なお、ネットワークN1の種別は特に限定されない。具体的な一例として、当該ネットワークN1は、Wi-Fi(登録商標)規格に基づくネットワークのような、所謂無線のネットワークにより構成されていてもよい。また、他の一例として、当該ネットワークN1は、インターネット、専用線、LAN(Local Area Network)、または、WAN(Wide Area Network)等により構成されていてもよい。また、当該ネットワークN1は、複数のネットワークを含んでもよく、一部が有線のネットワークとして構成されていてもよい。

20

【0018】

入出力装置30は、各種入力情報の取得や、当該入出力装置30を保持するユーザUaに対する各種出力情報の提示を行うための構成である。特に、本実施形態に係る入出力装置30は、例えば、所謂マイクロフォン等のような集音部を備え、ユーザUaの音声を集音可能に構成されている。例えば、入出力装置30は、当該集音部により集音した音声(音響)を、ネットワークN1を介して情報処理装置10に送信してもよい。また、入出力装置30は、所謂液晶ディスプレイや有機ELディスプレイ等のような表示部や、所謂スピーカ等のような音響出力部等の各種出力部を備え、当該出力部を介して情報を出力可能に構成されていてもよい。

30

【0019】

このような構成に基づき、入出力装置30は、情報処理装置10に対して送信した音声を入力とした処理の結果を示す出力情報を当該情報処理装置10からネットワークN1を介して取得し、出力部を介して取得した出力情報をユーザUaに提示してもよい。より具体的な一例として、入出力装置30は、当該音声に変換された文字情報を情報処理装置10から取得し、取得した当該文字情報を表示部に表示させることで、当該文字情報をユーザUaに視認可能に提示してもよい。また、他の一例として、入出力装置30は、情報処理装置10から取得した文字情報をTTS(Text to speech)等の技術に基づき音声に変換し、変換された当該音声を音響出力部に出力させることで、当該音声(即ち、文字情報が変換された音声)をユーザUaに聴取可能に提示してもよい。

40

【0020】

また、入出力装置30は、各種状態や状況を直接的または間接的に示すコンテキスト情報を取得可能に構成されていてもよい。この場合には、入出力装置30は、取得したコンテキスト情報を情報処理装置10に送信してもよい。これにより、情報処理装置10は、入出力装置30から取得したコンテキスト情報に基づき、当該状態や状況を認識することが可能となる。

【0021】

なお、コンテキスト情報としては、例えば、入出力装置30の外部環境(ひいては、ユーザUaの外部環境)の状態を示す情報が挙げられる。具体的な一例として、入出力装置30は、外部環境の音響を集音する集音部から、周囲の雑音の集音結果を取得することで

50

、当該雑音の影響（例えば、雑音の音量や、信号対雑音比（S N R : (signal - to - noise ratio) 等）を示す情報をコンテキスト情報として取得してもよい。

【 0 0 2 2 】

また、コンテキスト情報としては、発話時のユーザ（発話者）の状態や、当該ユーザが置かれた状況に関する情報が挙げられる。例えば、ユーザが、入出力装置 3 0 を介して他のユーザとコミュニケーションをとっている状況下では、当該コミュニケーションの相手に関する情報（例えば、発話者との関係を示す情報等）がコンテキスト情報として取得されてもよい。この場合には、例えば、入出力装置 3 0 は、通信を確立した外部装置の情報に基づき、コミュニケーションの相手に関する情報を特定してもよい。

【 0 0 2 3 】

また、各種入力デバイスへの情報の入力状況に関する情報がコンテキスト情報として取得されてもよい。例えば、キーボードやコントローラ等の入力デバイスに対する操作の頻度から、当該入力デバイスを操作するユーザの状態や、当該ユーザが置かれている状況を推定することも可能である。

【 0 0 2 4 】

また、コンテキスト情報としては、入出力装置 3 0 の状態を示す情報が取得されてもよい。具体的な一例として、入出力装置 3 0 が実行中のアプリケーションに関する情報が挙げられる。このようなコンテキスト情報から、例えば、入出力装置 3 0 を操作するユーザの状態や、当該ユーザが置かれている状況を推定することも可能である。

【 0 0 2 5 】

また、入出力装置 3 0 は、各種センサ等の検知部によるユーザの体温、呼吸、心拍、及び発汗量等の生体情報の検出結果を、当該ユーザの状態を示すコンテキスト情報として取得してもよい。

【 0 0 2 6 】

また、コンテキスト情報として、音声の発話速度や音量等のように、発話の態様に関する情報が取得されてもよい。例えば、入出力装置 3 0 は、音声の集音結果に基づき、発話速度や音量等を認識してもよい。また、集音された音声の音質に基づき、当該音声を発話したユーザの属性（例えば、性別や年齢）を示す情報（即ち、誰が発話している状況かを示す情報）をコンテキスト情報として取得することも可能である。また、音声の集音結果に基づき、当該音声の発話に係る長さに関する情報や、当該音声が変換された文字情報の長さ（即ち、ユーザが発話した文字列の長さ）に関する情報のような入力情報の状態に関する情報を、コンテキスト情報として取得することも可能である。もちろん、対象となる入力情報は音声のみには限定されず、他の入力情報について各種状態や状況に関するコンテキスト情報が取得されてもよい。具体的な一例として、キーボードやコントローラ等の入力デバイスに対する情報の入力速度を示す情報や、入力の強さ（例えば、圧力）を示す情報等が、コンテキスト情報として取得されてもよい。また、入出力装置 3 0 は、音声の集音結果は、上述した各種状態や状況を示す情報を間接的に取得するためのコンテキスト情報として取得してもよい。

【 0 0 2 7 】

なお、入出力装置 3 0 が上記コンテキスト情報を取得するタイミングは特に限定されない。例えば、入出力装置 3 0 は、ユーザが発話した場合（換言すると、音声を集音する場合）に、コンテキスト情報を取得してもよい。より具体的な一例として、入出力装置 3 0 は、所謂 P T T (Push to Talk) のように、所定の操作に基づきユーザが発話した音声の集音を開始する場合には、当該操作の検出に基づきコンテキスト情報を取得してもよい。このような構成により、入出力装置 3 0 は、ユーザが発話を開始したタイミング（即ち、音声の集音を開始したタイミング）における各種状態や状況に関するコンテキスト情報を取得することが可能となる。また、入出力装置 3 0 は、ユーザの発話を検出している期間中（即ち、音声の集音を継続している期間中）は、逐次コンテキスト情報を取得してもよい。

【 0 0 2 8 】

また、入出力装置 30 は、ユーザが発話を開始した場合（換言すると、音声の集音を開始した場合）に、当該発話の開始直前の状態または状況に関するコンテキスト情報を取得してもよい。具体的な一例として、ユーザが発話を開始する直前の期間における、当該ユーザによる他の入力デバイスへの操作状況を示すコンテキスト情報から、発話時における当該ユーザの状態や、当該ユーザが置かれた状況を推測することも可能である。このような場合には、例えば、入出力装置 30 は、所定の状態や状況を逐次監視し、所定のイベント（例えば、ユーザによる発話の開始）の検出をトリガとして、当該イベントの検出タイミングの前後の期間における監視結果に基づき、当該期間におけるコンテキスト情報を取得してもよい。このような構成により、入出力装置 30 は、例えば、ユーザの発話の開始を検出した場合に、当該発話の開始前後数秒（例えば、10 秒程度）の期間における各種状態や状況に関する情報を、当該発話の期間に関するコンテキスト情報として取得することが可能となる。

10

**【0029】**

なお、本開示では、「音声の集音に関する期間」と記載した場合には、当該音声を集音する期間のみに限らず、当該音声の集音を開始する直前の期間も含み得るものとする。同様に、「発話に関する期間」と記載した場合には、音声の発話期間のみに限らず、当該音声が発話される直前の期間も含み得るものとする。なお、一般的には、音声の集音開始前後の 10 秒程度の期間におけるコンテキスト情報を取得できれば、当該音声を発話したユーザの状態や、当該ユーザが置かれている状況を推定することが可能である。もちろん、ユースケースに応じて、コンテキスト情報の取得対象とする期間を変更してもよいことは言うまでもない。

20

**【0030】**

続いて、情報処理装置 10 について説明する。情報処理装置 10 は、入出力装置 30 から音声（音響）の集音結果を取得し、取得した音声を入力情報として各種処理を実行する。具体的には、情報処理装置 10 は、入出力装置 30 から音響の集音結果を取得した場合には、当該音響に含まれる音声を検出してもよいし、当該検出結果に基づき当該音響から音声を抽出してもよい。また、情報処理装置 10 は、取得した音声に対して音声認識処理（所謂音響解析）を施すことで当該音声を文字情報に変換してもよい。また、このとき情報処理装置 10 は、形態素解析、構文解析、及び意味解析等の自然言語処理（所謂言語解析）に基づき、音声に変換された文字情報を、意味内容を考慮した文字情報（例えば、漢字かな交じりの文章等）に変換してもよい。また、情報処理装置 10 は、音声に変換された文字情報の意味内容を認識することで、当該音声により指示された機能を実行してもよい。

30

**【0031】**

また、情報処理装置 10 は、音声の集音結果を入力情報とした各種処理の実行結果を、ネットワーク N1 を介して入出力装置 30 に送信してもよい。具体的な一例として、情報処理装置 10 は、音響の集音結果から音声を検出した場合に、当該検出結果を入出力装置 30 に通知してもよい。また、情報処理装置 10 は、音声の集音結果に対する音声認識処理や自然言語処理の結果に基づき、当該音声に変換された文字情報を入出力装置 30 に送信してもよい。また、このとき情報処理装置 10 は、入出力装置 30 の動作を制御することで、当該入出力装置 30 に音声に変換された文字情報を所定の出力部を介して出力させてもよい。

40

**【0032】**

また、情報処理装置 10 は、入出力装置 30 から上述したコンテキスト情報を取得してもよい。これにより情報処理装置 10 は、入出力装置 30 の周囲の環境の状態や、当該入出力装置 30 を保持するユーザの状態等のような各種状態や状況を認識することが可能となる。また、この場合には、情報処理装置 10 は、各種状態や状況の認識結果に応じて、各種処理を制御してもよい。具体的な一例として、情報処理装置 10 は、音声に変換された文字情報の提示態様を、取得したコンテキスト情報に応じて（即ち、当該コンテキスト情報が示す状態や状況に応じて）制御してもよい。

50



## 【 0 0 3 3 】

なお、上記に説明したシステム構成はあくまで一例であり、必ずしも図 1 に示す例のみには限定されない。具体的な一例として、情報処理装置 1 0 が、互いに連携する複数の装置（例えば、サーバ等）により構成されていてもよい。また、他の一例として、情報処理装置 1 0 と入出力装置 3 0 とが一体的に構成されていてもよい。

## 【 0 0 3 4 】

以上、図 1 を参照して、本開示の一実施形態に係る情報処理システムの概略的な構成について説明した。なお、以降の説明では、本開示に係る情報処理システムの各種実施形態についてそれぞれ説明する。

## 【 0 0 3 5 】

- < < 2 . 第 1 の実施形態 > >
- < 2 . 1 . 概要 >

まず、本開示の第 1 の実施形態に係る情報処理システムについて説明する。前述したように、音声認識技術や自然言語処理技術を応用することで、例えば、ユーザは、キーボード等の入力デバイスを介することなく、音声により文字情報を入力することが可能となる。一方で、音声認識処理や自然言語処理の精度が必ずしも高いとは限らず、音声が入って認識される場合が想定され得る。

## 【 0 0 3 6 】

具体的な一例として、音声を認識し文字情報に変換する場合には、音響解析時に生じる誤りや、言語解析時に生じる誤りが想定され得る。音響解析時に生じる誤りとして、例えば、音声入力の曖昧性に起因する誤りが挙げられる。具体的な一例として、音声の発音が入って認識されることで、音声として入力された語句が、発音が類似する他の語句として認識される場合が挙げられる。また、言語解析時に生じる誤りとしては、例えば、同音異義語間の変換の誤りが挙げられる。具体的な一例として、日本語の音声を認識するような状況下において、音声の発音が正しく認識されたとしても、当該発音を示す文字情報（例えば、ひらがなで示された文章）を漢字かな交じりの文字情報に変換する場合に、本来ユーザが意図していた語句とは意味の異なる語句に変換される場合が挙げられる。

## 【 0 0 3 7 】

このような状況から、音声の認識結果に誤りが生じたとしても、当該誤りをより簡易な操作で修正することが可能な UI が求められている。このような要望を満たす UI として、例えば、誤って認識された語句について修正候補を選択可能に提示する UI が挙げられる。このような UI により、ユーザは、提示された修正候補の中から所望の候補を選択することで、誤って認識された語句を修正することが可能となる。

## 【 0 0 3 8 】

一方で、上述したように、音声を認識し文字情報に変換する場合には、音声解析時に生じる誤りや、言語解析時に生じる誤り等が想定される。そのため、各段階で生じ得る誤りそれぞれを想定して単に修正候補を提示した場合には、修正候補の数がより多くなるため、当該修正候補の確認に係るユーザの負担が増加し、却って操作性が損なわれる場合もある。

## 【 0 0 3 9 】

このような状況を鑑み、本実施形態に係る情報処理システムでは、音声の集音に関する期間に取得されたコンテキスト情報に基づき、当該期間における各種状態や状況に応じた態様で修正候補を提示する。具体的な一例として、音声が集音される環境の状態に応じて、修正候補を提示する態様を制御する例について以下に説明する。

## 【 0 0 4 0 】

例えば、周囲の雑音の影響が小さいより静かな環境下においては、発音が類似する他の語句として認識されるような誤りが生じる可能性が高くなる傾向にある。そのため、本実施形態に係る情報処理システムは、例えば、雑音の影響が小さい環境下においては、音声が入って認識された文字情報の修正候補として、発音が類似する修正候補を優先的に提示する。

## 【 0 0 4 1 】

10

20

30

40

50

例えば、図2は、本実施形態に係る情報処理システムが提供するUIの一例について説明するための説明図であり、雑音の影響が小さい状況下における、音声に変換された文字情報の提示態様の一例を示している。具体的には、図2に示す例では、ユーザが、「Have a flower」と発話し、情報処理システムは、当該ユーザが発話した音声の認識結果として、「Have a floor」という文字情報V10をユーザに提示している。また、情報処理システムは、提示した文字情報V10のうち、誤って認識された「floor」という語句の選択をユーザから受けて、当該語句の修正候補が提示された表示情報V11を提示している。

#### 【0042】

図2に示す例では、情報処理システムは、音声の集音に関する期間に取得されたコンテキスト情報に基づき、音声の集音環境が雑音の影響が小さい状態であることを認識する。そして、情報処理システムは、当該認識結果に基づき、「floor」と発音が類似する他の語句の候補（例えば、「flower」、「four」、「float」、等）の優先度がより高くなるように各候補の優先度を制御し、当該優先度の制御結果に応じて少なくとも一部の修正候補をユーザに提示している。また、このとき、情報処理システムは、発音が類似する修正候補を優先して提示していることを示すアイコン等の表示情報V111aをあわせて提示してもよい。

#### 【0043】

また、他の一例として、図3は、本実施形態に係る情報処理システムが提供するUIの他の一例について説明するための説明図であり、雑音の影響が大きい状況下における、音声に変換された文字情報の提示態様の一例を示している。なお、図3に示す例では、ユーザが、「Have a flower」と発話し、情報処理システムは、当該ユーザが発話した音声の認識結果として、「Have a car」という文字情報V10をユーザに提示している。また、情報処理システムは、提示した文字情報V10のうち、誤って認識された「car」という語句の選択をユーザから受けて、当該語句の修正候補が提示された表示情報V11を提示している。

#### 【0044】

図3に示す例では、情報処理システムは、音声の集音に関する期間に取得されたコンテキスト情報に基づき、音声の集音環境が雑音の影響が大きい状態であることを認識する。そして、情報処理システムは、当該認識結果に基づき、「car」という語句の修正候補として、あわせて発話された「Have」という語句との間の共起関係に基づき、各候補の優先度を制御する。例えば、図3に示す例では、情報処理システムは、「Have」という語句とあわせて使われる可能性の高い語句の候補（例えば、「candy」、「cookie」、「flower」、等）の優先度がより高くなるように各候補の優先度を制御している。そして、情報処理システムは、優先度の制御結果に応じて少なくとも一部の修正候補をユーザに提示している。また、このとき、情報処理システムは、修正対象となる語句と他の語句との間の共起関係に基づき修正候補を提示していることを示すアイコン等の表示情報V111bをあわせて提示してもよい。

#### 【0045】

このように、本実施形態に係る情報処理システムは、音声の集音に関する期間に取得されたコンテキスト情報に応じた態様で、集音された音声に変換された文字情報の候補を提示する。より具体的には、本実施形態に係る情報処理システムは、音声の認識結果に誤りが生じたとしても、その時々々の状態や状況に応じて動的に、各修正候補の優先度を制御し、当該優先度の制御結果に基づき、より好適な修正候補をユーザに提示することが可能となる。このような構成により、ユーザは、限られた修正候補の中から、本来意図していた内容を示す候補を特定できる可能性がより高くなるため、多数の候補を確認するといった煩雑な操作を伴わずに、簡便な操作により音声の認識結果を修正することが可能となる。

#### 【0046】

以上、本開示の第1の実施形態に係る情報処理システムの概要について説明した。なお、以降では、本実施形態に係る情報処理システムについてさらに詳しく説明する。

10

20

30

40

50

## 【 0 0 4 7 】

## &lt; 2 . 2 . 構成 &gt;

続いて、図 4 を参照して、本実施形態に係る情報処理システムの機能構成の一例について、特に、図 1 に示した情報処理装置 1 0 及び入出力装置 3 0 の構成に着目して説明する。図 4 は、本実施形態に係る情報処理システムの機能構成の一例を示したブロック図である。

## 【 0 0 4 8 】

なお、図 4 に示す例では、情報処理装置 1 0 及び入出力装置 3 0 の構成をよりわかりやすくするために、当該情報処理装置 1 0 及び入出力装置 3 0 がネットワークを介して情報を送受信するための通信部の図示を省略している。なお、当該通信部の構成は、ネットワークの種別に応じて適宜変更されてもよい。例えば、当該ネットワークが、無線のネットワークの場合には、通信部は、ベースバンド ( B B ) プロセッサや R F 回路等を含んでもよい。なお、以降の説明では、情報処理装置 1 0 及び入出力装置 3 0 それぞれの内部の各構成が、ネットワークを介して外部装置と情報の送受信を行う場合には、特に説明が無い場合においても、当該通信部を介して情報の送受信を行うものとする。

## 【 0 0 4 9 】

まず、入出力装置 3 0 の構成について説明する。図 4 に示すように、入出力装置 3 0 は、入力部 3 1 0 と、出力部 3 3 0 とを含む。入力部 3 1 0 は、例えば、集音部 3 1 1 と、検知部 3 1 3 と、操作部 3 1 5 とを含む。また、出力部 3 3 0 は、例えば、表示部 3 3 1 を含む。

## 【 0 0 5 0 】

集音部 3 1 1 は、所謂マイクロフォン等の集音デバイスにより構成され、ユーザの音声等のような音響を集音する。集音部 3 1 1 による音響の集音結果に基づく音響信号は、例えば、ネットワークを介して情報処理装置 1 0 に送信される。

## 【 0 0 5 1 】

検知部 3 1 3 は、各種状態や状況を示すコンテキスト情報を取得するための構成である。例えば、検知部 3 1 3 は、各種センサを含み、当該センサにより周囲の環境の状態を示す情報をコンテキスト情報として取得してもよい。また、検知部 3 1 3 は、外部環境の音響を集音する集音部を備え、当該集音部による集音結果を、雑音の影響を示すコンテキスト情報として取得してもよい。また、検知部 3 1 3 は、入出力装置 3 0 の外部に設けられた各種デバイスからコンテキスト情報を取得してもよい。具体的な一例として、検知部 3 1 3 は、ユーザの身体に設置された各種センサから、当該ユーザの生体情報を示すコンテキスト情報を取得してもよい。なお、検知部 3 1 3 により取得されたコンテキスト情報は、例えば、ネットワークを介して情報処理装置 1 0 に送信される。

## 【 0 0 5 2 】

操作部 3 1 5 は、ユーザからの入力操作を受け付けるための構成であり、例えば、ボタン、タッチパネル、または所謂ポインティングデバイス等のような入力デバイスにより構成される。例えば、図 2 及び図 3 に示すように、音声に変換された文字情報の修正項目の選択は、当該操作部 3 1 5 を介した操作に基づき行われてもよい。即ち、入出力装置 3 0 は、操作部 3 1 5 に対するユーザからの入力操作の内容に基づき、操作対象 ( 即ち、選択対象 ) を認識すればよい。また、所謂 P T T 等のように音声の集音を開始するための操作が、当該操作部 3 1 5 を介して行われてもよい。

## 【 0 0 5 3 】

表示部 3 3 1 は、所謂液晶ディスプレイや有機 E L ディスプレイ等のような表示デバイスにより構成され、各種表示情報を表示することでユーザに対して情報を提示する。例えば、表示部 3 3 1 は、情報処理装置 1 0 から送信される出力情報を、表示情報として表示することで、当該出力情報をユーザに提示する。また、このとき表示部 3 3 1 は、情報処理装置 1 0 から送信される出力情報を、当該情報処理装置 1 0 により指定された態様で表示情報として表示してもよい。

## 【 0 0 5 4 】

なお、上述した入力部 3 1 0 及び出力部 3 3 0 に含まれる構成はあくまで一例であり、入出力装置 3 0 の利用用途に応じて、上述した構成以外の他の構成が含まれてもよい。具体的な一例として、出力部 3 3 0 は、所謂スピーカ等の音響デバイスにより構成された音響出力部を含んでもよい。このような構成により、入出力装置 3 0 は、情報処理装置 1 0 から送信された出力情報を、例えば、音声として出力することも可能となる。

【 0 0 5 5 】

続いて、情報処理装置 1 0 の構成について説明する。図 4 に示すように、情報処理装置 1 0 は、音声認識処理部 1 0 1 と、コンテキスト解析部 1 0 3 と、優先度制御部 1 0 5 と、出力制御部 1 0 7 とを含む。

【 0 0 5 6 】

音声認識処理部 1 0 1 は、入出力装置 3 0 の入力部 3 1 0 (集音部 3 1 1) による音響の集音結果に基づく音響信号を取得し、取得した当該音響信号に対して、音響解析や言語解析を施すことで、当該音響中に含まれる音声を文字情報に変換する。具体的には、音声認識処理部 1 0 1 は、V A D (Voice Activity Detection) 等の技術に基づき音響信号に含まれる音声信号を抽出する。また、音声認識処理部 1 0 1 は、抽出した音声信号に対して、音声認識処理 (音響解析) を施すことで、当該音声の発音を示す文字情報に変換する。また、音声認識処理部 1 0 1 は、音声認識処理の結果として得られた文字情報に対して自然言語処理 (言語解析) を施すことで、当該文字情報を、意味内容を考慮した文字情報に変換してもよい。

【 0 0 5 7 】

なお、音声認識処理や自然言語処理の結果が必ずしも一意的に特定されるとは限らない。そのため、音声認識処理部 1 0 1 は、音声を文字情報に変換する場合に、当該文字情報中に含まれる少なくとも一部の語句について、他の候補 (即ち、修正候補) を 1 以上特定してもよい。具体的な一例として、音声認識処理部 1 0 1 は、音声に変換された文字情報中に含まれる少なくとも一部の語句について、当該語句と発音が類似する他の候補を特定してもよい。また、他の一例として、音声認識処理部 1 0 1 は、音声に変換された文字情報中に含まれる少なくとも一部の語句について、当該語句の他の候補を、他の語句との間の共起関係に基づき特定してもよい。

【 0 0 5 8 】

そして、音声認識処理部 1 0 1 は、取得された音響信号に対する音声認識処理や自然言語処理の結果に基づき、音声に変換された文字情報や、当該文字情報中の語句の他の候補を示す情報を、出力制御部 1 0 7 及び優先度制御部 1 0 5 に出力する。

【 0 0 5 9 】

コンテキスト解析部 1 0 3 は、入出力装置 3 0 からコンテキスト情報を取得し、取得したコンテキスト情報を解析することで各種状態や状況を認識する。具体的な一例として、コンテキスト解析部 1 0 3 は、入出力装置 3 0 の周囲の環境における音響の集音結果を、コンテキスト情報として取得してもよい。この場合には、例えば、コンテキスト解析部 1 0 3 は、当該集音結果に基づき、音声の集音環境における雑音の影響の大きさ (例えば、雑音の音量等) を認識してもよい。

【 0 0 6 0 】

一般的には、騒音レベルが 6 0 d B 以上の場合には、うるさいと感じられる傾向にある。なお、「普通の会話」、「チャイム」、「時速 4 0 キロで走る自動車の内部」の音量が、約 6 0 d B に相当する。また、騒音レベルが 8 0 d B 以上の場合には、極めてうるさいと感じられる傾向にある。なお、「地下鉄の車内 (窓を開けた場合)」、「ピアノの音」の音量が、約 8 0 d B に相当する。そのため、例えば、コンテキスト解析部 1 0 3 は、雑音の音量が 6 0 d B 以上の場合に、雑音の影響が大きい (即ち、うるさい環境である) 環境であると認識してもよい。

【 0 0 6 1 】

また、他の一例として、コンテキスト解析部 1 0 3 は、ユーザが発話した音声の集音結果に基づく音響信号を、コンテキスト情報として取得してもよい。この場合には、例えば

10

20

30

40

50

、コンテキスト解析部 103 は、当該音響信号に対する音響解析（例えば、音声認識処理）の結果に基づき、発話速度や音声の音量等を認識してもよい。なお、コンテキスト解析部 103 は、当該音響信号に対する音響解析の結果を、音声認識処理部 101 から取得してもよい。

【0062】

また、他の一例として、コンテキスト解析部 103 は、ユーザが発話した音声の集音結果に基づく音響信号に対する音響解析や言語解析の結果を、コンテキスト情報として取得してもよい。この場合には、例えば、コンテキスト解析部 103 は、当該音響信号に対する音響解析や言語解析の結果に基づき、ユーザが発話した音声の長さや、ユーザが発話した文字列の長さ（即ち、音声に変換された文字情報の長さ）等を認識してもよい。

10

【0063】

以上のようにして、コンテキスト解析部 103 は、取得したコンテキスト情報に基づき各種状態や状況を認識し、認識結果を示す情報を、出力制御部 107 及び優先度制御部 105 に出力する。

【0064】

優先度制御部 105 は、音声認識処理部 101 から、取得された音響信号に対する音響解析や言語解析の結果として、音声に変換された文字情報や、当該文字情報中の語句の他の候補を示す情報を取得する。また、優先度制御部 105 は、コンテキスト解析部 103 から、取得されたコンテキスト情報に基づく各種状態や状況の認識結果を示す情報を取得する。優先度制御部 105 は、各種状態や状況の認識結果に応じて異なるアルゴリズムに基づき、音声に変換された文字情報中の語句の他の候補間の優先度を制御してもよい。なお、優先度の制御方法の一例については、図 2 及び図 3 を参照して前述したとおりである。また、図 2 及び図 3 に説明した例以外についても、別途変形例や実施例として後述する。そして、優先度制御部 105 は、音声に変換された文字情報中の語句の他の候補間の優先度の制御結果を示す情報を、出力制御部 107 に出力する。

20

【0065】

出力制御部 107 は、音声認識処理部 101 から、音声に変換された文字情報や、当該文字情報中の語句の他の候補を示す情報を取得する。このとき出力制御部 107 は、取得した当該文字情報を入出力装置 30 に提示させてもよい。これにより、入出力装置 30 を保持するユーザは、当該入出力装置 30 の出力部 330（例えば、表示部 331）を介して自身が発話した音声の認識結果を確認することが可能となる。

30

【0066】

また、出力制御部 107 は、音響解析や言語解析の結果に基づき、音声に変換された文字情報に含まれる各語句を選択可能に、当該文字情報を入出力装置 30 に提示させてもよい。この場合には、例えば、出力制御部 107 は、入出力装置 30 の入力部 310（例えば、操作部 335）を介した入力操作の結果に基づき、音声に変換された文字情報に含まれる各語句のうち、ユーザが選択した語句を認識する。そして、出力制御部 107 は、例えば、図 2 及び図 3 に示すように、ユーザにより選択された語句（即ち、修正対象となる語句）の他の候補を、修正候補として入出力装置 30 に提示させてもよい。

40

【0067】

また、このとき出力制御部 107 は、コンテキスト解析部 103 から、取得されたコンテキスト情報に基づく各種状態や状況の認識結果を示す情報を取得し、当該認識結果に応じて修正候補の提示態様を制御してもよい。例えば、図 5 は、本実施形態に係る情報処理システムが提供する UI の他の一例について説明するための説明図である。図 5 に示す例では、修正対象となる語句と他の語句との間の共起関係に基づき修正候補を提示する場合に、発音が類似する修正候補を提示する場合（例えば、図 2）とは異なる態様で修正候補を提示している。具体的には、図 5 に示す UI の例では、修正対象となる語句の修正候補が、当該修正候補と他の語句との間の共起関係を示す情報とあわせて提示された表示情報 V13 を提示されている。

【0068】

50

また、出力制御部 107 は、状態や状況に応じて、図 2 に示した発音の類似する修正候補を提示する場合の UI と、図 5 に示した修正対象となる語句と他の語句との間の共起関係に基づき修正候補を提示する場合の UI とを組み合わせ提示してもよい。例えば、図 6 は、本実施形態に係る情報処理システムが提供する UI の他の一例について説明するための説明図であり、図 2 に示す UI と図 5 に示す UI とを組み合わせ提示した場合の一例である。具体的には、状態や状況を必ずしも択一的に特定できる場合のみとは限らない。このような状況を鑑み、出力制御部 107 は、図 6 に示すように、異なるアルゴリズムに基づき優先度が制御された修正候補を、アルゴリズムごとに、対応する UI を組み合わせることで提示してもよい。なお、このとき出力制御部 107 は、アルゴリズムごとに提示された修正候補のうち、それぞれに共通で提示される修正候補については、強調表示して

10

【0069】

また、出力制御部 107 は、優先度制御部 105 から、音声に変換された文字情報中の語句の他の候補間の優先度の制御結果を示す情報を取得してもよい。この場合には、例えば、出力制御部 107 は、図 2 及び図 3 に示すように、当該優先度の制御結果に基づき、ユーザにより選択された語句の他の候補を、修正候補として入出力装置 30 に提示させて

20

【0070】

なお、出力制御部 107 は、入出力装置 30 の入力部 310 を介して修正候補の指定を受けた場合には、指定された当該修正候補を、音声に変換された文字情報中の対応する語句の修正候補として確定する。この場合には、出力制御部 107 は、音声に変換された文字情報のうち、対応する語句を指定された修正候補に置き換えたうえで、当該置き換え後の当該文字情報を入出力装置 30 に提示させてもよい。

【0071】

なお、図 4 に示した情報処理システムの機能構成はあくまで一例であり、上述した各構成の機能を実現することが可能であれば、当該情報処理システムの機能構成は必ずしも図 4 に示す例のみには限定されない。例えば、情報処理装置 10 の一部の構成が、当該情報処理装置 10 の外部に設けられていてもよい。具体的な一例として、出力制御部 107 に相当する構成が、入出力装置 30 側に設けられていてもよい。この場合には、入出力装置 30 は、音声に変換された文字情報、当該文字情報中の語句の他の候補を示す情報、及び当該他の候補間の優先度の制御結果を示す情報を、情報処理装置 10 から取得すればよい。また、入力部 310 及び出力部 330 に含まれる各構成のうち少なくとも一部の構成が、入出力装置 30 の外部に設けられていてもよい。また、前述したように、入出力装置 30 と情報処理装置 10 とが一体的に構成されていてもよい。

30

【0072】

以上、図 4 を参照して、本実施形態に係る情報処理システムの機能構成の一例について、特に、図 1 に示した情報処理装置 10 及び入出力装置 30 の構成に着目して説明した。

40

【0073】

< 2.3. 処理 >

続いて、図 7 及び図 8 を参照して、本実施形態に係る情報処理システムの一連の処理の流れの一例について説明する。図 7 及び図 8 は、本実施形態に係る情報処理システムの一連の処理の流れの一例について示したフローチャートである。なお、本説明では、特に、外部環境の騒音の影響を示すコンテキスト情報に基づき、集音された音声に変換された文字情報の候補を提示する場合に着目して説明する。

【0074】

まず、図 7 を参照して、全体的な処理の流れに着目して説明する。図 7 に示すように、入出力装置 30 (集音部 311) は、ユーザの音声等のような音響を集音し、当該集音結

50

果に基づく音響信号を、ネットワークを介して情報処理装置 10 に送信する (S 101)。また、入出力装置 30 (検知部 313) は、外部環境の状態に関するコンテキスト情報として、周囲の環境の音響 (雑音) を集音し、当該音響の集音結果を示す音響信号を、雑音の影響を示すコンテキスト情報として、ネットワークを介して情報処理装置 10 に送信する (S 103)。

【0075】

情報処理装置 10 (音声認識処理部 101) は、入出力装置 30 から音響の集音結果に基づく音響信号を取得し、取得した当該音響信号に対して、音響解析や言語解析を施すことで、当該音響中に含まれる音声を文字情報に変換する。また、このとき情報処理装置 10 は、音声に変換された文字情報中に含まれる少なくとも一部の語句について、他の候補 (即ち、修正候補) を 1 以上特定してもよい。また、情報処理装置 10 (コンテキスト解析部 103) は、入出力装置 30 からコンテキスト情報を取得し、取得したコンテキスト情報を解析することで各種状態や状況を認識する。次いで、情報処理装置 10 (優先度制御部 105) は、取得されたコンテキスト情報に基づく各種状態や状況の認識結果に応じて、音声に変換された文字情報中の語句の他の候補間の優先度を制御する。また、情報処理装置 10 (出力制御部 107) は、音声に変換された文字情報を、当該文字情報に含まれる各語句を選択可能に、入出力装置 30 に提示させる。そして、情報処理装置 10 は、音声に変換された文字情報に含まれる各語句の選択を受けて、選択された当該語句の他の候補を、上記優先度の制御結果に基づき、当該語句に関連付けて入出力装置 30 に提示させる (S 120)。

【0076】

ここで、図 8 を参照して、参照符号 S 120 で示した処理のうち、修正候補間の優先度の制御や、当該優先度の制御結果に基づく修正候補の提示に係る処理について、特に、雑音の影響の大きさに応じて制御を切り替える場合の例に着目して説明する。

【0077】

例えば、雑音の音量が閾値未満の場合等のように雑音の影響が小さい場合には (S 121、NO)、情報処理装置 10 (優先度制御部 105) は、例えば、発音の類似度に基づき修正候補間の優先度を制御する (S 123)。また、この場合には、情報処理装置 10 (出力制御部 107) は、図 2 に示すように、雑音の影響が小さい場合に対応する提示態様で各修正候補が提示された UI に基づき、当該修正候補を提示してもよい (S 125)。

【0078】

また、雑音の音量が閾値以上の場合等のように雑音の影響が大きい場合には (S 121、YES)、情報処理装置 10 (優先度制御部 105) は、例えば、修正対象となる語句と他の語句との間の共起関係に基づき修正候補間の優先度を制御する (S 127)。また、この場合には、情報処理装置 10 (出力制御部 107) は、図 5 に示すように、雑音の影響が大きい場合に対応する提示態様で各修正候補が提示された UI に基づき、当該修正候補を提示してもよい (S 129)。

【0079】

ここで、再度図 7 を参照して、以降の処理について説明する。情報処理装置 10 (出力制御部 107) は、入出力装置 30 を介して修正候補の指定を受けた場合には、指定された当該修正候補を、音声に変換された文字情報中の対応する語句の修正候補として確定する。この場合には、情報処理装置 10 は、音声に変換された文字情報のうち、対応する語句を指定された修正候補に置き換えたうえで、当該置き換え後の当該文字情報を入出力装置 30 に提示させてもよい (S 107)。

【0080】

一方で、提示した修正候補の中にユーザが所望する修正候補が存在しない場合等のように、修正候補の指定が受けられない場合には (S 105、NO)、情報処理装置 10 は、その他の修正手段をユーザに提示してもよい (S 140)。具体的な一例として、情報処理装置 10 は、優先度の制御結果に限らず全ての修正候補を、入出力装置 30 を介して提

10

20

30

40

50

示してもよい。また、他の一例として、情報処理装置 10 は、修正対象をユーザが直接修正するための UI を、入出力装置 30 を介して提示してもよい。また、他の一例として、情報処理装置 10 は、音声の再入力をユーザに促してもよい。

【0081】

以上、図 7 及び図 8 を参照して、本実施形態に係る情報処理システムの一連の処理の流れの一例について説明した。

【0082】

< 2.4. 変形例 >

続いて、本実施形態に係る情報処理システムの変形例について説明する。

【0083】

< 2.4.1. 変形例 1-1：修正候補の提示するための UI の一例 >

まず、変形例 1-1 として、音声に変換された文字情報を修正する場合における、修正候補を提示するための UI の一例について説明する。前述したように、音声を認識し文字情報に変換する場合には、音響解析時に生じる誤りや、言語解析時に生ずる誤りが想定され、修正候補の数がより多くなる傾向にある。そのため、全ての修正候補を提示すると、ユーザは、所望の修正候補を選択することが困難となる場合がある。このような状況を鑑み、変形例 1-1 に係る情報処理システムでは、まず、発音の類似する候補を提示することでユーザに正しい発音を選択させ、選択された発音を示す修正候補（即ち、同音異義語）を提示する。

【0084】

例えば、図 9 は、変形例 1-1 に係る情報処理システムが提供する UI の一例について説明するための説明図である。図 9 に示す例では、ユーザが、日本語で「ログはどこにあるの」と発話し、情報処理システムは、当該ユーザが発話した音声の認識結果として、「六はどこにあるの」という日本語の文字情報をユーザに提示している。また、情報処理システムは、提示した文字情報のうち、修正対象として、「ログは」という語句が誤って認識された「六は」という語句の選択をユーザから受けたものとする。

【0085】

この場合には、情報処理システムは、まず、図 9 の左側の図に示すように、選択された「六は」という語句と発音が類似する他の語句（例えば、「どくは」、「ろぐは」、「ろっくは」、等）が提示された表示情報 V11a を提示することで、ユーザに正しい発音を選択させる。ここでは、ユーザは、「ろぐは」という発音を示す候補を選択したものとする。なお、このとき情報処理システムは、表示情報 V11a に提示した候補が、発音の候補であることを示すアイコン等の表示情報 V111c をあわせて提示してもよい。

【0086】

次いで、情報処理システムは、図 9 の右側の図に示すように、ユーザにより選択された「ろぐは」という発音を示す修正候補（例えば、「ログは」、「ログハ」、「炉具は」、等）が提示された表示情報 V11b を提示することで、ユーザに正しい修正候補を選択させる。

【0087】

このような制御により、候補の選択に係る手順は増えるものの、各手順で提示される候補の数が制限されるため、膨大な数の修正候補の中から所望の修正候補を選択する場合に比べて、修正候補の探索に係るユーザへの負荷を軽減することが可能となる。また、正しい発音を選択させた後に、当該発音を示す修正候補を提示することで、音声認識処理に特有の誤りをより精度良く修正することが可能となる。

【0088】

なお、情報処理システムは、変形例 1-1 として説明した上記 UI（図 9 参照）と、上述した実施形態で説明した各 UI（例えば、図 1、図 5、図 6 参照）とを、取得されたコンテキスト情報が示す状態や状況に応じて選択的に切り替えてもよい。

【0089】

以上、変形例 1-1 として、音声に変換された文字情報を修正する場合における、修正

10

20

30

40

50



候補を提示するためのUIの一例について説明した。

【0090】

< 2.4.2. 変形例1-2：集音結果に基づく入力情報の状態に応じた制御 >

続いて、変形例1-2として、ユーザが発話した音声の集音結果に基づく入力情報の状態に応じて、当該音声に変換された文字情報を修正するためのUIの一例について説明する。

【0091】

例えば、ユーザが発話した語句（文字列）の長さがより短いほど、修正候補の数はより多くなる傾向にある。また、ユーザが発話した語句が、発音が類似する候補が大量にある語句の場合には、修正候補の数もより多くなる。例えば、図10は、変形例1-2に係る情報処理システムの概要について説明するための説明図であり、修正候補の数がより多くなる傾向にある語句の一例を示している。

10

【0092】

例えば、図10の上側の図は、ユーザが英語で発話した場合の一例である。具体的には、ユーザが「Oh」と発話し、情報処理システムは、当該ユーザが発話した音声の認識結果として、「No」という文字情報をユーザに提示している。この場合には、情報処理システムは、例えば、「Non」、「O」、「Ah」、...等のような発音が類似する修正候補が提示された表示情報V211をユーザに提示することとなる。

【0093】

また、図10の下側の図は、ユーザが日本語で発話した場合の一例である。具体的には、ユーザが「佑（ゆう）」と発話し、情報処理システムは、当該ユーザが発話した音声の認識結果として、「夕（ゆう）」という文字情報をユーザに提示している。この場合には、情報処理システムは、例えば、修正対象となる語句と発音が類似する修正候補として、「用」、「Yo」、「卯」、...等が提示された表示情報V213aや、同様の発音を示す修正候補（即ち、同音異義語）が提示された表示情報V213bをユーザに提示することとなる。

20

【0094】

このように、修正候補の数がより多くなるような状況下においては、当該修正候補の数に比例して、所望の修正候補を探索するために係るユーザへの負荷がより大きくなる傾向にある。

30

【0095】

このような状況を鑑み、変形例1-2に係る情報処理システムでは、コンテキスト情報として、ユーザが発話した音声の集音結果に基づく入力情報の状態に関する情報を取得し、取得した当該コンテキスト情報に基づき、修正候補の数がより多くなるか否かを認識する。そして、情報処理システムは、修正候補の数がより多くなる状態にあると認識した場合には、多くの修正候補の中から所望の修正候補の探索を補助するためのUIをユーザに提示する。例えば、図11及び図12は、変形例1-2に係る情報処理システムが提供するUIの一例について説明するための説明図である。

【0096】

例えば、図11に示すように、情報処理システムは、修正対象となる語句の修正候補とともに、修正候補の数を示す情報V221が提示された表示情報V22を提示してもよい。このような構成により、全ての修正候補が提示されていない場合においても、ユーザは、修正候補の数を認識することが可能となるため、所望の修正候補の探索に係る労力を見積もることが可能となる。

40

【0097】

また、図12に示す例では、情報処理システムは、例えば、右上の図に示すように、修正候補の数が多い場合には、日本語におけるひらがなやカタカナ等のように、発音を示す修正候補を優先して提示された表示情報V23aをユーザに提示している。また、このとき情報処理システムは、図11に示した例と同様に、修正候補の数を示す情報V231をあわせて提示してもよい。

50

## 【 0 0 9 8 】

また、情報処理システムは、提示した修正候補の中から、条件を指定して修正候補を絞り込むための機能を提供してもよい。

## 【 0 0 9 9 】

例えば、図 1 2 に示す例では、左上の図に示すように、ユーザが「佑（ゆう）」と発話し、情報処理システムは、発話された音声を「夕（ゆう）」と認識し、当該認識結果に応じて右上の図に示すように修正候補を提示している。このような状況の基で、情報処理システムは、例えば、左下の図に示すように、ユーザからの音声により、修正候補の画数の条件を示す情報の入力を受ける。例えば、図 1 2 に示す例では、修正候補の画数として「7」が指定されている。当該条件の指定を受けて、情報処理システムは、修正対象となる語句の修正候補のうち、指定された条件を満たす修正候補を抽出し、抽出した修正候補が提示された表示情報 V 2 3 b をユーザに提示している。例えば、図 1 2 に示す例では、情報処理システムは、画数が「7」修正候補を抽出し、抽出した当該修正候補をユーザに提示している。このような構成により、ユーザは、煩雑な操作を伴わずに、多数の修正候補の中から所望の修正候補を探索することが可能となる。なお、同機能は、日本語の場合に限らず、他の言語（例えば、英語）の場合についても適用可能である。具体的な一例として、英語の場合には、情報処理システムは、例えば、文字数の指定を受けて修正候補を抽出してもよい。

10

## 【 0 1 0 0 】

以上、変形例 1 - 2 として、図 1 0 ~ 図 1 2 を参照して、ユーザが発話した音声の集音結果に基づく入力情報の状態に応じて、当該音声を変換された文字情報を修正するための UI の一例について説明した。

20

## 【 0 1 0 1 】

< 2 . 4 . 3 . 変形例 1 - 3 : 情報入力の態様に応じた制御 >

続いて、変形例 1 - 3 として、ユーザによる情報処理システムに対する情報入力の態様に  
に応じて、音声を変換された文字情報を修正するための UI の一例について説明する。

## 【 0 1 0 2 】

例えば、図 1 3 ~ 図 1 6 は、変形例 1 - 3 に係る情報処理システムが提供する UI の一例について説明するための説明図である。なお、図 1 3 ~ 図 1 5 は、情報入力の態様として、特に、ユーザが音声を入力する場合における発話の態様（例えば、発話速度）に応じた制御の一例を示している。また、図 1 6 は、ユーザが音声を入力する際に、他の入力デバイス  
を介した操作を行っている場合における、当該操作の態様に  
応じた制御の一例を示している。

30

## 【 0 1 0 3 】

まず、図 1 3 に示す例に着目する。図 1 3 は、ユーザがゲームをプレイしている状況下において、音声により情報を入力する際に、当該音声をゆっくり発話している場合（即ち、発話速度が遅い場合）に提示される UI の一例を示している。具体的には、ユーザがゆっくり発話している状況下では、発話された語句は、例えば、固有名詞やゲーム内の専門用語等のような一般的に使用される語句に比べて認知度の低い語句の可能性が高い。このような状況下では、汎用的なコーパス参照に基づき修正候補を推定したとしても、ユーザの意図する修正候補の優先度がより低くなる可能性がある。そのため、変形例 1 - 3 に  
係る情報処理システムは、コンテキスト情報に基づきユーザがゆっくり発話している状況であることを認識した場合には、固有名詞のコーパスや、実行中の機能（例えば、ゲーム等）に固有のコーパスを参照することで修正候補を推定してもよい。

40

## 【 0 1 0 4 】

例えば、図 1 3 に示す例では、ユーザがゲーム内で使われる新しい語句である「cure」と発話し、情報処理システムは、発話された音声を「care」と認識し、当該認識結果に応じて修正候補が提示された表示情報 V 3 1 を提示している。なお、このとき情報処理システムは、実行中のゲームに固有のコーパスに基づき修正候補間の優先度を制御し、当該優先度の制御結果に基づき、修正候補を提示している。例えば、図 1 3 に示す例では、情報

50

処理システムは、修正対象となる語句「care」の修正候補として、「quea」、「cure」、「cua」、...等を提示している。また、情報処理システムは、上述した各種コーパス等のように、優先度を制御するために参照するデータを切り替えた場合には、参照したデータを切り替えたことを示す情報（例えば、参照したデータに応じた情報）V 3 1 1を、修正候補とあわせて提示してもよい。

**【 0 1 0 5 】**

なお、上記に説明した例では、主に発話速度に着目してコーパスを切り替える例について説明したが、必ずしも上記に説明した例のみには限定されない。即ち、発話速度に限らず、発話の態様（例えば、音量や抑揚等）に応じて、どのようなコーパスを使用し、どのように修正候補の優先度を制御するかについては、例えば、ユースケース等に応じて事前に設定しておけばよい。また、当該設定をユーザが所定の操作により適宜切り替えられるように、情報処理システムが構成されていてもよい。

10

**【 0 1 0 6 】**

次いで、図 1 4 に示す例に着目する。図 1 4 は、ユーザの音声を速く発話している場合（即ち、発話速度が速い場合）に提示されるUIの一例を示している。ユーザの発話速度が速い場合には、何らかの原因によりユーザが急いでいる状況や焦っている状況が想定され得る。そのため、変形例 1 - 3 に係る情報処理システムは、コンテキスト情報に基づきユーザが速く発話している状況であることを認識した場合には、よりユーザが修正候補を選択しやすいように、当該修正候補の提示態様を制御してもよい。

**【 0 1 0 7 】**

20

例えば、図 1 4 に示す例では、情報処理システムは、ユーザが速く発話している状況であると認識した場合には、参照符号 V 3 2 として示すように、修正候補の数を制限することで、当該修正候補の視認性を向上させている。なお、提示する修正候補の数については、例えば、ユースケース等を考慮してあらかじめ設定しておけばよい。また、他の一例として、情報処理システムは、ユーザが速く発話している状況下において、当該ユーザが発話した音声を一意的に認識できなかった場合には、修正候補を提示せずに、音声の再入力を当該ユーザに促してもよい。

**【 0 1 0 8 】**

次いで、図 1 5 に示す例に着目する。図 1 5 は、ユーザが音声を速く発話している場合（即ち、発話速度が速い場合）に提示されるUIの他の一例であり、特に、ユーザが発話した日本語の音声を認識する場合の一例を示している。例えば、図 1 5 に示す例では、ユーザが、「薬草はどこ」と発話し、情報処理システムは、発話された音声を「やくしょうはどこ」と認識している。このとき、情報処理システムは、ユーザが速く発話している状況下においては、音声の認識結果として、漢字等の意味内容を考慮した文字情報（文章等）よりも、ひらがな等のような発音を示す文字情報を優先して提示してもよい。このような制御により、例えば、ユーザが何らかの要因により急いでいるような状況下においても、音声に変換された文字情報を修正する際に、漢字への変換の誤りが発生しにくくなり、当該修正もより容易になる。

30

**【 0 1 0 9 】**

次いで、図 1 6 に示す例に着目する。図 1 6 は、ユーザが音声を入力する際に、他の入力デバイスを介した操作を行っている状況下において、ユーザが余裕を持って入力デバイスを操作している（例えば、ゆっくり操作している）場合に提示されるUIの一例である。具体的には、ユーザが余裕を持って操作を行っている状況下では、音声に変換された文字情報の確認に多くの時間を割ける可能性が高い。そのため、このような状況下では、情報処理システムは、音声に変換された文字情報の修正候補として、複数種類の修正候補を提示してもよい。

40

**【 0 1 1 0 】**

例えば、図 1 6 に示す例では、情報処理システムは、修正対象となる語句の修正候補として、互いに異なるアルゴリズムに基づき修正候補が提示された表示情報 V 1 1、V 1 3、及び V 3 1 を提示している。例えば、表示情報 V 1 1 は、図 2 を参照して説明したよう

50

に、発音の類似度に応じて優先度が制御された修正候補を提示するための表示情報である。また、表示情報 V 1 3 は、図 5 を参照して説明したように、修正対象となる語句と他の語句との間の共起関係に応じて優先度が制御された修正候補を提示するための表示情報である。また、表示情報 V 3 1 は、図 1 3 を参照して説明したように、実行中の機能（例えば、ゲーム等）に固有のコーパスを参照することで推定された修正候補を提示するための表示情報である。

【 0 1 1 1 】

以上、変形例 1 - 3 として、ユーザによる情報処理システムに対する情報入力の態様に  
応じて、音声に変換された文字情報を修正するための UI の一例について説明した。

【 0 1 1 2 】

< 2 . 5 . 実施例 >

続いて、上述した実施形態や変形例に係る情報処理システムのその他の実施例について  
具体的な例を挙げて説明する。

【 0 1 1 3 】

例えば、ユーザがゲームをプレイしているようなユースケースを想定した場合には、コ  
ントローラやタッチパネル等への入力の速度や強さ（圧力）を示す情報をコンテキスト情  
報として利用することが可能である。具体的には、入力デバイスに対する入力が速い場合  
や、入力時の圧力が強い場合には、何らかの原因によりユーザが急いでいる状況や焦っ  
ている状況が想定され得る。そのため、この場合には、情報処理システムは、例えば、ユー  
ザが急いでいる状況や焦っている状況の認識結果に応じて、提示する修正候補の優先度や  
、当該修正候補の提示態様を制御してもよい。また、前述したように、ユーザが余裕を持  
って入力デバイスを操作している（例えば、ゆっくり操作している）場合には、当該ユー  
ザは、各種作業を行うために比較的多くの時間を割ける状況であると認識することが可能  
である。

【 0 1 1 4 】

また、カメラ等の撮像部により撮像されたユーザの画像を、当該ユーザの状態や状況を  
認識するためのコンテキスト情報として利用することが可能である。具体的な一例として  
、撮像部により撮像されたユーザの顔の画像を解析することで、当該ユーザの表情を推定  
し、当該推定結果に基づき当該ユーザの状態を認識することが可能である。より具体的  
には、ユーザの顔の険しさから、当該ユーザの焦りなどを推定することが可能である。

【 0 1 1 5 】

また、撮像部により撮像されたユーザの画像の解析結果や、HMD（Head Mounted D  
isplay）に設けられた赤外線センサの検出結果等によりユーザの視線を検出することによ  
り、当該視線の検出結果をコンテキスト情報として利用することも可能である。具体的  
な一例として、ユーザがゲームをプレイしているような状況下において、ユーザの視線が、  
テキスト情報の入力エリアではなく、プレイ画面が提示されるエリアに移動している場  
合には、何らかの原因によりユーザが急いでいる状況や焦っている状況が想定され得る。  
これは、ユーザの視線が、テキスト情報の入力エリアの外部に移動している場合につい  
ても同様である。

【 0 1 1 6 】

また、前述したように、発汗量等の生体情報の検出結果をコンテキスト情報として利用  
することも可能である。具体的な一例として、発汗量が普段より多い場合には、何らか  
の原因によりユーザが急いでいる状況や焦っている状況が想定され得る。

【 0 1 1 7 】

以上、上述した実施形態や変形例に係る情報処理システムのその他の実施例について具  
体的な例を挙げて説明した。なお、上述した例はあくまで一例であり、必ずしも本実施  
形態に係る情報処理システムの利用形態を限定するものではない。

【 0 1 1 8 】

< 2 . 6 . 評価 >

以上、説明したように、本実施形態に係る情報処理システムは、音声の集音に関する期

10

20

30

40

50

間を取得されたコンテキスト情報に応じた態様で、集音された音声を変換された文字情報の候補を提示する。具体的には、情報処理システムは、コンテキスト情報が示す各種状態や状況に応じて、音声を変換された文字情報のうち少なくとも一部の語句について、他の候補（即ち、修正候補）間の優先度を制御する。また、情報処理システムは、コンテキスト情報が示す各種状態や状況に応じた提示態様で、音声を変換された文字情報のうち少なくとも一部の語句に対応する修正候補を提示する。

#### 【0119】

このような制御により、本実施形態に係る情報処理システムは、音声の認識結果に誤りが生じたとしても、その時々々の状態や状況に応じて動的に、各修正候補の優先度を制御し、当該優先度の制御結果に基づき、より好適な修正候補をユーザに提示することが可能となる。このような構成により、ユーザは、限られた修正候補の中から、本来意図していた内容を示す候補を特定できる可能性がより高くなるため、多数の候補を確認するといった煩雑な操作を伴わずに、簡便な操作により音声の認識結果を修正することが可能となる。

10

#### 【0120】

また、本実施形態に係る情報処理システムに依れば、その時々々の状態や状況に応じてより好適な提示態様（例えば、優先度制御のアルゴリズムに応じた提示態様）で、修正対象となる語句の修正候補が提示されることとなる。そのため、ユーザは、煩雑な確認作業を伴わずに修正対象となる語句の修正候補を、各種状態や状況に応じて直感的に認識することが可能となり、ひいては、所望の修正候補をより簡便な操作により選択することが可能となる。

20

#### 【0121】

<< 3 . 第2の実施形態 >>

< 3 . 1 . 概要 >

続いて、本開示の第2の実施形態に係る情報処理システムについて説明する。一般的に、音声により情報を入力する場合には、キーボード等の入力デバイスを介して情報を入力する場合に比べて、当該音声の入力を開始してから当該音声の認識結果が確認可能となるまでに時間を要する傾向にある。

#### 【0122】

ここで、図17を参照して具体的な例を挙げて説明する。図17は、本実施形態に係る情報処理システムの概要について説明するための説明図であり、発話された音声を認識し当該音声を文字情報に変換してユーザに提示するUIの一例を示している。例えば、図17に示す例では、ユーザが「Walk in the park」と発話し、情報処理システムは、当該ユーザが発話した音声の認識結果として、「Work in the park」という文字情報をユーザに提示している。即ち、図17に示す例では、情報処理システムは、ユーザが発話した「Walk」という語句を、誤って「Work」と認識したことになる。また、図17に示す例では、情報処理システムは、ユーザから誤って認識された「Work」という語句の指定を受けて、当該語句の修正候補を提示している。

30

#### 【0123】

一方で、図17に示す例では、ユーザが発話した「Walk in the park」という一連の音声認識された後に、当該音声の認識結果が提示されることとなる。このような場合には、例えば、入力した音声の認識結果が誤って認識されたとしても、ユーザは、一連の音声に対する認識処理（例えば、音声認識処理（音響解析）や自然言語処理（言語解析）等）が完了するまで、当該認識に誤りが生じていることや、当該誤りを修正する場合の修正候補が正しく特定されているか否かを認識することが困難となる。そのため、例えば、ユーザが、提示された修正候補の中に所望の修正候補が存在しないことを認識し、再度音声入力を行った方が望ましいか否かの意思決定を行うとしても、当該意思決定が行われるのは、一連の認識処理が完了した後となる。

40

#### 【0124】

このような状況を鑑み、本実施形態に係る情報処理システムでは、音声の認識時に誤りが生じた場合に、ユーザが、より早いタイミングで、再度音声入力を行った方が望ましい

50

か否かの意思決定を行うことが可能な仕組みを提供する。例えば、図 18 は、本実施形態に係る情報処理システムが提供する UI の概要について説明するための説明図である。図 18 に示す例では、ユーザが「Walk in the park」と発話し、情報処理システムは、当該音声を「Work in the park」と認識したものとする。

【0125】

本実施形態に係る情報処理システムは、逐次集音される音声に対して音声認識処理や自然言語処理等の各種認識処理を逐次実行し、集音される一連の音声のうち、当該認識処理が施された音声に対応する語句の文字情報を、当該語句の他の候補（即ち、修正候補）とあわせて逐次提示する。

【0126】

例えば、図 18 の 1 段目の図は、集音された「Walk in the park」という音声のうち、「Walk」という語句を認識したタイミングで提示される表示情報の一例を示している。このとき、情報処理システムは、「Walk」という音声の認識結果として「Work」という語句を提示するとともに、当該語句の他の候補として「Walk」及び「Weak」を提示している。これにより、ユーザは、このタイミングにおいて、「Walk」という音声が誤って「Work」と認識されていることと、他の候補として実際に発話した語句である「Walk」が特定されていることとを認識することが可能となる。そのため、ユーザは、例えば、このまま発話を継続し、後ほど誤った認識結果に基づく「Work」という語句を、候補として提示された「Walk」に置き換えればよいという意思決定を、このタイミングで行うことが可能となる。

【0127】

同様に、図 18 の 2 段目の図は、集音された「Walk in the park」という音声のうち、「in」という語句を認識したタイミングで提示される表示情報の一例を示している。また、図 18 の 3 段目の図は、集音された「Walk in the park」という音声のうち、「park」という語句を認識したタイミングで提示される表示情報の一例を示している。このように、本実施形態に係る情報処理システムは、集音される一連の音声を音声認識処理に基づき文字情報に変換する際に、当該音声を文字情報に変換する処理の経過を逐次提示する。

【0128】

そして、図 18 の 4 段目の図のように、集音された一連の音声の認識が完了すると、5 段目の図に示すように、当該音声に変換された文字情報を修正するための UI が提示される。

【0129】

以上のような制御により、ユーザは、例えば、実際に発話した「Walk」という語句を、提示された修正候補の中から選択することで、誤って認識された「Work」という語句を修正することが可能となる。

【0130】

また、例えば、発話した一連の音声のうち一部の語句が誤って認識され、かつ、当該語句の他の候補としても実際に発話した語句が特定されていなかった場合も想定され得る。このような状況下においても、本実施形態に係る情報処理システムに依れば、ユーザは、一連の音声の認識結果が提示される前に、対象となる語句の認識結果が提示されたタイミングで、例えば、発話を中断し再度音声入力を行うというような意思決定が可能となる。

【0131】

以上、本開示の第 2 の実施形態に係る情報処理システムの概要について説明した。なお、以降では、本実施形態に係る情報処理システムについてさらに詳しく説明する。

【0132】

< 3.2. 構成 >

続いて、図 19 を参照して、本実施形態に係る情報処理システムの機能構成の一例について説明する。図 19 は、本実施形態に係る情報処理システムの機能構成の一例を示したブロック図である。

10

20

30

40

50

## 【 0 1 3 3 】

図 1 9 に示すように、本実施形態に係る情報処理システム 2 は、入出力装置 3 0 と、情報処理装置 2 0 とを含む。なお、入出力装置 3 0 は、前述した第 1 の実施形態に係る情報処理システム 1 における入出力装置 3 0 に相当する。即ち、図 1 9 について、図 4 と同様の符号を付した構成は、当該図 4 に示す例と同様の構成を示すものとする。そのため、本説明では、主に、情報処理装置 2 0 の構成に着目して説明し、前述した第 1 の実施形態に係る情報処理システム 1 と同様の構成については詳細な説明は省略する。

## 【 0 1 3 4 】

また、図 1 9 に示す例では、情報処理装置 2 0 及び入出力装置 3 0 の構成をよりわかりやすくするために、当該情報処理装置 2 0 及び入出力装置 3 0 がネットワークを介して情報を送受信するための通信部の図示を省略している。なお、当該通信部の構成は、前述した第 1 の実施形態に係る情報処理システム 1 ( 図 4 参照 ) と同様である。また、以降の説明では、情報処理装置 2 0 及び入出力装置 3 0 それぞれの内部の各構成が、ネットワークを介して外部装置と情報の送受信を行う場合には、特に説明が無い場合においても、当該通信部を介して情報の送受信を行うものとする。

## 【 0 1 3 5 】

図 1 9 に示すように、情報処理装置 2 0 は、音声認識処理部 2 0 1 と、信頼度算出部 2 0 3 と、出力制御部 2 0 5 とを含む。

## 【 0 1 3 6 】

音声認識処理部 2 0 1 は、入出力装置 3 0 の入力部 3 1 0 ( 集音部 3 1 1 ) により集音された音響中に含まれる音声を文字情報に変換するための構成である。なお、音声を文字情報に変換するための仕組みについては、前述した第 1 の実施形態に係る情報処理装置 1 0 における音声認識処理部 1 0 1 と同様のため詳細な説明は省略する。

## 【 0 1 3 7 】

音声認識処理部 2 0 1 は、入力部 3 1 0 から音声の集音結果に基づく音響信号を逐次取得し、取得した音響信号からの音声信号の抽出、抽出した音声信号に対する音声認識処理 ( 音響解析 ) 、及び当該音声認識処理の結果に対する自然言語処理 ( 言語解析 ) を逐次実行する。以上のような制御により、音声認識処理部 2 0 1 は、逐次入力される音声を、逐次文字情報に変換する。

## 【 0 1 3 8 】

また、このとき音声認識処理部 2 0 1 は、音声認識処理や自然言語処理に基づき文字情報に変換された語句の他の候補を特定してもよい。例えば、音声認識処理部 2 0 1 は、音声認識処理等の音響解析の結果に基づき、音声から文字情報に変換された語句と発音が類似する他の候補を特定してもよい。また、他の一例として、音声認識処理部 2 0 1 は、自然言語処理等の言語解析の結果に基づき、音声に変換された文字情報中に含まれる少なくとも一部の語句について、当該語句の他の候補を、他の語句との間の共起関係に基づき特定してもよい。

## 【 0 1 3 9 】

そして、音声認識処理部 2 0 1 は、取得された音響信号に対する音声認識処理や自然言語処理の結果に基づき、音声に変換された文字情報や、当該文字情報中の語句の他の候補を示す情報を、出力制御部 2 0 5 に逐次出力する。また、音声認識処理部 2 0 1 は、取得された音響信号に対する音響解析 ( 例えば、音声認識処理 ) の結果や、当該音響解析の結果に対する言語解析 ( 例えば、自然言語処理 ) の結果を示す情報を、信頼度算出部 2 0 3 に逐次出力する。

## 【 0 1 4 0 】

信頼度算出部 2 0 3 は、音声認識処理部 2 0 1 から、取得された音響信号に対する音響解析 ( 例えば、音声認識処理 ) の結果や、当該音響解析の結果に対する言語解析 ( 例えば、自然言語処理 ) の結果を示す情報を逐次取得する。そして、信頼度算出部 2 0 3 は、取得した当該情報に基づき、音声に変換された文字情報に含まれる語句や当該語句の他の候補の信頼度を算出する。

10

20

30

40

50

## 【 0 1 4 1 】

具体的な一例として、信頼度算出部 2 0 3 は、音響解析の結果に基づき、発音の類似度に応じて、音声に変換された文字情報に含まれる語句や当該語句の他の候補の信頼度を算出してよい。また、他の一例として、信頼度算出部 2 0 3 は、言語解析の結果に基づき、音声に変換された文字情報のうち対象となる語句と既に認識された他の語句との間の共起関係に基づき、当該対象となる語句や当該語句の他の候補の信頼度を算出してよい。もちろん、上記に説明した信頼度の算出方法はあくまで一例であり、必ずしも上記に説明した例には限定されない。具体的な一例として、信頼度算出部 2 0 3 は、音響モデル、言語モデル、集音時の状態や状況、及び実行中のコンテンツ等の条件を複合的に判断することで、信頼度を算出してよい。

10

## 【 0 1 4 2 】

そして、信頼度算出部 2 0 3 は、音声に変換された文字情報に含まれる語句や当該語句の他の候補の信頼度の算出結果を示す情報を出力制御部 2 0 5 に出力する。

## 【 0 1 4 3 】

出力制御部 2 0 5 は、音声認識処理部 2 0 1 から、音声に変換された文字情報や、当該文字情報中の語句の他の候補を示す情報を逐次取得する。また、出力制御部 2 0 5 は、信頼度算出部 2 0 3 から、当該文字情報に含まれる語句や当該語句の他の候補の信頼度の算出結果を示す情報を取得する。

## 【 0 1 4 4 】

出力制御部 2 0 5 は、音声に変換された文字情報を取得すると、逐次、当該文字情報を入出力装置 3 0 に提示させる。また、このとき出力制御部 2 0 5 は、当該文字情報中の語句の他の候補のうち少なくとも一部の候補を、当該語句と当該他の候補とのそれぞれの信頼度の算出結果に基づき当該語句に関連付けて、入出力装置 3 0 に提示させる。以上のような制御に基づき、出力制御部 2 0 5 は、入出力装置 3 0 に、音声に変換された文字情報や当該文字情報に含まれる語句の修正候補を提示するための UI (例えば、図 1 8 参照) を逐次更新させる。

20

## 【 0 1 4 5 】

なお、出力制御部 2 0 5 は、入出力装置 3 0 に修正候補を提示させる際に、各修正候補の信頼度に応じて当該修正候補の提示態様を制御してもよい。具体的な一例として、出力制御部 2 0 5 は、修正候補が表示される際に、修正候補の色、大きさ、濃淡等を当該修正候補の信頼度に応じて制御してもよい。このような制御により、ユーザは、各修正候補の提示態様に応じて、当該修正候補の信頼度を直感的に認識することが可能となる。

30

## 【 0 1 4 6 】

なお、図 1 9 に示した情報処理システムの機能構成はあくまで一例であり、上述した各構成の機能を実現することが可能であれば、当該情報処理システムの機能構成は必ずしも図 1 9 に示す例のみには限定されない。例えば、情報処理装置 2 0 の一部の構成が、当該情報処理装置 2 0 の外部に設けられていてもよい。具体的な一例として、出力制御部 2 0 5 に相当する構成が、入出力装置 3 0 側に設けられていてもよい。この場合には、入出力装置 3 0 は、音声に変換された文字情報、当該文字情報中の語句の他の候補を示す情報、及び当該当該文字情報に含まれる語句や当該語句の他の候補の信頼度の算出結果を示す情報を、情報処理装置 2 0 から取得すればよい。

40

## 【 0 1 4 7 】

また、他の一例として、情報処理装置 2 0 の各構成が、前述した第 1 の実施形態に係る情報処理装置 1 0 (図 4 参照)の一部として構成されていてもよい。なお、この場合には、情報処理装置 2 0 の各構成の機能のうち、少なくとも一部の機能が、情報処理装置 1 0 に含まれるいずれかの構成により実現されてもよい。具体的な一例として、情報処理装置 1 0 に含まれる音声認識処理部 1 0 1 により、前述した音声認識処理部 2 0 1 の機能が実現されてもよい。

## 【 0 1 4 8 】

また、入力部 3 1 0 及び出力部 3 3 0 に含まれる各構成のうち少なくとも一部の構成が

50



、入出力装置 30 の外部に設けられていてもよい。また、入出力装置 30 と情報処理装置 10 とが一体的に構成されていてもよい。

【0149】

以上、図 19 を参照して、本実施形態に係る情報処理システムの機能構成の一例について説明した。

【0150】

< 3.3. 処理 >

続いて、図 20 を参照して、本実施形態に係る情報処理システムの一連の処理の流れの一例について説明する。図 20 は、本実施形態に係る情報処理システムの一連の処理の流れの一例について示したフローチャートである。

【0151】

図 20 に示すように、入出力装置 30 (集音部 311) は、ユーザの音声等のような音響を集音し、当該集音結果に基づく音響信号を、ネットワークを介して情報処理装置 20 に逐次送信する (S201)。

【0152】

情報処理装置 20 (音声認識処理部 201) は、入出力装置 30 から音響の集音結果に基づく音響信号を逐次取得し、取得した当該音響信号に対して、音響解析や言語解析を施すことで、当該音響中に含まれる音声を文字情報に変換する。また、このとき情報処理装置 20 (音声認識処理部 201) は、音声認識処理や自然言語処理に基づき文字情報に変換された語句の他の候補を特定してもよい。また、情報処理装置 20 (信頼度算出部 203) は、音響解析や言語解析の結果に基づき、音声に変換された文字情報に含まれる語句や当該語句の他の候補の信頼度を算出する (S203)。

【0153】

情報処理装置 20 (出力制御部 205) は、音声を文字情報に変換すると、当該文字情報を入出力装置 30 に提示させる。また、このとき情報処理装置 20 は、当該文字情報中の語句の他の候補のうち少なくとも一部の候補を、当該語句と当該他の候補とのそれぞれの信頼度の算出結果に基づき当該語句に関連付けて、入出力装置 30 に提示させる。以上のようにして、情報処理装置 20 は、入出力装置 30 に、音声に変換された文字情報や当該文字情報に含まれる語句の修正候補を提示するための UI を更新させる (S207)。

【0154】

参照符号 S201 ~ S205 に示した一連の処理が、発話の終了が検知される逐次実行される (S207、NO)。そして、発話の終了が検知されると (S207、YES)、情報処理システムは、当該一連の処理を終了する。なお、この後、情報処理装置 20 は、例えば、集音された一連の音声に変換された文字情報を修正するための UI を入出力装置 30 に提示させてもよい。

【0155】

以上、図 20 を参照して、本実施形態に係る情報処理システムの一連の処理の流れの一例について説明した。

【0156】

< 3.4. 変形例 >

続いて、本実施形態に係る情報処理システムの変形例について説明する。

【0157】

< 3.4.1. 変形例 2-1: 認識結果を確定するためのインタフェースの一例 >

まず、変形例 2-1 として、音声の認識結果に基づき逐次提示される語句を確定するためのインタフェースの一例について説明する。前述したように、集音された音声は、例えば、当該音声に対する音響解析 (例えば、音声認識処理) や言語解析 (例えば、自然言語処理) に基づき文字情報に変換される。そのため、例えば、音響解析の結果に基づく音声の認識結果が、その後に行われる言語解析の結果に基づき修正される場合も想定される。一方で、後から実行される言語解析の結果に基づく修正により、必ずしも正しい認識結果が提示されるとは限らない。

10

20

30

40

50

## 【 0 1 5 8 】

例えば、図 2 1 は、本実施形態に係る情報処理システムによる制御の一例について説明するための説明図である。図 2 1 に示す例では、ユーザが「Walk in the park」と発話している。また、情報処理システムは、発話された「Walk」という音声を、まず音響解析の結果に基づき「Walk」と認識し、その後に認識された「in」という語句との間の共起確率に基づき、認識した「Walk」という語句を「Work」に修正している。一方で、図 2 1 に示す例の場合には、ユーザは、「Walk」と発話しているため、当該ユーザにとっては、「Walk」と認識された語句を「Work」に修正する処理が、必ずしも望ましい処理とは限らない。

## 【 0 1 5 9 】

そこで、変形例 2 - 1 に係る情報処理システムは、音声の認識結果に基づき逐次提示される語句を確定するためのインタフェースをユーザに提供する。例えば、図 2 2 は、変形例 2 - 1 に係る情報処理システムが提供する UI の一例について説明するための説明図である。

## 【 0 1 6 0 】

図 2 1 に示す例では、ユーザが「Walk in the park」と発話している。これに対し、例えば、図 2 2 の左側の図に示すように、情報処理システムは、発話された「Walk」という音声を、音響解析の結果に基づき「Walk」と認識し、音声の認識結果として「Walk」という語句を提示している。このとき、変形例 2 - 1 に係る情報処理システムは、ユーザから所定の入力デバイス（例えば、コントローラ等）を介して、認識結果の確定を示す指示を受けると、提示した「Walk」という語句を対応する音声の認識結果として確定する。これにより、図 2 2 の右側の図に示すように、その後に「in」という語句が認識されたとしても、当該語句との共起確率により、「Walk」という語句の認識結果が「Work」という語句に修正されるといった事態の発生が抑止される。

## 【 0 1 6 1 】

なお、情報処理システムは、認識結果を確定するための指示を受け付けている状態か否かや、当該確定に係る指示の対象となる語句を、識別可能にユーザに提示してもよい。具体的な一例として、情報処理システムは、確定に係る指示の対象となる語句を、他の語句とは異なる態様（例えば、異なる色）で提示する等の制御により強調表示してもよい。

## 【 0 1 6 2 】

このような仕組みにより、ユーザは、例えば、音声の認識結果に基づき逐次提示された語句が意図するもの（即ち、実際に発話した語句）である場合には、情報処理システムに対して所定の指示を行うことで、当該語句が以降に実行される処理に基づき修正される事態を防止することが可能となる。

## 【 0 1 6 3 】

以上、変形例 2 - 1 として、図 2 1 及び図 2 2 を参照して、音声の認識結果に基づき逐次提示される語句を確定するためのインタフェースの一例について説明した。

## 【 0 1 6 4 】

< 3 . 4 . 2 . 変形例 2 - 2 : 状態や状況に応じた修正候補の制御 >

続いて、変形例 2 - 2 として、本実施形態に係る情報処理システムと、前述した第 1 の実施形態に係る情報処理システムとを組み合わせることで、各種状況や状態に応じて提示する修正候補を制御する場合の一例について説明する。例えば、図 2 3 は、変形例 2 - 2 に係る情報処理システムが提供する UI の一例について説明するための説明図である。

## 【 0 1 6 5 】

まず、図 2 3 の上側の図に着目する。図 2 3 の上側の図は、雑音の影響が小さい場合（即ち、静かな環境である場合）に提示される UI の一例について示しており、ユーザの「Have a flower in my garden」という発話に対して、情報処理システムは、「flower」という音声の認識結果として「floor」という語句をユーザに提示している。また、変形例 2 - 2 に係る情報処理システムは、コンテキスト情報に基づき雑音の影響が小さい環境下であることを認識した場合には、発音の類似度に基づき修正候補間の優先度を制御し

10

20

30

40

50

、当該優先度の制御結果に基づき、対象となる語句の修正候補を提示してもよい。例えば、図23の上側の図では、情報処理システムは、「floor」という語句の他の候補（即ち、修正候補）として、「floor」と発音が類似する「flower」及び「four」という語句を提示している。

【0166】

次いで、図23の下側の図に着目する。図23の下側の図は、雑音の影響が大きい（即ち、うるさい環境である場合）場合に提示されるUIの一例について示しており、ユーザの「Have a flower in my garden」という発話に対して、情報処理システムは、「flower」という音声の認識結果として「car」という語句をユーザに提示している。また、変形例2-2に係る情報処理システムは、コンテキスト情報に基づき雑音の影響が大きい環境下であることを認識した場合には、対象となる語句と他の語句との間の共起関係に基づき優先度を制御し、当該優先度の制御結果に基づき、対象となる語句の修正候補を提示してもよい。例えば、図23の下側の図では、情報処理システムは、「car」という語句の他の候補（即ち、修正候補）として、「Have」という語句との間の共起関係に基づき「flower」及び「cookie」という語句を提示している。

10

【0167】

以上のような制御により、変形例2-2に係る情報処理システムは、その時々々の状態や状況に応じて動的に、より好適な修正候補をユーザに提示することが可能となる。

【0168】

以上、変形例2-2として、図23を参照して、本実施形態に係る情報処理システムと、前述した第1の実施形態に係る情報処理システムとを組み合わせることで、各種状況や状態に応じて提示する修正候補を制御する場合の一例について説明する。

20

【0169】

< 3.5. 評価 >

以上説明したように、本実施形態に係る情報処理システムは、逐次集音される音声に対して音声認識処理を逐次実行し、集音される一連の音声のうち、当該音声認識処理が施された音声に対応する語句の文字情報を、当該語句の他の候補（即ち、修正候補）とあわせて逐次提示する。このような制御により、例えば、発話した一連の音声のうち一部の語句が誤って認識されたとしても、ユーザは、一連の音声の認識結果が提示される前に当該誤りを認識することが可能となる。また、誤って認識された語句の他の候補として、実際に発話した語句が情報処理システムにより特定されていない状況下においても、ユーザは、当該状況を、対象となる語句の修正候補が提示されたタイミングで認識することが可能となる。即ち、本実施形態に係る情報処理システムに依れば、ユーザは、一連の音声の認識結果が提示される前に、対象となる語句の認識結果が提示されたタイミングで、例えば、発話を中断し再度音声入力を行うという意思決定が可能となる。

30

【0170】

<< 4. ハードウェア構成 >>

次に、図24を参照しながら、前述した情報処理装置10及び20と、入出力装置30とのように、本実施形態に係る情報処理システム1を構成する情報処理装置900のハードウェア構成について、詳細に説明する。図24は、本開示の一実施形態に係る情報処理システム1を構成する情報処理装置900のハードウェア構成の一例を示す機能ブロック図である。

40

【0171】

本実施形態に係る情報処理システム1を構成する情報処理装置900は、主に、CPU901と、ROM903と、RAM905と、を備える。また、情報処理装置900は、更に、ホストバス907と、ブリッジ909と、外部バス911と、インタフェース913と、入力装置915と、出力装置917と、ストレージ装置919と、ドライブ921と、接続ポート923と、通信装置925とを備える。

【0172】

CPU901は、演算処理装置及び制御装置として機能し、ROM903、RAM90

50

5、ストレージ装置 919 又はリムーバブル記録媒体 927 に記録された各種プログラムに従って、情報処理装置 900 内の動作全般又はその一部を制御する。ROM 903 は、CPU 901 が使用するプログラムや演算パラメータ等を記憶する。RAM 905 は、CPU 901 が使用するプログラムや、プログラムの実行において適宜変化するパラメータ等を一次記憶する。これらは CPU バス等の内部バスにより構成されるホストバス 907 により相互に接続されている。なお、図 4 を参照して前述した、音声認識処理部 101、コンテキスト解析部 103、優先度制御部 105、及び出力制御部 107 は、例えば、CPU 901 により実現され得る。

#### 【0173】

ホストバス 907 は、ブリッジ 909 を介して、PCI (Peripheral Component Interconnect / Interface) バスなどの外部バス 911 に接続されている。また、外部バス 911 には、インタフェース 913 を介して、入力装置 915、出力装置 917、ストレージ装置 919、ドライブ 921、接続ポート 923 及び通信装置 925 が接続される。

#### 【0174】

入力装置 915 は、例えば、マウス、キーボード、タッチパネル、ボタン、スイッチ、レバー及びペダル等、ユーザが操作する操作手段である。また、入力装置 915 は、例えば、赤外線やその他の電波を利用したリモートコントロール手段（いわゆる、リモコン）であってもよいし、情報処理装置 900 の操作に対応した携帯電話や PDA 等の外部接続機器 929 であってもよい。さらに、入力装置 915 は、例えば、上記の操作手段を用いてユーザにより入力された情報に基づいて入力信号を生成し、CPU 901 に出力する入力制御回路などから構成されている。情報処理装置 900 のユーザは、この入力装置 915 を操作することにより、情報処理装置 900 に対して各種のデータを入力したり処理動作を指示したりすることができる。なお、図 4 を参照して前述した入力部 310 は、例えば、入力装置 915 により実現され得る。

#### 【0175】

出力装置 917 は、取得した情報をユーザに対して視覚的又は聴覚的に通知することが可能な装置で構成される。このような装置として、CRT ディスプレイ装置、液晶ディスプレイ装置、プラズマディスプレイ装置、EL ディスプレイ装置及びランプ等の表示装置や、スピーカ及びヘッドホン等の音声出力装置や、プリンタ装置等がある。出力装置 917 は、例えば、情報処理装置 900 が行った各種処理により得られた結果を出力する。具体的には、表示装置は、情報処理装置 900 が行った各種処理により得られた結果を、テキスト又はイメージで表示する。他方、音声出力装置は、再生された音声データや音響データ等からなるオーディオ信号をアナログ信号に変換して出力する。なお、図 6 を参照して前述した出力部 330 は、例えば、出力装置 917 により実現され得る。

#### 【0176】

ストレージ装置 919 は、情報処理装置 900 の記憶部の一例として構成されたデータ格納用の装置である。ストレージ装置 919 は、例えば、HDD (Hard Disk Drive) 等の磁気記憶部デバイス、半導体記憶デバイス、光記憶デバイス又は光磁気記憶デバイス等により構成される。このストレージ装置 919 は、CPU 901 が実行するプログラムや各種データ等を格納する。

#### 【0177】

ドライブ 921 は、記録媒体用リーダーライターであり、情報処理装置 900 に内蔵、あるいは外付けされる。ドライブ 921 は、装着されている磁気ディスク、光ディスク、光磁気ディスク又は半導体メモリ等のリムーバブル記録媒体 927 に記録されている情報を読み出して、RAM 905 に出力する。また、ドライブ 921 は、装着されている磁気ディスク、光ディスク、光磁気ディスク又は半導体メモリ等のリムーバブル記録媒体 927 に記録を書き込むことも可能である。リムーバブル記録媒体 927 は、例えば、DVD メディア、HD-DVD メディア又は Blu-ray (登録商標) メディア等である。また、リムーバブル記録媒体 927 は、コンパクトフラッシュ (登録商標) (CF: Compact

10

20

30

40

50

ctFlash)、フラッシュメモリ又はSDメモリカード(Secure Digital memory card)等であってもよい。また、リムーバブル記録媒体927は、例えば、非接触型ICチップを搭載したICカード(Integrated Circuit card)又は電子機器等であってもよい。

#### 【0178】

接続ポート923は、情報処理装置900に直接接続するためのポートである。接続ポート923の一例として、USB(Universal Serial Bus)ポート、IEEE1394ポート、SCSI(Small Computer System Interface)ポート等がある。接続ポート923の別の例として、RS-232Cポート、光オーディオ端子、HDMI(登録商標)(High-Definition Multimedia Interface)ポート等がある。この接続ポート923に外部接続機器929を接続することで、情報処理装置900は、外部接続機器929から直接各種のデータを取得したり、外部接続機器929に各種のデータを提供したりする。

10

#### 【0179】

通信装置925は、例えば、通信網(ネットワーク)931に接続するための通信デバイス等で構成された通信インタフェースである。通信装置925は、例えば、有線若しくは無線LAN(Local Area Network)、Bluetooth(登録商標)又はWUSB(Wireless USB)用の通信カード等である。また、通信装置925は、光通信用のルータ、ADSL(Asymmetric Digital Subscriber Line)用のルータ又は各種通信用のモデム等であってもよい。この通信装置925は、例えば、インターネットや他の通信機器との間で、例えばTCP/IP等の所定のプロトコルに則して信号等を送受信することができる。また、通信装置925に接続される通信網931は、有線又は無線によって接続されたネットワーク等により構成され、例えば、インターネット、家庭内LAN、赤外線通信、ラジオ波通信又は衛星通信等であってもよい。

20

#### 【0180】

以上、本開示の実施形態に係る情報処理システム1を構成する情報処理装置900の機能を実現可能なハードウェア構成の一例を示した。上記の各構成要素は、汎用的な部材を用いて構成されていてもよいし、各構成要素の機能に特化したハードウェアにより構成されていてもよい。従って、本実施形態を実施する時々の技術レベルに応じて、適宜、利用するハードウェア構成を変更することが可能である。なお、図24では図示しないが、本実施形態に係る情報処理システム1を構成する情報処理装置900に対応する各種の構成を当然備える。

30

#### 【0181】

なお、上述のような本実施形態に係る情報処理システム1を構成する情報処理装置900の各機能を実現するためのコンピュータプログラムを作製し、パーソナルコンピュータ等を実装することが可能である。また、このようなコンピュータプログラムが格納された、コンピュータで読み取り可能な記録媒体も提供することができる。記録媒体は、例えば、磁気ディスク、光ディスク、光磁気ディスク、フラッシュメモリなどである。また、上記のコンピュータプログラムは、記録媒体を用いずに、例えばネットワークを介して配信してもよい。また、当該コンピュータプログラムを実行させるコンピュータの数は特に限定されない。例えば、当該コンピュータプログラムを、複数のコンピュータ(例えば、複数のサーバ等)が互いに連携して実行してもよい。なお、単数のコンピュータ、または、複数のコンピュータが連携するものを、「コンピュータシステム」とも称する。

40

#### 【0182】

<<5. 適用例>>

続いて、本実施形態に係る情報処理システムの適用例について説明する。本開示の各実施形態に係る情報処理システムは、音声により情報を入力することが可能なシステムやデバイスに対して適用することが可能である。具体的な一例として、スマートフォン、タブ

50

レット端末、P C (Personal Computer) 等のような情報処理装置に対して、本実施形態に係る情報処理システムを適用することも可能である。この場合には、例えば、当該情報処理装置を、前述した入出力装置 30 として構成してもよいし、入出力装置 30 及び情報処理装置 10 が一体的に構成された装置として構成してもよい。

【0183】

また、他の一例として、所謂カーナビゲーションシステム等のような車載システムに対して、本実施形態に係る情報処理システムが適用されてもよい。また、所謂スマート家電に対して、本実施形態に係る情報処理システムが適用されてもよい。

【0184】

また、上述した適用例はあくまで一例であり、音声により情報を入力することが可能なシステムやデバイスであれば、本実施形態に係る情報処理システムの適用先は特に限定されない。

10

【0185】

以上、本実施形態に係る情報処理システムの適用例について説明した。

【0186】

<<6. むすび>>

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。

20

【0187】

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。

【0188】

なお、以下のような構成も本開示の技術的範囲に属する。

(1)

音声の集音に関する期間におけるコンテキスト情報を取得する取得部と、前記音声に変換された文字情報の候補を、前記コンテキスト情報に応じた態様で、所定の出力部に提示させる制御部と、を備える、情報処理装置。

30

(2)

前記コンテキスト情報は、前記音声が集音される環境の状態に関する情報を含む、前記(1)に記載の情報処理装置。

(3)

前記コンテキスト情報は、入力情報の状態または当該入力情報の入力状況に関する情報を含む、前記(1)に記載の情報処理装置。

(4)

前記コンテキスト情報は、前記音声の発話の態様に関する情報を含む、前記(1)に記載の情報処理装置。

40

(5)

前記発話の態様に関する情報は、当該発話の速度に関する情報であり、前記制御部は、前記発話の速度に応じて、前記音声に変換された前記文字情報として、当該音声の発音を示す文字情報を前記出力部に提示させる、前記(4)に記載の情報処理装置。

(6)

前記発話の態様に関する情報は、当該発話の速度に関する情報であり、前記制御部は、前記発話の速度に応じて、前記出力部に提示させる前記候補の数を制限する、

50

前記(4)に記載の情報処理装置。

(7)

前記コンテキスト情報は、前記音声が発話したユーザの状態に関する情報を含む、前記(1)に記載の情報処理装置。

(8)

前記制御部は、前記音声の集音に関する期間における前記コンテキスト情報に応じて、当該音声に変換された前記文字情報の1以上の前記候補のうち少なくとも一部の候補の優先度を制御する、前記(1)~(7)のいずれか一項に記載の情報処理装置。

(9)

前記コンテキスト情報は、雑音の影響に関する情報を含み、

前記制御部は、前記雑音の影響が閾値よりも小さい場合には、前記音声の発音に基づく類似度に応じて前記優先度を制御する、前記(8)に記載の情報処理装置。

10

(10)

前記コンテキスト情報は、雑音の影響に関する情報を含み、

前記制御部は、前記雑音の影響が閾値よりも大きい場合には、前記音声に変換された複数の文字情報間の共起関係に応じて前記優先度を制御する、前記(8)に記載の情報処理装置。

(11)

前記制御部は、

集音された前記音声である音響の解析結果に応じて、当該音声の発音を示す文字情報の1以上の候補を前記出力部に提示させ、

提示された当該発音を示す1以上の候補からの少なくとも一部の当該候補に対する選択の結果に基づき前記優先度を制御する、

前記(8)に記載の情報処理装置。

20

(12)

前記制御部は、前記コンテキスト情報に応じて、前記優先度を制御するために参照するデータを切り替える、前記(8)に記載の情報処理装置。

(13)

前記制御部は、

前記コンテキスト情報に応じて、互いに異なる複数のアルゴリズムに基づき、当該アルゴリズムごとに前記優先度を制御し、

前記アルゴリズムごとに、当該アルゴリズムに基づき制御された前記優先度に応じて、前記文字情報の前記候補を前記出力部に提示させる、

前記(8)に記載の情報処理装置。

30

(14)

前記制御部は、前記アルゴリズムごとに異なる態様で、前記文字情報の前記候補を提示する、前記(13)に記載の情報処理装置。

(15)

前記制御部は、前記音声に変換された文字情報に関する条件の指定を受けて、当該条件に基づき、前記出力部に提示させる当該文字情報の前記候補を制限する、前記(1)~(14)のいずれか一項に記載の情報処理装置。

40

(16)

前記制御部は、前記音声に変換された文字情報の前記候補の数を示す情報を、当該文字情報に関連付けて提示する、前記(1)~(15)のいずれか一項に記載の情報処理装置。

(17)

前記取得部は、逐次集音される前記音声に基づき実行される音声認識処理の結果を逐次取得し、

前記制御部は、逐次取得される前記音声認識処理の結果に基づき、集音された当該音声のうち少なくとも一部の音声に変換された文字情報の1以上の候補を、前記出力部に提示

50

させる、

前記(1)～(16)のいずれか一項に記載の情報処理装置。

(18)

所定の取得部により取得された音声の集音に関する期間におけるコンテキスト情報を外部装置に送信する送信部と、

外部装置から送信される前記音声に変換された文字情報の候補を、前記コンテキスト情報に応じた態様で提示する出力部と、

を備える、情報処理装置。

(19)

コンピュータシステムが、

音声の集音に関する期間におけるコンテキスト情報を取得することと、

前記音声に変換された文字情報の候補を、取得された前記コンテキスト情報に応じた態様で、所定の出力部に提示させることと、

を含む、情報処理方法。

10

(20)

コンピュータシステムが、

所定の取得部により取得された音声の集音に関する期間におけるコンテキスト情報を外部装置に送信することと、

外部装置から送信される前記音声に変換された文字情報の候補を、前記コンテキスト情報に応じた態様で提示することと、

を含む、情報処理方法。

20

【符号の説明】

【0189】

1、2 情報処理システム

10 情報処理装置

101 音声認識処理部

103 コンテキスト解析部

105 優先度制御部

107 出力制御部

20 情報処理装置

201 音声認識処理部

203 信頼度算出部

205 出力制御部

30 入出力装置

310 入力部

311 集音部

313 検知部

315 操作部

330 出力部

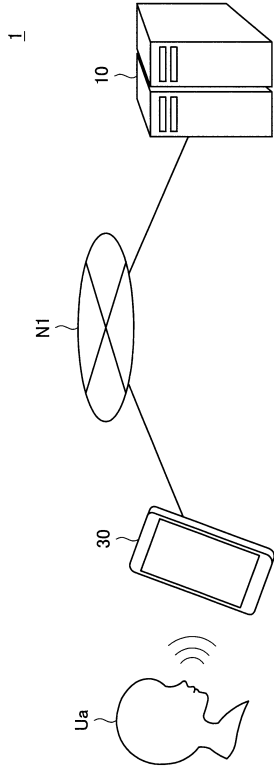
331 表示部

30

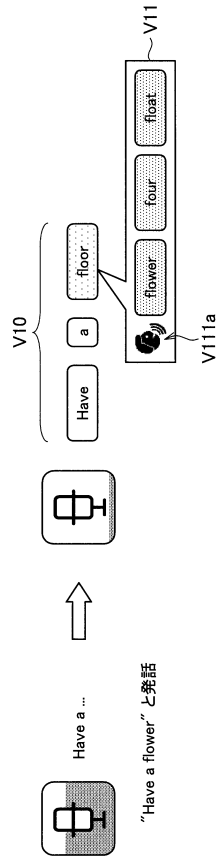
40



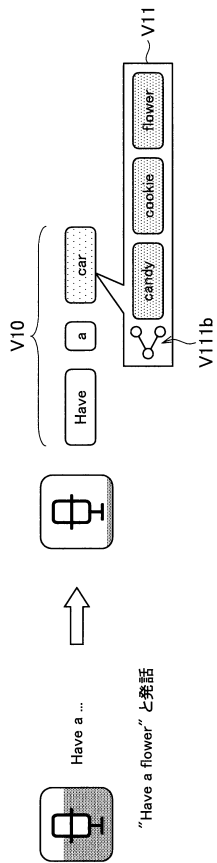
【図1】



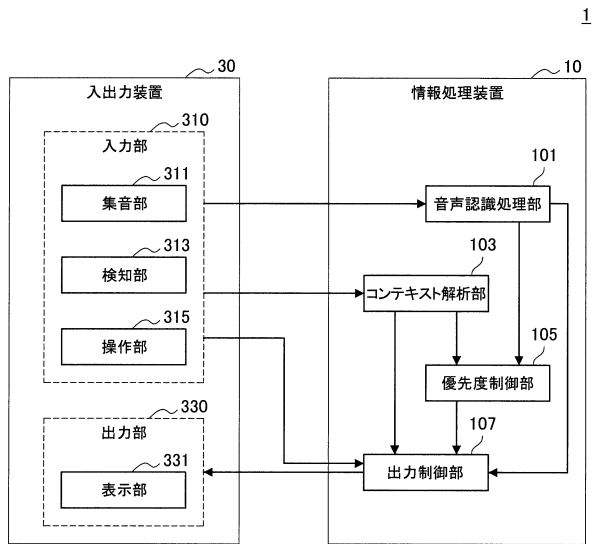
【図2】



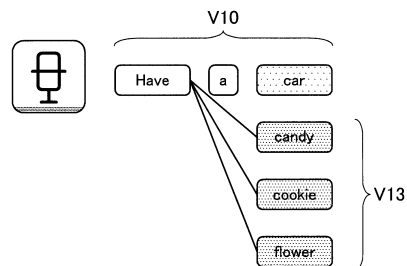
【図3】



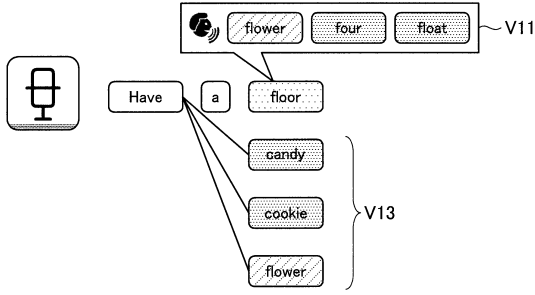
【図4】



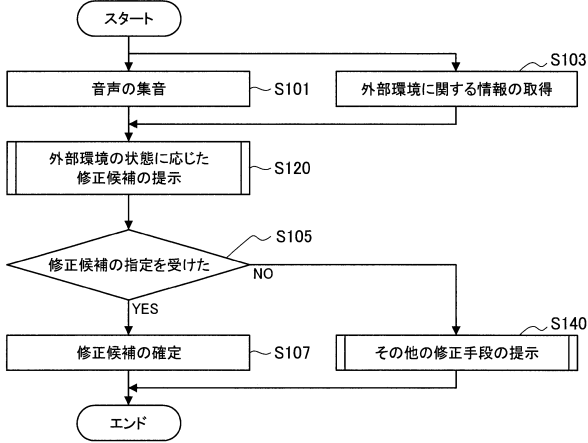
【図5】



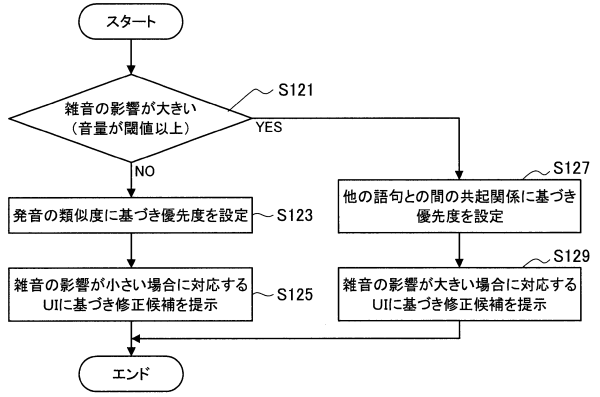
【図6】



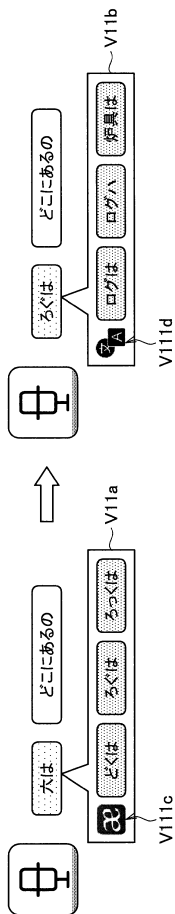
【図7】



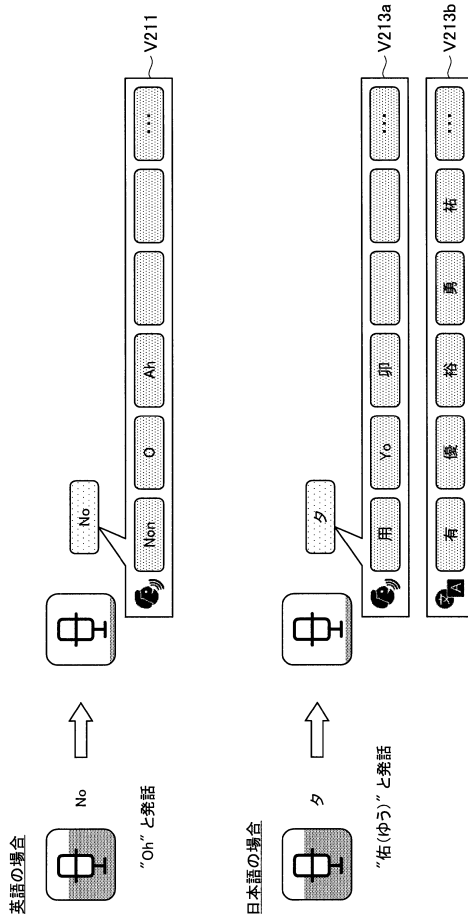
【図8】



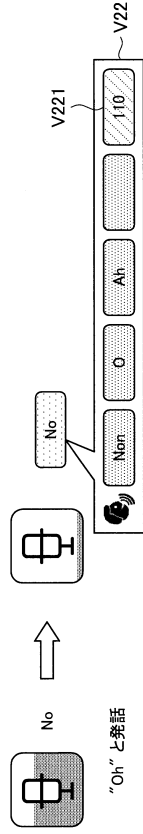
【図9】



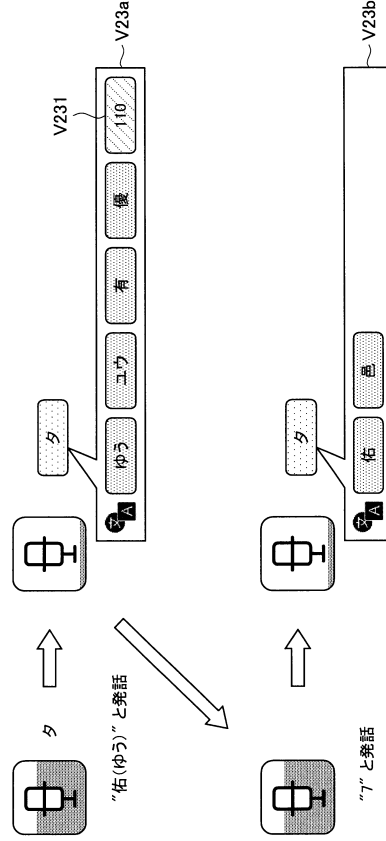
【図10】



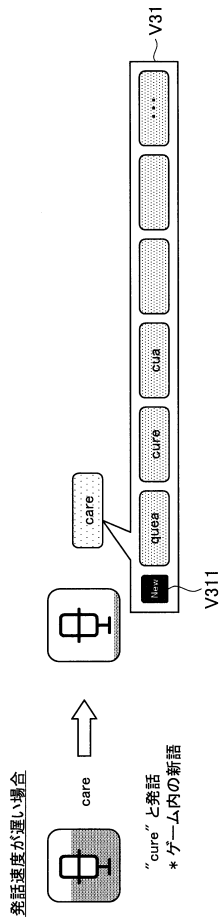
【図 1 1】



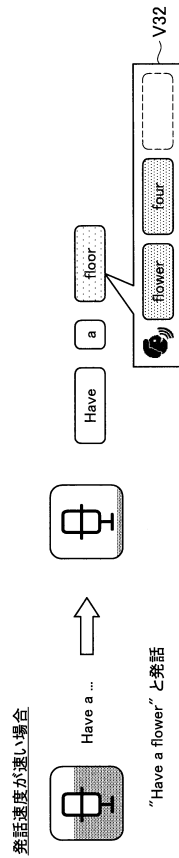
【図 1 2】



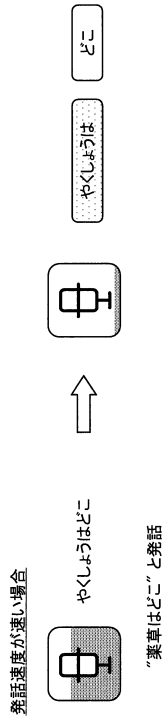
【図 1 3】



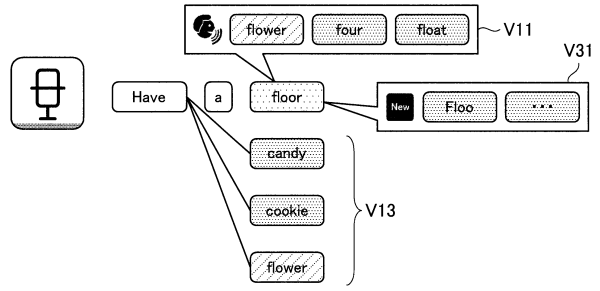
【図 1 4】



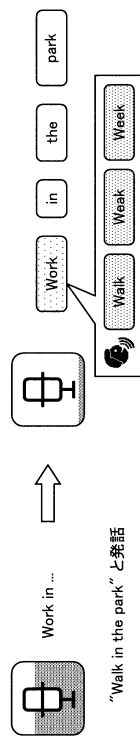
【図 15】



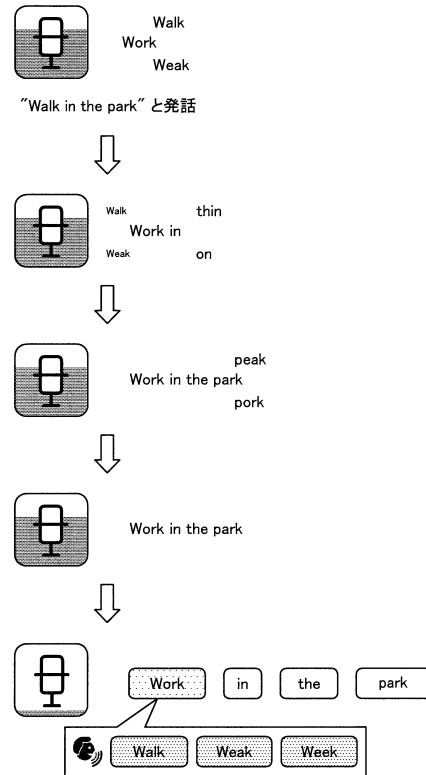
【図 16】



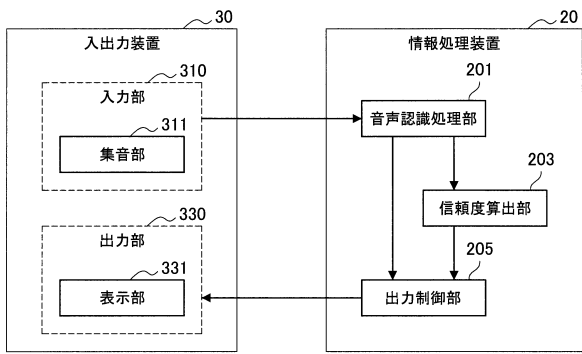
【図 17】



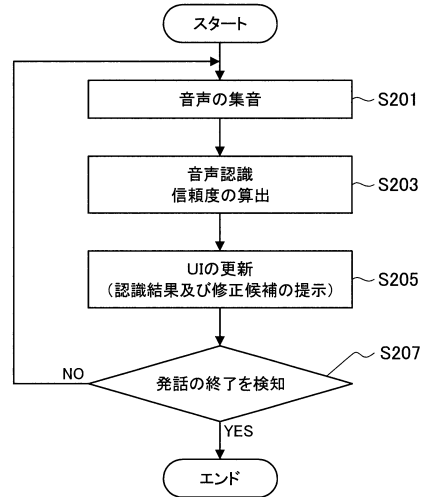
【図 18】



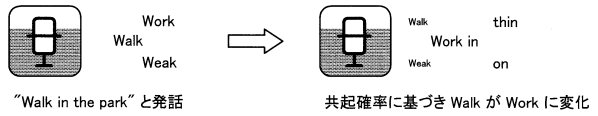
【図19】



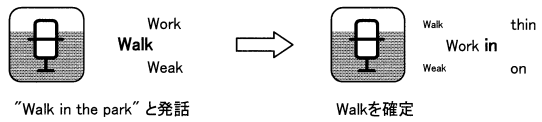
【図20】



【図21】

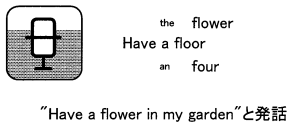


【図22】

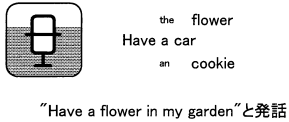


【図23】

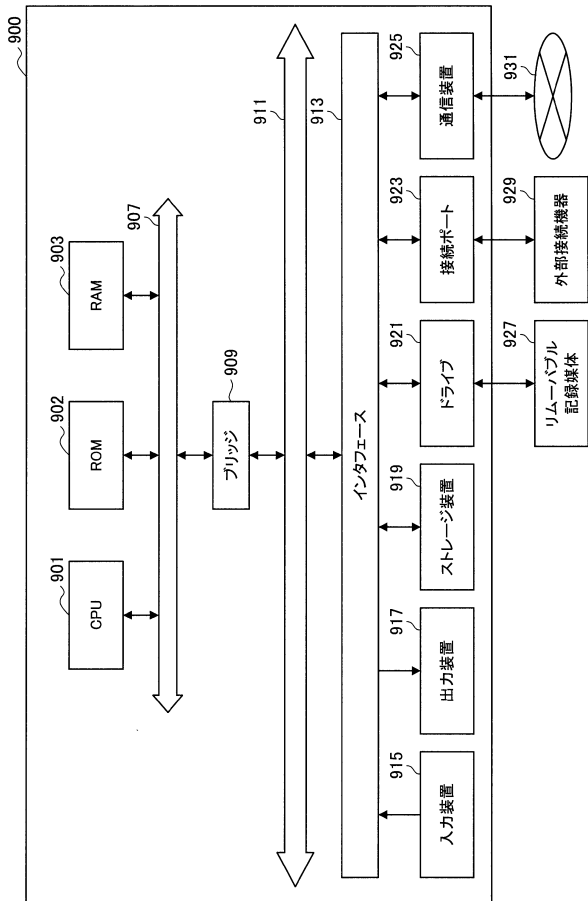
雑音の影響が小さい場合



雑音の影響が大きい場合



【図24】



---

フロントページの続き

(72)発明者 中川 佑輔  
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 上田 雄

(56)参考文献 特開平09 - 274497 (JP, A)  
特開2010 - 204637 (JP, A)  
特開2003 - 066991 (JP, A)  
特開2004 - 117063 (JP, A)  
特開昭64 - 028699 (JP, A)  
特開2000 - 338991 (JP, A)  
欧州特許出願公開第02339576 (EP, A2)

(58)調査した分野(Int.Cl., DB名)  
G10L 15/00 - 15/34