

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2018年2月15日(15.02.2018)



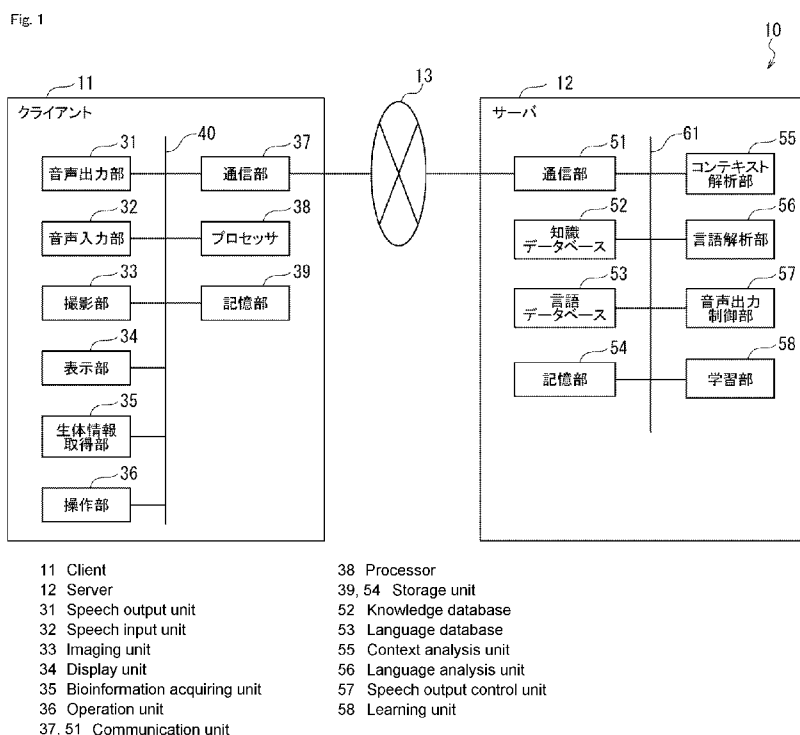
(10) 国際公開番号

WO 2018/030149 A1

- (51) 国際特許分類:
G10L 13/10 (2013.01) *G10L 13/02* (2013.01)
- (21) 国際出願番号: PCT/JP2017/026961
- (22) 国際出願日: 2017年7月26日(26.07.2017)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2016-156120 2016年8月9日(09.08.2016) JP
- (71) 出願人: ソニー株式会社 (SONY CORPORATION) [JP/JP]; 〒1080075 東京都港区港南1丁目7番1号 Tokyo (JP).
- (72) 発明者: 河野 真一 (KAWANO Shinichi); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP). 岩瀬 広 (IWASE Hiro); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP). 斎藤 真里 (SAITO Mari); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP).
- (74) 代理人: 西川 孝, 外 (NISHIKAWA Takashi et al.); 〒1600023 東京都新宿区西新宿7丁目5番25号 西新宿木村屋ビルディング9階 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,

(54) Title: INFORMATION PROCESSING DEVICE AND INFORMATION PROCESSING METHOD

(54) 発明の名称: 情報処理装置及び情報処理方法



(57) Abstract: The present invention relates to an information processing device and an information processing method whereby the probability of a user becoming aware of a synthesized speech can be increased. The information processing device according to the present invention is provided with a speech output control unit for controlling the output form of synthesized speech obtained by converting text to speech, on the basis of a context of when the synthesized speech is outputted. Alternatively, the information processing device is provided with: a communication unit for transmitting,



CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH,
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,
NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 国際調査報告(条約第21条(3))

to another information processing device, context data relating to the context of when synthesized speech obtained by converting text to speech is outputted, and receiving speech control data used to generate the synthesized speech, the output form of which is controlled on the basis of the context data, from the other information processing device; and a speech synthesizing unit for generating the synthesized speech on the basis of the speech control data. The present invention can be applied in a server or a client for performing output control or outputting of synthesized speech, for example.

(57) 要約: 本技術は、ユーザの意識が合成音声に向かう確率を高くすることができるようにする情報処理装置及び情報処理方法に関する。情報処理装置は、テキストを音声に変換した合成音声を出力するときのコンテキストに基づいて、前記合成音声の出力態様を制御する音声出力制御部を備える。或いは、情報処理装置は、テキストを音声に変換した合成音声を出力するときのコンテキストに関するコンテキストデータを他の情報処理装置に送信し、前記コンテキストデータに基づいて出力態様が制御された前記合成音声の生成に用いる音声制御データを前記他の情報処理装置から受信する通信部と、前記音声制御データに基づいて前記合成音声を生成する音声合成部とを備える。本技術は、例えば、合成音声の出力制御又は出力を行うサーバ又はクライアントに適用できる。

明 細 書

発明の名称： 情報処理装置及び情報処理方法

技術分野

[0001] 本技術は、情報処理装置及び情報処理方法に関し、特に、テキストを音声に変換して出力する場合に用いて好適な情報処理装置及び情報処理方法に関する。

背景技術

[0002] TTS (Text to Speech) によりテキストを音声に変換して出力する場合、声色や抑揚等が単調なため、ユーザの注意力が散漫になりテキストの内容が伝わらなくなることがある。

[0003] これに対して、従来、発声開始からの経過時間に応じて、発声速度を段階的に上げるとともに、発声速度以外のピッチ、音量、声質などの各種のパラメータをランダムに変更することにより、ユーザが合成音声に飽きるのを防ぐことが提案されている（例えば、特許文献1 参照）。

先行技術文献

特許文献

[0004] 特許文献1：特開平11-161298号公報

発明の概要

発明が解決しようとする課題

[0005] しかしながら、特許文献1に記載の技術では、経過時間のみに応じて合成音声のパラメータが変更されるため、必ずしもパラメータの変更がいつも有効であるとは言えず、ユーザの意識が合成音声に向かうとは限らない。

[0006] そこで、本技術は、ユーザの意識が合成音声に向かう確率を高くすることができるようにするものである。

課題を解決するための手段

[0007] 本技術の第1の側面の情報処理装置は、テキストを音声に変換した合成音声を出力するときのコンテキストに基づいて、前記合成音声の出力態様を制

御する音声出力制御部を備える。

- [0008] 前記音声出力制御部には、前記コンテキストが所定の条件を満たした場合、前記合成音声の出力態様を変更させることができる。
- [0009] 前記合成音声の出力態様の変更には、前記合成音声の特性、前記合成音声に対する効果、前記合成音声の背景のBGM (Back Ground Music)、前記合成音声により出力される前記テキスト、及び、前記合成音声を出力する装置の動作のうち少なくとも1つの変更を含ませることができる。
- [0010] 前記合成音声の特性には、速度、ピッチ、音量、及び、抑揚のうち少なくとも1つを含ませ、前記合成音声に対する効果には、前記テキスト内の特定の言葉の繰り返し、及び、前記合成音声への間の挿入のうち少なくとも1つを含ませることができる。
- [0011] 前記音声出力制御部には、ユーザの意識が前記合成音声に向いていない状態が検出された場合、前記合成音声の出力態様を変更させることができる。
- [0012] 前記音声出力制御部には、前記合成音声の出力態様を変更した後、ユーザの意識が前記合成音声に向いている状態が検出された場合、前記合成音声の出力態様を元に戻させることができる。
- [0013] 前記音声出力制御部には、前記合成音声の特性の変化量が所定の範囲内である状態が所定の時間以上継続した場合、前記合成音声の出力態様を変更させることができる。
- [0014] 前記音声出力制御部には、前記コンテキストに基づいて、前記合成音声の出力態様の変更方法を選択させることができる。
- [0015] 前記合成音声の出力態様の変更方法に対するユーザの反応を学習する学習部をさらに設け、前記音声出力制御部には、前記ユーザの反応の学習結果に基づいて、前記合成音声の出力態様の変更方法を選択させることができる。
- [0016] 前記音声出力制御部には、さらに前記テキストの特性に基づいて、前記合成音声の出力態様を制御させることができる。
- [0017] 前記音声出力制御部には、前記テキストの特徴量が第1の閾値以上である場合、又は、前記テキストの特徴量が第2の閾値未満である場合、前記合成

音声の出力態様を変更させることができる。

- [0018] 前記音声出力制御部には、前記合成音声の生成に用いる音声制御データを他の情報処理装置に供給することにより、前記他の情報処理装置における前記合成音声の出力態様を制御させることができる。
- [0019] 前記音声出力制御部には、前記他の情報処理装置から取得した前記コンテキストに関するコンテキストデータに基づいて、前記音声制御データを生成させることができる。
- [0020] 前記コンテキストデータには、ユーザの周囲を撮影した画像に基づくデータ、前記ユーザの周囲の音声に基づくデータ、及び、前記ユーザの生体情報に基づくデータのうち少なくとも1つを含ませることができる。
- [0021] 前記コンテキストデータに基づいて、前記コンテキストを解析するコンテキスト解析部をさらに設けることができる。
- [0022] 前記コンテキストには、ユーザの状態、前記ユーザの特性、前記合成音声出力される環境、及び、前記合成音声の特性のうち少なくとも1つを含ませることができる。
- [0023] 前記合成音声出力される環境には、前記ユーザの周囲の環境、前記合成音声出力する装置、及び、前記合成音声出力するアプリケーションプログラムのうち少なくとも1つを含ませることができる。
- [0024] 本技術の第1の側面の情報処理方法は、テキストを音声に変換した合成音声出力するときのコンテキストに基づいて、前記合成音声の出力態様を制御する音声出力制御ステップを含む。
- [0025] 本技術の第2の側面の情報処理装置は、テキストを音声に変換した合成音声出力するときのコンテキストに関するコンテキストデータを他の情報処理装置に送信し、前記コンテキストデータに基づいて出力態様が制御された前記合成音声の生成に用いる音声制御データを前記他の情報処理装置から受信する通信部と、前記音声制御データに基づいて前記合成音声生成する音声合成部とを備える。
- [0026] 本技術の第2の側面の情報処理方法は、テキストを音声に変換した合成音

声を出力するときのコンテキストに関するコンテキストデータを他の情報処理装置に送信し、前記コンテキストデータに基づいて出力態様が制御された前記合成音声の生成に用いる音声制御データを前記他の情報処理装置から受信する通信ステップと、前記音声制御データに基づいて前記合成音声を生成する音声合成ステップとを含む。

[0027] 本技術の第1の側面においては、テキストを音声に変換した合成音声を出力するときのコンテキストに基づいて、前記合成音声の出力態様が制御される。

[0028] 本技術の第2の側面においては、テキストを音声に変換した合成音声を出力するときのコンテキストに関するコンテキストデータが他の情報処理装置に送信され、前記コンテキストデータに基づいて出力態様が制御された前記合成音声の生成に用いる音声制御データが前記他の情報処理装置から受信され、前記音声制御データに基づいて前記合成音声が生成される。

発明の効果

[0029] 本技術の第1又は第2の側面によれば、ユーザの意識を合成音声に向けさせることができる。特に、本技術の第1又は第2の側面によれば、ユーザの意識が合成音声に向かう確率が高くなる。

[0030] なお、ここに記載された効果は必ずしも限定されるものではなく、本開示中に記載されたいずれかの効果であってもよい。

図面の簡単な説明

[0031] [図1]本技術を適用した情報処理システムの一実施の形態を示すブロック図である。

[図2]クライアントのプロセッサにより実現される機能の構成例の一部を示すブロック図である。

[図3]合成音声により出力するテキストの例を示す図である。

[図4]合成音声出力処理を説明するためのフローチャートである。

[図5]合成音声出力制御処理を説明するためのフローチャートである。

[図6]TTSデータ生成処理の詳細を説明するためのフローチャートである。

[図7]通常モード時とアテンションモード時のTTSデータの具体例を示す図である。

[図8]アテンションモードの具体例を説明するための図である。

[図9]アテンションモードの具体例を説明するための図である。

[図10]アテンションモードの具体例を説明するための図である。

[図11]アテンションモードの具体例を説明するための図である。

[図12]アテンションモードの具体例を説明するための図である。

[図13]アテンションモードの具体例を説明するための図である。

[図14]アテンションモードの具体例を説明するための図である。

[図15]コンピュータの構成例を示すブロック図である。

発明を実施するための形態

[0032] 以下、発明を実施するための形態（以下、「実施形態」と記述する）について図面を用いて詳細に説明する。なお、説明は以下の順序で行う。

1. 実施の形態

2. 変形例

[0033] <<1. 実施の形態>>

<1-1. 情報処理システムの構成例>

まず、図1を参照して、本技術を適用した情報処理システム10の構成例について説明する。

[0034] 情報処理システム10は、TTSによりテキストを音声に変換して出力するシステムである。情報処理システム10は、クライアント11、サーバ12、及び、ネットワーク13により構成される。クライアント11とサーバ12は、ネットワーク13を介して相互に接続されている。

[0035] なお、図中、クライアント11が1つのみ示されているが、実際には複数のクライアント11がネットワーク13に接続され、複数のユーザがクライアント11を介して情報処理システム10を利用する。

[0036] クライアント11は、サーバ12から提供されるTTSデータに基づいて、テキストを音声に変換して出力する。ここで、TTSデータとは、合成音

声の生成に用いられる音声制御データである。

[0037] 例えば、クライアント 11 は、スマートフォン、タブレット、携帯電話機、ノート型のパーソナルコンピュータ等の携帯情報端末、ウェアラブルデバイス、デスクトップ型のパーソナルコンピュータ、ゲーム機、動画再生装置、音楽再生装置等により構成される。また、ウェアラブルデバイスには、例えば、眼鏡型、腕時計型、ブレスレット型、ネックレス型、ネックバンド型、イヤフォン型、ヘッドセット型、ヘッドマウント型等の各種の方式を採用することができる。

[0038] クライアント 11 は、音声出力部 31、音声入力部 32、撮影部 33、表示部 34、生体情報取得部 35、操作部 36、通信部 37、プロセッサ 38、及び、記憶部 39 を備える。音声出力部 31、音声入力部 32、撮影部 33、表示部 34、生体情報取得部 35、操作部 36、通信部 37、プロセッサ 38、及び、記憶部 39 は、バス 40 を介して相互に接続されている。

[0039] 音声出力部 31 は、例えばスピーカにより構成される。スピーカ数は、任意に設定することができる。音声出力部 31 は、プロセッサ 38 から供給される音声データに基づく音声を出力する。

[0040] 音声入力部 32 は、例えばマイクロフォンにより構成される。マイクロフォンの数は、任意に設定することができる。音声入力部 32 は、ユーザの周囲の音声（ユーザが発する音声を含む）を収集する。音声入力部 32 は、収集した音声を示す音声データをプロセッサ 38 に供給したり、記憶部 39 に記憶させたりする。

[0041] 撮影部 33 は、例えばカメラにより構成される。カメラの数は、任意に設定することができる。撮影部 33 は、ユーザの周囲（ユーザを含む）を撮影し、撮影の結果得られる画像を示す画像データをプロセッサ 38 に供給したり、記憶部 39 に記憶させたりする。

[0042] 表示部 34 は、例えばディスプレイにより構成される。ディスプレイの数は、任意に設定することができる。表示部 34 は、プロセッサ 38 から供給される画像データに基づく画像を表示する。

- [0043] 生体情報取得部３５は、各種の人の生体情報を取得するデバイスやセンサ等により構成される。生体情報取得部３５が取得する生体情報には、例えば、ユーザの体調、集中度及び緊張度等を検出するのに用いられるデータ（例えば、心拍、脈拍、発汗量、脳波、体温等）が含まれる。生体情報取得部３５は、取得した生体情報をプロセッサ３８に供給したり、記憶部３９に記憶させたりする。
- [0044] 操作部３６は、各種の操作部材により構成され、クライアント１１の操作に用いられる。
- [0045] 通信部３７は、各種の通信デバイスにより構成される。通信部３７の通信方式は特に限定されるものではなく、無線通信又は有線通信のいずれでもよい。また、通信部３７が、複数の通信方式に対応していてもよい。通信部３７は、ネットワーク１３を介してサーバ１２と通信を行う。通信部３７は、サーバ１２から受信したデータをプロセッサ３８に供給したり、記憶部３９に記憶させたりする。
- [0046] プロセッサ３８は、クライアント１１の各部の制御を行ったり、通信部３７及びネットワーク１３を介して、サーバ１２とデータの送受信を行ったりする。また、プロセッサ３８は、サーバ１２から取得したＴＴＳデータに基づいて音声合成を行い（ＴＴＳデータの再生処理を行い）、得られた合成音声を示す音声データを生成し、音声出力部３１に供給する。
- [0047] 記憶部３９は、クライアント１１の処理に必要なプログラム、データ等を記憶する。
- [0048] サーバ１２は、クライアント１１からの要求に従ってＴＴＳデータを生成し、生成したＴＴＳデータを、ネットワーク１３を介してクライアント１１に送信する。サーバ１２は、通信部５１、知識データベース５２、言語データベース５３、記憶部５４、コンテキスト解析部５５、言語解析部５６、音声出力制御部５７、及び、学習部５８を備える。
- [0049] 通信部５１は、各種の通信デバイスにより構成される。通信部５１の通信方式は特に限定されるものではなく、無線通信又は有線通信のいずれでもよ

い。また、通信部 5 1 が、複数の通信方式に対応していてもよい。通信部 5 1 は、ネットワーク 1 3 を介してクライアント 1 1 と通信を行う。通信部 5 1 は、クライアント 1 1 から受信したデータをサーバ 1 2 内の各部に供給したり、記憶部 5 4 に記憶させたりする。

[0050] 知識データベース 5 2 は、各種の知識に関するデータを格納する。例えば、知識データベース 5 2 は、T T S データに基づく合成音声の背景で流す B G M に使用可能な楽曲に関するデータを格納する。

[0051] 言語データベース 5 3 は、各種の言語に関するデータを格納する。例えば、言語データベース 5 3 は、言葉の言い回しに関するデータ等を格納する。

[0052] 記憶部 5 4 は、サーバ 1 2 の処理に必要なプログラム、データ等を記憶する。例えば、記憶部 5 4 は、クライアント 1 1 に提供する T T S データにより再生される対象となるテキストを記憶する。

[0053] コンテキスト解析部 5 5 は、クライアント 1 1 から取得したデータ等に基づいて、視線認識処理、画像認識処理、音声認識処理等を行い、クライアント 1 1 において合成音声を出力するときのコンテキストの解析を行う。コンテキスト解析部 5 5 は、解析結果を音声出力制御部 5 7 及び学習部 5 8 に供給したり、記憶部 5 4 に記憶させたりする。

[0054] なお、合成音声を出力するときのコンテキストには、例えば、合成音声を出力する対象となるユーザの状態、ユーザの特性、合成音声が出力される環境、合成音声の特性等が含まれる。

[0055] 言語解析部 5 6 は、合成音声により出力するテキストの言語解析を行い、テキストの特性等を検出する。言語解析部 5 6 は、解析結果を音声出力制御部 5 7 及び学習部 5 8 に供給したり、記憶部 5 4 に記憶させたりする。

[0056] 音声出力制御部 5 7 は、知識データベース 5 2 及び言語データベース 5 3 に記憶されているデータ、コンテキスト解析部 5 5 による解析結果、言語解析部 5 6 による解析結果、並びに、学習部 5 8 による学習結果等に基づいて、クライアント 1 1 で使用する T T S データを生成する。音声出力制御部 5 7 は、生成した T T S データを、通信部 5 1 を介してクライアント 1 1 に送

信する。

[0057] また、音声出力制御部 57 は、コンテキスト及びテキストの特性に基づいて、合成音声の出力モードを通常モード又はアテンションモードに設定する。通常モードとは、標準的な出力態様で合成音声を出力するモードである。アテンションモードとは、通常モードとは異なる出力態様で合成音声を出力し、ユーザの意識を合成音声に向けるように促すモードである。音声出力制御部 57 は、各出力モードに対応した TTS データを生成し、クライアント 11 に提供することにより、各出力モードで合成音声出力されるようにクライアント 11 を制御する。

[0058] 学習部 58 は、コンテキスト解析部 55 による解析結果、及び、言語解析部 56 による解析結果等に基づいて、各ユーザの特性等を学習する。学習部 58 は、学習結果を記憶部 54 に記憶させる。

[0059] なお、以下、クライアント 11（通信部 37）とサーバ 12（通信部 51）がネットワーク 13 を介して通信を行う場合の”ネットワーク 13 を介して”の記載は省略する。

[0060] <1-2. クライアントのプロセッサにより実現される機能の一部の例>

図 2 は、クライアント 11 のプロセッサ 38 により実現される機能の一部の例を示している。例えば、プロセッサ 38 が所定の制御プログラムを実行することにより、音声合成部 101 及びコンテキストデータ取得部 102 を含む機能が実現される。

[0061] 音声合成部 101 は、サーバ 12 から取得した TTS データに基づいて音声合成を行い、得られた合成音声を示す音声データを音声出力部 31 に供給する。

[0062] コンテキストデータ取得部 102 は、合成音声出力するときのコンテキストに関するデータを取得し、取得したデータに基づいて、コンテキストデータを生成する。コンテキストデータには、例えば、音声入力部 32 により収集される音声に基づくデータ、撮影部 33 により撮影される画像に基づくデータ、及び、生体情報取得部 35 により取得される生体情報に基づくデー

タが含まれる。

[0063] ここで、音声に基づくデータには、例えば、音声データ自身、音声データから抽出される特徴量を表すデータ、及び、音声データの解析結果を示すデータのうちいくつかが含まれる。画像に基づくデータには、例えば、画像データ自身、画像データから抽出される特徴量を表すデータ、及び、画像データの解析結果を示すデータのうちいくつかが含まれる。生体情報に基づくデータには、例えば、生体情報自身、生体情報から抽出される特徴量を表すデータ、及び、生体情報の解析結果を示すデータのうちいくつかが含まれる。

[0064] コンテキストデータ取得部 102 は、生成したコンテキストデータを、通信部 37 を介してサーバ 12 に送信する。

[0065] <1-3. 情報処理システム 10 の処理>

次に、図 3 乃至図 14 を参照して、情報処理システム 10 の処理について説明する。なお、以下、適宜、図 3 に示されるテキストを音声に変換して出力する場合を例に挙げながら説明する。このテキストは、2008 年 11 月 4 日にアメリカのイリノイ州シカゴで行われたオバマ大統領のスピーチの一部である。

[0066] (合成音声出力処理)

まず、図 4 のフローチャートを参照して、クライアント 11 により実行される合成音声出力処理について説明する。この処理は、例えば、ユーザがクライアント 11 の操作部 36 を介して、合成音声の出力を開始する操作を行ったとき開始される。

[0067] ステップ S1 において、音声合成部 101 は、TTS データの送信を要求する。具体的には、音声合成部 101 は、TTS データの送信を要求するためのコマンドである TTS データ送信要求を生成し、通信部 37 を介してサーバ 12 に送信する。

[0068] TTS データ送信要求には、例えば、TTS データの再生に用いるクライアント 11 の種類及びアプリケーションプログラム（以下、APP と称する）の種類、ユーザ属性情報等が含まれる。ユーザ属性情報には、例えば、ユ

ーザの年齢、性別、住所、職業、国籍等が含まれる。或いは、ユーザ属性情報には、例えば、ユーザを一意に識別し、サーバ１２が保持するユーザの属性情報と関連付けることが可能なユーザＩＤが含まれる。

[0069] サーバ１２は、後述する図５のステップＳ５１において、ＴＴＳデータ送信要求を受信し、ステップＳ５９において、ＴＴＳデータを送信する。

[0070] ステップＳ２において、クライアント１１は、コンテキストに関するデータの取得を開始する。具体的には、音声入力部３２は、ユーザの周囲の音声を収集し、得られた音声データをコンテキストデータ取得部１０２に供給する処理を開始する。撮影部３３は、ユーザの周囲を撮影し、得られた画像データをコンテキストデータ取得部１０２に供給する処理を開始する。生体情報取得部３５は、ユーザの生体情報を取得し、コンテキストデータ取得部１０２に供給する処理を開始する。

[0071] コンテキストデータ取得部１０２は、取得した音声データ、画像データ、及び、生体情報に基づいてコンテキストデータを生成し、通信部３７を介してサーバ１２に送信する処理を開始する。

[0072] サーバ１２は、後述する図５のステップＳ５２において、コンテキストデータを受信する。

[0073] ステップＳ３において、音声合成部１０１は、ＴＴＳデータを受信したか否かを判定する。音声合成部１０１は、サーバ１２から送信されたＴＴＳデータを、通信部３７を介して受信した場合、ＴＴＳデータを受信したと判定し、処理はステップＳ４に進む。

[0074] ステップＳ４において、クライアント１１は、ＴＴＳデータに基づいて合成音声を出力する。具体的には、音声合成部１０１は、ＴＴＳデータに基づいて音声合成を行い、得られた合成音声を示す音声データを音声出力部３１に供給する。音声出力部３１は、音声データに基づいて合成音声を出力する。

[0075] ステップＳ５において、音声合成部１０１は、合成音声の出力の停止が指令されたか否かを判定する。合成音声の出力の停止が指令されていないと判

定された場合、処理はステップS 3に戻る。

[0076] その後、ステップS 3において、T T Sデータを受信しなかったと判定されるか、ステップS 5において、合成音声の出力の停止が指令されたと判定されるまで、ステップS 3乃至S 5の処理が繰り返し実行される。

[0077] 一方、ステップS 5において、音声合成部1 0 1は、例えば、ユーザがクライアント1 1の操作部3 6を介して、合成音声の出力を停止する操作を行ったとき、合成音声の出力の停止が指令されたと判定し、処理はステップS 6に進む。

[0078] ステップS 6において、音声合成部1 0 1は、T T Sデータの送信の停止を要求する。具体的には、音声合成部1 0 1は、T T Sデータの送信の停止を要求するコマンドであるT T Sデータ送信停止要求を生成し、通信部3 7を介してサーバ1 2に送信する。

[0079] サーバ1 2は、後述する図5のステップS 6 2において、クライアント1 1からT T Sデータ送信停止要求を受信する。

[0080] その後、合成音声出力処理は終了する。

[0081] 一方、ステップS 3において、T T Sデータを受信しなかったと判定された場合、合成音声出力処理は終了する。

[0082] (合成音声出力制御処理)

次に、図5のフローチャートを参照して、図4の合成音声出力処理に対応して、サーバ1 2により実行される合成音声出力制御処理について説明する。

[0083] ステップS 5 1において、通信部5 1は、T T Sデータの送信の要求を受ける。すなわち、通信部5 1は、クライアント1 1から送信されたT T Sデータ送信要求を受信する。通信部5 1は、T T Sデータ送信要求をコンテキスト解析部5 5及び音声出力制御部5 7に供給する。

[0084] ステップS 5 2において、サーバ1 2は、コンテキストの解析を開始する。具体的には、通信部5 1は、クライアント1 1からコンテキストデータを受信し、コンテキスト解析部5 5に供給する処理を開始する。

- [0085] コンテキスト解析部 55 は、TTS データ送信要求及びコンテキストデータに基づいて、コンテキストの解析を開始する。
- [0086] 例えば、コンテキスト解析部 55 は、TTS データ送信要求に含まれるユーザ属性情報、並びに、コンテキストデータに含まれる音声データ及び画像データに基づいて、ユーザの特性の解析を開始する。ユーザの特性には、例えば、ユーザの属性、嗜好、能力、集中力等が含まれる。ユーザの属性には、例えば、性別、年齢、住所、職業、国籍等が含まれる。
- [0087] また、コンテキスト解析部 55 は、コンテキストデータに含まれる音声データ、画像データ、及び、生体情報に基づいて、ユーザの状態の解析を開始する。ユーザの状態には、例えば、視線の方向、行動、表情、緊張度、集中度、発話内容、体調等が含まれる。
- [0088] さらに、コンテキスト解析部 55 は、TTS データ送信要求に含まれるクライアント 11 の種類及びAPP の種類に関する情報、並びに、コンテキストデータに含まれる音声データ及び画像データに基づいて、合成音声が出力される環境の解析を開始する。合成音声が出力される環境には、例えば、ユーザの周囲の環境、合成音声を出力するクライアント 11、合成音声を出力するAPP 等が含まれる。ユーザの周囲の環境には、例えば、ユーザの現在位置、ユーザの周囲の人物や物体の状態、ユーザの周囲の明るさ、ユーザの周囲の音声等が含まれる。
- [0089] また、コンテキスト解析部 55 は、解析結果を音声出力制御部 57 及び学習部 58 に供給したり、記憶部 54 に記憶させたりする処理を開始する。
- [0090] ステップ S 53 において、音声出力制御部 57 は、通常モードに設定する。
- [0091] ステップ S 54 において、サーバ 12 は、アテンションモードへの移行条件を設定する。具体的には、コンテキスト解析部 55 は、例えば、ユーザの集中力、及び、必要とされる集中力を推定する。
- [0092] 例えば、コンテキスト解析部 55 は、ユーザの集中力に対する学習結果が記憶部 54 に記憶されている場合、その学習結果に基づいて、ユーザの集中

力を推定する。

[0093] 一方、コンテキスト解析部 55 は、ユーザの集中力に対する学習結果が記憶部 54 に記憶されていない場合、ユーザの属性等に基づいて、ユーザの集中力を推定する。例えば、コンテキスト解析部 55 は、ユーザの年齢に基づいて、ユーザの集中力を推定する。例えば、ユーザが子供である場合、集中力は低く推定される。また、例えば、コンテキスト解析部 55 は、ユーザの職業に基づいて、ユーザの集中力を推定する。例えば、ユーザが集中力を要する仕事に従事している場合、集中力は高く推定される。

[0094] さらに、コンテキスト解析部 55 は、コンテキストの解析結果等に基づいて、集中力の推定結果を補正する。例えば、コンテキスト解析部 55 は、ユーザの体調が良好である場合、集中力を高く補正し、ユーザの体調が悪い場合、集中力を低く補正する。また、例えば、コンテキスト解析部 55 は、ユーザの周囲が集中しやすい環境（例えば、静かな場所、周囲に人がいない場所等）である場合、集中力を高く補正する。一方、コンテキスト解析部 55 は、ユーザの周囲が集中しづらい環境（例えば、騒々しい場所、周囲に人がいる場所等）である場合、集中力を低く補正する。さらに、例えば、コンテキスト解析部 55 は、ユーザが使用している A P P がユーザの嗜好度が高いコンテンツを扱う場合（例えば、ユーザの趣味のコンテンツを扱う A P P の場合）、集中力を高く補正する。一方、コンテキスト解析部 55 は、ユーザが使用している A P P がユーザの嗜好度が低いコンテンツを扱う場合（例えば、各種の資格や教科の学習用の A P P である場合）、集中力を低く補正する。

[0095] また、例えば、コンテキスト解析部 55 は、ユーザが使用している A P P に基づいて、必要とされる集中力を推定する。例えば、ユーザが天気予報の A P P を使用している場合、天気予報は内容を把握するのにそれほど集中力を要さず、一部を聞き逃してもあまり影響がない。従って、この場合、必要とされる集中力は低く推定される。一方、ユーザが各種の資格や教科の学習用の A P P を使用している場合、学習には集中力が必要とされ、一部を聞き

逃すと内容を理解できなくなる可能性が高い。従って、この場合、必要とされる集中力は高く推定される。

[0096] コンテキスト解析部 55 は、ユーザの集中力、及び、必要とされる集中力の推定結果を音声出力制御部 57 に供給する。

[0097] 音声出力制御部 57 は、ユーザの集中力、及び、必要とされる集中力の推定結果に基づいて、アテンションモードへの移行条件を設定する。例えば、音声出力制御部 57 は、ユーザの集中力が高いほど、又は、必要とされる集中力が低いほど、アテンションモードへの移行条件を厳しくして、アテンションモードに移行しにくくする。一方、音声出力制御部 57 は、ユーザの集中力が低いほど、又は、必要とされる集中力が高いほど、アテンションモードへの移行条件を緩くして、アテンションモードに移行しやすくする。

[0098] なお、ステップ S 54 において、音声出力制御部 57 は、ユーザやコンテキスト等によらずに、一律に標準的な移行条件に設定するようにしてもよい。

[0099] ステップ S 55 において、音声出力制御部 57 は、出力するテキストがあるか否かを判定する。例えば、音声出力制御部 57 は、1 回目のステップ S 55 の処理において、合成音声により出力されるテキスト（以下、出力対象テキストと称する）を、記憶部 54 に記憶されているテキストの中から検索する。音声出力制御部 57 は、出力対象テキストが見つかった場合、出力するテキストがあると判定し、処理はステップ S 56 に進む。

[0100] また、例えば、音声出力制御部 57 は、2 回目以降のステップ S 55 の処理において、出力対象テキストのうち未出力の部分が残っている場合、出力するテキストがあると判定し、処理はステップ S 56 に進む。

[0101] ステップ S 56 において、音声出力制御部 57 は、新たに出力する部分を設定する。具体的には、音声出力制御部 57 は、出力対象テキストの未出力の部分の先頭から所定の位置までの部分を新たに出力する部分（以下、新規出力部分と称する）に設定する。なお、新規出力部分は、例えば、文章単位、節単位、又は、単語単位で設定される。

- [0102] ステップS 5 7において、言語解析部5 6は、テキストの解析を行う。具体的には、音声出力制御部5 7は、出力対象テキストの新規出力部分を言語解析部5 6に供給する。言語解析部5 6は、新規出力部分の言語解析を行う。例えば、言語解析部5 6は、形態素解析、自立語解析、複合語解析、句解析、係り受け解析、意味解析等を行う。このとき、言語解析部5 6は、必要に応じて出力対象テキストの出力済みの部分や、新規出力部分の後の部分を参照してもよい。これにより、言語解析部5 6は、出力対象テキストの内容や特徴量等を把握する。
- [0103] また、言語解析部5 6は、解析結果に基づいて、出力対象テキストの難易度の解析を行う。出力対象テキストの難易度には、内容の難易度、及び、使用している単語や文章の長さ等に基づく文章の難易度が含まれる。
- [0104] なお、言語解析部5 6は、出力対象テキストの難易度について、ユーザの能力等に応じて相対的な評価を行うようにしてもよいし、ユーザの能力等に関わらず絶対的な評価を行うようにしてもよい。
- [0105] 例えば、前者の場合、ユーザの専門分野やユーザの好きな分野に関するテキストの難易度は低くなり、ユーザの専門以外の分野やユーザの嫌いな分野に関するテキストの難易度は高くなる。また、例えば、ユーザの年齢や学歴等に応じて、テキストの難易度は変化する。さらに、例えば、ユーザの母国語と同じ言語のテキストの難易度は低くなり、ユーザの母国語と異なる言語のテキストの難易度は高くなる。
- [0106] 言語解析部5 6は、解析結果を音声出力制御部5 7及び学習部5 8に供給する。
- [0107] ステップS 5 8において、音声出力制御部5 7は、T T Sデータ生成処理を行う。ここで、図6のフローチャートを参照して、T T Sデータ生成処理の詳細について説明する。
- [0108] ステップS 1 0 1において、音声出力制御部5 7は、アテンションモードに設定されているか否かを判定する。アテンションモードに設定されていないと判定された場合、処理はステップS 1 0 2に進む。

[0109] ステップS 1 0 2において、音声出力制御部5 7は、アテンションモードに移行するか否かを判定する。音声出力制御部5 7は、アテンションモードへの移行条件が満たされている場合、アテンションモードに移行すると判定し、処理はステップS 1 0 3に進む。

[0110] なお、アテンションモードへの移行条件は、例えば、コンテキスト及びテキストの特性のうち少なくとも一方に基づいて設定される。例えば、以下は、アテンションモードへの移行条件の例である。

[0111] ・ユーザの意識が合成音声に向いていない。

・合成音声の変化が少ない。これは、合成音声の変化が少なく単調である場合、ユーザの注意が散漫になる可能性が高いためである。

・テキストの特徴量が大きい。これは、テキストの特徴量が大きい部分は、情報量が多かったり、重要度が高かったりする可能性が高く、ユーザが合成音声に注意を向ける必要性が高いからである。

・テキストの特徴量が小さい状態が継続している。これは、テキストの特徴量が小さい部分は、情報量が少なかったり、重要度が低かったりする可能性が高く、ユーザの注意が散漫になる可能性が高いためである。

[0112] 例えば、音声出力制御部5 7は、ユーザの集中度又は緊張度が所定の閾値以下の状態が所定の規定時間以上継続している場合、ユーザの意識が合成音声に向いていないと判定する。或いは、音声出力制御部5 7は、ユーザの視線がクライアント1 1の方向に向いていない状態が所定の規定時間以上継続している場合、ユーザの意識が合成音声に向いていないと判定する。或いは、音声出力制御部5 7は、ユーザの居眠りが検出された場合、ユーザの意識が合成音声に向いていないと判定する。

[0113] また、例えば、音声出力制御部5 7は、T T Sデータにより生成される合成音声の特性（例えば、速度、ピッチ、抑揚、音量等）を表す各パラメータ（以下、特性パラメータと称する）の変化量がそれぞれ所定の範囲内である状態が所定の規定時間以上継続している場合、合成音声の変化が少ないと判定する。或いは、音声出力制御部5 7は、単純に通常モードが所定の規定時

間以上継続している場合、合成音声の変化が少ないと判定する。

[0114] さらに、例えば、音声出力制御部 57 は、出力対象テキストの新規出力部分の特徴量が所定の第 1 の閾値以上である場合、テキストの特徴量が大きいと判定する。

[0115] なお、テキストの特徴量は、例えば、以下の場合に高くなる。

[0116] ・ "The glistening snow covered the field." の "The glistening snow" のような名詞句が含まれる場合

・ "What are those birds?" のように、5 W 1 H を表す単語が含まれる疑問文である場合

・ "The Thames is the river which flows through London" の "river" と "which flows through London" のように係り受けの関係がある場合

・ 話している内容に対する話し手の判断や感じ方を表す言語表現（モダリティ）を含む場合

[0117] また、例えば、音声出力制御部 57 は、出力対象テキストの新規出力部分の特徴量が所定の第 2 の閾値未満である場合、テキストの特徴量が小さいと判定する。なお、この第 2 の閾値は、上述した第 1 の閾値より小さい値に設定される。また、音声出力制御部 57 は、出力対象テキストの特徴量が第 2 の閾値未満である状態が所定の時間以上継続した場合に、テキストの特徴量が小さいと判定するようにしてもよい。

[0118] なお、上記の判定条件の閾値、規定時間、変動範囲が、例えば、上述したステップ S 54、及び、後述するステップ S 61 において調整される。

[0119] また、以上のアテンションモードへの移行条件は、その一例であり、他の移行条件を追加したり、いくつかを削除したりすることも可能である。また、複数の移行条件を用いる場合、そのうちのいくつかの移行条件が満たされた場合に、アテンションモードに移行するようにしてもよいし、少なくとも 1 つの移行条件が満たされた場合に、アテンションモードに移行するようにしてもよい。

[0120] ステップ S 103 において、音声出力制御部 57 は、アテンションモード

に設定する。

[0121] その後、処理はステップS 1 0 6に進む。

[0122] 一方、ステップS 1 0 2において、音声出力制御部5 7は、アテンションモードへの移行条件が満たされていない場合、アテンションモードに移行しないと判定し、ステップS 1 0 3の処理はスキップされ、通常モードのまま、処理はステップS 1 0 6に進む。

[0123] また、ステップS 1 0 1において、アテンションモードに設定されていると判定された場合、処理はステップS 1 0 4に進む。

[0124] ステップS 1 0 4において、音声出力制御部5 7は、アテンションモードを解除するか否かを判定する。例えば、音声出力制御部5 7は、コンテキスト解析部5 5による解析結果に基づいて、ユーザの意識が合成音声に向いている状態が検出された場合、アテンションモードを解除すると判定し、処理はステップS 1 0 5に進む。

[0125] 例えば、ユーザの意識が合成音声に向いている状態として、以下の例が挙げられる。

[0126] ・ユーザが、「ん？」や「何だ？」のようなアテンションモードに対する反応を示す音声を発した場合

・ユーザの視線がクライアント1 1（例えば、音声出力部3 1や表示部3 4）の方向に向いた場合

[0127] なお、音声出力制御部5 7は、ユーザの意識が合成音声に向いている状態が検出されなくても、アテンションモードが所定の規定時間以上継続した場合、アテンションモードを解除すると判定し、処理はステップS 1 0 5に進む。

[0128] ステップS 1 0 5において、音声出力制御部5 7は、通常モードに設定する。これにより、アテンションモードが解除される。

[0129] その後、処理はステップS 1 0 6に進む。

[0130] 一方、ステップS 1 0 4において、音声出力制御部5 7は、ユーザの意識が合成音声に向いている状態が検出されておらず、かつ、アテンションモー

ドがまだ所定の規定時間以上継続していない場合、アテンションモードを解除しないと判定する。そして、ステップS 1 0 5の処理はスキップされ、アテンションモードに設定されたまま、処理はステップS 1 0 6に進む。

[0131] ステップS 1 0 6において、音声出力制御部5 7は、アテンションモードに設定されているか否かを判定する。アテンションモードに設定されていないと判定された場合、処理はステップS 1 0 7に進む。

[0132] ステップS 1 0 7において、音声出力制御部5 7は、通常モードに従ってT T Sデータを生成する。具体的には、音声出力制御部5 7は、出力対象テキストの新規出力部分の合成音声を生成するためのT T Sデータを生成する。このとき、音声出力制御部5 7は、速度、ピッチ、抑揚、音量等の特性パラメータを所定の標準値に設定する。

[0133] その後、処理はステップS 1 0 9に進む。

[0134] 一方、ステップS 1 0 6において、アテンションモードに設定されていると判定された場合、処理はステップS 1 0 8に進む。

[0135] ステップS 1 0 8において、音声出力制御部5 7は、アテンションモードに従ってT T Sデータを生成する。具体的には、音声出力制御部5 7は、出力対象テキストの新規出力部分の合成音声を生成するためのT T Sデータを生成する。このとき、音声出力制御部5 7は、通常モードとは異なる出力態様で合成音声出力されるようにT T Sデータを生成する。これにより、アテンションモードにおける合成音声の出力態様が通常モードから変化し、ユーザの注意を引くことができる。

[0136] ここで、合成音声の出力態様の変更方法の例を示す。合成音声の出力態様の変更方法には、例えば、合成音声の特性を変更する方法、合成音声に対する効果を変更する方法、合成音声の背景のB G Mを変更する方法、合成音声により出力されるテキストを変更する方法、合成音声を出力するクライアントの動作を変更する方法等がある。

[0137] 以下は、合成音声の特性の変更方法の例である。

[0138] ・合成音声の速度、ピッチ、音量、抑揚等の特性を変更する。

[0139] 以下は、合成音声に対する効果の変更方法の例である。

[0140] ・合成音声にエコーをかける。

・合成音声の不協和音で出力する。

・合成音声の話者の設定（例えば、性別、年齢、声質等）を変更する。

・合成音声に間（ポーズ）を挿入する。例えば、名詞句の途中や接続詞の後に、一定時間の間を挿入する。

・合成音声により出力されるテキスト内の特定の言葉を繰り返す。

[0141] なお、特定の言葉を繰り返す場合、出力対象テキストの新規出力部分において繰り返す言葉の数の上限値（以下、最大リピート対象数と称する）、繰り返す言葉（以下、リピート対象と称する）、リピート対象を繰り返す回数（以下、リピート回数と称する）、リピート対象を繰り返す方法（以下、リピート方法と称する）等が設定される。

[0142] 最大リピート対象数は、例えば、ユーザ、出力対象テキストの言語解析の結果、合成音声の出力時間等に基づいて設定される。例えば、最大リピート対象数は、3個を上限として、出力対象テキストの新規出力部分の主要品詞の数に応じて設定される。或いは、最大リピート対象数は、出力対象テキストの新規出力部分の合成音声の出力時間が30秒以上である場合、3個に設定される。或いは、最大リピート対象数は、出力対象テキストの新規出力部分において、10秒に1個の頻度で設定される。

[0143] 或いは、ユーザが所定の年齢以下の子供である場合、最大リピート対象数は無制限に設定され、出力対象テキストの新規出力部分の全ての名詞がリピート対象に設定される。また、ユーザが所定の年齢以上の老人である場合、最大リピート対象数は、1個に設定される。さらに、ユーザが子供及び老人以外である場合、最大リピート対象数は3個に設定される。

[0144] リピート対象は、例えば、名詞、固有名詞、動詞、自立語等の中から設定される。或いは、リピート対象は、合成音声に間が挿入される場合、間の直後の言葉に設定される。

[0145] リピート回数は、例えば、リピート対象の品詞等に基づいて設定される。

例えば、リピート対象が名詞句又は固有名詞を含む場合、リピート回数は3回に設定される。或いは、リピート回数は、例えば、ユーザが子供の場合、2回に設定され、ユーザが老人の場合、3回に設定され、ユーザが子供及び老人以外の場合、1回に設定される。

[0146] リピート方法は、例えば、以下のように設定される。

[0147] ・リピート対象の前に間を挿入する。

・リピート対象の後に言葉を付加する。例えば、“山田さん”がリピート対象である場合、“山田さんだよ”、“山田さんね”のように、“山田さん”の後に、助詞、助動詞、感嘆詞等が付加される。

・リピート対象を前後の言葉と異なる特性で出力する。例えば、リピート対象の音量が大きくされたり、ピッチが高くされたり、速度が遅くされたりする。

[0148] 以下は、BGMの変更方法の例である。

[0149] ・BGMを開始又は停止する。

・BGMを変更する。

[0150] なお、BGMを開始又は変更する場合、例えば、ユーザの嗜好や属性等に応じてBGMが選択される。例えば、ユーザの好きなアーティストの楽曲や、ユーザが良く聴く楽曲がBGMに選択される。或いは、ユーザの年代に応じた楽曲がBGMに選択される。或いは、ユーザの青春時代の流行歌がBGMに選択される。或いは、ユーザが子供の場合、流行している番組のテーマソングがBGMに選択される。

[0151] 以下は、合成音声により出力されるテキストの変更方法の例である。

[0152] ・ユーザに応じた言い回しを付加する。例えば、ユーザが子供であれば、名詞に擬声語や擬態語が付加される。例えば、“かわいい犬がいます。”というテキストに対して、“かわいいぷよぷよの犬がいます。”といったように、“ぷよぷよの”という擬態語が付加される。なお、例えば、付加した擬態語や擬声語の音量を上げるようにしてもよい。

[0153] 以下は、合成音声を出力するクライアント11の動作の変更方法の例であ

る。

[0154] ・合成音声を出力するクライアント 11 の本体又は付属物（例えば、コントローラ等）を振動させる。

[0155] なお、上述した合成音声の出力態様の変更方法は、個別に実施してもよいし、複数の変更方法を組み合わせて実施してもよい。また、1 回のアテンションモードにおいて、変更方法を切り替えたり、各変更方法におけるパラメータ（例えば、特性パラメータ）を変化させたりしてもよい。

[0156] また、実施する出力態様の変更方法は、例えば、コンテキストに基づいて選択される。

[0157] 例えば、ユーザの周囲がざわついている場合、合成音声の音量を大きくする方法が選択される。例えば、ユーザの周囲がざわついていて、ユーザが他の人と話している場合、クライアント 11 を振動させる方法が選択される。例えば、ユーザの周囲が静かであるが、ユーザがクライアント 11 の方向を向いていない場合、出力対象テキストの新規出力部分の名詞句の途中まで出力した時点で一定時間の間を挿入する方法が選択される。例えば、ユーザが小学生である場合、出力対象テキストの新規出力部分の名詞を含む部分をリピート対象に設定し、リピート対象の発話速度を遅くする方法が選択される。

[0158] 或いは、実施する出力態様の変更方法は、各変更方法に対するユーザの反応の学習結果に基づいて選択される。例えば、過去の学習処理によりユーザの反応が高い変更方法ほど、優先して選択される。これにより、各ユーザに対してより効果的な変更方法が選択される。

[0159] 或いは、実施する出力態様の変更方法は、アテンションモードの移行回数や移行頻度に基づいて選択される。例えば、アテンションモードの移行回数や移行頻度が大きく、ユーザの意識がなかなか合成音声に向かない場合、出力態様の变化がより大きく、ユーザがより気づきやすい方法が選択される。

[0160] 図 7 は、図 3 のテキストのうちの "I miss them tonight. I know that my debt to them is beyond measure." の部分の通常モード時とアテンションモ

ード時の S S L M (Speech Synthesis Markup Language) 形式の T T S データの具体例を示している。図内の上側が通常モード時の T T S データを示し、下側がアテンションモード時の T T S データを示している。

[0161] prosody rate (速さ) は、通常モード時に 1 に設定されているのに対し、アテンションモード時に 0.8 に設定され、遅くなっている。pitch は、通常モード時に 100 に設定されているのに対し、アテンションモード時に 130 に設定され、高くなっている。volume (音量) は、通常モード時に 48 に設定されているのに対し、アテンションモード時に 20 に設定され、低くなっている。また、アテンションモード時に、“I miss them” と “tonight” の間に 3000 ms の break time が設定され、3 秒の間が挿入されている。さらに、“tonight” に phoneme が設定され、抑揚が付けられている。

[0162] なお、アテンションモード時の合成音声の出力態様の変更は、ユーザの意識を合成音声に向けさせることが目的であるので、合成音声が不自然であったり、ユーザが聞き取りにくかったりしても構わない。

[0163] その後、処理はステップ S 109 に進む。

[0164] ステップ S 109 において、音声出力制御部 57 は、T T S データの特性パラメータを記憶する。すなわち、音声出力制御部 57 は、生成した T T S データの新規出力部分の速度、ピッチ、抑揚、音量を含む特性パラメータの値を記憶部 54 に記憶させる。

[0165] その後、T T S データ生成処理は終了する。

[0166] 図 5 に戻り、ステップ S 59 において、音声出力制御部 57 は、T T S データを送信する。すなわち、音声出力制御部 57 は、出力対象テキストの新規出力部分の T T S データを、通信部 51 を介して、クライアント 11 に送信する。

[0167] ステップ S 60 において、音声出力制御部 57 は、アテンションモードへの移行条件を変更するか否かを判定する。

[0168] 例えば、音声出力制御部 57 は、上述したステップ S 54 において推定したユーザの集中力と、現在のユーザの集中力との間の誤差が所定の閾値以上

である場合、アテンションモードへの移行条件を変更すると判定する。ユーザの集中力の推定誤差が大きくなる要因としては、例えば、ユーザの体調や、時間の経過によるユーザの集中力の低下等が想定される。また、例えば、出力対象テキストの内容に対するユーザの嗜好度が想定される。例えば、出力対象テキストの内容に対するユーザの嗜好度が高いほど、ユーザの集中力は高くなり、ユーザの嗜好度が低いほど、ユーザの集中力は低くなる。

[0169] また、例えば、音声出力制御部 57 は、上述したステップ S 54 において推定した必要とされる集中力と、実際に必要とされる集中力との間の誤差が所定の閾値以上である場合、アテンションモードへの移行条件を変更すると判定する。必要とされる集中力の推定誤差が大きくなる要因としては、例えば、出力対象テキストの難易度等が想定される。例えば、出力対象テキストの難易度が高いほど、必要とされる集中力は高くなり、出力対象テキストの難易度が低いほど、必要とされる集中力は低くなる。

[0170] また、例えば、音声出力制御部 57 は、アテンションモードへの移行回数又は移行頻度が所定の閾値を超えた場合、アテンションモードへの移行条件を変更すると判定し、処理はステップ S 61 に進む。すなわち、アテンションモードへの移行が頻繁に行われると、ユーザに不快感を与えてしまうおそれがある。そのため、例えば、アテンションモードへの移行回数の上限値を設定し、アテンションモードに移行しないようにしたり、時間の経過につれてアテンションモードへの移行条件を緩くし、移行頻度を下げたりすることが考えられる。

[0171] そして、アテンションモードの移行条件を変更すると判定された場合、処理はステップ S 61 に進む。

[0172] ステップ S 61 において、音声出力制御部 57 は、アテンションモードへの移行条件を変更する。例えば、音声出力制御部 57 は、ユーザの集中力、及び、必要とされる集中力の再推定を行い、その推定結果に基づいて、アテンションモードへの移行条件を再設定する。

[0173] また、例えば、音声出力制御部 57 は、アテンションモードへの移行回数

又は移行頻度に基づいて、アテンションモードへの移行条件を変更する。例えば、アテンションモードへの移行回数が所定の閾値（例えば、50回／日）を超えている場合、以降、アテンションモードへの移行が禁止される。

[0174] その後、処理はステップS62に進む。

[0175] 一方、ステップS60において、アテンションモードへの移行条件を変更しないと判定された場合、ステップS61の処理はスキップされ、処理はステップS62に進む。

[0176] ステップS62において、音声出力制御部57は、TTSデータの送信の停止が要求されたか否かを判定する。TTSデータの送信の停止が要求されていないと判定された場合、処理はステップS55に戻る。

[0177] その後、ステップS55において、出力するテキストがないと判定されるか、ステップS61において、TTSデータの送信の停止が要求されたと判定されるまで、ステップS55乃至S61の処理が繰り返し実行される。

[0178] これにより、TTSデータを生成し、クライアント11に送信する処理が継続される。また、アテンションモードへの移行条件が満たされたとき、アテンションモードに移行し、アテンションモードの解除条件が満たされるとき、通常モードに移行しながら、TTSデータが生成される。

[0179] 一方、1回目のステップS55の処理において、音声出力制御部57は、出力対象テキストが見つからなかった場合、出力するテキストがないと判定し、処理はステップS63に進む。また、例えば、音声出力制御部57は、2回目以降のステップS55の処理において、出力対象テキストのうち未出力の部分が残っていない場合、出力するテキストがないと判定し、処理はステップS63に進む。

[0180] また、ステップS62において、音声出力制御部57は、クライアント11から送信されたTTSデータ送信停止要求を、通信部51を介して受信した場合、TTSデータの送信の停止が要求されたと判定し、処理はステップS63に進む。

[0181] ステップS63において、学習部58は、学習処理を行う。例えば、学習

部 5 8 は、合成音声出力中のユーザの視線の方向、行動、緊張度及び集中度の履歴等に基づいて、ユーザの集中力の学習を行う。ユーザの集中力は、例えば、集中力の高さや継続時間等により表される。

[0182] また、例えば、学習部 5 8 は、出力対象テキストの特性、並びに、合成音声出力中のユーザの視線の方向、表情、行動、緊張度及び集中度の履歴等に基づいて、ユーザの嗜好の学習を行う。例えば、ユーザの集中力が高かった場合、そのときの出力対象テキストのコンテンツに対するユーザの嗜好度は高いと推定される。一方、ユーザの集中力が低かった場合、そのときの出力対象テキストのコンテンツに対するユーザの嗜好度は低いと推定される。

[0183] さらに、例えば、学習部 5 8 は、アテンションモード移行時のユーザの反応の有無や反応時間等に基づいて、合成音声の出力態様の各変更方法に対するユーザの反応の学習を行う。例えば、各変更方法に対して、ユーザがどの程度の確率及び時間で反応するかが学習される。

[0184] 学習部 5 8 は、学習結果を記憶部 5 4 に記憶させる。

[0185] その後、T T S データ提供処理は終了する。

[0186] ここで、図 8 乃至図 1 4 を参照して、アテンションモードの具体的なイメージについて説明する。

[0187] まず、図 8 に示されるように、音声出力部 3 1 から図 3 に示されるテキストの初めから合成音声出力される。

[0188] 次に、図 9 に示されるように、“who I am”の部分までの合成音声出力されたとき、ユーザ 2 0 1 が、T V 2 0 2 に映っている T V プログラムに興味に移り、T V 2 0 1 の方向を向いたとする。このとき、サーバ 1 2 のコンテンツ解析部 5 5 は、撮影部 3 3 からの画像データに基づいて、ユーザ 2 0 1 の視線が T V 2 0 1 の方向を向いていることを検出する。これにより、音声出力制御部 5 7 は、出力モードを通常モードからアテンションモードに移行させる。

[0189] そして、図 1 0 に示されるように、次に音声出力部 3 1 から出力される“I miss them tonight.”の“I miss them”の部分の合成音声の音量が小さくなる

。さらに、名詞句”them”が出力されてから次の”tonight”が出力されるまでの間に、所定の間隔が設けられる。このように合成音声の出力態様が変化することにより、ユーザ２０１が違和感を覚える。

[0190] そして、次の”tonight”が引き続き小さな音で出力されたとき、ユーザ２０１が”what?”と呟いたとする。このとき、サーバ１２のコンテキスト解析部５５は、音声入力部３２からの音声データに基づいて、ユーザ２０１がアテンションモードに対して反応したことを検出する。その結果、音声出力制御部５７は、出力モードをアテンションモードから通常モードに戻す。

[0191] 次に、”I know that my debt to them is beyond measure.”の部分までの合成音声が通常モードにより出力されたとき、再びユーザ２０１がＴＶプログラムに興味に移り、ＴＶ２０２の方向を向いたとする。その結果、再び出力モードが通常モードからアテンションモードに移行する。

[0192] そして、次に、図１１に示されるように、出力されるテキストが変更されるとともに、テキスト内の特定の言葉が繰り返される。具体的には、”Maya”の前に”Mrs”が付加され、”Alma”の前に”Miss”が付加される。また、”Mrs Maya”及び”Miss Alma”の部分が２回繰り返される。さらに、繰り返される部分の音量が大きくなる。また、図１２に示されるように、１回目の”Mrs Maya”を出力してから２回目の”Mrs Maya”を出力するまでの間に、所定の間隔が設けられる。同様に、図１３に示されるように、１回目の”Miss Alma”を出力してから２回目の”Miss Alma”を出力するまでの間に、所定の間隔が設けられる。このように合成音声の出力態様が変化することにより、ユーザ２０１が再び違和感を覚える。

[0193] そして、図１４に示されるように、”all my other brothers and sisters”が出力されたとき、ユーザ２０１の視線がＴＶ２０１からクライアント１１（音声出力部３１）の方向に向いたとする。このとき、サーバ１２のコンテキスト解析部５５は、撮影部３３からの画像データに基づいて、ユーザ２０１の視線がクライアント１１の方向に向いたことを検出する。これにより、音声出力制御部５７は、出力モードをアテンションモードから通常モードに

戻す。

[0194] そして、次のテキストの合成音声は通常モードで出力される。

[0195] このように、コンテンツやテキストの特性等に基づいて、アテンションモードに移行し、通常モードと異なる出力態様で合成音声を出力することにより、ユーザの注意が喚起され、ユーザの意識が合成音声に向かう確率が高くなる。

[0196] <<2. 変形例>>

以下、上述した本技術の実施の形態の変形例について説明する。

[0197] <2-1. システムの構成例に関する変形例>

図1の情報処理システム10の構成例は、その一例であり、必要に応じて変更することが可能である。

[0198] 例えば、クライアント11の機能の一部をサーバ12に設けたり、サーバ12の機能の一部をクライアント11に設けたりすることが可能である。

[0199] また、例えば、クライアント11とサーバ12を一体化し、1つの装置で上記の処理を行うことも可能である。

[0200] <2-2. その他の変形例>

例えば、音声出力制御部57が、記憶部54又は外部からTTLデータを取得し、通常モードの場合、取得したTTLデータを加工せずにクライアント12に送信し、アテンションモードの場合、合成音声の出力態様を変更するように、取得したTTLデータを加工して、クライアント12に送信するようにしてもよい。

[0201] <2-3. コンピュータの構成例>

上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、コンピュータにインストールされる。ここで、コンピュータには、専用のハードウェアに組み込まれているコンピュータや、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコン

ピュータなどが含まれる。

[0202] 図15は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すブロック図である。

[0203] コンピュータにおいて、CPU (Central Processing Unit) 401, ROM (Read Only Memory) 402, RAM (Random Access Memory) 403は、バス404により相互に接続されている。

[0204] バス404には、さらに、入出力インタフェース405が接続されている。入出力インタフェース405には、入力部406、出力部407、記憶部408、通信部409、及びドライブ410が接続されている。

[0205] 入力部406は、キーボード、マウス、マイクロフォンなどよりなる。出力部407は、ディスプレイ、スピーカなどよりなる。記憶部408は、ハードディスクや不揮発性のメモリなどよりなる。通信部409は、ネットワークインタフェースなどよりなる。ドライブ410は、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリなどのリムーバブルメディア411を駆動する。

[0206] 以上のように構成されるコンピュータでは、CPU401が、例えば、記憶部408に記憶されているプログラムを、入出力インタフェース405及びバス404を介して、RAM403にロードして実行することにより、上述した一連の処理が行われる。

[0207] コンピュータ (CPU401) が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブルメディア411に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の伝送媒体を介して提供することができる。

[0208] コンピュータでは、プログラムは、リムーバブルメディア411をドライブ410に装着することにより、入出力インタフェース405を介して、記憶部408にインストールすることができる。また、プログラムは、有線または無線の伝送媒体を介して、通信部409で受信し、記憶部408にイン

ストールすることができる。その他、プログラムは、ROM 402や記憶部 408に、あらかじめインストールしておくことができる。

[0209] なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

[0210] また、複数のコンピュータが連携して上述した処理を行うようにしてもよい。そして、上述した処理を行う単数又は複数のコンピュータにより、コンピュータシステムが構成される。

[0211] また、本明細書において、システムとは、複数の構成要素（装置、モジュール（部品）等）の集合を意味し、すべての構成要素が同一筐体中にあるか否かは問わない。したがって、別個の筐体に収納され、ネットワークを介して接続されている複数の装置、及び、1つの筐体の中に複数のモジュールが収納されている1つの装置は、いずれも、システムである。

[0212] さらに、本技術の実施の形態は、上述した実施の形態に限定されるものではなく、本技術の要旨を逸脱しない範囲において種々の変更が可能である。

[0213] 例えば、本技術は、1つの機能をネットワークを介して複数の装置で分担、共同して処理するクラウドコンピューティングの構成をとることができる。

[0214] また、上述のフローチャートで説明した各ステップは、1つの装置で実行する他、複数の装置で分担して実行することができる。

[0215] さらに、1つのステップに複数の処理が含まれる場合には、その1つのステップに含まれる複数の処理は、1つの装置で実行する他、複数の装置で分担して実行することができる。

[0216] また、本明細書に記載された効果はあくまで例示であって限定されるものではなく、他の効果があってもよい。

[0217] また、例えば、本技術は以下のような構成も取ることができる。

(1)

テキストを音声に変換した合成音声を出力するときのコンテキストに基づいて、前記合成音声の出力態様を制御する音声出力制御部を備える情報処理装置。

(2)

前記音声出力制御部は、前記コンテキストが所定の条件を満たした場合、前記合成音声の出力態様を変更する

前記(1)に記載の情報処理装置。

(3)

前記合成音声の出力態様の変更は、前記合成音声の特性、前記合成音声に対する効果、前記合成音声の背景のBGM (Back Ground Music)、前記合成音声により出力される前記テキスト、及び、前記合成音声を出力する装置の動作のうち少なくとも1つの変更を含む

前記(2)に記載の情報処理装置。

(4)

前記合成音声の特性は、速度、ピッチ、音量、及び、抑揚のうち少なくとも1つを含み、

前記合成音声に対する効果は、前記テキスト内の特定の言葉の繰り返し、及び、前記合成音声への間の挿入のうち少なくとも1つを含む

前記(3)に記載の情報処理装置。

(5)

前記音声出力制御部は、ユーザの意識が前記合成音声に向いていない状態が検出された場合、前記合成音声の出力態様を変更する

前記(2)乃至(4)のいずれかに記載の情報処理装置。

(6)

前記音声出力制御部は、前記合成音声の出力態様を変更した後、ユーザの意識が前記合成音声に向いている状態が検出された場合、前記合成音声の出力態様を元に戻す

前記(2)乃至(5)のいずれかに記載の情報処理装置。

(7)

前記音声出力制御部は、前記合成音声の特性の変化量が所定の範囲内である状態が所定の時間以上継続した場合、前記合成音声の出力態様を変更する

前記(2)乃至(6)のいずれかに記載の情報処理装置。

(8)

前記音声出力制御部は、前記コンテキストに基づいて、前記合成音声の出力態様の変更方法を選択する

前記(2)乃至(7)のいずれかに記載の情報処理装置。

(9)

前記合成音声の出力態様の変更方法に対するユーザの反応を学習する学習部を

さらに備え、

前記音声出力制御部は、前記ユーザの反応の学習結果に基づいて、前記合成音声の出力態様の変更方法を選択する

前記(2)乃至(8)のいずれかに記載の情報処理装置。

(10)

前記音声出力制御部は、さらに前記テキストの特性に基づいて、前記合成音声の出力態様を制御する

前記(1)乃至(9)のいずれかに記載の情報処理装置。

(11)

前記音声出力制御部は、前記テキストの特徴量が第1の閾値以上である場合、又は、前記テキストの特徴量が第2の閾値未満である場合、前記合成音声の出力態様を変更する

前記(10)に記載の情報処理装置。

(12)

前記音声出力制御部は、前記合成音声の生成に用いる音声制御データを他の情報処理装置に供給することにより、前記他の情報処理装置における前記合成音声の出力態様を制御する

前記（１）乃至（１１）のいずれかに記載の情報処理装置。

（１３）

前記音声出力制御部は、前記他の情報処理装置から取得した前記コンテキストに関するコンテキストデータに基づいて、前記音声制御データを生成する

前記（１２）に記載の情報処理装置。

（１４）

前記コンテキストデータは、ユーザの周囲を撮影した画像に基づくデータ、前記ユーザの周囲の音声に基づくデータ、及び、前記ユーザの生体情報に基づくデータのうち少なくとも１つを含む

前記（１３）に記載の情報処理装置。

（１５）

前記コンテキストデータに基づいて、前記コンテキストを解析するコンテキスト解析部を

さらに含む前記（１３）又は（１４）に記載の情報処理装置。

（１６）

前記コンテキストは、ユーザの状態、前記ユーザの特性、前記合成音声出力される環境、及び、前記合成音声の特性のうち少なくとも１つを含む

前記（１）乃至（１５）のいずれかに記載の情報処理装置。

（１７）

前記合成音声出力される環境は、前記ユーザの周囲の環境、前記合成音声出力する装置、及び、前記合成音声出力するアプリケーションプログラムのうち少なくとも１つを含む

前記（１６）に記載の情報処理装置。

（１８）

テキストを音声に変換した合成音声出力するときのコンテキストに基づいて、前記合成音声の出力態様を制御する音声出力制御ステップを

含む情報処理方法。

(19)

テキストを音声に変換した合成音声を出力するときのコンテキストに関するコンテキストデータを他の情報処理装置に送信し、前記コンテキストデータに基づいて出力態様が制御された前記合成音声の生成に用いる音声制御データを前記他の情報処理装置から受信する通信部と、

前記音声制御データに基づいて前記合成音声を生成する音声合成部とを備える情報処理装置。

(20)

テキストを音声に変換した合成音声を出力するときのコンテキストに関するコンテキストデータを他の情報処理装置に送信し、前記コンテキストデータに基づいて出力態様が制御された前記合成音声の生成に用いる音声制御データを前記他の情報処理装置から受信する通信ステップと、

前記音声制御データに基づいて前記合成音声を生成する音声合成ステップとを含む情報処理方法。

符号の説明

[0218] 10 情報処理システム, 11 クライアント, 12 サーバ, 31 音声出力部, 32 音声入力部, 33 撮影部, 34 表示部, 35 生体情報取得部, 38 プロセッサ, 55 コンテキスト解析部, 56 言語解析部, 57 音声出力制御部, 58 学習部, 101 音声合成部, 102 コンテキストデータ取得部

請求の範囲

- [請求項1] テキストを音声に変換した合成音声を出力するときのコンテキストに基づいて、前記合成音声の出力態様を制御する音声出力制御部を備える情報処理装置。
- [請求項2] 前記音声出力制御部は、前記コンテキストが所定の条件を満たした場合、前記合成音声の出力態様を変更する
請求項 1 に記載の情報処理装置。
- [請求項3] 前記合成音声の出力態様の変更は、前記合成音声の特性、前記合成音声に対する効果、前記合成音声の背景の B G M (Back Ground Music)、前記合成音声により出力される前記テキスト、及び、前記合成音声を出力する装置の動作のうち少なくとも 1 つの変更を含む
請求項 2 に記載の情報処理装置。
- [請求項4] 前記合成音声の特性は、速度、ピッチ、音量、及び、抑揚のうち少なくとも 1 つを含み、
前記合成音声に対する効果は、前記テキスト内の特定の言葉の繰り返し、及び、前記合成音声への間の挿入のうち少なくとも 1 つを含む
請求項 3 に記載の情報処理装置。
- [請求項5] 前記音声出力制御部は、ユーザの意識が前記合成音声に向いていない状態が検出された場合、前記合成音声の出力態様を変更する
請求項 2 に記載の情報処理装置。
- [請求項6] 前記音声出力制御部は、前記合成音声の出力態様を変更した後、ユーザの意識が前記合成音声に向いている状態が検出された場合、前記合成音声の出力態様を元に戻す
請求項 2 に記載の情報処理装置。
- [請求項7] 前記音声出力制御部は、前記合成音声の特性の変化量が所定の範囲内である状態が所定の時間以上継続した場合、前記合成音声の出力態様を変更する
請求項 2 に記載の情報処理装置。

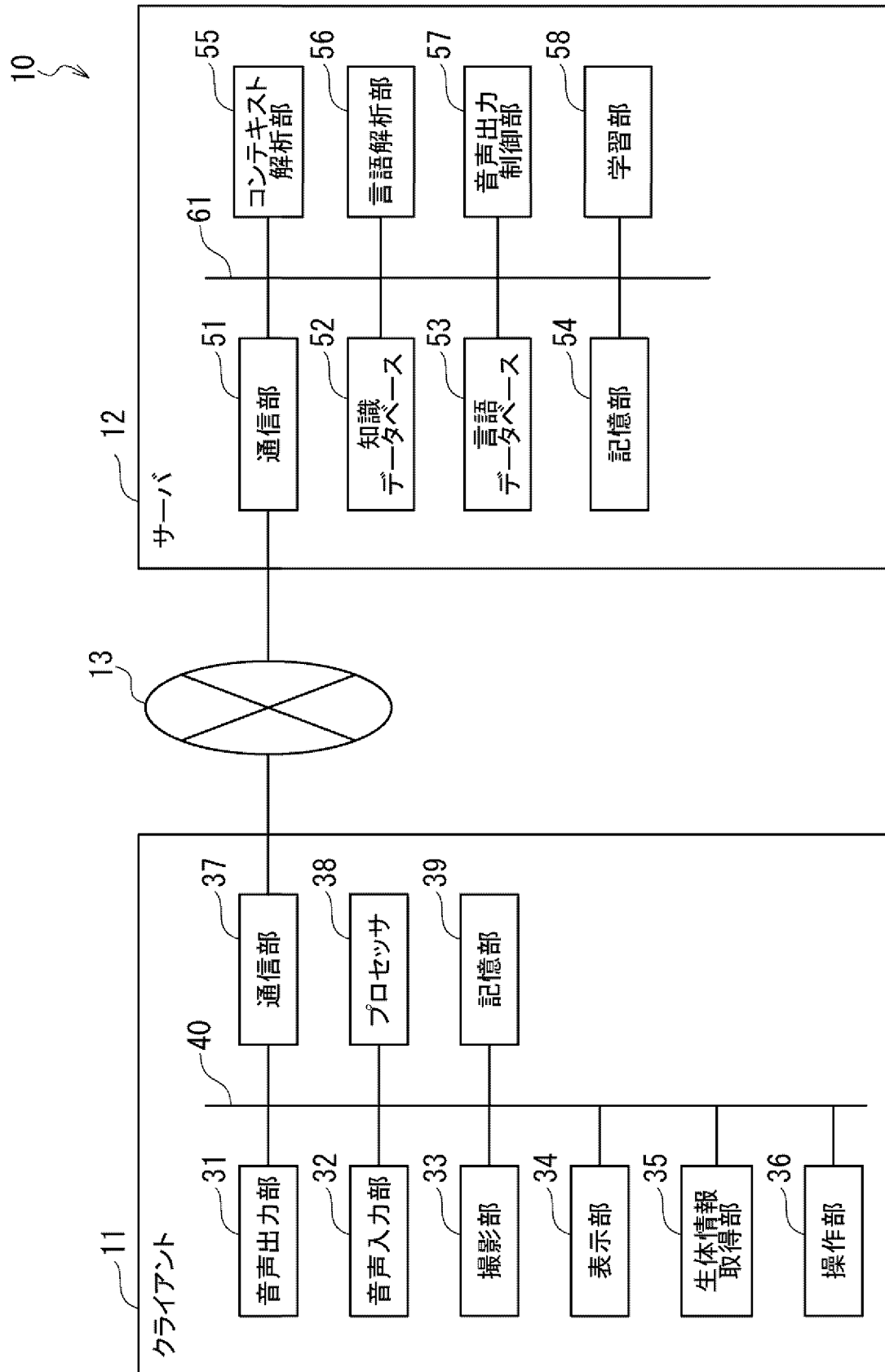
- [請求項8] 前記音声出力制御部は、前記コンテキストに基づいて、前記合成音声の出力態様の変更方法を選択する
請求項2に記載の情報処理装置。
- [請求項9] 前記合成音声の出力態様の変更方法に対するユーザの反応を学習する学習部を
さらに備え、
前記音声出力制御部は、前記ユーザの反応の学習結果に基づいて、前記合成音声の出力態様の変更方法を選択する
請求項2に記載の情報処理装置。
- [請求項10] 前記音声出力制御部は、さらに前記テキストの特性に基づいて、前記合成音声の出力態様を制御する
請求項1に記載の情報処理装置。
- [請求項11] 前記音声出力制御部は、前記テキストの特徴量が第1の閾値以上である場合、又は、前記テキストの特徴量が第2の閾値未満である場合、前記合成音声の出力態様を変更する
請求項10に記載の情報処理装置。
- [請求項12] 前記音声出力制御部は、前記合成音声の生成に用いる音声制御データを他の情報処理装置に供給することにより、前記他の情報処理装置における前記合成音声の出力態様を制御する
請求項1に記載の情報処理装置。
- [請求項13] 前記音声出力制御部は、前記他の情報処理装置から取得した前記コンテキストに関するコンテキストデータに基づいて、前記音声制御データを生成する
請求項12に記載の情報処理装置。
- [請求項14] 前記コンテキストデータは、ユーザの周囲を撮影した画像に基づくデータ、前記ユーザの周囲の音声に基づくデータ、及び、前記ユーザの生体情報に基づくデータのうち少なくとも1つを含む
請求項13に記載の情報処理装置。

- [請求項15] 前記コンテキストデータに基づいて、前記コンテキストを解析するコンテキスト解析部を
さらに含む請求項13に記載の情報処理装置。
- [請求項16] 前記コンテキストは、ユーザの状態、前記ユーザの特性、前記合成音声出力される環境、及び、前記合成音声の特性のうち少なくとも1つを含む
請求項1に記載の情報処理装置。
- [請求項17] 前記合成音声出力される環境は、前記ユーザの周囲の環境、前記合成音声出力する装置、及び、前記合成音声出力するアプリケーションプログラムのうち少なくとも1つを含む
請求項16に記載の情報処理装置。
- [請求項18] テキストを音声に変換した合成音声出力するときのコンテキストに基づいて、前記合成音声出力態様を制御する音声出力制御ステップを
含む情報処理方法。
- [請求項19] テキストを音声に変換した合成音声出力するときのコンテキストに関するコンテキストデータを他の情報処理装置に送信し、前記コンテキストデータに基づいて出力態様が制御された前記合成音声の生成に用いる音声制御データを前記他の情報処理装置から受信する通信部と、
前記音声制御データに基づいて前記合成音声生成する音声合成部と
を備える情報処理装置。
- [請求項20] テキストを音声に変換した合成音声出力するときのコンテキストに関するコンテキストデータを他の情報処理装置に送信し、前記コンテキストデータに基づいて出力態様が制御された前記合成音声の生成に用いる音声制御データを前記他の情報処理装置から受信する通信ステップと、

前記音声制御データに基づいて前記合成音声を生成する音声合成ステップと
を含む情報処理方法。

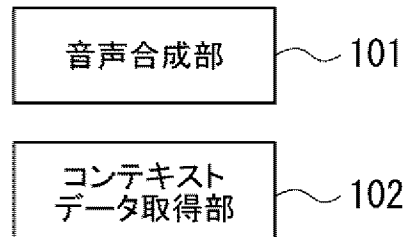
[図1]

Fig. 1



[図2]

Fig. 2

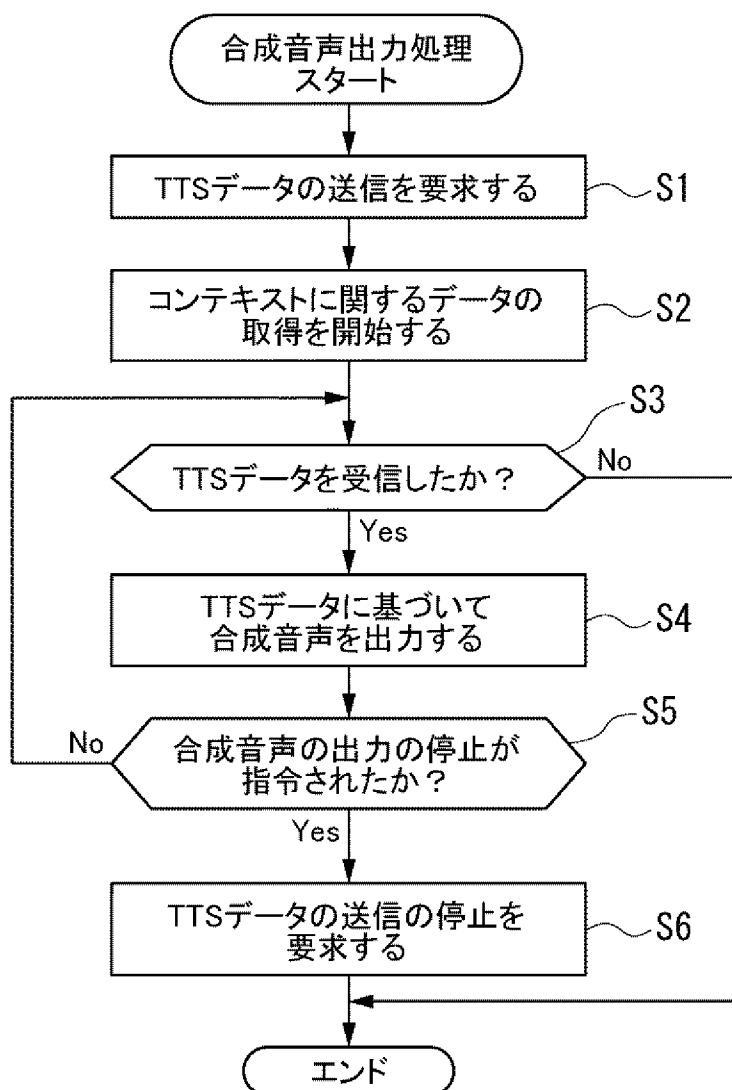


[図3]

Fig. 3

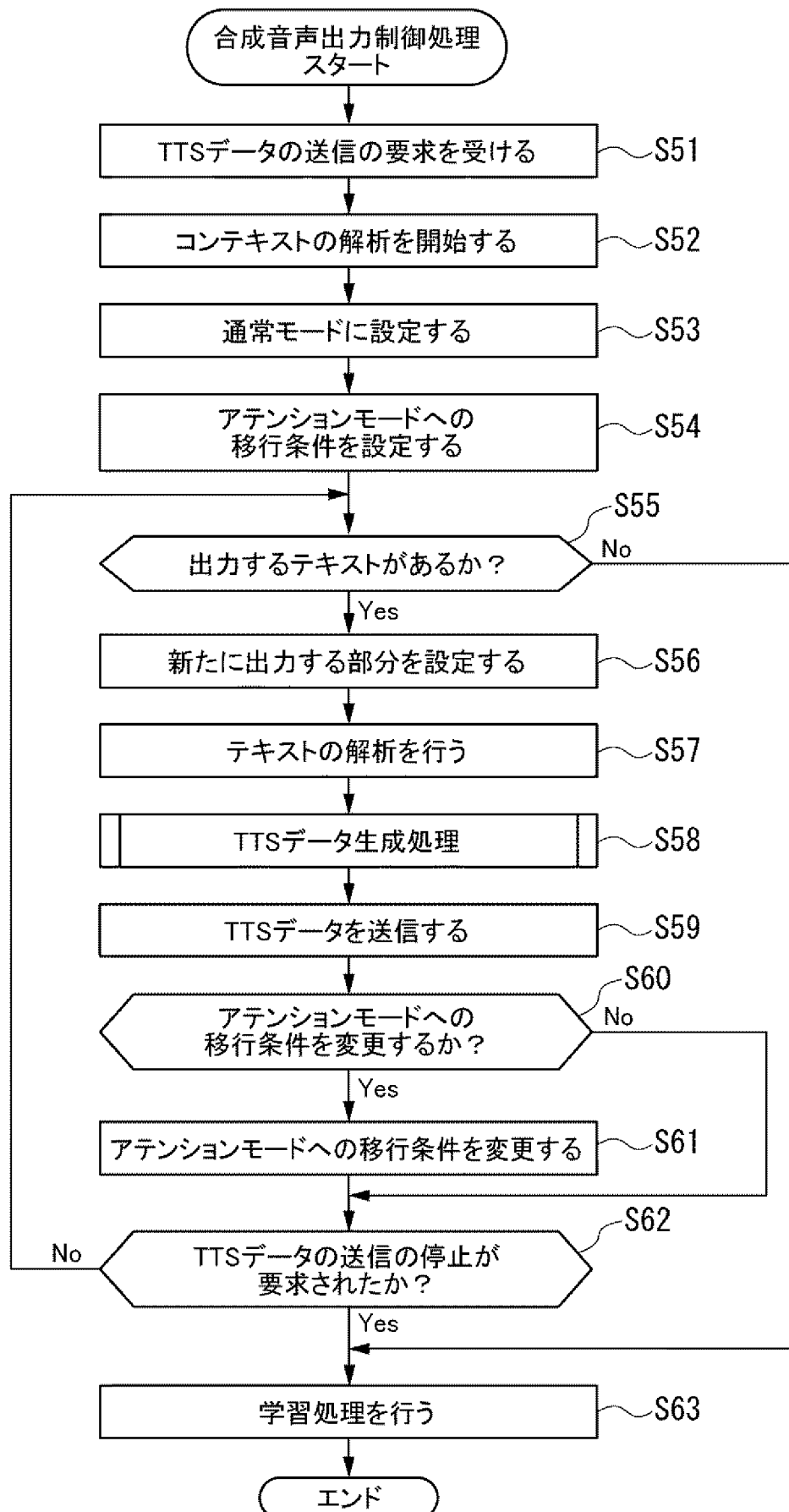
And while she's no longer with us, I know my grandmother's watching, along with the family that made me who I am. I miss them tonight. I know that my debt to them is beyond measure. To my sister Maya, my sister Alma, all my other brothers and sisters, thank you so much for all the support that you've given me.

[図4]
Fig. 4



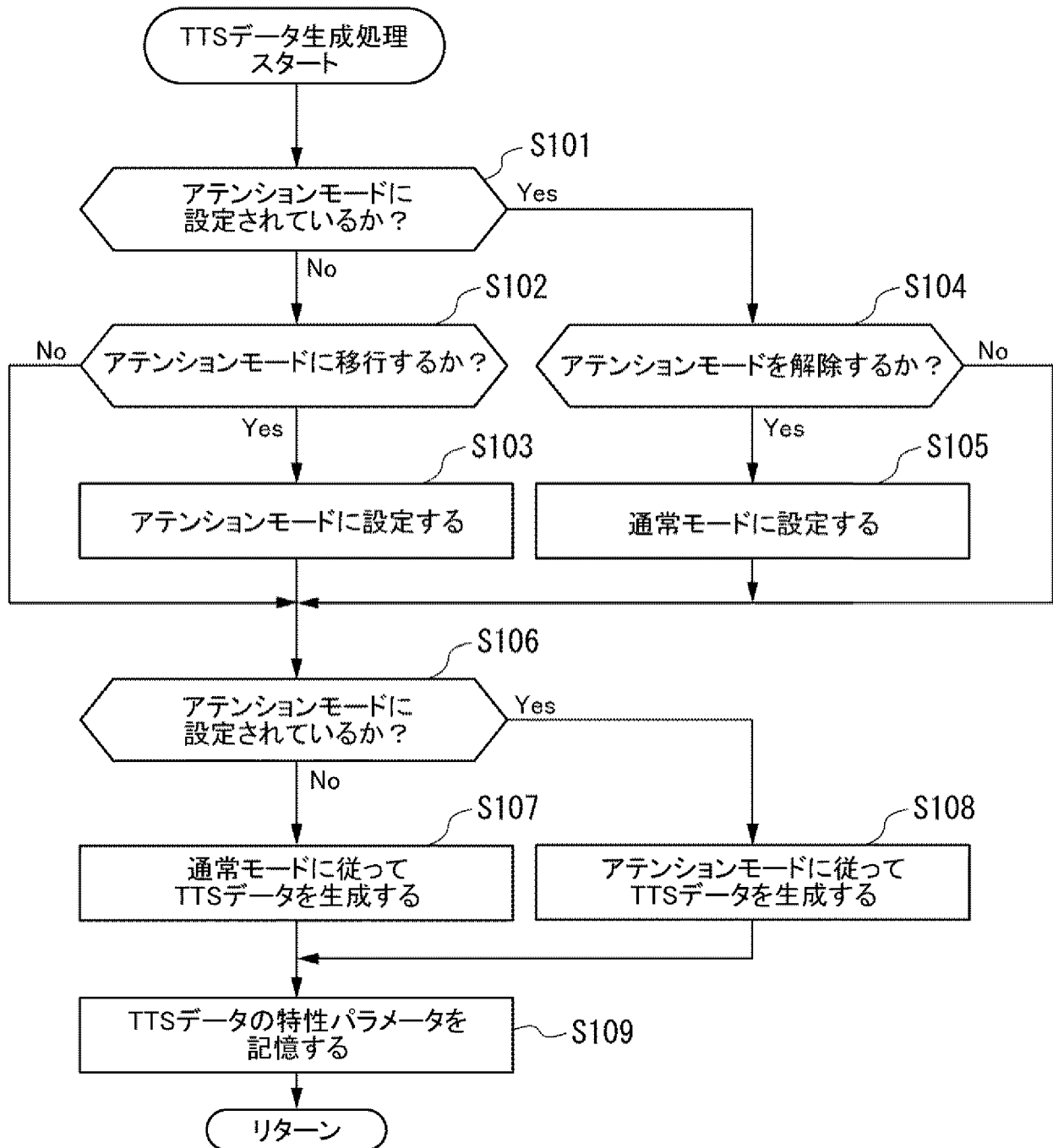
[図5]

Fig. 5



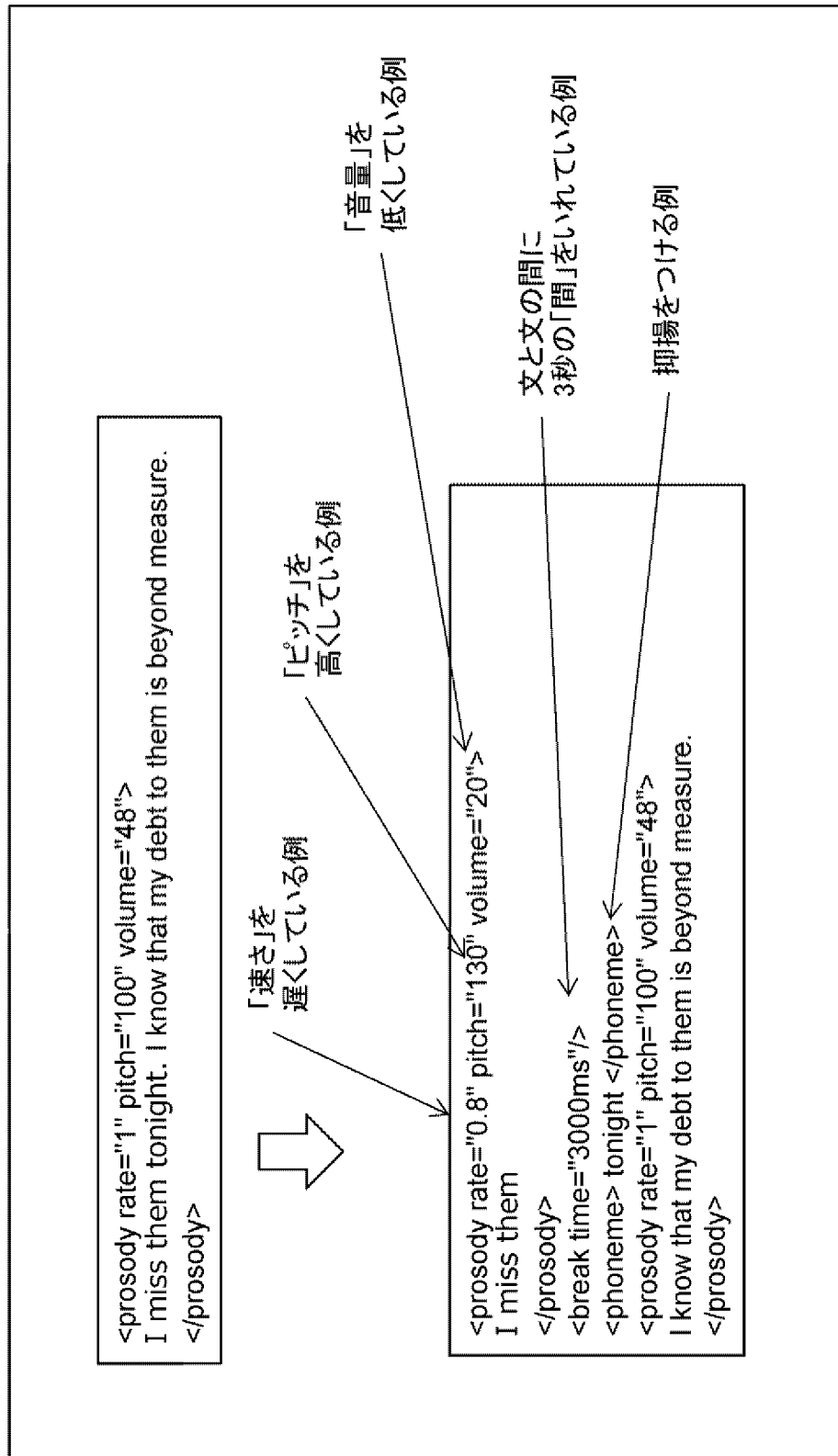
[図6]

Fig. 6



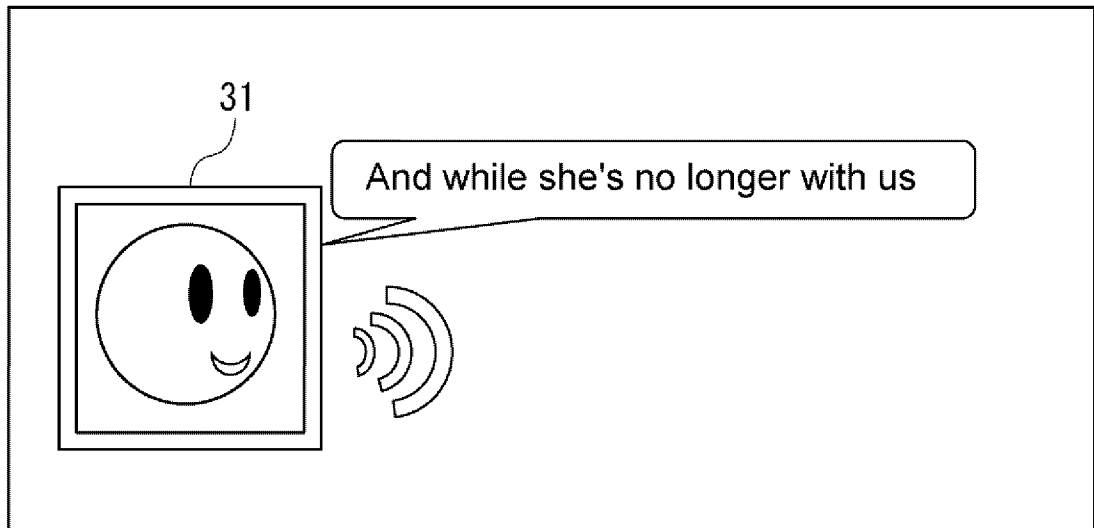
[図7]

Fig. 7



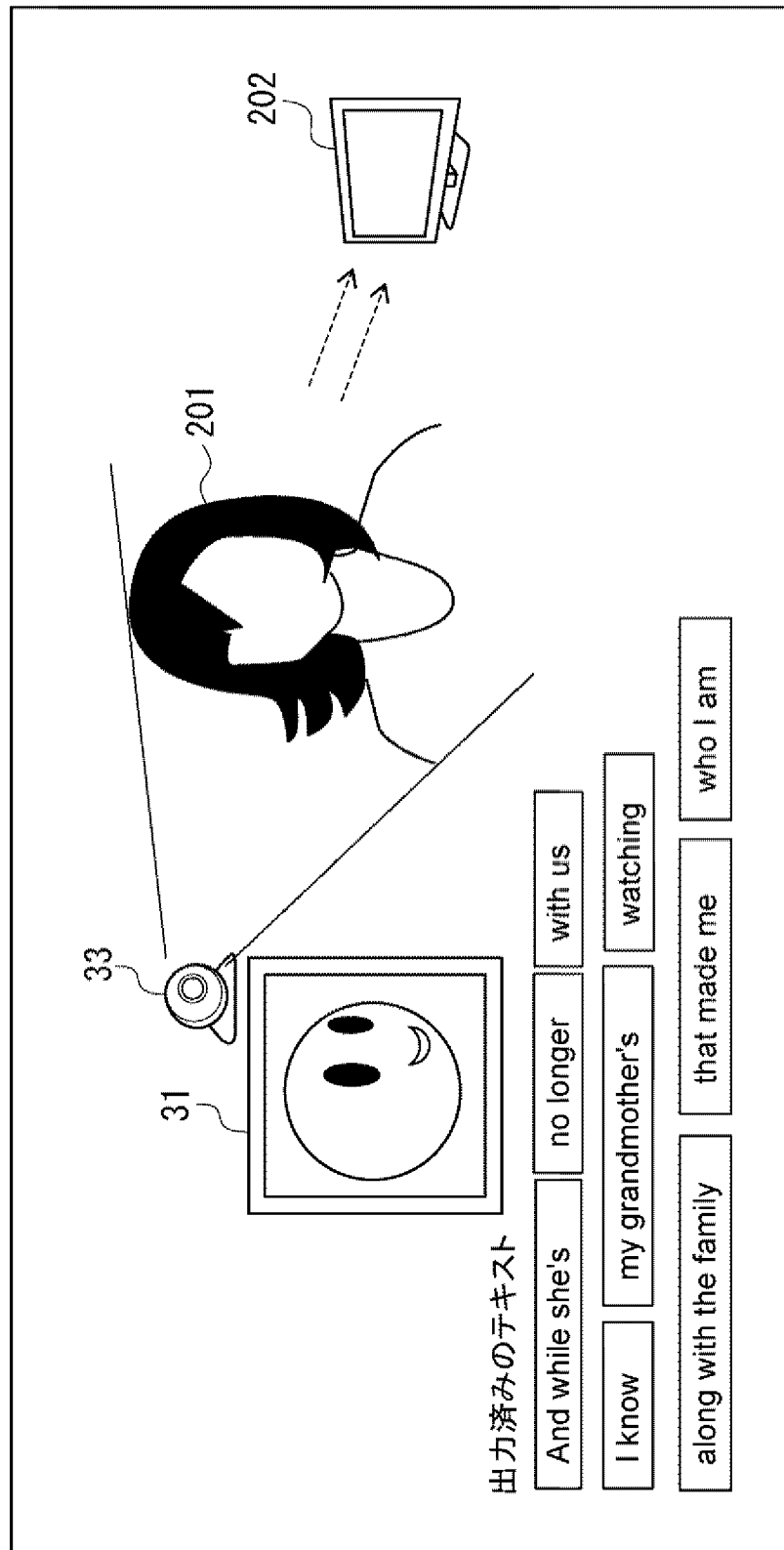
[図8]

Fig. 8



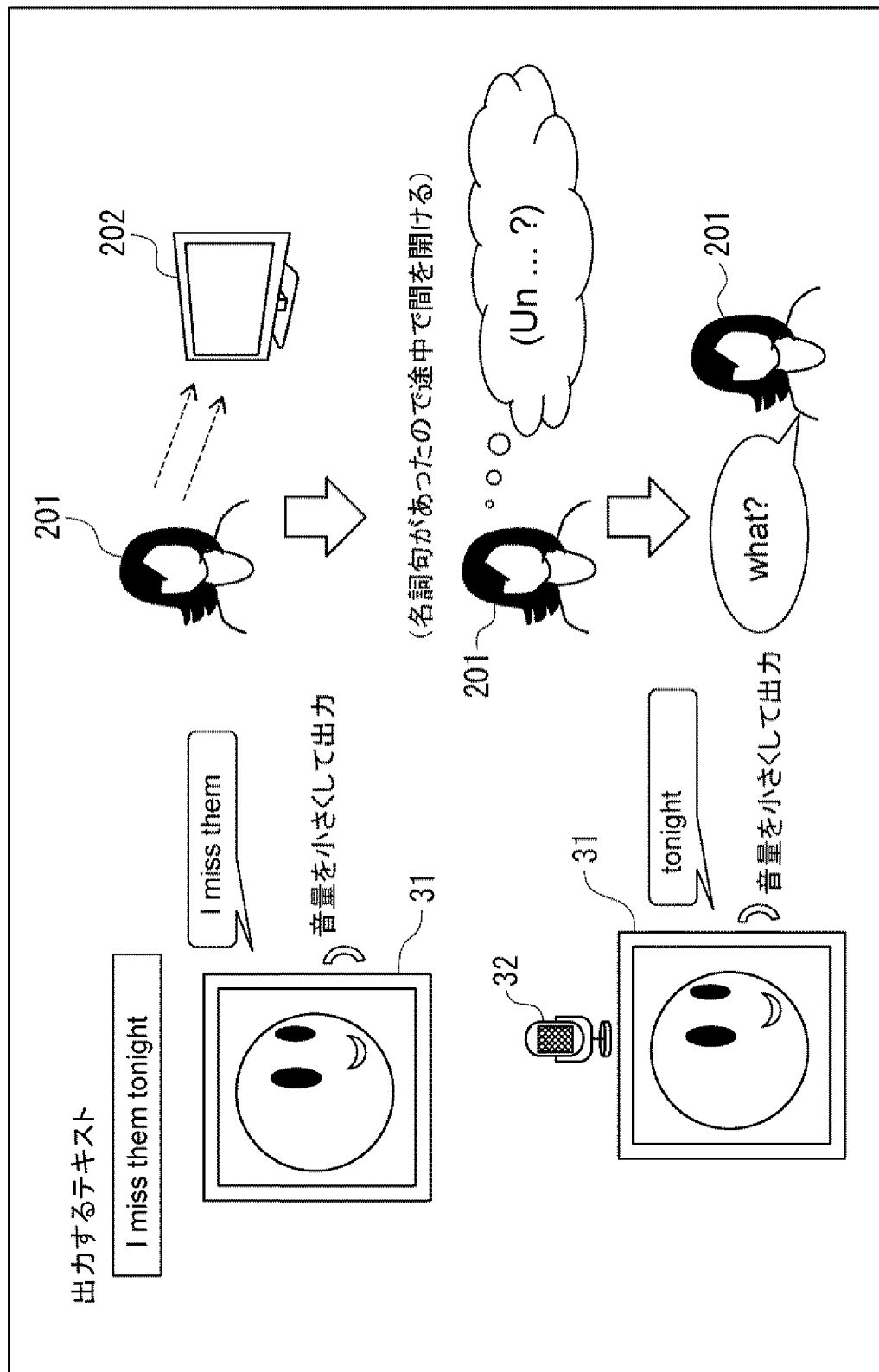
[図9]

Fig. 9



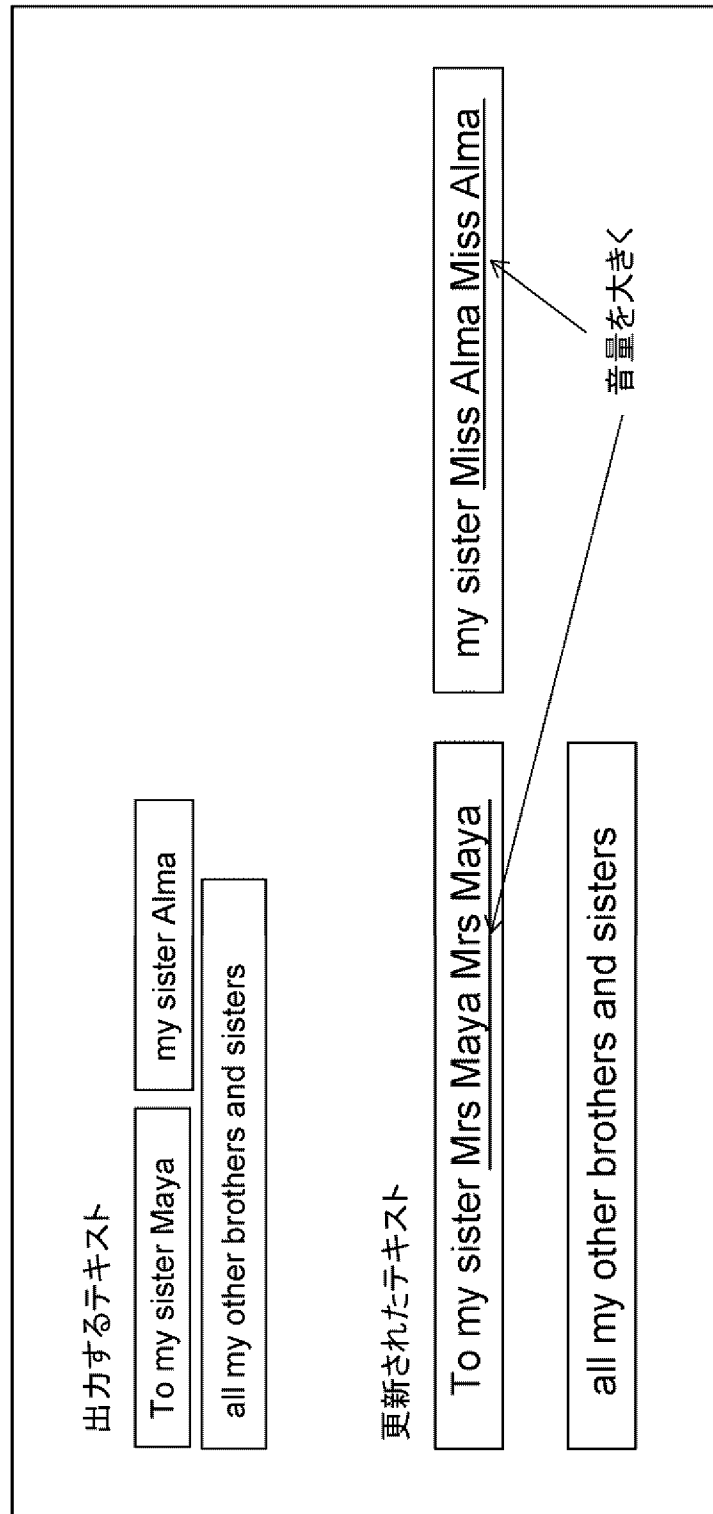
[図10]

Fig. 10

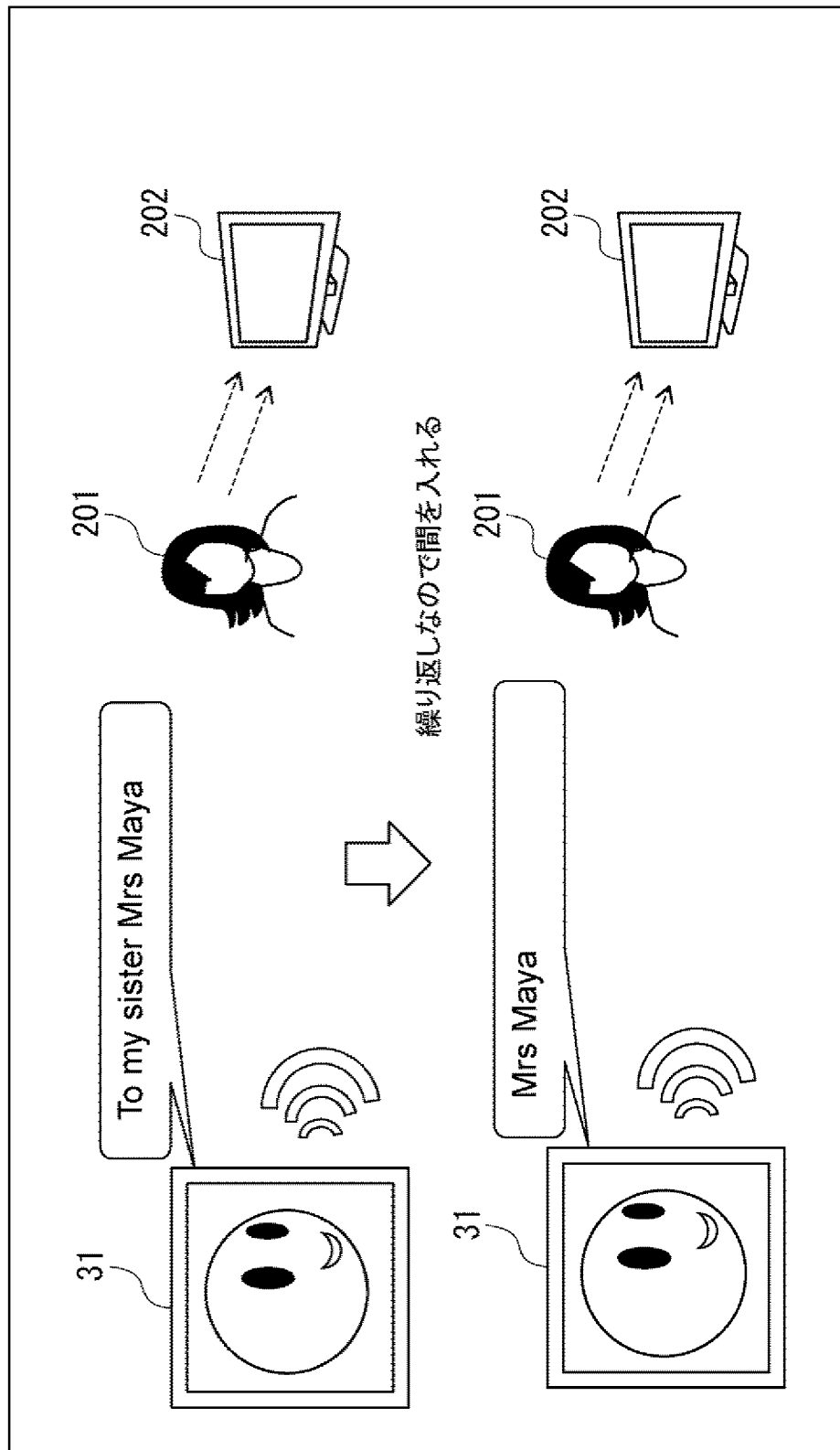


[図11]

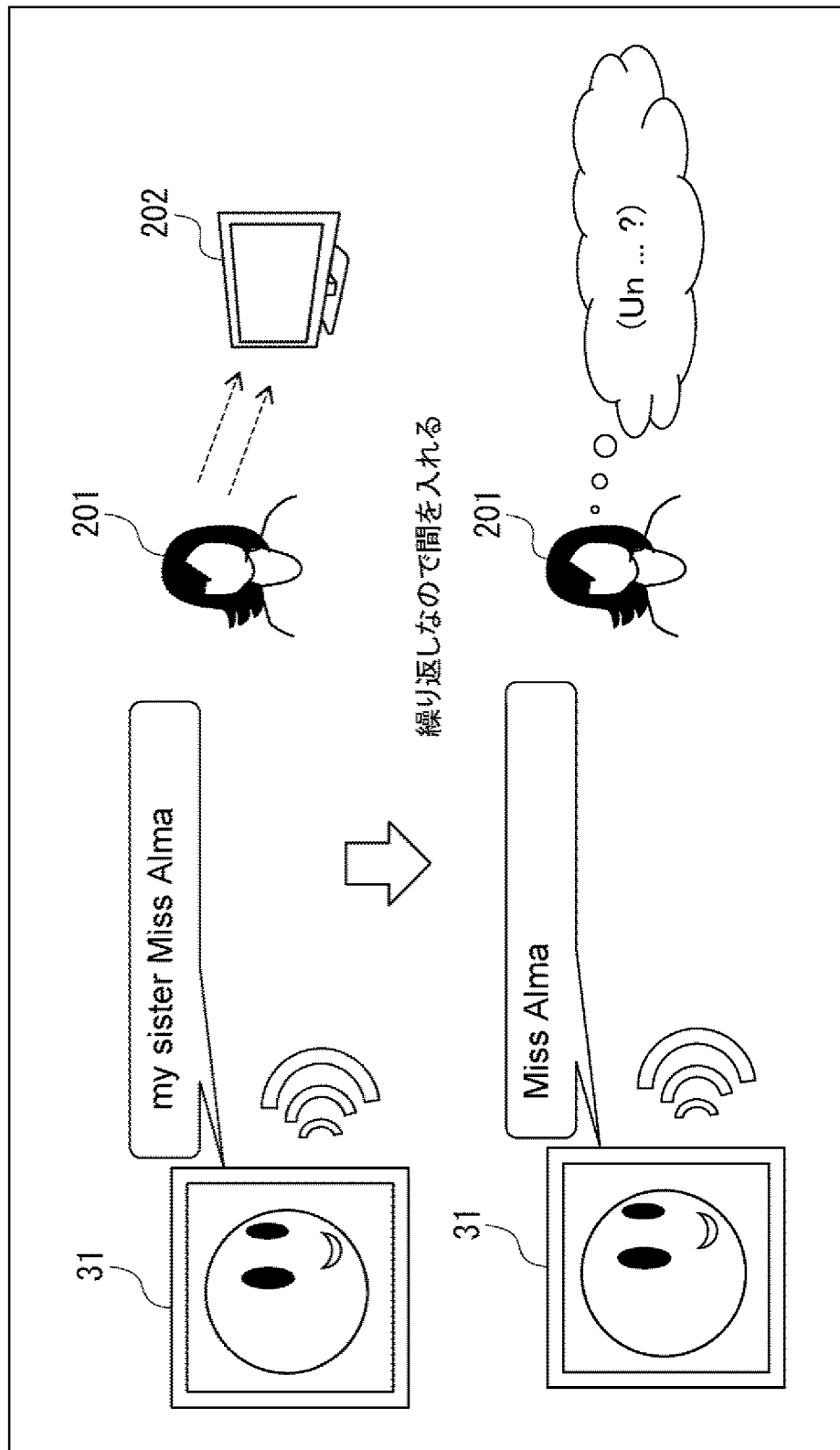
Fig. 11



[図12]
Fig. 12

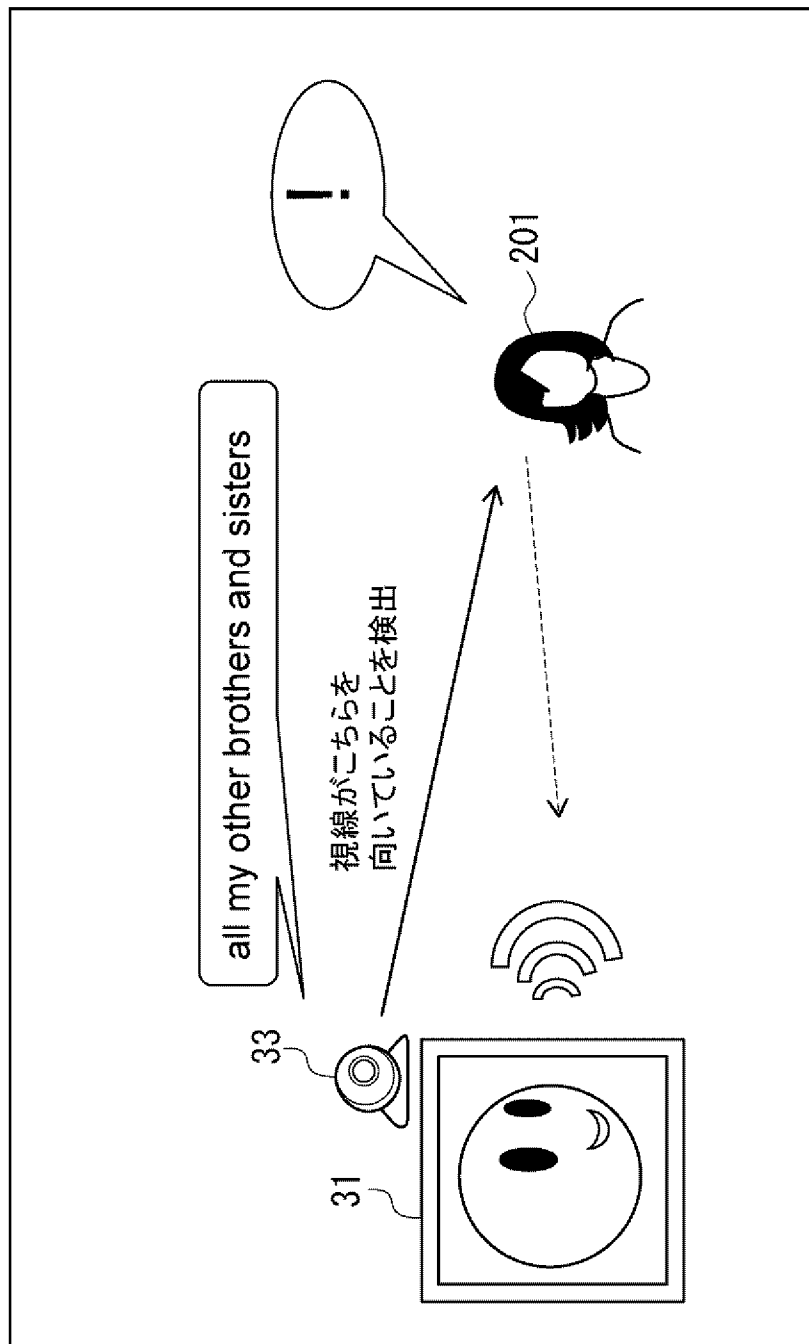


[図13]
Fig. 13



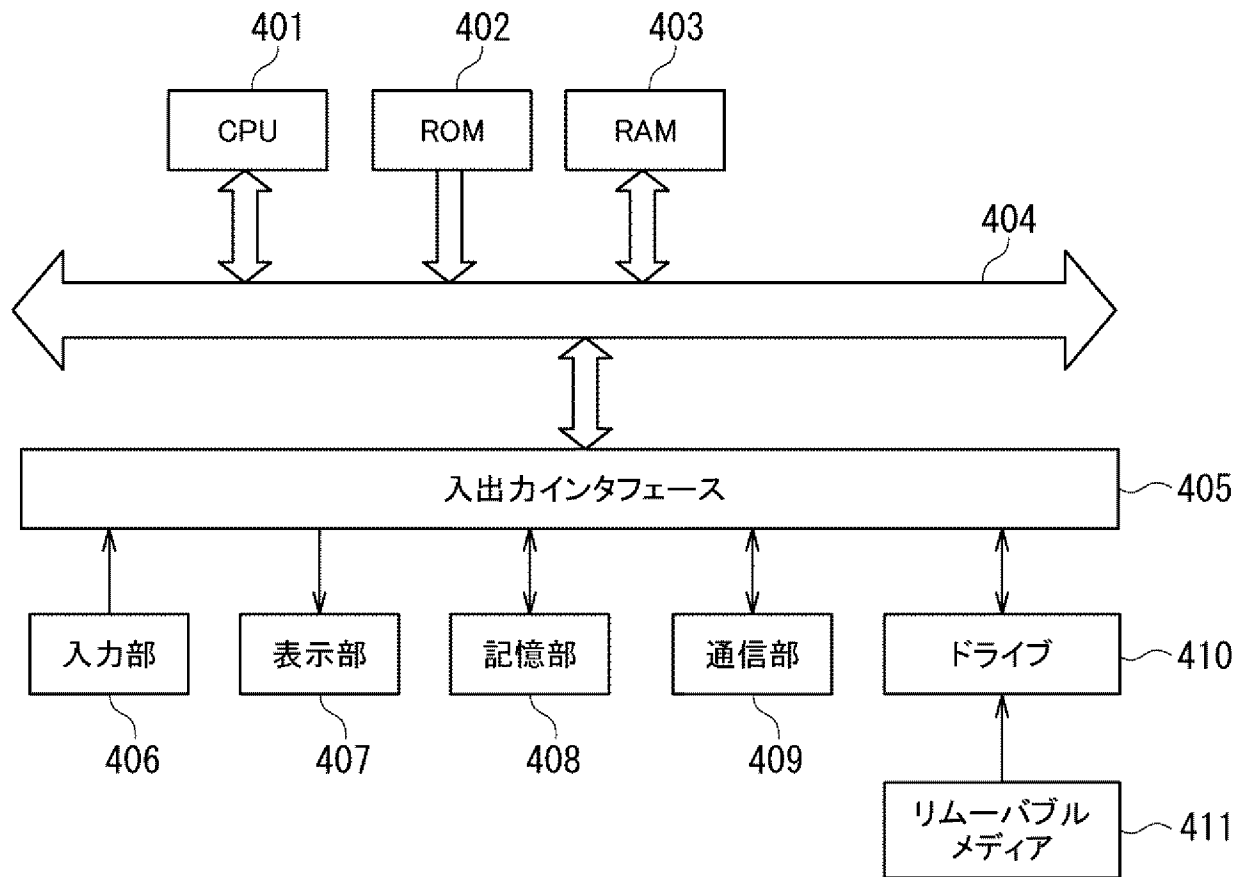
[図14]

Fig. 14



[図15]

Fig. 15



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/026961

A. CLASSIFICATION OF SUBJECT MATTER

G10L13/10(2013.01)i, G10L13/02(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L13/10, G10L13/02

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2017
Kokai Jitsuyo Shinan Koho	1971-2017	Toroku Jitsuyo Shinan Koho	1994-2017

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2011-186143 A (Hitachi, Ltd.), 22 September 2011 (22.09.2011), paragraphs [0021] to [0077] (Family: none)	1-3, 5-9, 12-20 4, 10, 11
Y	JP 2003-131700 A (Matsushita Electric Industrial Co., Ltd.), 09 May 2003 (09.05.2003), paragraphs [0014] to [0105] (Family: none)	4, 10, 11
A	JP 2010-128099 A (Toyota InfoTechnology Center, Co., Ltd.), 10 June 2010 (10.06.2010), paragraphs [0022] to [0053] (Family: none)	1-20

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
08 September 2017 (08.09.17)

Date of mailing of the international search report
19 September 2017 (19.09.17)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/026961

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 11-161298 A (Toshiba Corp.), 18 June 1999 (18.06.1999), paragraphs [0015] to [0061] (Family: none)	1-20
A	JP 5-224688 A (NEC Corp.), 03 September 1993 (03.09.1993), paragraphs [0013] to [0030] (Family: none)	1-20
A	JP 2008-299135 A (NEC Corp.), 11 December 2008 (11.12.2008), paragraphs [0046] to [0119] (Family: none)	1-20
A	JP 2001-022370 A (Fujitsu Ten Ltd.), 06 December 2001 (06.12.2001), paragraphs [0019] to [0090] (Family: none)	1-20

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G10L13/10(2013.01)i, G10L13/02(2013.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G10L13/10, G10L13/02

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2017年
日本国実用新案登録公報	1996-2017年
日本国登録実用新案公報	1994-2017年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X	JP 2011-186143 A（株式会社日立製作所） 2011.09.22, 段落[0021]-[0077]	1-3, 5-9, 12-20
Y	（ファミリーなし）	4, 10, 11
Y	JP 2003-131700 A（松下電器産業株式会社） 2003.05.09, 段落[0014]-[0105] （ファミリーなし）	4, 10, 11

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

08.09.2017

国際調査報告の発送日

19.09.2017

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

鈴木 圭一郎

5Z

3141

電話番号 03-3581-1101 内線 3591

C (続き) . 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2010-128099 A (株式会社トヨタ I T開発センター) 2010. 06. 10, 段落[0022]-[0053] (ファミリーなし)	1-20
A	JP 11-161298 A (株式会社東芝) 1999. 06. 18, 段落[0015]-[0061] (ファミリーなし)	1-20
A	JP 5-224688 A (日本電気株式会社) 1993. 09. 03, 段落[0013]-[0030] (ファミリーなし)	1-20
A	JP 2008-299135 A (日本電気株式会社) 2008. 12. 11, 段落[0046]-[0119] (ファミリーなし)	1-20
A	JP 2001-022370 A (富士通テン株式会社) 2001. 12. 06, 段落 [0019]-[0090] (ファミリーなし)	1-20