

## (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  
2019年7月25日(25.07.2019)



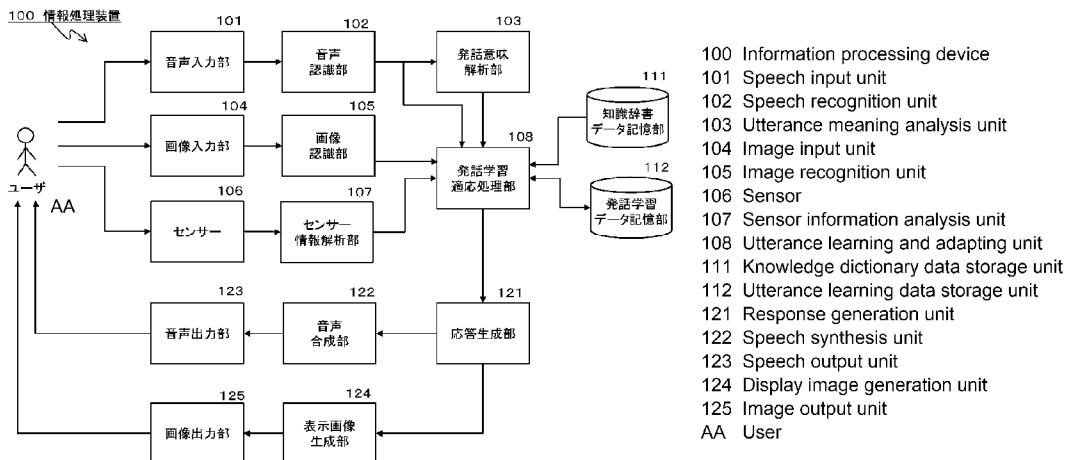
(10) 国際公開番号

WO 2019/142427 A1

- (51) 国際特許分類:  
**G10L 15/10** (2006.01)      **G06F 17/27** (2006.01)  
**G06F 3/16** (2006.01)      **G10L 15/22** (2006.01)
- (21) 国際出願番号 :      PCT/JP2018/039826
- (22) 国際出願日 :      2018年10月26日(26.10.2018)
- (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 2018-005061 2018年1月16日(16.01.2018) 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: 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). 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). 澤井 邦仁(SAWAI, Kunihito); 〒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: 宮田 正昭, 外(MIYATA, Masaaki et al.);  
 〒1040032 東京都中央区八丁堀三丁目25番9号 Daiwa 八丁堀駅前ビル西館8階 特許業務法人 大同特許事務所 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, 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,

(54) Title: INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING SYSTEM, INFORMATION PROCESSING METHOD, AND PROGRAM

(54) 発明の名称: 情報処理装置、情報処理システム、および情報処理方法、並びにプログラム



(57) Abstract: The present invention provides a device and a method capable of accurately estimating the intent of a speech uttered by a user. An utterance learning and adapting unit analyzes a plurality of user utterances input by a user to generate learning data wherein actual information of unclear intent included in a user utterance is mapped to the correct intent, and stores the learning data in a storage unit. The utterance learning and adapting unit generates learning data wherein the intent acquired from an utterance from a user in response to an utterance from the device after the input of a first user utterance of unclear intent is mapped to the actual information included in the first user utterance and is recorded. The higher-level concept of the actual information is included and recorded in the learning data. When estimating the intent of a new user utterance, learning data with a similar higher-level concept is used.

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：

一 国際調査報告（条約第21条(3)）

---

(57) 要約：ユーザ発話の高精度な意図推定を可能とした装置、方法を実現する。発話学習適応処理部が、ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報と、正しい意図を対応付けた学習データを生成して記憶部に格納する。発話学習適応処理部は、意図の不明確な第1のユーザ発話の入力後の装置発話に対するユーザからの応答発話から取得した意図を、第1のユーザ発話に含まれる実体情報に対応付けて記録した学習データを生成する。学習データには実体情報の上位意味概念情報を含めて記録する。新たなユーザ発話に対する意図推定時には、上位意味概念情報の類似する学習データを利用する。

## 明 細 書

### 発明の名称 :

情報処理装置、情報処理システム、および情報処理方法、並びにプログラム

### 技術分野

[0001] 本開示は、情報処理装置、情報処理システム、および情報処理方法、並びにプログラムに関する。さらに詳細には、ユーザ発話の音声認識処理を行う情報処理装置、情報処理システム、および情報処理方法、並びにプログラムに関する。

### 背景技術

[0002] 昨今、ユーザ発話の音声認識を行い、認識結果に基づく様々な処理や応答を行う音声対話システムの利用が増大している。

この音声認識システムにおいては、マイクを介して入力するユーザ発話を認識理解して、それに応じた処理を行う。

例えばユーザが、「明日の天気を教えて」と発話した場合、天気情報提供サーバから天気情報を取得して、取得情報に基づく応答を生成してスピーカーから出力する処理を行う。あるいは、ユーザが、「テレビを消して」と発話した場合、テレビをオフにする動作を行う。

[0003] 一般的な音声対話システムは、例えば、NLU (Natural Language Understanding) 等の自然言語理解機能を有しており、自然言語理解 (NLU) 機能を適用して、ユーザ発話の意図 (Intent) を理解する。

[0004] しかし、言い回しの個人差等により、自然言語理解 (NLU) 処理を適用してもユーザ発話の意図を正しく理解できない場合がある。

発話ユーザには、様々な方言、年代を持つユーザが含まれ、また、システムの対応言語を母国語としないノンネイティブ (non-native) の人なども含まれる。

このような様々なタイプのユーザの発話は特徴的であり、システムが理解できない場合がある。

[0005] 近年の音声認識システムでは、発話の多様性に対応するために、様々な発話データの学習を行い、学習データを用いて様々な発話の意図理解精度を高める処理が行われている。

しかし、このような学習処理を利用して、多様なユーザ発話の全てについて、正確な意図理解を行うことは困難である。

システムが発話意図を理解してくれない場合、ユーザはシステムに対する話しかけを止めてしまうことが多い。

[0006] 音声対話システムの一般的な処理は以下のように行われる。

音声入力部（マイク）から入力したユーザ発話音声を自動音声認識（A S R : A u t o m a t i c S p e e c h R e c o g n i t i o n ）機能を有する音声認識部に入力して、音声データをテキストデータに変換する。

音声認識部が生成したテキストデータを意図理解部に入力し、意図理解部が自然言語理解（N L U）処理を実行して、テキストデータから発話意図を解析する。

一般的な音声対話システムは、このような処理を実行する。

[0007] 例えば、上記処理に含まれる音声認識（A S R）処理において、ユーザ発話を正しくテキスト化できない（誤認識）場合、後段の意図理解部における自然言語理解（N L U）処理も正しく実行することができず、誤った意図理解や、意図理解不能という結果を出力することになる。

[0008] 例えば、ユーザ個人の滑舌には個人差が存在し、また自動音声認識（A S R）の性能にも限界がある。従って、自動音声認識（A S R）処理や自然言語理解（N L U）処理におけるエラー（誤認識）を完全に防止することは困難である。

[0009] なお、音声認識システムの意図理解精度を高める構成を開示した従来技術として、例えば特許文献1（国際公開WO 2016/151699号公報）や、特許文献2（特表2017-516153号公報）がある。

[0010] 特許文献1は、発話から得られたテキストデータの構成要素に基づくベクトル表現データを生成して、様々なテキストデータに基づくベクトル表現データとの類似度判定を行うことでユーザ発話の意図理解精度を向上させる構成を開示している。

特許文献2は、音声認識システムがユーザに対して複数の発話を実行させて、第1発話と第2発話との意味論的なつながりに基づいて、ユーザ発話の意図把握の精度を高める構成を開示している。

## 先行技術文献

### 特許文献

[0011] 特許文献1：国際公開WO 2016/151699号公報

特許文献2：特表2017-516153号公報

## 発明の概要

### 発明が解決しようとする課題

[0012] 上述したように、ユーザ発話の意図理解精度を向上させるための構成として、複数の技術が開示されている。

しかし、これらの従来技術の構成は、ユーザ固有の独特な言い回しに対する正しい意図理解を実現するには不十分である。また、ユーザが音声認識システム側に解釈の修正を求めるといった処理を実行してユーザ発話の正しい意図理解を行うといったこともできない。

[0013] 本開示は、例えば、上記問題点に鑑みてなされたものであり、ユーザ発話の意図理解を、より確実に行う構成を実現する情報処理装置、情報処理システム、および情報処理方法、並びにプログラムを提供することを目的とする。

## 課題を解決するための手段

[0014] 本開示の第1の側面は、

ユーザ発話の意図（インテント）を解析する発話学習適応処理部を有し、前記発話学習適応処理部は、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する情報処理装置にある。

- [0015] さらに、本開示の第2の側面は、  
ユーザ端末と、データ処理サーバを有する情報処理システムであり、  
前記ユーザ端末は、  
ユーザ発話を入力する音声入力部を有し、  
前記データ処理サーバは、  
前記ユーザ端末から受信する前記ユーザ発話の意図（インテント）を解析する発話学習適応処理部を有し、  
前記発話学習適応処理部は、  
ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する情報処理システムにある。
- [0016] さらに、本開示の第3の側面は、  
情報処理装置において実行する情報処理方法であり、  
音声入力部が、ユーザ発話を入力し、  
発話学習適応処理部が、前記ユーザ発話の意図（インテント）を解析する発話学習適応処理を実行し、  
前記発話学習適応処理において、  
ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する処理を実行する情報処理方法にある。
- [0017] さらに、本開示の第4の側面は、  
ユーザ端末と、データ処理サーバを有する情報処理システムにおいて実行する情報処理方法であり、  
前記ユーザ端末が、  
ユーザ発話を入力する音声入力処理を実行し、

前記データ処理サーバが、

前記ユーザ端末から受信する前記ユーザ発話の意図（インテント）を解析する発話学習適応処理を実行し、

前記発話学習適応処理において、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する処理を実行する情報処理方法にある。

[0018] さらに、本開示の第5の側面は、

情報処理装置において情報処理を実行させるプログラムであり、

音声入力部に、ユーザ発話を入力させる音声入力ステップと、

発話学習適応処理部に、前記ユーザ発話の意図（インテント）を解析する発話学習適応処理を実行させ、

前記発話学習適応処理において、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する処理を実行させるプログラムにある。

[0019] なお、本開示のプログラムは、例えば、様々なプログラム・コードを実行可能な情報処理装置やコンピュータ・システムに対して、コンピュータ可読な形式で提供する記憶媒体、通信媒体によって提供可能なプログラムである。このようなプログラムをコンピュータ可読な形式で提供することにより、情報処理装置やコンピュータ・システム上でプログラムに応じた処理が実現される。

[0020] 本開示のさらに他の目的、特徴や利点は、後述する本開示の実施例や添付する図面に基づくより詳細な説明によって明らかになるであろう。なお、本明細書においてシステムとは、複数の装置の論理的集合構成であり、各構成の装置が同一筐体内にあるものには限らない。

## 発明の効果

[0021] 本開示の一実施例の構成によれば、ユーザ発話の高精度な意図推定を可能

とした装置、方法が実現される。

具体的には、例えば、発話学習適応処理部が、ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報と、正しい意図を対応付けた学習データを生成して記憶部に格納する。発話学習適応処理部は、意図の不明確な第1のユーザ発話の入力後の装置発話に対するユーザからの応答発話から取得した意図を、第1のユーザ発話に含まれる実体情報に対応付けて記録した学習データを生成する。学習データには実体情報の上位意味概念情報を含めて記録する。新たなユーザ発話に対する意図推定時には、上位意味概念情報の類似する学習データを利用する。

本構成により、ユーザ発話の高精度な意図推定を可能とした装置、方法が実現される。

なお、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、また付加的な効果があってもよい。

### 図面の簡単な説明

[0022] [図1]ユーザ発話を認識して応答を行う情報処理装置の具体的な処理例について説明する図である。

[図2]情報処理装置の構成例と利用例について説明する図である。

[図3]情報処理装置の具体的な構成例について説明する図である。

[図4]情報処理装置の実行する処理について説明するフローチャートを示す図である。

[図5]「学習処理A＝ユーザ発話意図の学習対話フロー」の具体例について説明する図である。

[図6]具体的な学習データのデータ構成例について説明する図である。

[図7]ユーザ発話に基づく「学習処理」と、学習処理結果を利用したユーザ発話の認識、意図推定処理の具体例について説明する図である。

[図8]ユーザ発話に基づく「学習処理」と、学習処理結果を利用したユーザ発話の認識、意図推定処理の具体例について説明する図である。

[図9]ユーザ発話実行時の対話状態（ステート）を区分し、区分ステート単位

の個別の学習データを生成して記録する構成例について説明する図である。

[図10]学習データ忘却スコア  $f(t, c)$  について説明する図である。

[図11]「学習処理 B = ユーザ発話意図の修正対話フロー」の具体例について説明する図である。

[図12]学習データを用いたユーザ発話の意図（インテント）と実体情報（エンティティ）の推定処理の詳細シーケンスについて説明するフローチャートを示す図である。

[図13]学習データを用いたユーザ発話の意図（インテント）と実体情報（エンティティ）の推定処理の詳細シーケンスについて説明するフローチャートを示す図である。

[図14]ユーザからの言い換えによる発話学習フローの例について説明する図である。

[図15]ユーザからの言い換えによる発話学習を行う例について説明する図である。

[図16]学習時にユーザに何を学習した（している）のかを明示して伝えるための表示例について説明する図である。

[図17]学習時にユーザに何を学習した（している）のかを明示して伝えるための表示例について説明する図である。

[図18]本開示の情報処理装置の状態遷移を説明する図である。

[図19]情報処理システムの構成例について説明する図である。

[図20]情報処理装置のハードウェア構成例について説明する図である。

## 発明を実施するための形態

[0023] 以下、図面を参照しながら本開示の情報処理装置、情報処理システム、および情報処理方法、並びにプログラムの詳細について説明する。なお、説明は以下の項目に従って行なう。

1. 情報処理装置の構成例について
2. 発話学習適応処理部が実行する学習対話フローの詳細について
3. 様々な意味を有する同一発話データである多義語に対する処理について

て

4. 誤った学習処理に対する対策について
5. 学習データの自動忘却処理について
6. 発話学習適応処理部が実行する学習対話フローの詳細について
7. 学習データを適用したユーザ発話の意図推定処理の詳細について
8. その他の実施例について
9. 本開示の情報処理装置の状態遷移と処理の効果について
10. 情報処理装置、および情報処理システムの構成例について
11. 情報処理装置のハードウェア構成例について
12. 本開示の構成のまとめ

[0024] [1. 情報処理装置の構成例について]

まず、図1以下を参照して、本開示の一実施例の情報処理装置の構成例について説明する。

[0025] 図1は、ユーザ1の発するユーザ発話を認識して応答を行う情報処理装置10の一処理例を示す図である。

情報処理装置10は、ユーザの発話、例えば、  
ユーザ発話=「大阪の明日、午後の天気を教えて」  
このユーザ発話の音声認識処理を実行する。

[0026] さらに、情報処理装置10は、ユーザ発話の音声認識結果に基づく処理を実行する。

図1に示す例では、ユーザ発話=「大阪の明日、午後の天気を教えて」に応答するためのデータを取得し、取得データに基づいて応答を生成して生成した応答を、スピーカー14を介して出力する。

図1に示す例では、情報処理装置10は、以下の装置応答を行っている。  
装置応答=「大阪の明日、午後の天気は晴れですが、夕方、にわか雨がある可能性があります。」

[0027] 情報処理装置10は、装置内の記憶部から取得した知識データ、またはネットワークを介して取得した知識データを利用して応答を生成して出力する

。

図1に示す情報処理装置10は、カメラ11、マイク12、表示部13、スピーカー14を有しており、音声入出力と画像入出力が可能な構成を有する。

[0028] 図1に示す情報処理装置10は、例えばスマートスピーカーあるいはエージェント機器と呼ばれる。

本開示の情報処理装置10は、図2に示すように、エージェント機器10aに限らず、スマホ10bやPC10c等のような様々な装置形態とすることが可能である。

[0029] 情報処理装置10は、ユーザ1の発話を認識して、ユーザ発話に基づく応答を行う他、例えば、ユーザ発話に応じて図2に示すテレビ、エアコン等の外部機器30の制御も実行する。

例えばユーザ発話が「テレビのチャンネルを1に変えて」、あるいは「エアコンの設定温度を20度にして」といった要求である場合、情報処理装置10は、このユーザ発話の音声認識結果に基づいて、外部機器30に対して制御信号(Wi-Fi、赤外光など)を出力して、ユーザ発話に従った制御を実行する。

[0030] なお、情報処理装置10は、ネットワークを介してサーバ20と接続され、サーバ20から、ユーザ発話に対する応答を生成するために必要となる情報を取得することが可能である。また、音声認識処理や意味解析処理をサーバに行わせる構成としてもよい。

[0031] 次に、図3を参照して、情報処理装置の具体的な構成例について説明する。

。

図3は、ユーザ発話を認識して応答を行う情報処理装置100の一構成例を示す図である。

[0032] 図3に示すように、情報処理装置100は、音声入力部101、音声認識部102、発話意味解析部103、画像入力部104、画像認識部105、センサー106、センサー情報解析部107、発話学習適応処理部108、

知識辞書データ記憶部 111、発話学習データ記憶部 112、応答生成部 121、音声合成部 122、音声出力部 123、表示画像生成部 124、画像出力部 125 を有する。

なお、これらの構成要素は、全てを 1 つの情報処理装置 100 内部に構成することも可能であるが、一部の構成や機能を他の情報処理装置や外部サーバに備える構成としてもよい。

[0033] ユーザの発話音声はマイクなどの音声入力部 101 に入力される。

音声入力部（マイク）101 は、入力したユーザ発話音声を音声認識部 102 に入力する。

音声認識部 102 は、例えば ASR (A u t o m a t i c S p e e c h R e c o g n i t i o n) 機能を有し、音声データを複数の単語から構成されるテキストデータに変換する。

[0034] 音声認識部 102において生成されたテキストデータは、発話意味解析部 103 に入力される。

発話意味解析部 103 は、テキストに含まれるユーザの意図候補を選択して出力する。

発話意味解析部 103 は、例えば、NLU (N a t u r a l L a n g u a g e U n d e r s t a n d i n g) 等の自然言語理解機能を有し、テキストデータから、ユーザ発話の意図（インテント： I n t e n t ）や、発話に含まれる意味のある要素（有意要素）である実体情報（エンティティ： E n t i t y ）を推定する。

[0035] 具体例について説明する。例えば以下のユーザ発話が入力されたとする。

ユーザ発話＝明日の大阪の午後の天気を教えて

このユーザ発話の、

意図（インテント）は、天気を知りたいであり、

実体情報（エンティティ）は、大阪、明日、午後、これらのワードである

。

[0036] ユーザ発話から、意図（エンティティ）と、実体情報（エンティティ）を

正確に推定、取得することができれば、情報処理装置 100 は、ユーザ発話に対する正確な処理を行うことができる。

例えば、上記の例では、明日の大坂の午後の天気を取得して、応答として出力することができる。

[0037] ただし、実際には、ユーザ発話の滑舌が悪い、方言が含まれるといった場合があり、このような場合には、ユーザ発話の意図（エンティティ）と、実体情報（エンティティ）を正確に推定、取得することができない場合がある。

[0038] 本開示の情報処理装置 100 は、このような事態の発生を低減させるための構成を有する。

その一つが、発話学習適応処理部 108 である。

[0039] 発話学習適応処理部 108 には、以下の各データが入力される。

(1) 音声認識部 102 が、ユーザ発話に基づいて生成したテキストデータ、

(2) 発話意味解析部 103 が、テキストデータに対する自然言語理解（NLU）を実行して生成したユーザ発話の意図（インテント）と実体情報（エンティティ）、

[0040] また、発話学習適応処理部 108 には、以下の各情報も入力される。

(3) カメラなどの画像入力部 104 が取得した発話ユーザおよびその周囲の画像に対する画像認識部 105 の画像認識結果情報、

(4) センサー 106 が取得した発話ユーザおよびその周囲の状態の検出情報に基づいて、センサー情報解析部 107 が解析したセンサー解析情報、

[0041] 発話学習適応処理部 108 は、上記 (1) ~ (4) の情報を入力して、ユーザ発話の学習データの生成、修正、破棄処理を含むユーザ発話の学習処理を実行する。

発話学習適応処理部 108 において実行された学習処理の結果データである発話学習データは、発話学習データ記憶部 112 に格納される。

この学習処理の詳細については後述する。

[0042] 発話学習データ記憶部112に所定の学習データが記憶された後においては、発話学習適応処理部108は、新たに入力するユーザ発話に対する学習データ適応処理を行う。すなわち、ユーザ発話対応のテキストデータを新たに入力した場合、発話学習データ記憶部112に格納された学習データを参照して、ユーザ固有の発話に対する意図（インテント）と実体情報（エンティティ）を推定する。

この推定処理は、学習データを利用することで、より高精度な処理として実行することができる。

[0043] なお、発話学習適応処理部108が実行するユーザ発話の学習データ生成処理や、適応処理においては、ユーザ発話を構成する単語の上位概念などを参照するために知識辞書データ記憶部111に格納された知識辞書データが利用される。

知識辞書データ記憶部111には、様々な事実や、常識、習慣、経験等の一般的な知識をコンピュータ可読形式のデータとしてデータベース化した知識辞書データが格納されている。知識辞書データは、ナレッジベース辞書データとも呼ばれる。

[0044] このように、発話学習適応処理部108は、発話学習データ記憶部112に格納された学習データや、知識辞書データ記憶部111に格納された知識辞書データを用いて、ユーザ固有の発話に対する意図（インテント）と実体情報（エンティティ）を推定する。

この推定結果は、応答生成部121に入力される。

[0045] 応答生成部121は、発話学習適応処理部108において推定されたユーザ発話の意図（インテント）と実体情報（エンティティ）に基づいて、ユーザに対する応答を生成する。応答は、音声、または画像の少なくともいずれかによって構成される。

応答音声を出力する場合は、音声合成部122において生成した音声情報が、スピーカー等の音声出力部123を介して出力される。

応答画像を出力する場合は、表示画像合成部124において生成した表示

画像情報が、ディスプレイ等の画像出力部 125 を介して出力される。

[0046] 次に、図 4 に示すフローチャートを参照して、発話学習適応処理部 108 の実行する処理のシーケンスについて説明する。

図 4 に示すフローチャートに従った処理は、例えば情報処理装置 100 の記憶部に格納されたプログラムに従って実行することが可能である。

発話学習適応処理部 108 における処理は、例えばプログラム実行機能を有する CPU 等のプロセッサによるプログラム実行処理として行うことができる。

以下、図 4 に示すフローの各ステップの処理について説明する。

[0047] (ステップ S101)

まず、ステップ S101において、発話学習適応処理部 108 は、ユーザ発話に対して発話意味解析部 103 が実行した自然言語理解 (NLU) 結果を取得し、発話学習データ記憶部 112 に格納された学習データを適用して、ユーザ発話の意図 (インテント) と実体情報 (エンティティ) を推定する。

[0048] なお、先に説明したように、発話学習適応処理部 108 には、ユーザ発話に対する以下の解析データが入力されている。

(1) 音声認識部 102 が、ユーザ発話に基づいて生成したテキストデータ、

(2) 発話意味解析部 103 が、テキストデータに対する自然言語理解 (NLU) を実行して生成したユーザ発話の意図 (インテント) と実体情報 (エンティティ) 、

(3) カメラなどの画像入力部 104 が取得した発話ユーザおよびその周囲の画像に対する画像認識部 105 の画像認識結果情報、

(4) センサー 106 が取得した発話ユーザおよびその周囲の状態の検出情報に基づいて、センサー情報解析部が解析したセンサー解析情報、

[0049] 発話学習適応処理部 108 は、ユーザ発話に対応するテキストデータに対して、発話学習データ記憶部 112 に格納済みの発話学習データから適用で

きるデータが存在するか検索する。

なお、この学習データ適応処理の詳細については後述する。

[0050] 適用できるデータが発話学習データ記憶部112に格納済みの発話学習データに存在する場合は、学習データを利用して、ユーザ発話の意図（インテント）と実体情報（エンティティ）を推定する。

[0051] 一方、適用できるデータが発話学習データ記憶部112に格納済みの発話学習データに存在しない場合は、発話意味解析部103がテキストデータに対する自然言語理解（NLU）を実行して生成したユーザ発話の意図（インテント）と実体情報（エンティティ）を、そのままユーザ発話の意図と実体情報とする。

なお、初期状態で、学習データ自体が存在しない場合も、同様の処理を行う。

[0052] このステップS101の意図推定処理に際しては、発話学習データ記憶部112に格納された学習データを適用する。

この学習データの生成、修正、破棄等の処理は、次に説明するステップS103やステップS104において実行される。

すなわち、発話学習データ記憶部112に格納される学習データは、次に説明するステップS103やステップS104において実行される学習処理により、逐次更新されることになる。

[0053] ステップS101における処理は、発話学習データ記憶部112に格納された最新の学習データを用いて実行される。

この最新の学習データを用いたユーザ発話の意図（インテント）と実体情報（エンティティ）の推定処理の詳細シーケンスについては、ステップS103やステップS104において実行される学習処理の詳細説明を行った後に、図12、図13に示すフローチャートを参照して説明する。

[0054] (ステップS102)

ステップS102において、発話学習適応処理部108は、

ステップS101で実行したユーザ発話の意図（インテント）と実体情報

(エンティティ) の推定処理の結果を以下の 3 つに分類する。

(推定結果 1) ユーザ発話は、意図解釈不可（OOD）発話である。

(推定結果 2) ユーザ発話は、意図解釈間違指摘発話である。

(推定結果 3) ユーザ発話は、情報処理装置 100 への依頼、応答発話である。

[0055] なお、ユーザ発話の意図（インテント）や実体情報（エンティティ）を理解できなかった発話を、意図解釈不可 [OOD : Out Of Domain (アウト オブ ドメイン)] 発話と呼ぶ。

上記（推定結果 1）に示す「意図解釈不可（OOD）発話」とは、ユーザ発話の意図（インテント）や実体情報（エンティティ）を理解できなかったことを意味する。

[0056] 発話意味解析部 103 が実行した自然言語理解（NLU）によってユーザ発話の意図（インテント）や実体情報（エンティティ）が推定できなかった場合も、ユーザ発話は、（推定結果 1）「意図解釈不可（OOD）発話」に分類される。

発話意味解析部 103 が実行した自然言語理解（NLU）によって推定したユーザ発話の意図（インテント）や実体情報（エンティティ）に設定される属性情報としての信頼度スコアが既定のしきい値より低い場合も同様である。

[0057] ステップ S101において実行されたユーザ発話の意図（インテント）と実体情報（エンティティ）の推定処理の結果が、

(推定結果 1) ユーザ発話は、意図解釈不可（OOD）発話である。

と判定された場合は、ステップ S103 に進む。

また、

(推定結果 2) ユーザ発話は、意図解釈間違指摘発話である。

と判定された場合は、ステップ S104 に進む。

さらに、

(推定結果 3) ユーザ発話は、情報処理装置 100 への依頼、応答発話で

ある。

と判定された場合は、ステップS105に進む。

[0058] (ステップS103)

上述したように、ステップS101において実行されたユーザ発話の意図（インテント）と実体情報（エンティティ）の推定処理の結果が、  
(推定結果1) ユーザ発話は、意図解釈不可（OOD）発話である。

と判定された場合は、ステップS103に進む。

[0059] この場合、発話学習適応処理部108は、ステップS103において、「学習処理A」を実行する。

この学習処理Aは、ユーザ発話意図の学習対話フローの実行処理である。

この学習対話フローの実行により、ユーザ発話の真の意図を獲得すると学習データを生成して、発話学習データ記憶部112に記憶する。

ステップS103において実行する「学習処理A=ユーザ発話意図の学習対話フロー」の詳細については後述する。

[0060] (ステップS104)

また、ステップS101において実行されたユーザ発話の意図（インテント）と実体情報（エンティティ）の推定処理の結果が、

(推定結果2) ユーザ発話は、意図解釈間違い指摘発話である。

と判定された場合は、ステップS104に進む。

[0061] この場合、発話学習適応処理部108は、ステップS104において、「学習処理B」を実行する。

この学習処理Bは、ユーザ発話意図の修正対話フローの実行処理である。

このユーザ発話意図の修正対話フローの実行により、ユーザ発話の真の意図を獲得すると、発話学習データ記憶部112に記録する学習データの生成、または発話学習データ記憶部112に格納済みの学習データの修正を行う。  
。

なお、このフローの実行によってユーザ発話の真の意図を得られなかった場合には前回のユーザ発話に対する学習データは破棄される。

このステップS104において実行する「学習処理B＝ユーザ発話意図の修正対話フロー」の詳細については後述する。

[0062] (ステップS105)

さらに、ステップS101において実行されたユーザ発話の意図（インテント）と実体情報（エンティティ）の推定処理の結果が、

(推定結果3) ユーザ発話は、情報処理装置への依頼、応答発話である。  
と判定された場合は、ステップS105に進む。

[0063] この場合、発話学習適応処理部108は、ステップS105において、ユーザ発話意図に応じた処理や応答処理を実行する。

なお、ステップS103の「学習処理A＝ユーザ発話意図の学習対話フロー」の実行、あるいはステップS104における「学習処理B＝ユーザ発話意図の修正対話フロー」の実行によって、ユーザ発話の真の意図を獲得できた場合にも、ステップS105に進み、ユーザ発話意図に応じた処理や応答処理を実行する。

[0064] 具体的には、例えば、以下のようなユーザ発話がつあった場合、  
ユーザ発話＝大阪の明日、午後の天気を教えて  
情報処理装置100は、音声出力部123を介して大阪の明日、午後の天  
気情報を音声出力する。

[0065] [2. 発話学習適応処理部が実行する学習対話フローの詳細について]  
次に、図4のフローチャートのステップS103において、発話学習適応処理部108が実行する「学習処理A＝ユーザ発話意図の学習対話フロー」について説明する。

このステップS103では、学習データの生成処理や学習データの適応処理が行われる。

[0066] 先に図4を参照して説明したように、図4のフローチャートのステップS103の「学習処理A＝ユーザ発話意図の学習対話フロー」は、情報処理装置100が、ユーザ発話意図を解釈することができなかった（OOD発話の）場合に実行される。

- [0067] 図5を参照して、ステップS103において実行する「学習処理A＝ユーザ発話意図の学習対話フロー」の具体例について説明する。
- [0068] 図5に示すように、ユーザは、まず、ステップS121において、鈴木さんから来ているからメッセージをシステム（情報処理装置100）から読んでもらう意図で、  
「鈴木さんの読んで」と発話する。
- [0069] 情報処理装置100は、音声認識処理に際して「読んで」を「4で」とテキスト化してしまい、ユーザ発話の意図が理解できないと判定する。  
すなわち、ステップS122において、ユーザ発話の自然言語理解（NLU）結果を「意図解釈不可（OOD）」と判定する。
- [0070] なお、図5に示す例では音声認識エンジンの誤認識の例を示しているが、ユーザ発話が方言やスラング、ノンネイティブ（Non-Native）のユーザによるカタコト発話などの場合も、情報処理装置100は、ユーザ発話を意図解釈不可（OOD）と判定する場合がある。
- [0071] 情報処理装置100の発話学習適応処理部108は、このようにユーザ発話を意図解釈不可（OOD）と判定した場合、先に説明した図4のステップS103において「学習処理A＝ユーザ発話意図の学習対話フロー」を実行する。
- [0072] なお、ユーザ発話は文法など言語的に正しい発話になっている必要は無く、ユーザ固有の言い回しであっても、ユーザ発話を意図解釈不可（OOD）と判定した場合には、この「学習処理A＝ユーザ発話意図の学習対話フロー」を実行する。  
ここで実行する学習処理では、ユーザ発話の発話文言（発話テキスト）そのものをユーザ発話の意図（インテント）に紐づけて、学習データとして記憶部（発話学習データ記憶部112）に格納するからである。
- [0073] 図5に示すステップS122において、情報処理装置100の発話学習適応処理部108は、ユーザ発話を意図解釈不可（OOD）と判定すると、ユ

ーザの発話意図を獲得するために、発話ユーザに対する問い合わせを行う。すなわち、意図解釈不可（OOD）と判定した「ユーザのOOD発話内容」についての問い合わせ文を、音声合成処理（TTS：Text → Speech）により生成して、

「”ユーザのOOD発話内容”の意味を教えてください」

上記のような問い合わせ応答を出力する。

[0074] なお、意図解釈不可（OOD）と判定した場合に情報処理装置100が、

ユーザ発話意図を獲得するための問い合わせ文言としては、

「言い換えてください」

「分からなかったので、違う言い方で言ってください」

など、様々な文言の利用が可能である。

図5に示す例では、情報処理装置100は、ステップS123において、

「「鈴木さんの4で」の意味を教えてください」

上記の問い合わせ応答を出力している。

[0075] 情報処理装置100の発話学習適応処理部108は、この問い合わせ実行

後、予め規定した時間内（例えば5秒以内）に実行されたユーザ発話（図5の例ではステップS124のユーザ発話＝「鈴木さんのメッセージを読み上げて」）を解析する。

具体的には、ユーザ発話「鈴木さんのメッセージを読み上げて」の自然言語理解（NLU）結果（または学習データの適用結果）を「鈴木さんの4で」の真の発話意図として記憶する。

すなわち、学習データとして、発話学習データ記憶部112に格納する。

[0076] ただし、上記のステップS123の問い合わせ実行後、予め規定した時間

内（例えば5秒以内）にユーザ発話が無かった場合は、学習対話フローを終了する。すなわち、学習データ生成を行わない。

[0077] また、問い合わせ後のユーザの再発話も、意図解釈不可（OOD）と判定

された場合には、再度、ユーザ発話意図獲得のための問い合わせを行い、真の発話意図が獲得できるまで繰り返す。

繰り返しは、予め規定した上限回数まで行い、上限階数に達した場合には学習対話フローを終了するようにしてもよい。

また、繰り返しが発生し、真の発話意図が獲得できた場合には、最初の意図解釈不可（OOD）と判定したユーザ発話のみを学習データとして記憶部に格納（ユーザ発話＝繰り返しにより獲得できた真の発話意図として記録）してもよいし、問い合わせの繰り返しによって得られた全てのユーザ発話を学習データとして記憶部に格納（ユーザ発話＝繰り返しにより獲得できた真の発話意図として記録）してもよい。

- [0078] 図5の例では、情報処理装置100が、ステップS123において、「「鈴木さんの4で」の意味を教えてください」上記の問い合わせを出力し、その後、規定時間（5秒）以内のユーザ発話としてステップS124の発話、すなわち、「鈴木さんのメッセージを読み上げて」というユーザ発話がなされている。

- [0079] 情報処理装置100は、このステップS124のユーザ発話、すなわち、「鈴木さんのメッセージを読み上げて」このユーザ発話の自然言語理解（NLU）結果（または学習データの適用結果）を先に意図解釈不可（OOD）と判定したユーザ発話「鈴木さんの4で」の真の発話意図として記憶する。

すなわち、学習データとして、発話学習データ記憶部112に格納する。

- [0080] 情報処理装置100は、図5のステップS125において、ステップS125に示すデータを学習データをとして発話学習データ記憶部112に格納する。

- [0081] 図5のステップS125に示すデータは、ユーザ発話の意図（インテント）と、発話に含まれる意味のある要素（有意要素）である実体情報（エンティティ）を記録したデータである。このデータが発話テキスト（発話文言）「鈴木さんの4で」に対応付けて発話学習データ記憶部112に記録される。

[0082] すなわち、情報処理装置100の発話学習適応処理部108は、図5に示すステップS121～S125の処理により、図5の下に示す(a)学習処理を行うことになる。すなわち、

ユーザ発話「鈴木さんの4で」は、

意図(インテント)=メッセージ読み上げ(R E A D \_ M E S S A G E)

、

実体情報(エンティティ)=鈴木

これらの意図と実体情報を有するユーザ発話であることの学習データを生成して発話学習データ記憶部112に記録する。

[0083] この学習処理の後、情報処理装置100の発話学習適応処理部108は、ステップS126において、応答生成部121等を適用してユーザ発話に対応した処理、すなわち、

鈴木さんのメッセージの読み上げ処理を実行する。

[0084] 図5(a)には、学習データの一例として、

意図(インテント)=メッセージ読み上げ(R E A D \_ M E S S A G E)

、

実体情報(エンティティ)=鈴木

これらの意図、実体情報データの構成例を示している。

学習データは、このように、「ユーザ発話」と「意図、実体情報データ」との対応データとなる。

[0085] 発話学習データ記憶部112に記録する具体的な学習データのデータ構成例について、図6を参照して説明する。

図6(a)学習データ例は、1つのユーザ発話に対して設定される学習データの例である。

図6(b)実体情報(エンティティ)構成例は、図6(a)学習データの構成データである実体情報(エンティティ)の詳細データ例である。

[0086] 図6(a)学習データに示すように、学習データは、以下の構成データを有する。

- (1) ユーザ発話テキスト (S p e e c h)
- (2) 意図 (インテント)
- (3) 実体情報 (エンティティ)
- (4) その値の属性情報 (ユーザ発話日時 (d a t e)、ユーザ発話の回数 (S p e e c h C o u n t))

[0087] 図6 (a) に示すデータは、

- (1) ユーザ発話テキスト (S p e e c h) = 「桑原圭祐にして」  
このユーザ発話に対応付けられた学習データである。  
このユーザ発話テキスト (S p e e c h) = 「桑原圭祐にして」の、
- (2) 意図 (インテント) は、音楽再生 (P L A Y \_ M U S I C) であり、  
(3) 実体情報 (エンティティ) は、図6 (b) に示す各データであることが学習データとして記録される。  
なお、(4) その値の属性情報 (ユーザ発話日時 (d a t e)、ユーザ発話の回数 (S p e e c h C o u n t))  
についても記録されることになる。

[0088] 学習データに含まれる「実体情報 (エンティティ)」としては、例えば、図6 (b) に示す各データによって構成される。

- なお、図6 (b) に示すデータは、ユーザ発話に含まれる1つのエンティティ = 「桑原圭祐」に関する情報である。  
ユーザ発話に複数のエンティティがある場合、各エンティティについて、図6 (b) に示すデータが登録される。

[0089] 図6 (b) に示すように、1つのエンティティ = 「桑原圭祐」に関する情報として以下のデータが記録される。

- (1) エンティティのユーザ発話文字 (L i t e r a l) = 「桑原圭祐」
- (2) エンティティのタイプ (t y p e) = 再生対象 (T A R G E T \_ M U S I C)
- (3) エンティティの自然言語処理情報 (n l p l n f o) として、

- (3 a) エンティティを含む文節の句の種別：名詞句
- (3 b) エンティティを含む文節の格（主格、時間格、場所格など 5 W 1 H や修飾節を表す）：連用修飾節
- (3 c) エンティティを含む文節の言語的な意味判定結果：対象、依頼、願望
- (3 d) エンティティ単語の品詞と種別：名詞：一般名詞：固有名詞：名：人物
- (3 e) エンティティ単語の上位意味概念：コンテンツ：音楽：人物

[0090] 図 6 (a) 学習データの実体情報（エンティティ）記録領域には、この図 6 (b) に示すように、各エンティティ単位で様々な実体情報（エンティティ）が記録される。

なお、これらの実体情報（エンティティ）に記録するための情報は、例えば、知識辞書データ記憶部 111 に格納された辞書データから取得可能である。

[0091] このように、図 6 に示す学習データ、すなわち、

- (1) ユーザ発話テキスト (S p e e c h)
- (2) 意図 (インテント)
- (3) 実体情報 (エンティティ)
- (4) その値の属性情報 (ユーザ発話日時 (d a t e)、ユーザ発話の回数 (S p e e c h C o u n t))

これらの構成データからなる学習データが、発話学習適応処理部 108 において生成されて、発話学習データ記憶部 112 に格納される。

[0092] 例えば、図 4 に示すステップ S 103 において実行される学習処理について説明した図 5 (a) に示すデータの例は、発話学習データ記憶部 112 に格納される学習データの一部のみのデータである。

図 5 に示す例においても、図 6 を参照して説明したと同様の構成データからなる学習データが生成されて記憶部に格納される。例えば、以下のデータからなる学習データである。

- (1) ユーザ発話テキスト (S p e e c h) = 鈴木さんの4で
- (2) 意図 (インテント) = R E A D \_ M E S S A G E
- (3) 実体情報 (エンティティ) = 鈴木
- (4) その値の属性情報 (ユーザ発話日時 (d a t e) 、ユーザ発話の回数 (S p e e c h C o u n t))

図4に示すステップS 1 0 3の学習処理Aでは、これらの全てのデータを有する学習データが発話学習適応処理部1 0 8において生成されて、発話学習データ記憶部1 1 2に格納される。

[0093] 次に、図7、図8を参照して、あるユーザ発話に基づく「学習処理」と、学習処理結果を利用したユーザ発話の認識、意図推定処理、すなわち学習結果データの「適応処理」の具体例について説明する。

[0094] 図7は、あるユーザ発話に基づく「学習処理」の具体例を説明する図である。

図7に示すように、ユーザは、まず、ステップS 1 3 1において、ミュージシャンである桑原圭祐の曲を聴きたいという意図で、「桑原圭祐にして」と発話する。

[0095] 情報処理装置1 0 0は、「桑原圭祐にして」の意図が理解できないと判定する。

すなわち、ステップS 1 3 2において、ユーザ発話の自然言語理解 (N L U) 結果を「意図解釈不可 (O O D)」と判定する。

[0096] 情報処理装置1 0 0の発話学習適応処理部1 0 8は、このようにユーザ発話を意図解釈不可 (O O D) と判定した場合、先に説明した図4のステップS 1 0 3において「学習処理A=ユーザ発話意図の学習対話フロー」を実行する。

[0097] 図7に示すステップS 1 3 2において、情報処理装置1 0 0の発話学習適応処理部1 0 8は、ユーザ発話を意図解釈不可 (O O D) と判定すると、ユーザの発話意図を獲得するために、発話ユーザに対する問い合わせを行う。

すなわち、ステップS133において、意図解釈不可（OOD）と判定した「ユーザのOOD発話内容」についての問い合わせ文を、音声合成処理（TTS：Text To Speech）により生成して、  
 「「桑原圭祐にして」の意味を教えてください」  
 上記の問い合わせ応答を出力する。

- [0098] 情報処理装置100の発話学習適応処理部108は、この問い合わせ実行後、予め規定した時間内（例えば5秒以内）に実行されたユーザ発話、図7の例ではステップS134のユーザ発話、すなわち、  
 「桑原圭祐の曲を再生して」  
 このユーザ発話の自然言語理解（NLU）結果（または学習データの適用結果）を、最初のOOD判定とされたユーザ発話、すなわち、  
 「桑原圭祐にして」  
 このユーザ発話の意図として記憶する。

- [0099] すなわち、発話学習適応処理部108は、  
 「桑原圭祐の曲を再生して」というユーザ発話（教師発話）から、真の意図、  
 意図（インテント）＝音楽再生（PLAY\_MUSIC）、  
 実体情報（エンティティ）＝桑原圭祐、  
 これらのデータを得る。

- [0100] この実体情報（エンティティ）＝桑原圭祐、  
 このデータが、ステップS131の「意図解釈不可（OOD）発話」と判定されたユーザ発話「桑原圭祐にして」に含まれることから、学習データにはOOD発話を、知識辞書を用いて自然言語処理した”桑原圭祐”の部分の結果としての【学習意味概念A】が記憶される。

- [0101] 図7に示すように、【学習意味概念A】は、  
 エンティティの自然言語処理情報（nlpinfo）、すなわち、  
 （a）エンティティを含む文節の句の種別：名詞句  
 （b）エンティティを含む文節の格（主格、時間格、場所格など5W1H

や修飾節を表す）：連用修飾節

（c）エンティティを含む文節の言語的な意味判定結果：対象、依頼、願望

（d）エンティティ単語の品詞と種別：名詞：一般名詞：固有名詞：名：人物

（e）エンティティ単語の上位意味概念：コンテンツ：音楽：人物  
これらの情報の少なくともいずれかを含む。

[0102] これらのエンティティの自然言語処理情報（n | p | n f o）を〔学習意味概念A〕として、意図（インテント）に対応付けた学習データを生成して、発話学習データ記憶部112に格納する。

[0103] なお（e）エンティティ単語の上位意味概念とは、実体情報（エンティティ）の種類を示す情報である。具体的には、例えば、  
実体情報（エンティティ）単語が、音楽関係のアーティストを示す単語、  
実体情報（エンティティ）単語が、映画タイトルを示す単語、  
実体情報（エンティティ）単語が、ゲームタイトルを示す単語、  
実体情報（エンティティ）単語が、都市名を示す単語、  
等、実体情報（エンティティ）単語の属するカテゴリー等を識別可能とした情報である。

[0104] なお、この上位意味概念は、知識辞書データ記憶部111に格納された知識辞書データから取得できる。

このように、発話学習適応処理部108は、実体情報（エンティティ）の上位意味概念を学習データとして、発話学習データ記憶部112に格納する。このようなデータを学習データに含めて記録することで、発話学習適応処理部108は、その後の新たなユーザ発話の意図解析処理に際して、その新たなユーザ発話に含まれる実体情報（エンティティ）の上位意味概念と同様の上位意味概念を持つ学習データを選択取得し、その学習データに記録された意図（インテント）を、その新たなユーザ発話の意図であると推定することが可能となる。

この処理については、後段で、図8を参照して説明する。

- [0105] 上述したように、発話学習適応処理部108は、ユーザ発話に含まれる実体情報（エンティティ）の上位意味概念を含む自然言語処理情報（nlp！nf！）、すなわち「学習意味概念A」を、意図（インテント）に対応付けた学習データを生成して、発話学習データ記憶部112に格納する。
- [0106] すなわち、発話学習適応処理部108は、図7を参照して説明したように、ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して発話学習データ記憶部112に格納する。
- [0107] 図7を参照して説明したように、最初の意図の不明確な第1のユーザ発話の入力後に、意図確認のための装置発話を実行して、該装置発話に対するユーザからの応答発話である第2のユーザ発話を入力し、第2のユーザ発話の解析結果として取得した意図を、第1のユーザ発話に含まれる実体情報に対応付けて記録した学習データを生成して、発話学習データ記憶部112に記憶する。

具体的には、例えば、第1のユーザ発話と第2のユーザ発話に共通に含まれる実体情報（桑原圭祐）と、正しい意図（インテント）を対応付けた学習データを生成して、発話学習データ記憶部112に記憶する。

- [0108] 図7（a）に示すように、  
意図（インテント）＝音楽再生（PLAY\_MUSIC）  
実体情報（エンティティ）＝【学習意味概念A】を含むデータ  
発話学習適応処理部108は、このような意図（インテント）と、実体情報（エンティティ）の対応データを有する学習データを生成して、発話学習データ記憶部112に格納する。

なお、この学習データは、

ユーザ発話テキスト（Speech）＝「桑原圭祐にして」  
に対応付けて格納される。

学習データの全体構成は、先に図6を参照して説明した通りである。

[0109] 次に、この学習処理によって生成された学習データを利用したユーザ発話の認識、意図推定処理、すなわち学習結果データの「適応処理」の具体例について、図8を参照して説明する。

[0110] 図8は、図7において説明した学習処理において生成した学習データ、すなわち、

意図（インテント）＝音楽再生（PLAY\_MUSIC）

実体情報（エンティティ）＝【学習意味概念A】を含むデータ

このような意図（インテント）と、実体情報（エンティティ）の対応データを有する学習データが発話学習データ記憶部112に格納された後のユーザ発話に対する処理の実行例である。

[0111] 図8に示すように、ユーザは、まず、ステップS141において、ミュージシャンである星野ゆかりの曲を聴きたいという意図で、「星野ゆかりにして」と発話する。

[0112] 情報処理装置100の発話学習適応処理部108は、  
ユーザ発話＝「星野ゆかりにして」  
このユーザ発話に含まれる  
実体情報（エンティティ）＝星野ゆかり  
この「星野ゆかり」に関する情報を知識辞書データ記憶部111から取得する。取得する情報は、  
エンティティの自然言語処理情報（nlpinfo）、すなわち、  
(a) エンティティを含む文節の句の種別：名詞句  
(b) エンティティを含む文節の格（主格、時間格、場所格など5W1Hや修飾節を表す）：連用修飾節  
(c) エンティティを含む文節の言語的な意味判定結果：対象、依頼、願望  
(d) エンティティ単語の品詞と種別：名詞：一般名詞：固有名詞：名：人物

(e) エンティティ単語の上位意味概念：コンテンツ：音楽：人物

これらの情報を含む。

[0113] すなわち、先に説明した【学習意味概念A】に相当する情報である。

発話学習適応処理部108は、

ユーザ発話=「星野ゆかりにして」

このユーザ発話に含まれる

実体情報（エンティティ）=星野ゆかりに対応する【学習意味概念A】と  
、類似する【学習意味概念A】を登録した学習データを発話学習データ記憶  
部112から検索する。

[0114] この検索結果、

意図（インテント）=音楽再生（PLAY\_MUSIC）

実体情報（エンティティ）=【学習意味概念A】を含むデータ

この意図（インテント）と、実体情報（エンティティ）の対応データを有  
する学習データが抽出される。

この学習データは、先に図7を参照して説明した学習処理によって発話学  
習データ記憶部112に格納された学習データである。

なお、この学習データは、

ユーザ発話テキスト（Speech）=「桑原圭祐にして」

に対応付けて格納されている。

[0115] 発話学習適応処理部108は、この学習データに基づいて、

ユーザ発話に含まれる”星野ゆかり”は、学習データとして記録された1つ  
の学習データ、すなわち、

意図（インテント）=音楽再生（PLAY\_MUSIC）

実体情報（エンティティ）=【学習意味概念A】を含むデータ

この対応データを記録した学習データと一致すると判定する。

[0116] さらに、

ユーザ発話=「星野ゆかりにして」

このユーザ発話に含まれる実体情報（エンティティ）である「星野ゆかり

」以外の発話テキスト（発話文言）の構成部分「にして」が学習データに記録された。

ユーザ発話テキスト（S p e e c h）＝「桑原圭祐にして」と一致すると判定する。

- [0117] 発話学習適応処理部108は、これらの結果に基づいて、図8に示すステップS141において実行されたユーザ発話、すなわち、「星野ゆかりにして」このユーザ発話の意図（インテント）を、意図（インテント）＝音楽再生（P L A Y \_ M U S I C）、と判定する。

- [0118] ユーザ発話の全文言を学習して、文言の完全一致時の学習結果を適用すると、実体情報（エンティティ）の中身だけが異なる発話に対応できない。しかし、本開示の発話学習適応処理部108は、知識ベース辞書を使用して、ユーザ発話に含まれる実体情報（エンティティ）の意味概念を学習して、学習データとして記憶部（発話学習データ記憶部112）に格納する構成としている。、

- [0119] この結果、ユーザ発話に含まれる実体情報（エンティティ）の意味概念が共通する学習データを利用したユーザ意図の推定を行うことができる。なお、上述したように、実体情報（エンティティ）の意味概念には、エンティティ単語の上位意味概念、すなわち、実体情報（エンティティ）の種類を示す情報も含まれる。

- [0120] 従って、学習データとして記憶部に記録された「学習意味概念」の比較による処理を可能とした構成では、例えば、以下のような処理が可能となる。「<学習意味概念＝アーティスト名>にして」で音楽再生、「<学習意味概念＝食べ物>にして」でレシピ提示、このような処理を行うことができる。

- [0121] このように、学習データとしてユーザ発話と、ユーザ発話に含まれる実体情報（エンティティ）に実体情報（エンティティ）の種類を示す上位意味概

念を含む「学習意味概念」を対応付けて記録することで、その後の新たなユーザ発話に対して、実体情報（エンティティ）の意味概念が共通する学習データを抽出し、その学習データに設定された意図（インテント）を、その新たなユーザ発話の意図（インテント）と解釈することが可能となる。

[0122] すなわち、本開示の情報処理装置100の発話学習適応処理部108は、例えば、新たに入力した新たなユーザ発話の意図解析処理に際して、その新たなユーザ発話に含まれる実体情報の上位意味概念情報を取得し、取得情報と類似する上位意味概念情報を記録した学習データを記憶部から抽出し、抽出した学習データに記録された意図を新たなユーザ発話の意図として推定する処理を実行することができる。

[0123] [3. 様々な意味を有する同一発話データである多義語に対する処理について]

次に、様々な意味を有する同一発話データである多義語に対する処理について説明する。

[0124] 上述したように、学習データとして記憶部に記録された「学習意味概念」の比較による処理を可能とした構成では、例えば、以下の処理が可能となる。

「<学習意味概念＝アーティスト名>にして」で音楽再生、  
「<学習意味概念＝食べ物>にして」でレシピ提示、  
このような処理を行うことができる。

[0125] しかし、ユーザ発話に含まれる実体情報（エンティティ）が、様々な意味を有する同一発話データである多義語である場合がある。

例えば、

食べ物名＝ブルーベリー

が存在し、さらに、

アーティスト名＝ブルーベリー

が存在するとする。

[0126] この場合、ユーザ発話として、

ユーザ発話=「ブルーベリーにして」

このようなユーザ発話があった場合、情報処理装置 100 は、音楽再生と、レシピ提示のどちらを行るべきかを判断することが困難になる。

[0127] 以下では、このような問題を解決する以下の 2 つの構成例について説明する。

(多義語対策構成例 1)

ユーザ発話実行時の対話状態（ステート）を区分し、区分ステート単位の個別の学習データを生成して記録する。

(多義語対策構成例 2)

ユーザ発話の意図（インテント）に併せてユーザ発話時の状況情報（コンテキスト）を記録した学習データを生成して記録する。

以下、これらの構成例について、順次、説明する。

[0128] (多義語対策構成例 1)

ユーザ発話実行時の対話状態（ステート）を区分し、区分ステート単位の個別の学習データを生成して記録する。

まず、多義語対策構成例 1 として、ユーザ発話実行時の対話状態（ステート）を区分し、区分ステート単位の個別の学習データを生成して記録する構成例について、図 9 を参照して説明する。

[0129] 図 9 に示すように、発話学習データ記憶部 112 に格納される学習データを、ユーザ発話実行時の対話状態（ステート）に応じて区分し、区分ステート単位の個別の学習データとする。

[0130] ユーザ発話実行時の対話状態（ステート）の区分は、例えば、図 9 に示すように、以下の 5 つの状態（ステート）区分とする。

ステート 1 = メッセージ通知状態ステート

ステート 2 = メッセージ読み上げステート

ステート 3 = メッセージ送信ステート

ステート 4 = 単機能処理依頼ステート

ステート 5 = シーケンス機能処理依頼ステート

なお、このステート区分は一例であり、その他の区分設定としてもよい。

- [0131] 例えば、ユーザと装置の対話内容に応じたステート区分を行う構成としてもよい。

具体的には、ユーザと装置の対話内容が食べ物や料理の話をしているときは食べ物ステート、音楽の話をしているときは音楽ステート、スポーツの話をしているときはスポーツステート、旅行の話をしているときは旅行ステート等である。

各対話の内容に応じたステート単位のステート対応の学習データやNLUを設定して利用する。

- [0132] 発話学習適応処理部108は、ユーザ発話が実行された際の状況に応じて、ステートを判定する。さらに、発話学習データ記憶部112に対して格納する学習データもステート単位で区分した記憶領域に格納する。

学習処理を実行して学習データの生成、更新等を実行する際には、判定したステートに応じて、利用する記憶領域を決定する。

- [0133] 学習処理に際して利用する自然言語理解（NLU）処理適用データ131についても、ステート単位のデータに区分して、各ステートに最適なデータを利用する。

- [0134] このように、学習データと、自然言語理解（NLU）処理適用データを、いずれもユーザ発話が実行された際の状況に応じたステート単位のデータとすることで、各ステートに応じたデータを利用したユーザ発話の意図推定が可能となる。

- [0135] 具体例として、例えば、

ユーザ発話=「ブルーベリーにして」

このようなユーザ発話があった場合の対応例について考察する。

- [0136] 情報処理装置100が、

まず、装置発話として、

「何を食べたいですか」

上記発話後に、

ユーザ発話=「ブルーベリーにして」  
が入力された場合、  
情報処理装置100の発話学習適応処理部108は、この対話ステートが  
、食べ物ステートであると判定し、食べ物ステート対応の学習データと、N  
LUデータを適用した処理を実行する、  
この処理により、情報処理装置100は、食べ物であるブルーベリーを使  
ったレシピを提示することが可能となる。

- [0137] また、情報処理装置100が、  
まず、装置発話として、  
「何の曲をかけますか」  
上記発話後に、  
ユーザ発話=「ブルーベリーにして」  
が入力された場合、  
情報処理装置100の発話学習適応処理部108は、この対話ステートが  
、音楽ステートであると判定し、音楽ステート対応の学習データと、NLU  
データを適用した処理を実行する、  
この処理により、情報処理装置100は、ミュージシャンであるブルーベ  
リーの曲を再生することが可能となる。

[0138] このように、対話ステートに応じて学習データやNLUデータを区分する  
ことで、多義語を含むユーザ発話に対するユーザ意図の正確な把握が可能と  
なる。

[0139] (多義語対策構成例2)

ユーザ発話の意図（インテント）に併せてユーザ発話時の状況情報（コン  
テキスト）を記録した学習データを生成して記録する。

次に、多義語対策構成例2として、ユーザ発話の意図（インテント）に併  
せてユーザ発話時の状況情報（コンテキスト）を記録した学習データを生成  
して記録する処理例について説明する。

[0140] 情報処理装置100の発話学習適応処理部108は、ユーザ発話に対する

学習処理を実行して学習データを生成するが、この学習データは、先に図6を参照して説明したデータ、すなわち、以下の各データを含むデータとなる。

- - (1) ユーザ発話テキスト (S p e e c h)
  - (2) 意図 (インテント)
  - (3) 実体情報 (エンティティ)
  - (4) その値の属性情報 (ユーザ発話日時 (d a t e) 、ユーザ発話の回数 (S p e e c h C o u n t))

[0141] さらに、多義語に対する正確な意図推定を実現するため。学習データに、以下の環境情報や、ユーザ情報等のコンテキスト情報を記録する。

- (a) 時刻情報 (朝／午前中／昼／午後／夕方／夜／夜中)
- (b) 機器の使用状態情報 (音楽再生中／動画再生中／ゲーム中)
- (c) ユーザ状態情報 (ユーザの位置／顔向き／視線方向)
- (d) 外部機器使用情報 (テレビのON, OFF／ラジオのON, OFF／エアコンのON, OFF等)

[0142] これらの環境情報や、ユーザ情報等のコンテキスト情報を、図6を参照して説明したユーザ発話対応の学習データ中に追加記録する。

[0143] 情報処理装置100の発話学習適応処理部108は、このようなコンテキスト情報を記録した学習データを用いてユーザ発話の意図推定を実行する。

[0144] ユーザ発話に含まれる実体情報 (エンティティ) の多義性により、学習データ内の複数の学習データが適用候補として選択された場合、各候補の学習データ生成時のコンテキスト情報と、意図推定対象である新たなユーザ発話時のコンテキスト情報の類似度を算出し、最も類似度が高い学習データを選択して、その選択学習データに記録された意図 (インテント) を意図推定対象である新たなユーザ発話の意図 (インテント) と判定する。

[0145] なお、コンテキスト情報の類似度の算出においては、各コンテキスト情報の種類数に応じた次元数のベクトルを生成し、ベクトル距離に基づいて類似

度を算出（距離が近いほど類似）するといった処理が可能である。

- [0146] コンテキスト情報を学習データに記録して、これを利用してユーザ発話の意図推定を行う具体例について説明する。

音楽のタイトルと、ゲームタイトルとで同じタイトル、例えば、「ブルースカイ X」が利用されているとする。

ユーザ発話として、以下の発話がなされたとする。

ユーザ発話=「ブルースカイ X お願い」

- [0147] ここで、ユーザ発話のなされた時点のコンテキストを取得する。

具体的には、

ユーザが何をしている状態か（音楽再生中、ゲーム中）、

機器の使用状態（ゲーム機が ON/OFF）

これらのコンテキスト情報を取得する。

- [0148] 例えば、情報処理装置 100 の発話学習適応処理部 108 は、ユーザ発話時にユーザがゲーム中であり、ゲーム機が利用状態（ON）であるというコンテキスト情報が得られた場合、情報処理装置 100 の発話学習適応処理部 108 は、このコンテキスト情報と類似するコンテキスト情報を記録した学習データを選択する。

さらに、選択した学習データに記録された意図（インテント）を取得する。

この取得した意図（インテント）には、ゲーム「ブルースカイ X」の再生と記録されている。

- [0149] この結果、情報処理装置 100 の発話学習適応処理部 108 は、

ユーザ発話=「ブルースカイ X お願い」

このユーザ発話の意図は、ゲーム「ブルースカイ X」の再生を行ってほしいという意図であると理解して、ゲーム再生を開始する。

- [0150] 一方、例えば、ユーザ発話時にユーザが音楽再生中であり、ゲーム機が非利用状態（OFF）であるというコンテキスト情報が得られた場合、情報処理装置 100 の発話学習適応処理部 108 は、このコンテキスト情報と類似

するコンテキスト情報を記録した学習データを選択する。

さらに、選択した学習データに記録された意図（インテント）を取得する。  
。

この取得した意図（インテント）には、音楽「ブルースカイX」の再生と記録されている。

[0151] この結果、情報処理装置100の発話学習適応処理部108は、  
ユーザ発話＝「ブルースカイXお願い」  
このユーザ発話の意図は、音楽「ブルースカイX」の再生を行ってほしい  
という意図であると理解して、音楽再生を開始する。

[0152] このように、学習データに、ユーザ発話時の環境情報、ユーザ情報等のコンテキストを併せて記録することで、コンテキストの類似度判定に基づくユーザ意図推定の精度を高めることが可能となる。

[0153] なお、コンテキスト情報を利用した処理例としては、例えば以下のような様々な処理例がある。

[0154] (具体例1)  
ユーザ発話＝「目覚まし見せて」  
利用コンテキスト＝時刻情報、外部機器情報  
朝、いつもの時間に発話ユーザが、テレビを見ていたら”目覚まし（番組）”のチャンネルに変える。

それ以外の時間だったら、アラームの時刻を提示する。

[0155] (具体例2)  
ユーザ発話＝「ブルーベリーでなんかある？」  
利用コンテキスト＝ユーザ状態  
ユーザが、キッチンで調理中ならブルーベリー（食材）を使ったレシピを提示  
ユーザがリビングにいたら、ブルーベリー（アーティスト）の曲を再生。

[0156] (具体例3)  
ユーザ発話＝「父さんにメール」

利用コンテキスト＝ユーザ状態

発話ユーザが家の子供である場合、パパにメールする。

発話ユーザが、子供のお父さんである場合、（子供の）おじいちゃんにメールする。

[0157] (具体例4)

ユーザ発話＝「音量あげて」

利用コンテキスト＝ユーザ状態

発話ユーザが見ている方向（テレビ等）の音量を上げる。

[0158] (具体例5)

ユーザ発話＝「切り方（焼き方）教えて」

利用コンテキスト＝ユーザ状態

発話ユーザが、キッチンにいたら「調理法としての切り方（焼き方）」を提示する。

発話ユーザが、PCなどの機器を操作していたら「機器操作法としての電源の切り方、メディアへの焼き方」を提示する。

[0159] (具体例6)

ユーザ発話＝「たこのあげかた教えて」

利用コンテキスト＝ユーザ状態

発話ユーザが、キッチンにいたら調理法として「蛸」の揚げ方を提示する。

発話ユーザの近くに子供が写っていたら遊び方としての「廁」のあげ方を提示する。

[0160] このように、学習データにユーザ発話時の環境情報やユーザ情報からなるコンテキスト情報を併せて記録し、新たなユーザ発話時に取得したコンテキストと、学習データに記録されたコンテキストとの照合を行うことで、ユーザ発話に多義語がふくまれていた場合でも、より正しい意図推定を行うことができる。

[0161] [4. 誤った学習処理に対する対策について]

本開示の情報処理装置 100 の発話学習適応処理部 108 は、ユーザ発話の学習処理により、ユーザ発話の意図推定を正しく実行するための学習データを生成し、またこの学習データを利用して発話の意図推定を行う。

[0162] しかし、学習処理が誤って実行され、誤った学習データが生成されてしまう可能性もある。

例えば、情報処理装置 100 の発話学習適応処理部 108 が、ユーザ発話の意図を確認するために、ユーザに対して問い合わせ発話として、  
装置発話=「～～の意味を教えてください」  
と発話したとする。

[0163] この装置発話に対して、悪意のあるユーザが、全くでたらめな発話を行ってしまうことがある。

例えば、

ユーザ発話=「桑原圭祐にして」  
装置発話=「「桑原圭祐にして」の意味を教えてください」  
このような対話がなされた後、ユーザが、  
ユーザ発話=「午前 2 時にアラームをセットして」  
このような発話をを行うことがある。

[0164] このような場合、情報処理装置 100 の発話学習適応処理部 108 は誤った学習データを生成してしまう可能性がある。

以下、このような誤学習による誤った学習データの生成や、記憶部（発話学習データ記憶部 112）に対する格納処理を防止可能とした構成について説明する。

以下の 2 つの構成例について説明する。

(誤学習対策構成例 1)

言語解析による実体情報（エンティティ）の意味概念の相違度を検証して誤学習を防止する。

(誤学習対策構成例 2)

画像による個人識別を使った学習データの適用範囲を判定して誤学習を防

止する。

以下、これらの構成例について、順次、説明する。

[0165] (誤学習対策構成例 1)

言語解析による実体情報（エンティティ）の意味概念の相違度を検証して誤学習を防止する。

まず、誤学習対策構成例 1 として、言語解析による実体情報（エンティティ）の意味概念の相違度を検証して誤学習を防止する構成について説明する。

[0166] この誤学習防止処理は、情報処理装置 100 の発話学習適応処理部 108において、ユーザ発話が、「意図解釈不可（OOD）発話」と判定された場合、この「意図解釈不可（OOD）発話」を言語解析し、ユーザ発話内に含まれる実体情報（エンティティ）の候補となる対象格や主格の意味概念を抽出する。

[0167] この処理は、例えば、先に説明した図 4 に示すフローのステップ S102において、ユーザ発話が「意図解釈不可（OOD）発話」と判定され、ステップ S103 に進んだ時点で実行される。

[0168] 情報処理装置 100 の発話学習適応処理部 108 は、

(1) ユーザの最初の発話、すなわち、「意図解釈不可（OOD）発話」と判定されたユーザ発話と、

(2) その後の情報処理装置 100 からの問い合わせに対するユーザの言い直し発話、

これら 2 つの発話内に含まれる実体情報（エンティティ）の候補となる対象格や主格の意味概念を抽出する。

[0169] さらに、これら 2 つの発話の実体情報（エンティティ）の候補の意味概念の意味の相違度を意味ベクトル距離などで判定する。

2 つの発話のエンティティのベクトル距離が近いほど、2 つの発話は類似すると判定する。

一方、2 つの発話のエンティティのベクトル距離が遠いほど、2 つの発話

は非類似であると判定する。

[0170] 2つの発話のエンティティのベクトル距離を算出した後、情報処理装置100の発話学習適応処理部108は、予め規定したしきい値を用いて、2つの発話のエンティティのベクトル距離と比較する。

2つの発話のエンティティのベクトル距離がしきい値以上である場合、2つの発話は類似しない、すなわち、無関係の発話であると判断し、ユーザの言い直しの発話を学習処理の対象から除外する。

[0171] 先に説明した発話例では、2つの発話の意味概念は、以下のようになる。

(1) ユーザの最初の発話、すなわち、「意図解釈不可（OOD）発話」と判定されたユーザ発話内の実体情報（エンティティ）＝桑原圭祐の意味概念＝<アーティスト>

(2) その後の情報処理装置100からの問い合わせに対するユーザの言い直し発話内の実体情報（エンティティ）＝午前2時の意味概念＝<時刻>

。

[0172] 情報処理装置100の発話学習適応処理部108は、これら2つの発話の意味概念、

<アーティスト>

<時刻>

これらの意味概念の類似度を、例えばベクトル距離により算出する。

ベクトル距離は、規定のしきい値より大きく、類似度が低いと判定する。

その結果、ユーザの言い直しの発話を学習処理の対象から除外する。

[0173] (誤学習対策構成例2)

画像による個人識別を使った学習データの適用範囲を判定して誤学習を防止する。

次に、誤学習対策構成例2として、画像による個人識別を使った学習データの適用範囲を判定して誤学習を防止する構成について説明する。

[0174] この処理は、ユーザ発話の実行された時点の画像を、画像入力部104によって取得し、この取得画像による個人識別を行い、識別結果を利用して学

習データの適用範囲を判定する処理である。

[0175] この処理においては、以下の処理を実行する。

「意図解釈不可（OOD）発話」に対する言い直し発話をって学習対話を行ったユーザ本人を、この「意図解釈不可（OOD）発話」の学習データに対する適用対象ユーザとして学習データ内に記録する。また、学習対話時に一緒にいることが画像から確認されたユーザも適用対象ユーザとして学習データ内に記録する。

その後の発話解析時には画像から確認した発話ユーザが、学習データの適用対象ユーザに含まれている場合のみ適用を行う。

[0176] 上記処理を行うことで、例えば、子供が親と一緒にいない状況においてイタズラで学習させたデータは、その後の親の発話には適用されない設定が可能となる。また、別の例として、使用方法に詳しくないおばあちゃんが「意図解釈不可（OOD）発話」を行い、次に一緒に画像に写っているお父さんの言い直し発話が行われると、学習データには適用対象ユーザとしておばあちゃんとお父さんが記録され、その後のおばあちゃんの発話の解析に利用される設定が可能となる効果もある。

[0177] ただし、ユーザ（A）が、一緒にいる特定の人（B）を指定して学習発話を許可すると、その後はAと一緒にいない状況でも、Bが行った学習対話による学習データの適用対象ユーザとしてAを含めて記録する。

具体的には、お父さん（B）と一緒にいるおばあちゃん（A）が、「この人の教えてもらうよ」と情報処理装置100に話しかける。その後、おばあちゃん（A）がいない時にお父さん（B）が行った学習発話による学習データは、おばあちゃん（A）の発話の解析に利用される。

[0178] 上記処理を行うことで、お父さん（B）が一人でおばあちゃん（A）のために学習データを生成することも可能である。

例えば、お父さん（B）が、おばあちゃん（A）の「意図解釈不可（OOD）発話」を真似た発話を実行して、その後、言い直し発話をって学習処理を実行するといった処理が可能である。

[0179] このように、（誤学習対策構成例2）は、画像による個人識別を使った学習データの適用範囲を判定して誤学習データの適用を防止する構成であり、学習データの発話毎に、学習者と適用範囲の情報を記憶する。これは、いわゆるファイルのアクセス権限設定処理と同様の処理である。

[0180] [5. 学習データの自動忘却処理について]

次に、学習データの自動忘却処理について説明する。

すなわち、発話学習データ記憶部112に格納された学習データを自動的に破棄する処理である。

上記の（誤学習対策構成例1，2）を利用して、すべての誤学習を防止することは困難である。

従って、発話学習データ記憶部112には、正しい学習データに併せて誤学習データも蓄積されてしまうことになる。

[0181] このような誤学習データの蓄積量の増加を低減するため、例えば、継続的に適用範囲の人が使用しなければ誤学習データを破棄する。すなわち、発話学習データ記憶部112に格納された学習データは、予め規定された条件に従って自動忘却される。

[0182] この学習データの自動忘却処理について説明する。

情報処理装置100の発話学習適応処理部108は、発話学習データ記憶部112に格納された学習データ各々について、以下の（式1）に従って学習データ忘却スコア $f(t, c)$ を算出する。

$$f(t, c) = (k_c / c) \log(k_t t + 1) \dots \text{(式1)}$$

[0183] なお、上記（式1）において、各パラメータは以下の意味を持つ。

$t$ ：学習データ対応の発話をユーザが最後に発話してから経過した時間（日数（Days））

$c$ ：学習データ対応の発話をユーザが繰り返した回数

$k_t$ ：時間の重み係数

$k_c$ ：繰り返し回数の重み係数

[0184] 重み係数： $k_t = 2$ ， $k_c = 2$ とした時の忘却スコアの時間推移例を図10

に示す。

図10には、学習データの発話をユーザが繰り返した回数： $c = 1, 2, 3, 4, 5$ の5つの例の学習データ忘却スコア  $f(t, c)$  を示している。

横軸が、時間 ( $t$ ) 、縦軸が学習データ忘却スコア  $f(t, c)$  である。

横軸の時間 ( $t$ ) は、上記の定義の  $t$  であり、

$t$ ：学習データの発話をユーザが最後に発話してから経過した時間（日数  
(Days)）

に相当する。

[0185] 図10には、学習データの発話をユーザが繰り返した回数： $c = 1, 2, 3, 4, 5$ の5つの忘却スコア  $f(t, c)$  例を示している。繰り返し回数が少ないほど、忘却スコア  $f(t, c)$  は早く増加し、繰り返し回数が多いほど、忘却スコア  $f(t, c)$  の増加スピードは低下する。

[0186] なお、パラメータ  $c$ 、すなわち、

$c$ ：学習データの発話をユーザが繰り返した回数

このパラメータ  $c$  の更新処理は、以下の「発話繰り返し回数  $c$  更新ルール」に従って行う。

1日以上、時間が空いて同じ発話を入力された場合に更新 (+1) する。

1日以下の時間間隔で同じ発話を入力された場合は、

（発話間隔間隔／1日の総時間）を、現在の  $c$  の値に加算する。

情報処理装置100の発話学習適応処理部108は、上記の「発話繰り返し回数  $c$  更新ルール」に従ってパラメータ  $c$  の更新処理を行う。

[0187] 上記処理は、短時間に何回も繰り返して実行される発話より、時間を空けて繰り返し実行される発話の方を優先して学習データとして残すことを可能とするための処理である。

[0188] 図10の縦軸には、以下の2つのしきい値を示している。

（1）再認可能忘却しきい値、

（2）完全忘却しきい値、

図10には、一例として、

(1) 再認可能忘却しきい値=忘却スコア  $f(t, c) = 1$ 、

(2) 完全忘却しきい値=忘却スコア  $f(t, c) = 1.4$ 、

この設定とした例を示している。

このしきい値設定は、様々な設定が可能である。

[0189] 例えば、発話学習データ記憶部112に格納されているある学習データの忘却スコア  $f(t, c)$  が、再認可能忘却閾値（図10の例では1.0）を超えたなら、その学習データは、その後のユーザ発話の解析処理に適用しない。

ユーザ発話の解析処理に適用する学習データは、忘却スコア  $f(t, c)$  が、再認可能忘却閾値以下の学習データのみとする。

[0190] ただし、ある学習データの忘却スコア  $f(t, c)$  が、再認可能忘却閾値（図10の例では1.0）を超えた場合でも、この時点では、その学習データは、発話学習データ記憶部112から消去することはせず、残しておく。

[0191] さらに、発話学習データ記憶部112に格納されているある学習データの忘却スコア  $f(t, c)$  が、完全忘却閾値（図10の例では1.4）を超えたなら、その学習データを、発話学習データ記憶部112から消去する。すなわち破棄する。

[0192] なお、学習データの破棄処理は、ユーザ発話に対する学習処理や、学習データの適応処理時に使う構成としてもよいし、定期的に決まった時間に行う構成としてもよい。

また、情報処理装置100の起動／終了時や各対話ステートの開始／終了時に行うようにしてもよい。

[0193] このような処理を行うことで、例えば、誤学習やユーザがたまたま1回使っただけの言い回しは不要な学習データとして自動で破棄される。

情報処理装置100の発話学習適応処理部108は、発話学習データ記憶部112に格納されている1つの学習データを適用した際、その学習データの忘却スコアが再認可能忘却閾値より低い場合には、その学習データの日時情報を適応時の日時に更新し、発話繰り返し回数cを更新する。発話繰り返

し回数  $c$  の更新は、上述した「発話繰り返し回数  $c$  更新ルール」に従って行う。

この更新処理により、その学習データの忘却スコアは 0 にリセットされ、新たな更新後の発話繰り返し回数  $c$  による曲線に遷移する。

[0194] このような処理を行うことで、繰り返し発話された発話は時間がたっても忘却スコアが上がりにくくなる。すなわち、発話学習データ記憶部 112 に長期記憶される。

すなわち、誤学習などで 1 回のみ記憶した発話はすぐに忘却するが、繰り返されるユーザ固有の言い回し等は長期に渡って記憶部に記憶される。

[0195] 学習時に該当発話の学習データが既にあって忘却スコアが再認可能忘却閾値より高く、完全忘却閾値より低い場合（適用はされないので「意図解釈不可（OOD）発話」になるが破棄されずにデータは残っている状態＝再認可能状態）には、その学習データの日時情報を学習時の日時に更新し、上述した「発話繰り返し回数  $c$  更新ルール」に従って発話繰り返し回数  $c$  の更新を行う。

[0196] 忘却スコアが再認可能忘却閾値より高く、完全忘却閾値より低い学習データは、ユーザ発話への適用はしないが再学習で再認識することで、再学習前より発話学習データ記憶部 112 には長期記憶されることになる。

[0197] なお、図 4 に示すステップ S104 において実行される学習処理 B（修正対話フロー）では、発話学習データ記憶部 112 に格納された学習データの意図（インテント）を上書きして修正する処理が行われるが、この場合、忘却スコアが再認可能忘却閾値より低い場合には、学習データの日時情報を修正学習時の日時に更新し、上述した「発話繰り返し回数  $c$  更新ルール」に従って発話繰り返し回数  $c$  の更新を行う。

図 4 に示すステップ S104 において実行される学習処理 B（修正対話フロー）の詳細については後述する。

[0198] 本開示の処理では、上述したように、学習データ忘却スコア  $f(t, c)$  の構成パラメータの 1 つであるパラメータ  $c$ 、すなわち、

c : 学習データの発話をユーザが繰り返した回数

このパラメータ c の更新処理を、以下の「発話繰り返し回数 c 更新ルール」に従って行う。

1日以上、時間が空いて同じ発話を入力された場合に更新 (+1) する。

1日以下の時間間隔で同じ発話を入力された場合は、

(発話間隔間隔 / 1日の総時間) を、現在の c の値に加算する。

情報処理装置 100 の発話学習適応処理部 108 は、上記の「発話繰り返し回数 c 更新ルール」に従ってパラメータ c の更新処理を行う。

[0199] このようなパラメータ c の更新処理を適用することで、時間を空けて繰り返された発話は指数的に長期記憶対象に設定される。

また、短時間内 (1日以下) の繰り返し発話は発話回数 c がほぼ増加しないため早期に (数日で) 破棄される。

[0200] このパラメータ c の更新処理を適用することによりメリットが発生する複数の具体例について説明する。

[0201] (具体例 1)

1年から数年に1回のイベント関連の言葉に関連する学習データは、発話学習データ記憶部 112 に長期記憶される。

例えば、「運動会の日は晴れ?」「今日金メダル撮った?」を数日間に渡り、毎日、発話されると、この発話関連の学習データが発話学習データ記憶部 112 に長期記憶される。

この結果、1~数年後も発話学習データ記憶部 112 に学習データが残る。

[0202] これに対して、例えば、数か月単位の発話頻度に基づく自動破棄を行う構成の場合、1~数年後には、学習データが発話学習データ記憶部 112 に存在しなくなるため、1~数年後の同じユーザ発話に対して新たな学習処理が必要となる。

なお、発話時期により忘却しやすさを変えてもよい。例えば、12月は「クリスマス」を含む文言を忘れにくくする等の処理である。

これは、例えば、知識辞書データ記憶部 111 に格納された知識辞書データを参照して行うことができる。

[0203] (具体例 2)

定期的に来る親戚の発話は覚え、単発の客との発話は忘れる。

例えば、情報処理装置 100 に発話をを行うユーザが、週 2 回来る祖母であり、関西弁の発話ユーザである場合、上記のパラメータ c の更新処理を適用することで長期記憶の対象となる。すなわち、上述したように時間を空けて繰り返された発話は指数的に長期記憶となる。

一方、めったに来ないが短期間で繰り返し発話をされた友人の九州弁は数日で忘れる。

[0204] 単に頻度による破棄処理を行う構成とした場合、数か月間はメインユーザが使い慣れない九州弁に反応して誤解釈が発生する可能性があるが、本開示の処理、すなわち、時間を空けて繰り返された発話を優先的に長期記憶させる構成とすることで、このような誤解釈の発生を低減できる。

[0205] (具体例 3)

単発の流行り言葉を破棄しやすい。

例えば、「品川の事件の犯人捕まった?」など単発ニュースに関連するワードは、短時間で繰り返し発話される傾向があるが、このような短時間間隔の繰り返し発話に関連する学習データは、本開示の処理では、長期記憶対象とはなりにくい。

[0206] 本開示の処理では、このような短時間間隔の繰り返し発話に関する学習データは、記憶部から早期に破棄されるので、古い発話による誤発を防ぎ、記憶リソースも節約できる。

[0207] (具体例 4)

環境による音声認識の誤認識。

例えば、情報処理装置 100 が設置された部屋でたまたまついていたテレビ番組の音を入力して、情報処理装置 100 がその入力ワードに基づく学習処理を行って学習データを発話学習データ記憶部 112 に格納した場合、早

期に破棄される可能性が高くなる。

一方、近くを通り電車の騒音など定期・定常的に家の中で発生しているノイズについては、ノイズ環境下での音声認識の誤認識を含むユーザ発話による学習と適用が定期的に繰り返し行われ、定期・定常的なノイズ環境下での学習データが発話学習データ記憶部112に長期記憶される。

- [0208] [6. 発話学習適応処理部が実行する学習対話フローの詳細について]
- 次に、図4のフローチャートのステップS104において、発話学習適応処理部108が実行する「学習処理B=ユーザ発話意図の修正対話フロー」による学習データの生成処理の詳細について説明する。
- [0209] 先に図4を参照して説明したように、図4のフローチャートのステップS104の「学習処理B=ユーザ発話意図の修正対話フロー」は、情報処理装置100が、ユーザ発話として、意図解釈間違指摘発話を入力した場合に実行される。
- [0210] 図11を参照して、ステップS104において実行する「学習処理B=ユーザ発話意図の修正対話フロー」の具体例について説明する。
- [0211] 図11に示す例は、情報処理装置100がユーザ発話を入力して実行した自然言語理解(NLU)結果や、学習データから推定した意図がユーザの発話意図とは異なる場合の修正対話フローの例である。
- 図11に示すように、まず、ユーザはメッセージの一覧を見せてほしい意図を持ち、ステップS151において、
- ユーザ発話=「見せて」
- と発話する。
- [0212] 情報処理装置100は、ステップS152において、このユーザ発話を入力して自然言語理解(NLU)処理、や学習データから意図推定を行い、ユーザ発話の意図(インテント)が、「タスクを見せて(SHOW\_TASK)」であると判定する。
- さらに、この意図判定結果に基づく処理として、ステップS153において、情報処理装置100の音声出力部123を介して、

装置発話=「5件のタスクがあります」

との装置発話を実行し、さらに、画像出力部125にタスクの一覧表示を行う。

[0213] ユーザはメッセージの一覧を見せてほしい意図で「見せて」と発話したので、ステップS154において、間違い指摘発話として、

ユーザ発話=「ちがうよ」  
を発話する。

[0214] 情報処理装置100の発話学習適応処理部108は、このユーザからの間違い指摘発話の意図を自然言語理解（NLU）処理や学習データから推定する。

[0215] なお、ユーザによる間違い指摘発話の例としては、「違うよ」の他、「間違ってるよ」、「ダメ」、「やり直し」、「そうじゃないよ」などがある。

発話学習適応処理部108は、これらのユーザ発話を間違い指摘発話として認識する。

[0216] 発話学習適応処理部108は、ユーザの間違い指摘発話を認識すると、ステップS155において、再処理ユーザ発話「見せて」を学習候補に登録し、さらに、ステップS156において、最初のユーザ発話=「見せて」の発話意図を獲得するための問い合わせを行う。すなわち、間違い指摘がなされた「ユーザのOOD発話内容」についての問い合わせ文を、音声合成処理（TTS：Text To Speech）により生成して、

「”ユーザの最初の発話内容”の意味を教えてください」  
上記のような問い合わせ応答を出力する。

図11に示す例では、情報処理装置100は、ステップS156において  
、

「「見せて」の意味を教えてください」  
上記の問い合わせ応答を出力している。

[0217] なお、ユーザ発話意図を獲得するためのTTS発話文言としては、「言い換えてください」「違う言い方で言ってください」などでもよい。

上記のユーザ発話意図獲得のための問い合わせ発話後、予め規定した時間内（例えば5秒以内）に実行されたユーザ発話（図11の例ではステップS157のユーザ発話）の自然言語理解（NLU）結果（または学習データの適用結果）を「鈴木さんの4で」の真の発話意図として記憶する。

すなわち、学習データとして、発話学習データ記憶部112に格納する。

- [0218] ただし、上記のステップS156の問い合わせ実行後、予め規定した時間内（例えば5秒以内）にユーザ発話が無かった場合は、学習対話フローを終了する。すなわち、学習データ生成を行わない。

もし、最初のユーザ発話=「見せて」発話の意図（インテント）を、「タスクを見せて（SHOW\_TASK）」であると判定する処理に、学習データを適用している場合は、その学習データを破棄、または修正する。

- [0219] 具体的には、最初の「見せて」発話の意図（インテント）を「タスクを見せて（SHOW\_TASK）」であると判定した処理が、学習データに基づく処理だった場合にはその学習データの記録データである意図（インテント）を上書きして修正する。

また、最初の「見せて」発話の意図（インテント）を「タスクを見せて（SHOW\_TASK）」であると判定した処理が、自然言語理解（NLU）の処理結果に基づくものだった場合には新規の学習データを生成して記憶する。

なお、学習データの適用時には、自然言語理解（NLU）結果より学習データが優先して適用される。

- [0220] ステップS156の情報処理装置100の処理、すなわち、  
「「見せて」の意味を教えてください」  
上記の問い合わせの出力に対するユーザ応答として、ステップS157において、  
「メッセージの一覧を見て」  
このようなユーザ応答があると、図11のステップS158において、ステップS158に示すデータを学習データとして発話学習データ記憶部11

2に格納する。

[0221] 図11のステップS158に示すデータは、

ユーザ発話の意図（インテント）＝「メッセージ出力（SHOW\_MESSAGE）」を記録したデータである。このデータがユーザテキスト（発話文言）＝「見せて」に対応付けて発話学習データ記憶部112に記録される。

[0222] すなわち、情報処理装置100の発話学習適応処理部108は、図11に示すステップS151～S158の処理により、図11の下に示す（a）学習処理を行うことになる。すなわち、

発話テキスト（発話文言）＝「見せて」は、

意図（インテント）＝メッセージ出力（SHOW\_MESSAGE）、

この意図を有するユーザ発話であることの学習データを生成して発話学習データ記憶部112に記録する。

[0223] この学習処理の後、情報処理装置100の発話学習適応処理部108は、ステップS159において、応答生成部121等を適用してユーザ発話に対応した処理、すなわち、

メッセージ出力に関連する装置発話処理を実行する。図に示す例では、

装置発話＝「3件のメッセージがあります」

を実行している。

[0224] このように、情報処理装置100の発話学習適応処理部108は、ユーザの発話意図をユーザの指摘に基づいて修正する。

この機能を備えた構成とすることで、例えば、誤学習のユーザによる訂正や、音声認識や自然言語理解（NLU）の性能限界による装置誤動作を、ユーザが修正することができる。

[0225] [7. 学習データを適用したユーザ発話の意図推定処理の詳細について]  
]

次に、図4に示すフローのステップS101において実行する処理、すなわち、学習データを適用したユーザ発話の意図推定処理の詳細について説明

する

[0226] 図4に示すフローのステップS101において、発話学習適応処理部108は、ユーザ発話に対応するテキストデータからの自然言語理解(NLU)結果を取得するとともに、発話学習データ記憶部112に格納された学習データを適用して、ユーザ発話の意図(インテント)と実体情報(エンティティ)を推定する。

[0227] なお、先に説明したように、発話学習適応処理部108には、ユーザ発話に対する以下の解析データが入力されている。

(1) 音声認識部102が、ユーザ発話に基づいて生成したテキストデータ、

(2) 発話意味解析部103が、テキストデータに対する自然言語理解(NLU)を実行して生成したユーザ発話の意図(インテント)と実体情報(エンティティ)、

(3) カメラなどの画像入力部104が取得した発話ユーザおよびその周囲の画像に対する画像認識部105の画像認識結果情報、

(4) センサー106が取得した発話ユーザおよびその周囲の状態の検出情報に基づいて、センサー情報解析部が解析したセンサー解析情報、

[0228] 発話学習適応処理部108は、ユーザ発話に対応するテキストデータに対して、発話学習データ記憶部112に格納済みの発話学習データから適用できるデータが存在するか検索する。

適用できるデータが発話学習データ記憶部112に格納済みの発話学習データに存在する場合は、学習データを利用して、ユーザ発話の意図(インテント)と実体情報(エンティティ)を推定する。

[0229] 一方、適用できるデータが発話学習データ記憶部112に格納済みの発話学習データに存在しない場合は、発話意味解析部103がテキストデータに対する自然言語理解(NLU)を実行して生成したユーザ発話の意図(インテント)と実体情報(エンティティ)を、そのままユーザ発話の意図と実体情報とする。

なお、初期状態で、学習データ自体が存在しない場合も、同様の処理を行う。

[0230] このステップS101の意図推定処理に際しては、発話学習データ記憶部112に格納された学習データを適用する。

先に説明したように、学習データの生成、修正、破棄等の処理は、次に説明するステップS103やステップS104において実行される。

すなわち、発話学習データ記憶部112に格納される学習データは、ステップS103やステップS104において実行される学習処理により、逐次更新されたデータである。

[0231] ステップS101における処理は、発話学習データ記憶部112に格納された最新の学習データを用いて実行される。

この発話学習データ記憶部112に格納された学習データを用いたユーザ発話の意図（インテント）と実体情報（エンティティ）の推定処理の詳細シーケンスについて、図12、図13に示すフローチャートを参照して説明する。

[0232] 図12、図13に示すフローチャートに従った処理は、例えば情報処理装置100の記憶部に格納されたプログラムに従って実行される。

発話学習適応処理部108における処理は、例えばプログラム実行機能を有するCPU等のプロセッサによるプログラム実行処理として実行可能である。

以下、図12、図13に示すフローの各ステップの処理について説明する。

[0233] (ステップS201)

まず、発話学習適応処理部108は、ステップS201において、ユーザ発話入力に応じて、記憶部に格納された学習データの検索を行う。

すなわち、図3に示す情報処理装置100の発話学習データ記憶部112に格納されている学習データの検索を行う。

[0234] なお、学習データの各々は、図6のデータ構造を持つ。

なお、学習データの各々は、先に図9を参照して説明したように、対話状態（ステート）毎に異なるデータセットを持っている。

ステップS201ではステップ実行時点の対話状態（ステート）を解析して、その対話状態（ステート）に対応の学習データを検索対象とした検索を行う。

この処理を行うことで、対話の文脈に合わせた学習データの適用が可能となり、誤解釈を低減できる。

[0235] (ステップS202)

次に、発話学習適応処理部108は、ステップS202において、ユーザ発話の意図推定に適用可能な学習データが発話学習データ記憶部112に格納されているか否かを判定する。

[0236] ユーザ発話の意図推定に適用可能な学習データが発話学習データ記憶部112に格納されていない場合は、ステップS225に進む。

一方、ユーザ発話の意図推定に適用可能な学習データが発話学習データ記憶部112に格納されている場合は、ステップS203に進む。

[0237] (ステップS203)

次に、発話学習適応処理部108は、ステップS203において、音声認識部102が、ユーザ発話に基づいて生成したテキストデータを入力する。

[0238] (ステップS204)

次に、発話学習適応処理部108は、ステップS204において、知識辞書データ記憶部111に格納された知識辞書データを参照して、テキストデータに対する自然言語処理を実行して生成したユーザ発話に含まれる実体情報（エンティティ）対応の意味概念の抽出を行う。

[0239] [意味概念]は、先に、図7、図8を参照して説明した情報である。

すなわち、ユーザ発話から取得される実体情報（エンティティ）の自然言語処理情報（n | p | n f o）、すなわち、

(a) エンティティを含む文節の句の種別

(b) エンティティを含む文節の格（主格、時間格、場所格など5W1H）

や修飾節を表す)

(c) エンティティを含む文節の言語的な意味判定結果

(d) エンティティ単語の品詞と種別

(e) エンティティ単語の上位意味概念

これらの情報を含む。

[0240] 発話学習適応処理部108は、ステップS204において、ユーザ発話に含まれる実体情報（エンティティ）対応の意味概念の抽出を行う。

[0241] (ステップS205～S212)

ステップS205～S212は、発話学習データ記憶部112に格納されている学習データ各々（学習データ要素）について、順次、実行するループ処理である。

なお、学習データ各々（学習データ要素）は、先に図6を参照して説明したデータ構成を有する。すなわち、以下の構成データを有する。

(1) ユーザ発話テキスト（S p e e c h）

(2) 意図（インテント）

(3) 実体情報（エンティティ）

(4) その値の属性情報（ユーザ発話日時（d a t e）、ユーザ発話の回数（S p e e c h C o u n t））

[0242] さらに、図6（b）に示すように、1つのエンティティに関する情報として以下のデータが記録される。

(1) エンティティのユーザ発話文字（L i t e r a l）、

(2) エンティティのタイプ（t y p e）、

(3) エンティティの自然言語処理情報（n l p l n f o）として、

(3 a) エンティティを含む文節の句の種別

(3 b) エンティティを含む文節の格（主格、時間格、場所格など5W1Hや修飾節を表す）

(3 c) エンティティを含む文節の言語的な意味判定結果

(3 d) エンティティ単語の品詞と種別

### (3 e) エンティティ単語の上位意味概念

これらの情報が、学習データ各々に記録されている。

- [0243] 発話学習適応処理部 108 は、ステップ S206において、ユーザ発話内のエンティティの意味概念と一致する意味概念を有する学習データを抽出する。

この処理は、先に図 8、図 9 を参照して説明した処理と同様の処理である。  
。

- [0244] なお、先に説明した図 9 ではユーザ発話内に意味概念が一致する単語が一つの場合を示しているが、2つ以上あってもよい。例としてユーザ発話「今日の東京の天気は？」の実体情報（エンティティ）は「今日」と「東京」の2つとなる。

- [0245] 次に、発話学習適応処理部 108 は、ステップ S207～S208において、ユーザ発話のエンティティ以外の発話部分が、抽出学習データの発話データ（speech）に一致しているか否かを判定する。

この処理は、具体的には、例えば、先に図 7、図 8 を参照して説明した処理における

「桑原圭祐にして」と、

「星野ゆかりにして」

のエンティティ（桑原圭祐、星野ゆか）以外の「にして」の一致性の判定処理である。

- [0246] ユーザ発話のエンティティ以外の発話部分が、抽出学習データの発話データ（speech）に一致していると判定した場合はステップ S209に進み、一致していない場合は、現学習データのループ処理を終了して、次の学習データ要素についてのループ処理に移行する。

- [0247] ユーザ発話のエンティティ以外の発話部分が、抽出学習データの発話データ（speech）に一致していると判定した場合はステップ S209において、学習データ内にある発話回数と発話日時情報から忘却スコアを算出する。

この処理は、先に図10を参照して説明した学習データ忘却スコア  $f(t, c)$  の算出処理である。

[0248] 具体的には、以下の（式1）に従ってスコア  $f(t, c)$  を算出する。

$$f(t, c) = (k_c / c) \log(k_t t + 1) \dots \text{ (式1)}$$

なお、上記（式1）において、各パラメータは以下の意味を持つ。

$t$ ：学習データ対応の発話をユーザが最後に発話してから経過した時間（日数（Days））

$c$ ：学習データ対応の発話をユーザが繰り返した回数

$k_t$ ：時間の重み係数

$k_c$ ：繰り返し回数の重み係数

[0249] 次に、発話学習適応処理部108は、ステップS210において、算出した忘却スコアが、再認可能忘却閾値以下であるか否かを判定する。

忘却スコアが再認可能忘却閾値以下である場合は、ステップS211に進む。

忘却スコアが再認可能忘却閾値以下でない場合は、現学習データのループ処理を終了して、次の学習データ要素についてのループ処理に移行する。

[0250] 忘却スコアが再認可能忘却閾値以下である場合は、ステップS211に進む。

発話学習適応処理部108は、ステップS211において、ループ中の今の学習データ要素と意味概念一致で抽出したエンティティを、意図推定候補の一つとしてメモリに一時記憶する。

[0251] 全ての学習データ要素についてステップS205～S212のループ処理が完了すると、次のステップS221に進む。

[0252] （ステップS221）

発話学習適応処理部108は、ステップS221において、ステップS208～S212の学習データ要素サーチで抽出、一時記憶した意図推定候補となる学習データ要素をメモリから取出す。

[0253] （ステップS222）

発話学習適応処理部 108 は、ステップ S222において、メモリから取得した意図推定候補がいくつ有るかを判定する。

2つ以上ある場合は、ステップ S223 に進む。

1つのみの場合は、ステップ S224 に進む。

1つもない場合は、ステップ S225 に進む。

[0254] (ステップ S223)

発話学習適応処理部 108 は、メモリから取得した意図推定候補が2つ以上ある場合は、ステップ S223 において、意図推定候補の中で学習データ要素のコンテキスト情報が現在のコンテキストと最も類似している意図推定候補を、最終意図推定用学習データ要素として選択する。

[0255] なお、メモリから取得した意図推定候補が2つ以上ある場合とは、例えばユーザ発話に含まれるエンティティが多義語である場合などに発生する。

この場合は、先に説明したコンテキストの解析を行い、ユーザ発話の実行された状況を示すコンテキストと類似するコンテキストが記録された1つの学習データ（要素）を選択する。

[0256] (ステップ S224)

また、発話学習適応処理部 108 は、メモリから取得した意図推定候補が1つのみである場合は、ステップ S224 において、取得した1つの意図推定候補を、最終意図推定用学習データ要素とする。

[0257] (ステップ S225)

また、発話学習適応処理部 108 は、メモリから取得した意図推定候補が1つもない場合は、ステップ S225 において、発話意味解析部 103 から、解析結果として入力された意図（Intent）と実体情報（Entity）を推定結果として後段フローに出力して処理を終了する。

すなわち、図4に示すフローのステップ S101 の処理を終了して、ステップ S102 に進む。

[0258] (ステップ S226)

一方、ステップ S223、またはステップ S224 において、最終意図推

定用学習データ要素を決定した場合は、ステップS226に進む。

発話学習適応処理部108は、ステップS226において、選択した最終意図推定学習データ要素の発話の回数cを更新する。

[0259] なお、回数パラメータcは、先に図10を参照して説明したように、「発話繰り返し回数c更新ルール」に従って更新処理を行う。

回数パラメータcは、学習データの発話をユーザが繰り返した回数であり、以下の「発話繰り返し回数c更新ルール」に従って更新を行う。

1日以上、時間が空いて同じ発話が入力された場合に更新(+1)する。

1日以下の時間間隔で同じ発話が入力された場合は、

(発話間隔間隔／1日の総時間)を、現在のcの値に加算する。

このように、発話学習適応処理部108は、上記の「発話繰り返し回数c更新ルール」に従ってパラメータcの更新処理を行う。

また、ステップS226において、選択した最終意図推定学習データ要素のユーザ発話日時情報を、ステップを実行している適応時の日時に更新してもよい。

[0260] (ステップS227)

次に、発話学習適応処理部108は、ステップS227において、選択した最終意図推定学習データ要素の意図(Intent)と、意味概念一致で抽出したエンティティ(Entity)を後段フローに出力して処理を終了する。

すなわち、図4に示すフローのステップS101の処理を終了して、ステップS102に進む。

[0261] [8. その他の実施例について]

次に、その他の実施例について説明する。

以下の実施例について説明する。

(a) 汎用的な自然言語理解(NLU)の学習データの生成を行う構成、

(b) ユーザからの言い換えによる発話学習を行う構成

(c) 学習処理の内容をユーザに提示する構成

## (d) その他の変形例

[0262] 以下、これらの実施例について、順次、説明する。

(a) 汎用的な自然言語理解（NLU）の学習データの生成を行う構成、  
多数の情報処理装置100を、ネットワークを介してサーバに接続する  
。

サーバは、各情報処理装置100において実行された学習データを収集し、  
発話文言と、意図（インテント）と実体情報（エンティティ）との対応データ  
の解析を行う。

[0263] サーバは、例えば、地域（エリア）や年齢などの属性分類毎に学習データ  
を分類することで、属性分類対応の学習データの区分設定を行うことができる。  
また、これらを集積して汎用的なNLUエンジンの学習データとしても  
よい。

[0264] (b) ユーザからの言い換えによる発話学習を行う構成

次に、ユーザからの言い換えによる発話学習を行う構成例について説明す  
る。

図14にユーザからの言い換えによる発話学習フローの例を示す。

[0265] 図14に示すように、ユーザは、まず、ステップS311において、以下の  
の発話をを行う。

ユーザ発話=「品川駅から横浜に行く電車の時間を教えて」

[0266] 情報処理装置100は、ステップS312において、ユーザ発話に対する  
自然言語理解（NLU）結果として、

ユーザ発話の意図（インテント）=電車の時間確認（CHECK\_TIME）、

ユーザ発話の実体情報（エンティティ）=品川駅、横浜方面、  
これらを取得する。

[0267] 情報処理装置100は、この解析結果に基づいて、ステップS313にお  
いて、

装置発話=「8時32分発の東海道線があります」

と装置発話を出力する。

- [0268] これに対して、ユーザは、装置発話出力から、規定時間  $t$  (例えば5秒) 以内に、ステップS 3 1 4において、以下の言い換えを指示する発話、

ユーザ発話=「言い換える」  
を発話する。

- [0269] 情報処理装置100は、装置発話出力から規定時間  $t$  (例えば5秒) 以内に、ユーザからの言い換え指示を入力すると、ステップS 3 1 5において、意図 (インテント) =発話学習 (L E A R N \_ S P E E C H) としたユーザ発話に対する自然言語理解 (N L U) 結果を取得し、さらに、ステップS 3 1 6において、具体的な言い換え文言の発話を求める質問を行う。すなわち、

「何て言われた時に今の操作をしますか？」

上記の装置発話をを行う。

- [0270] これに対して、ユーザは、ステップS 3 1 7において、言い換え発話、  
ユーザ発話=「電車の時間は」  
を実行する。

- [0271] このユーザ発話を入力した情報処理装置100の発話学習適応処理部108は、

意図 (インテント) =電車の時間確認 (C H E C K \_ T R A I N \_ T I M E) 、

ユーザ発話の実体情報 (エンティティ) =品川駅、横浜方面、  
これらを、

ユーザ発話テキスト (s p e e c h) =「電車の時間は」に対して記録した学習データを生成して、発話学習データ記憶部112に格納する。

- [0272] この学習データの生成、記憶後は、ユーザ発話=「電車の時間は」がなされた場合、この登録された学習データを参照して、ユーザ発話=「電車の時間は」が、

意図 (インテント) =電車の時間確認 (C H E C K \_ T R A I N \_ T I M E)

E)、

ユーザ発話の実体情報（エンティティ）＝品川駅、横浜方面、

これらの意図（インテント）と、実体情報（エンティティ）を含む発話であると判定する。

[0273] ユーザがシステム（情報処理装置100）に対して繰り返し（定期的に）行う発話がある場合、毎回同じ内容を全て言わなくてはならないのが面倒になる。このような場合により短く簡単な発話文言をユーザからシステムに覚えさせることにより、繰り返し行う発話の煩わしさを感じにくくすることができる。

[0274] さらに、図15を参照して、ユーザからの言い換えによる発話学習を行う例のもう一つの例について説明する。

[0275] 図15に示すように、ユーザは、まず、ステップS351において、以下の発話を行う。

ユーザ発話＝「明日7時にアラームをセットして」

情報処理装置100の発話学習適応処理部108は、ステップS352において、ユーザ発話に対する自然言語理解（NLU）結果として、

ユーザ発話の意図（インテント）＝アラームセット（ALARM\_SET）、

ユーザ発話の実体情報（エンティティ）＝明日の午前7時、

これらを取得する。

さらに、情報処理装置100の発話学習適応処理部108は、ステップS353において、以下の応答発話を行う。

装置発話＝「明日の午前7時にアラームをセットしました」

[0276] 次に、ユーザは、ステップS353の装置応答から規定時間（t秒）以内にステップS354において、以下の発話を行う。

ユーザ発話＝「玄関のカギ閉めて」

情報処理装置100の発話学習適応処理部108は、ステップS355において、ユーザ発話に対する自然言語理解（NLU）結果として、

ユーザ発話の意図（インテント）＝カギロック（KEY\_LOCK）、

ユーザ発話の実体情報（エンティティ）＝玄関、

これらを取得する。

さらに、情報処理装置100の発話学習適応処理部108は、ステップS356において、以下の応答発話をを行う。

装置発話＝「玄関の鍵を閉めました」

[0277] 次に、ユーザは、ステップS356の装置応答から規定時間（t秒）以内にステップS357において、以下の発話をを行う。

ユーザ発話＝「この部屋の電気を消して」

情報処理装置100の発話学習適応処理部108は、ステップS358において、ユーザ発話に対する自然言語理解（NLU）結果として、

ユーザ発話の意図（インテント）＝電気オフ（LIGHT\_OFF）、

ユーザ発話の実体情報（エンティティ）＝寝室、

これらを取得する。

さらに、情報処理装置100の発話学習適応処理部108は、ステップS359において、以下の応答発話をを行う。

装置発話＝「電気を消しました」

[0278] 次に、ユーザは、ステップS359の装置応答から規定時間（t秒）以内にステップS360において、以下の発話をを行う。

ユーザ発話＝「まとめて覚えて」

情報処理装置100の発話学習適応処理部108は、ステップS361において、ユーザ発話に対する自然言語理解（NLU）結果として、

ユーザ発話の意図（インテント）＝発話学習（LEARN\_SPEECH）

これらを取得し、さらに、ステップS362において、具体的な言い換え文言の発話を求める質問を行う。すなわち、

装置発話＝「何て言われた時に今の操作をしますか？」

上記の装置発話をを行う。

[0279] これに対して、ユーザは、ステップS363において、言い換え発話、  
ユーザ発話=「おやすみ」  
を実行する。

[0280] このユーザ発話を入力した情報処理装置100の発話学習適応処理部108は、

ユーザ発話=「おやすみ」  
に対応付けた学習データとして、以下の3つの意図（インテント）と、実  
体情報（エンティティ）を持つ学習データを生成して記憶部に格納する。

[0281] すなわち、図15の下段に示すように、

(1 a) ユーザ発話の意図（インテント）=アラームセット（A L E R M  
SET）、

(1 b) ユーザ発話の実体情報（エンティティ）=明日の午前7時、

(2 a) ユーザ発話の意図（インテント）=カギロック（K E Y \_ L O C  
K）、

(2 b) ユーザ発話の実体情報（エンティティ）=玄関、

(3 a) ユーザ発話の意図（インテント）=電気オフ（L I G H T \_ O F  
F）、

(3 b) ユーザ発話の実体情報（エンティティ）=寝室、

[0282] 3つの意図（インテント）と、実体情報（エンティティ）を、ユーザ発話  
=「おやすみ」に対応付けた学習データを生成して、記憶部、すなわち発話  
学習データ記憶部112に格納する。

[0283] 最後に、情報処理装置100の発話学習適応処理部108は、ステップS  
364において、以下の処理完了発話をを行う。

装置発話=「「おやすみ」を覚えました」

[0284] この学習以降は「おやすみ」に対して学習した3つの機能が一括実行され  
る。

複数の機能をマクロ的に短く簡単な発話で一括実行できるようになり利便  
性が上がる。

[0285] (c) 学習処理の内容をユーザに提示する構成

次に、学習処理の内容をユーザに提示する構成例について説明する。

情報処理装置100が実行する学習処理の態様は様々であり、ユーザは、情報処理装置がどのような学習処理を実行して、どのような学習データが生成されているかを確認したい場合がある。

[0286] このようなユーザの要望に応じて、学習時にユーザに何を学習した（している）のかを明示して伝えるための表示を情報処理装置100の画像出力部125を介して実行する。

具体的な表示データの例について、図16、図17を参照して説明する。

[0287] 図16(1)に示す学習内容説明表示例1

この学習内容説明表示例1は、

ユーザ発話が、意図解釈不可(OOD)発話であり、図4に示すステップS103のユーザ発話意図の学習対話フローを実行する場合、または、

ユーザ発話が、意図解釈間違い指摘発話であり、図4に示すステップS104のユーザ発話意図の修正対話フローを実行する場合、

これらいずれかの場合の学習内容説明表示例である。

[0288] 具体的には、これらの処理に際して、システム（情報処理装置100）からユーザに、真の発話意図を聞く時の画面表示例である。

この場合、画像出力部125には、

ユーザの意図解釈不可(OOD)発話の発話内容

「鈴木さんの選んで」

を表示するとともに、

学習候補であることを示すアイコン「？」を表示する。

[0289] このような表示を行うことで、ユーザはシステム（情報処理装置100）が何の発話に対する学習をしようとしているのかを理解することができる。

[0290] 図16(2)に示す学習内容説明表示例2

この学習内容説明表示例2は、例えば先に図14、図15を参照して説明したユーザからの言い換え指示の発話があった場合の表示例である。

システム（情報処理装置100）が、ユーザに言い換え文言を聞く時の画面表示例に相当する。

- [0291] この場合、画像出力部125には、  
    言い換えの対象となる（複数の）発話内容と、  
    その意図（インテント）や、実体情報（エンティティ）を明示する。  
    さらに、実行順を示す番号（のアイコン）を明示する。
- [0292] この表示により、ユーザは何を一括実行するのかを確認できるため、言い換え文言を言いやすくなる。
- [0293] 図17（3）に示す学習内容説明表示例3は、例えば図4に示すステップS103のユーザ発話意図の学習対話フローを実行する場合、または、図14、図15を参照して説明したユーザからの言い換え指示の発話があった場合などにおいて、学習データを発話学習データ記憶部112に格納する場合の表示例である。
- [0294] この場合、画像出力部125には、  
    学習したデータの発話文言と、  
    その意図（インテント）や、実体情報（エンティティ）を表示する。  
    なお、この表示処理においては、  
        既存学習済みのデータに対して今回学習したデータを、輝度、またはフォントを区別する、またはアニメーション表示等を行うことによって強調表示することが好ましい。
- [0295] このような表示処理を実行することで、ユーザはシステム（情報処理装置100）が学習データを発話学習データ記憶部112に記憶した発話文言とその機能を確認することが可能となり、以降その発話文言を使うようになる。
- [0296] (d) その他の変形例  
    次に、その他の変形例について説明する。  
    上述した本開示の情報処理装置100の発話学習適応処理部108は、さらに、以下の処理を行う構成としてもよい。

## [0297] (d 1) 変形例 1

ユーザ発話の自然言語理解（NLU）結果として複数の意図が取得できた場合、システム（情報処理装置100）からユーザに複数の意図のうちどれがユーザの真の発話意図であるかを質問し、その回答を最初のユーザ発話の意図として学習してもよい。

[0298] 例えば、先に説明した図11のシーケンスのように、ユーザの「見せて」という発話に対し、自然言語理解（NLU）結果として [SHOW\_TASK] と [SHOW\_MESSAGE] の2つの意図（インテント）が取得できた場合、以下の処理を行う。

[0299] システム（情報処理装置100）からユーザに、  
「タスク一覧とメッセージ一覧のどちらを見ますか？」  
という質問発話をを行い、ユーザから、  
「メッセージ見せて」  
という回答があったら、最初のユーザ発話「見せて」を、  
意図（インテント） = SHOW\_MESSAGE  
として学習する。

[0300] 一方、ユーザ発話が、「タスク見せて」という回答なら、  
意図（インテント） = SHOW\_TASK  
として学習する。

## [0301] (d 2) 変形例 2

ユーザ発話の自然言語理解（NLU）結果として取得した意図の信頼度スコアが低い場合、システム（情報処理装置100）からユーザに信頼度スコアの低い意図がユーザの真の発話意図であるかを質問し、その回答を最初のユーザ発話の意図として学習してもよい。

[0302] 例えば、先に図7を参照して説明したでユーザの  
「桑原圭祐にして」  
という発話に対し、自然言語理解（NLU）結果として  
意図（インテント） = 音楽再生（PLAY\_MUSIC）、

実体情報（エンティティ）＝桑原圭祐、

これらのデータの信頼度スコアが、予め規定したしきい値より低い場合、機能実行（音楽再生）を行う前にシステム（情報処理装置100）からユーザに、

「桑原圭祐の音楽再生でよろしいですか？」

という質問発話をを行う。

[0303] この質問に対して、ユーザから「いいよ」という回答があつたら最初のユーザ発話「桑原圭祐にして」を、

意図（インテント）＝音楽再生（PLAY\_MUSIC）、

実体情報（エンティティ）＝桑原圭祐、

として学習する。

[0304] 一方、ユーザからの回答が、「違うよ」という回答なら、

システム（情報処理装置100）から、

「桑原圭祐にしての意味を教えてください」

と質問を行い、図7を参照して説明した処理を実行する。

[0305] (d3) 変形例3

ユーザからの発話指示により、学習を行わないモード（非学習モード）に入つてもよい。

例えば、ユーザから「言ったこと覚えないで」と発話されると、非学習モードに入り、以降は学習データの生成や修正を行わない。

[0306] この非学習モードでは、ユーザの「意図解釈不可（OOD）発話」や、間違い指摘発話時に「～～の意味を教えてください」というシステム（情報処理装置100）からの発話をを行い、ユーザの真の発話意図を取得して機能実行までは行う。

しかし、学習データを発話学習データ記憶部112に記憶しない。

また、ユーザから「言ったこと覚えて」と発話されると非学習モードから抜けて、その後は、通常の学習処理を行う。

これは、例えば発話ユーザのプライバシーに関するデータを記憶部に記憶

しないために有効な構成である。

[0307] (d 4) 変形例4

あるユーザ発話の学習時に外部連携機器の使用状況を含めて学習し、新たなユーザ発話の意図解析時にも外部連携機器の使用状況を取得し、学習データに記録された外部連携機器使用状況と比較して、同じ使用状況のデータを持つ学習データを適用するようにしてもよい。

[0308] このように、機器使用状況をユーザ発話の学習データに記録する構成とすることで、以下のような処理が可能となる。

例えば、テレビが電源OFFの時に学習したユーザ発話「つけて」の、意図（インテント）＝テレビ電源ON（POWERON\_TV）は、テレビが電源ONの時のユーザ発話「つけて」には適用しない。テレビが電源OFFの時のユーザ発話「つけて」にのみ適用する。このような処理が可能となる。

[0309] また、仕事用のPCを使用中（仕事中）に学習したユーザ発話の意図（インテント）は、仕事用のPCを使用していない時（プライベート）には適用しない。

このような設定とすることで、外部連携機器の使用状況に応じた最適な処理（機能実行）が可能となる。

[0310] (d 5) 変形例5

画像やセンサーなどの情報から学習時のユーザ状況を認識し、ユーザが余裕のある時に真の意図を取得して学習を行う。

ユーザの動きが大きい場合や、部屋出て行こうとしている急いでいる状況の時はユーザの「意図解釈不可（OOD）発話」、例えば、

「見せて」に対して、

「分かりません」とシステム（情報処理装置100）が応答を行う。

[0311] 一方、ユーザが部屋でくつろいでいる（急いでなく余裕がある）時に真の意図を取得するため、システム（情報処理装置100）は、以下の発話をを行う。

「朝言われた「見せて」の意味を教えてください」

この質問発話に基づくユーザ発話を入力して学習を行う。

[0312] このような状況に応じた学習を行うことで、ユーザの余裕のある時間のみにユーザ対話に基づく学習を行うことが可能となる。

[0313] [9. 本開示の情報処理装置の状態遷移と処理の効果について]

次に、本開示の情報処理装置の状態遷移と処理の効果について説明する。

[0314] 図18は、本開示の情報処理装置100の、個々のユーザ発話に対する学習データのライフサイクル処理に着目した状態遷移の概念を説明する図である。

本開示の情報処理装置100は図18に示すように、以下の5つの状態を持ち、実行する処理に応じて各状態間を遷移する。

状態1 = 対話状態

状態2 = 学習状態

状態3 = 適応状態

状態4 = 修正状態

状態5 = 忘却状態

これらの状態は、情報処理装置100の発話学習適応処理部108の状態に相当し、発話学習適応処理部108の実行する処理に応じて、各状態を遷移する。

[0315] 状態1 = 対話状態は、情報処理装置100とユーザ間の対話をを行っている状態である。

情報処理装置100が対話によりユーザ意図（インテント）を獲得すると

、

状態2 = 学習状態

に移行する。

[0316] 状態2 = 学習状態では、処理に応じて状態3 = 適応状態との状態遷移を行う。

新たなユーザ発話が入力されると発話学習データ記憶部112に格納され

た学習データを利用するための記憶照会を行うことにより、状態3＝適応状態に移行する。

状態3＝適応状態で新たなユーザ発話に対して学習データの適用が行われると、学習データ内のユーザ発話回数情報を更新して長期記憶となる記憶強化が行われる。

- [0317] また、状態2＝学習状態では先に図10を参照して説明した学習データ忘却スコア $f(t, c)$ に従った学習データの破棄判定処理が適宜行われ、自動破棄と判定されると

状態5＝忘却状態

に移行し学習データは破棄される。

なお、発話学習適応処理部108は、先に図10を参照して説明したように、発話学習データ記憶部112に格納された学習データの各々について、以下の（式1）に従って学習データ忘却スコア $f(t, c)$ を算出する。

$$f(t, c) = (k_c / c) \log(k_t t + 1) \dots \text{(式1)}$$

- [0318] この算出値と、完全忘却しきい値を比較して、比較結果に応じて、学習データの破棄を判定する。

- [0319] 状態4＝修正状態は、例えば、図4に示すフロー中のステップS104の処理の実行状態に相当する。

すなわち図11に示すシーケンスに従って実行される学習データの修正処理を行う状態である。

この処理により、発話学習データ記憶部112に格納された学習データの修正や、破棄が行われる。

- [0320] 例えば、図4のフローのステップS104の処理、具体的には、図11を参照して説明した処理中、ステップS154のユーザ発話「ちがうよ」を入力し5秒以上経過すると、情報処理装置100は、違うと指摘された学習データを記憶部から削除する。

例えば、この処理が、図18に示す状態3＝適応状態から状態4＝修正状態への移行（ユーザ指摘）と、状態4＝修正状態から状態5＝忘却状態への

移行（手動破棄）に相当する。

[0321] また、新たなユーザ発話によって正しい学習データに更新されると、状態4＝修正状態から状態2＝学習状態に移行し、記憶部に正しい学習データの格納（記憶更新）が実行される。

[0322] このように、本開示の情報処理装置100は、複数の異なる状態を有し、実行する処理に応じて各状態間を遷移する。

情報処理装置100が、様々なユーザ発話を入力し、ユーザとの対話を実行することで、図18に示す状態遷移が継続して実行される。その結果、発話学習データ記憶部112に格納される学習データは、ユーザ発話の正確な意図把握が可能となるように最適化される。

[0323] 本開示の情報処理装置100の実行する処理による効果には、例えば以下の効果がある。

(1) システム（情報処理装置100）がユーザ固有の言い回しとその意図を使用中に学習し適応することで、ユーザは従来の汎用的に発話を学習したNLUエンジンではカバーしきれない多様な自然発話による音声入力が可能となる。

(2) ユーザは人に話しかけるように自然発話でシステム（情報処理装置100）に話しかけることができるようになるため、音声入力の負荷が下がり継続的に音声対話システム（情報処理装置100）を使用できるようになる。

[0324] (3) システム（情報処理装置100）は、使い込むほどに発話を学習して解釈できる固有表現が増えていくため、ユーザはシステムが自分に合わせて成長している感覚を得ることができ、使用頻度の向上や長期間の使用につながる。

(4) システム（情報処理装置100）は、ユーザが定常に使用しない（誤学習データを含む）発話学習データを自動で破棄し、またユーザ発話によっても学習データを修正・破棄が可能となるため、システム（情報処理装置100）のユーザ発話意図とは異なる誤動作の低減ができる。

[0325] (5) 使用される見込みのない発話学習データは自動破棄されるため、不要なデータの保持によるシステムリソースの消費を抑えることができる。

(6) 個人化学習データを汎用的な自然言語理解（NLU）エンジンの学習データとして使用することにより、その時々でよく使用されている発話（ある程度定着している流行りの言い回し）の学習データを効率的に自動収集することが可能となる。

[0326] [10. 情報処理装置、および情報処理システムの構成例について]

複数の実施例について説明したが、これらの実施例において説明した様々な処理機能、例えば、図3に示す情報処理装置100の各構成要素の処理機能は、すべて一つの装置、例えばユーザの所有するエージェント機器、あるいはスマホやPC等の装置内に構成することも可能であるが、その一部をサーバ等において実行する構成とすることも可能である。

[0327] 図19にシステム構成例を示す。

図19(1) 情報処理システム構成例1は、図3に示す情報処理装置のほぼ全ての機能を一つの装置、例えばユーザの所有するスマホやPC、あるいは音声入出力と画像入出力機能を持つエージェント機器等のユーザ端末である情報処理装置410内に構成した例である。

ユーザ端末に相当する情報処理装置410は、例えば応答文生成時に外部アプリを利用する場合にのみ、アプリ実行サーバ420と通信を実行する。

[0328] アプリ実行サーバ420は、例えば天気情報提供サーバ、交通情報提供サーバ、医療情報提供サーバ、観光情報提供サーバ等であり、ユーザ発話に対する応答を生成するための情報を提供可能なサーバ群によって構成される。

[0329] 一方、図19(2) 情報処理システム構成例2は、図3に示す情報処理装置の機能の一部をユーザの所有するスマホやPC、エージェント機器等のユーザ端末である情報処理装置410内に構成し、一部を情報処理装置と通信可能なデータ処理サーバ460において実行する構成としたシステム例である。

[0330] 例えば、図3に示す装置中の音声入力部101、画像入力部104、セン

サー106、音声出力部123、画像出力部125のみをユーザ端末側の情報処理装置410側に設け、その他の機能をすべてサーバ側で実行するといった構成等が可能である。

なお、ユーザ端末側の機能と、サーバ側の機能の機能分轄態様は、様々な異なる設定が可能であり、また、1つの機能を双方で実行する構成も可能である。

[0331] [11. 情報処理装置のハードウェア構成例について]

次に、図20を参照して、情報処理装置のハードウェア構成例について説明する。

図20を参照して説明するハードウェアは、先に図3を参照して説明した情報処理装置のハードウェア構成例であり、また、図19を参照して説明したデータ処理サーバ460を構成する情報処理装置のハードウェア構成の一例である。

[0332] CPU (Central Processing Unit) 501は、ROM (Read Only Memory) 502、または記憶部508に記憶されているプログラムに従って各種の処理を実行する制御部やデータ処理部として機能する。例えば、上述した実施例において説明したシーケンスに従った処理を実行する。RAM (Random Access Memory) 503には、CPU501が実行するプログラムやデータなどが記憶される。これらのCPU501、ROM502、およびRAM503は、バス504により相互に接続されている。

[0333] CPU501はバス504を介して入出力インターフェース505に接続され、入出力インターフェース505には、各種スイッチ、キーボード、マウス、マイクロホン、センサーなどによる入力部506、ディスプレイ、スピーカーなどによる出力部507が接続されている。CPU501は、入力部506から入力される指令に対応して各種の処理を実行し、処理結果を例えば出力部507に出力する。

[0334] 入出力インターフェース505に接続されている記憶部508は、例えばハ

ードディスク等からなり、CPU501が実行するプログラムや各種のデータを記憶する。通信部509は、Wi-Fi通信、ブルートゥース（登録商標）（BT）通信、その他インターネットやローカルエリアネットワークなどのネットワークを介したデータ通信の送受信部として機能し、外部の装置と通信する。

[0335] 入出力インターフェース505に接続されているドライブ510は、磁気ディスク、光ディスク、光磁気ディスク、あるいはメモリカード等の半導体メモリなどのリムーバブルメディア511を駆動し、データの記録あるいは読み取りを実行する。

[0336] [12. 本開示の構成のまとめ]

以上、特定の実施例を参照しながら、本開示の実施例について詳解してきた。しかしながら、本開示の要旨を逸脱しない範囲で当業者が実施例の修正や代用を成し得ることは自明である。すなわち、例示という形態で本発明を開示してきたのであり、限定的に解釈されるべきではない。本開示の要旨を判断するためには、特許請求の範囲の欄を参照すべきである。

[0337] なお、本明細書において開示した技術は、以下のような構成をとることができる。

(1) ユーザ発話の意図（インテント）を解析する発話学習適応処理部を有し、

前記発話学習適応処理部は、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する情報処理装置。

[0338] (2) 前記発話学習適応処理部は、

意図の不明確な第1のユーザ発話の入力後に、

意図確認のための装置発話を実行して、該装置発話に対するユーザからの応答発話である第2のユーザ発話を入力し、

前記第2のユーザ発話の解析結果として取得した意図を、前記第1のユ

ザ発話の発話テキストと実体情報に対応付けて記録した学習データを生成する（1）に記載の情報処理装置。

[0339] （3） 前記発話学習適応処理部は、

意図の不明確な第1のユーザ発話の入力後に、

意図確認のための装置発話を実行して、該装置発話に対するユーザからの応答発話である第2のユーザ発話を入力し、

（a）前記第1のユーザ発話の発話テキストと、

（b）前記第1のユーザ発話と前記第2のユーザ発話に共通に含まれる実体情報と、

（c）前記第2のユーザ発話の解析結果として取得した正しい意図、

これらを対応付けた学習データを生成する（1）または（2）に記載の情報処理装置。

[0340] （4） 前記発話学習適応処理部は、

前記学習データに前記実体情報の上位意味概念情報を含めて記録する（1）

）～（3）いずれかに記載の情報処理装置。

[0341] （5） 前記上位意味概念情報は、

前記実体情報の種類を示す情報である（4）に記載の情報処理装置。

[0342] （6） 前記発話学習適応処理部は、

前記学習データに前記実体情報の上位意味概念情報を含む自然言語処理情報を作成する（1）～（5）いずれかに記載の情報処理装置。

[0343] （7） 前記発話学習適応処理部は、

新たなユーザ発話の意図解析処理に際して、該新たなユーザ発話に含まれる実体情報の上位意味概念情報を取得し、取得情報と類似する上位意味概念情報を記録した学習データを記憶部から抽出し、

前記新たなユーザ発話に含まれる実体情報以外の発話テキストと、抽出した学習データに記録された発話テキストとの一致が確認された場合、抽出した学習データに記録された意図を、前記新たなユーザ発話の意図として推定する（1）～（6）いずれかに記載の情報処理装置。

[0344] (8) 前記発話学習適応処理部は、

ユーザから意図解釈間違いを指摘するユーザ発話を入力した場合、  
前記ユーザからの再発話に基づいて正しい意図を取得し、  
意図推定に適用した学習データに記録された誤った意図情報を、正しい意  
図情報に修正する処理を行う（1）～（7）いずれかに記載の情報処理装置  
。

[0345] (9) 前記学習データは、

ユーザ発話実行時の対話状態（ステート）に応じて区分されたステート単  
位の複数の学習データから構成されている（1）～（8）いずれかに記載の  
情報処理装置。

[0346] (10) 前記発話学習適応処理部は、

記憶部に格納された学習データの利用状況に応じて破棄、または利用停止  
を行う（1）～（9）いずれかに記載の情報処理装置。

[0347] (11) 前記発話学習適応処理部は、

記憶部に格納された学習データ各々について、  
学習データ対応の発話をユーザが最後に発話してから経過した時間：t と  
、  
学習データ対応の発話をユーザが繰り返した回数：c、  
これらのパラメータに基づいて、学習データ忘却スコア  $f(t, c)$  を算  
出し、  
算出した学習データ忘却スコア  $f(t, c)$  の値と規定しきい値とを比較  
して、各学習データの破棄、または利用停止を行う（10）に記載の情報処  
理装置。

[0348] (12) 前記学習データは、

ユーザ発話実行時のコンテキスト情報を記録した学習データであり、  
前記発話学習適応処理部は、  
新たなユーザ発話入力時のコンテキスト情報を取得し、取得情報と類似す  
るコンテキスト情報を記録した学習データを記憶部から抽出し、抽出した学

習データに記録された意図を、前記新たなユーザ発話の意図として推定する

(1) ~ (11) いずれかに記載の情報処理装置。

[0349] (13) 前記発話学習適応処理部は、

ユーザ発話が意図解釈不可発話である場合に実行されたユーザからの言い直し発話を解析し、

前記ユーザ発話と、前記言い直し発話との意味概念の相違度が予め規定したしきい値以上である場合、前記言い直し発話に基づく学習処理を実行しない (1) ~ (12) いずれかに記載の情報処理装置。

[0350] (14) 前記発話学習適応処理部は、

意図解釈不可発話に対する言い直し発話を行った第1のユーザと、第2のユーザが一緒にいることがカメラ撮影画像から確認された場合、または、

前記第1のユーザが、前記第2のユーザから学習発話を許可されたユーザであることが確認された場合、

前記第1のユーザによる言い直し発話に基づく学習データを、前記第2のユーザの新たな発話に適応して意図推定処理を実行する (1) ~ (13) いずれかに記載の情報処理装置。

[0351] (15) 前記発話学習適応処理部は、

先行して実行されたユーザ発話の意図に対するユーザからの発話の対応付けの指示として、ユーザ発話の言い換え要求を入力した場合、

言い換え要求後のユーザ発話に、先行して実行されたユーザ発話に対応する意図および実体情報を対応付けた学習データを生成する (1) ~ (14) いずれかに記載の情報処理装置。

[0352] (16) 前記発話学習適応処理部は、

複数のユーザ発話に対応する複数の意図を、1つのユーザ発話に対応付けた学習データを生成する (1) ~ (15) いずれかに記載の情報処理装置。

[0353] (17) ユーザ端末と、データ処理サーバを有する情報処理システムであり、

前記ユーザ端末は、

ユーザ発話を入力する音声入力部を有し、

前記データ処理サーバは、

前記ユーザ端末から受信する前記ユーザ発話の意図（インテント）を解析する発話学習適応処理部を有し、

前記発話学習適応処理部は、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する情報処理システム。

[0354] (18) 情報処理装置において実行する情報処理方法であり、

音声入力部が、ユーザ発話を入力し、

発話学習適応処理部が、前記ユーザ発話の意図（インテント）を解析する発話学習適応処理を実行し、

前記発話学習適応処理において、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する処理を実行する情報処理方法。

[0355] (19) ユーザ端末と、データ処理サーバを有する情報処理システムに

おいて実行する情報処理方法であり、

前記ユーザ端末が、

ユーザ発話を入力する音声入力処理を実行し、

前記データ処理サーバが、

前記ユーザ端末から受信する前記ユーザ発話の意図（インテント）を解析する発話学習適応処理を実行し、

前記発話学習適応処理において、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する処理を実行する情報処理方法。

[0356] (20) 情報処理装置において情報処理を実行させるプログラムであり

、  
音声入力部に、ユーザ発話を入力させる音声入力ステップと、  
発話学習適応処理部に、前記ユーザ発話の意図（インテント）を解析する  
発話学習適応処理を実行させ、  
前記発話学習適応処理において、  
ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ  
発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習  
データを生成して記憶部に格納する処理を実行させるプログラム。

[0357] また、明細書中において説明した一連の処理はハードウェア、またはソフトウェア、あるいは両者の複合構成によって実行することが可能である。ソフトウェアによる処理を実行する場合は、処理シーケンスを記録したプログラムを、専用のハードウェアに組み込まれたコンピュータ内のメモリにインストールして実行させるか、あるいは、各種処理が実行可能な汎用コンピュータにプログラムをインストールして実行させることができる。例えば、プログラムは記録媒体に予め記録しておくことができる。記録媒体からコンピュータにインストールする他、LAN（Local Area Network）、インターネットといったネットワークを介してプログラムを受信し、内蔵するハードディスク等の記録媒体にインストールすることができる。

[0358] なお、明細書に記載された各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。また、本明細書においてシステムとは、複数の装置の論理的集合構成であり、各構成の装置が同一筐体内にあるものには限らない。

## 産業上の利用可能性

[0359] 以上、説明したように、本開示の一実施例の構成によれば、ユーザ発話の高精度な意図推定を可能とした装置、方法が実現される。

具体的には、例えば、発話学習適応処理部が、ユーザから入力する複数の

ユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報と、正しい意図を対応付けた学習データを生成して記憶部に格納する。発話学習適応処理部は、意図の不明確な第1のユーザ発話の入力後の装置発話に対するユーザからの応答発話から取得した意図を、第1のユーザ発話に含まれる実体情報に対応付けて記録した学習データを生成する。学習データには実体情報の上位意味概念情報を含めて記録する。新たなユーザ発話に対する意図推定時には、上位意味概念情報の類似する学習データを利用する。

本構成により、ユーザ発話の高精度な意図推定を可能とした装置、方法が実現される。

## 符号の説明

[0360] 10 情報処理装置

11 カメラ

12 マイク

13 表示部

14 スピーカー

20 サーバ

30 外部機器

100 情報処理装置

101 音声入力部

102 音声認識部

103 発話意味解析部

104 画像入力部

105 画像認識部

106 センサー

107 センサー情報解析部

108 発話学習適応処理部

111 知識辞書データ記憶部

112 発話学習データ記憶部

- 1 2 1 応答生成部
- 1 2 2 音声合成部
- 1 2 3 音声出力部
- 1 2 4 表示画像生成部
- 1 2 5 画像出力部
- 4 1 0 情報処理装置
- 4 2 0 アプリ実行サーバ
- 4 6 0 データ処理サーバ
- 5 0 1 C P U
- 5 0 2 R O M
- 5 0 3 R A M
- 5 0 4 バス
- 5 0 5 入出力インターフェース
- 5 0 6 入力部
- 5 0 7 出力部
- 5 0 8 記憶部
- 5 0 9 通信部
- 5 1 0 ドライブ
- 5 1 1 リムーバブルメディア

## 請求の範囲

[請求項1] ユーザ発話の意図（インテント）を解析する発話学習適応処理部を有し、

前記発話学習適応処理部は、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する情報処理装置。

[請求項2] 前記発話学習適応処理部は、

意図の不明確な第1のユーザ発話の入力後に、

意図確認のための装置発話を実行して、該装置発話に対するユーザからの応答発話である第2のユーザ発話を入力し、

前記第2のユーザ発話の解析結果として取得した意図を、前記第1のユーザ発話の発話テキストと実体情報に対応付けて記録した学習データを生成する請求項1に記載の情報処理装置。

[請求項3] 前記発話学習適応処理部は、

意図の不明確な第1のユーザ発話の入力後に、

意図確認のための装置発話を実行して、該装置発話に対するユーザからの応答発話である第2のユーザ発話を入力し、

(a) 前記第1のユーザ発話の発話テキストと、

(b) 前記第1のユーザ発話と前記第2のユーザ発話に共通に含まれる実体情報と、

(c) 前記第2のユーザ発話の解析結果として取得した正しい意図、

これらを対応付けた学習データを生成する請求項1に記載の情報処理装置。

[請求項4] 前記発話学習適応処理部は、

前記学習データに前記実体情報の上位意味概念情報を含めて記録する請求項1に記載の情報処理装置。

- [請求項5] 前記上位意味概念情報は、  
前記実体情報の種類を示す情報である請求項4に記載の情報処理装置。
- [請求項6] 前記発話学習適応処理部は、  
前記学習データに前記実体情報の上位意味概念情報を含む自然言語処理情報を含めて記録する請求項1に記載の情報処理装置。
- [請求項7] 前記発話学習適応処理部は、  
新たなユーザ発話の意図解析処理に際して、該新たなユーザ発話に含まれる実体情報の上位意味概念情報を取得し、取得情報と類似する上位意味概念情報を記録した学習データを記憶部から抽出し、  
前記新たなユーザ発話に含まれる実体情報以外の発話テキストと、抽出した学習データに記録された発話テキストとの一致が確認された場合、抽出した学習データに記録された意図を、前記新たなユーザ発話の意図として推定する請求項1に記載の情報処理装置。
- [請求項8] 前記発話学習適応処理部は、  
ユーザから意図解釈間違いを指摘するユーザ発話を入力した場合、前記ユーザからの再発話に基づいて正しい意図を取得し、意図推定に適用した学習データに記録された誤った意図情報を、正しい意図情報に修正する処理を行う請求項1に記載の情報処理装置。
- [請求項9] 前記学習データは、  
ユーザ発話実行時の対話状態（ステート）に応じて区分されたステート単位の複数の学習データから構成されている請求項1に記載の情報処理装置。
- [請求項10] 前記発話学習適応処理部は、  
記憶部に格納された学習データの利用状況に応じて破棄、または利用停止を行う請求項1に記載の情報処理装置。
- [請求項11] 前記発話学習適応処理部は、  
記憶部に格納された学習データ各々について、

学習データ対応の発話をユーザが最後に発話してから経過した時間 : t と、

学習データ対応の発話をユーザが繰り返した回数 : c、

これらのパラメータに基づいて、学習データ忘却スコア  $f(t, c)$  を算出し、

算出した学習データ忘却スコア  $f(t, c)$  の値と規定しきい値とを比較して、各学習データの破棄、または利用停止を行う請求項 10 に記載の情報処理装置。

[請求項12] 前記学習データは、

ユーザ発話実行時のコンテキスト情報を記録した学習データであり、

前記発話学習適応処理部は、

新たなユーザ発話入力時のコンテキスト情報を取得し、取得情報と類似するコンテキスト情報を記録した学習データを記憶部から抽出し、抽出した学習データに記録された意図を、前記新たなユーザ発話の意図として推定する請求項 1 に記載の情報処理装置。

[請求項13] 前記発話学習適応処理部は、

ユーザ発話が意図解釈不可発話である場合に実行されたユーザからの言い直し発話を解析し、

前記ユーザ発話と、前記言い直し発話との意味概念の相違度が予め規定したしきい値以上である場合、前記言い直し発話に基づく学習処理を実行しない請求項 1 に記載の情報処理装置。

[請求項14] 前記発話学習適応処理部は、

意図解釈不可発話に対する言い直し発話を行った第 1 のユーザと、第 2 のユーザが一緒にいることがカメラ撮影画像から確認された場合、または、

前記第 1 のユーザが、前記第 2 のユーザから学習発話を許可されたユーザであることが確認された場合、

前記第1のユーザによる言い直し発話に基づく学習データを、前記第2のユーザの新たな発話に適応して意図推定処理を実行する請求項1に記載の情報処理装置。

- [請求項15] 前記発話学習適応処理部は、  
先行して実行されたユーザ発話の意図に対するユーザからの発話の対応付けの指示として、ユーザ発話の言い換え要求を入力した場合、言い換え要求後のユーザ発話に、先行して実行されたユーザ発話に対応する意図および実体情報を対応付けた学習データを生成する請求項1に記載の情報処理装置。
- [請求項16] 前記発話学習適応処理部は、  
複数のユーザ発話に対応する複数の意図を、1つのユーザ発話に対応付けた学習データを生成する請求項1に記載の情報処理装置。
- [請求項17] ユーザ端末と、データ処理サーバを有する情報処理システムであり、  
前記ユーザ端末は、  
ユーザ発話を入力する音声入力部を有し、  
前記データ処理サーバは、  
前記ユーザ端末から受信する前記ユーザ発話の意図（インテント）を解析する発話学習適応処理部を有し、  
前記発話学習適応処理部は、  
ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する情報処理システム。
- [請求項18] 情報処理装置において実行する情報処理方法であり、  
音声入力部が、ユーザ発話を入力し、  
発話学習適応処理部が、前記ユーザ発話の意図（インテント）を解析する発話学習適応処理を実行し、  
前記発話学習適応処理において、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する処理を実行する情報処理方法。

[請求項19] ユーザ端末と、データ処理サーバを有する情報処理システムにおいて実行する情報処理方法であり、

前記ユーザ端末が、

ユーザ発話を入力する音声入力処理を実行し、

前記データ処理サーバが、

前記ユーザ端末から受信する前記ユーザ発話の意図（インテント）を解析する発話学習適応処理を実行し、

前記発話学習適応処理において、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する処理を実行する情報処理方法。

[請求項20] 情報処理装置において情報処理を実行させるプログラムであり、

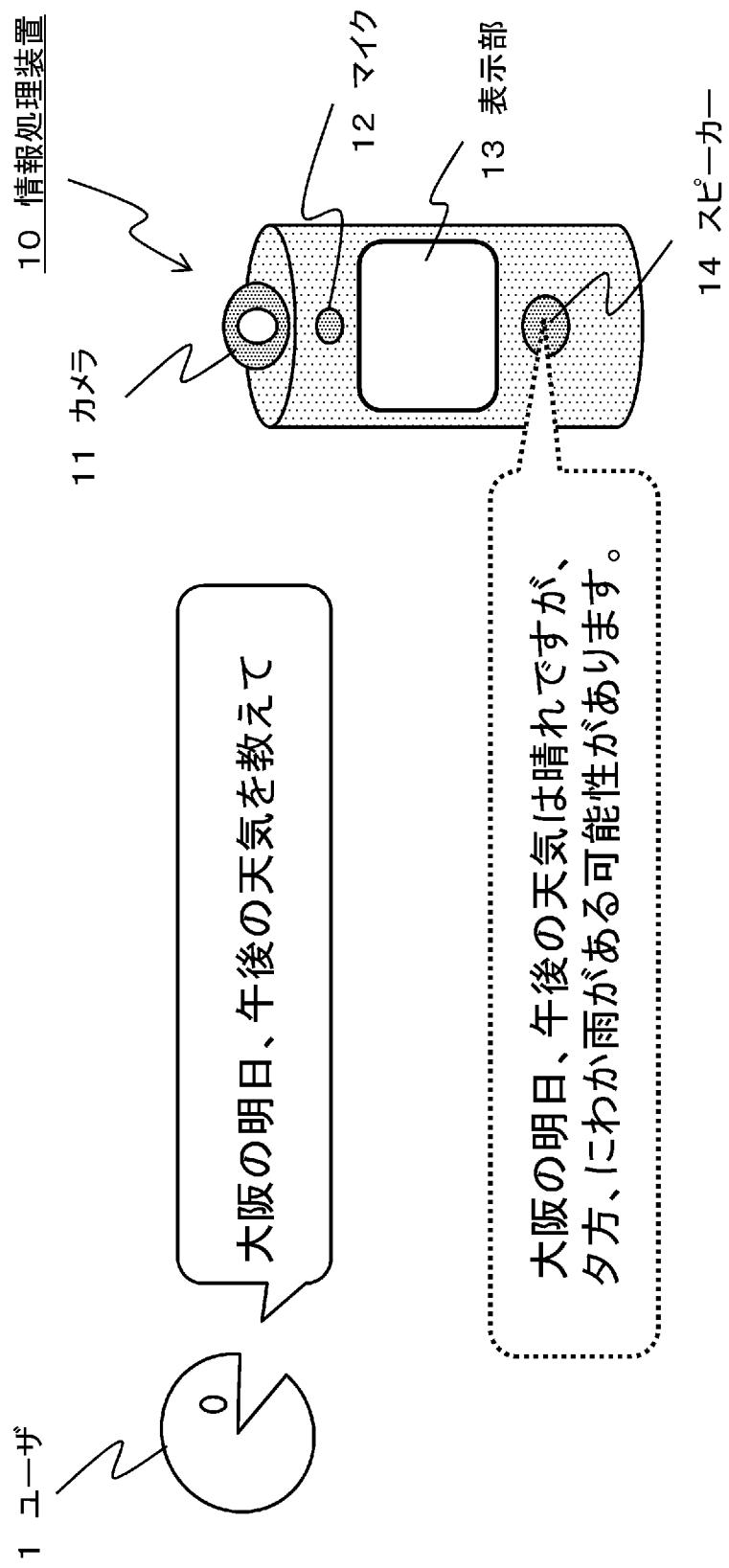
音声入力部に、ユーザ発話を入力させる音声入力ステップと、

発話学習適応処理部に、前記ユーザ発話の意図（インテント）を解析する発話学習適応処理を実行させ、

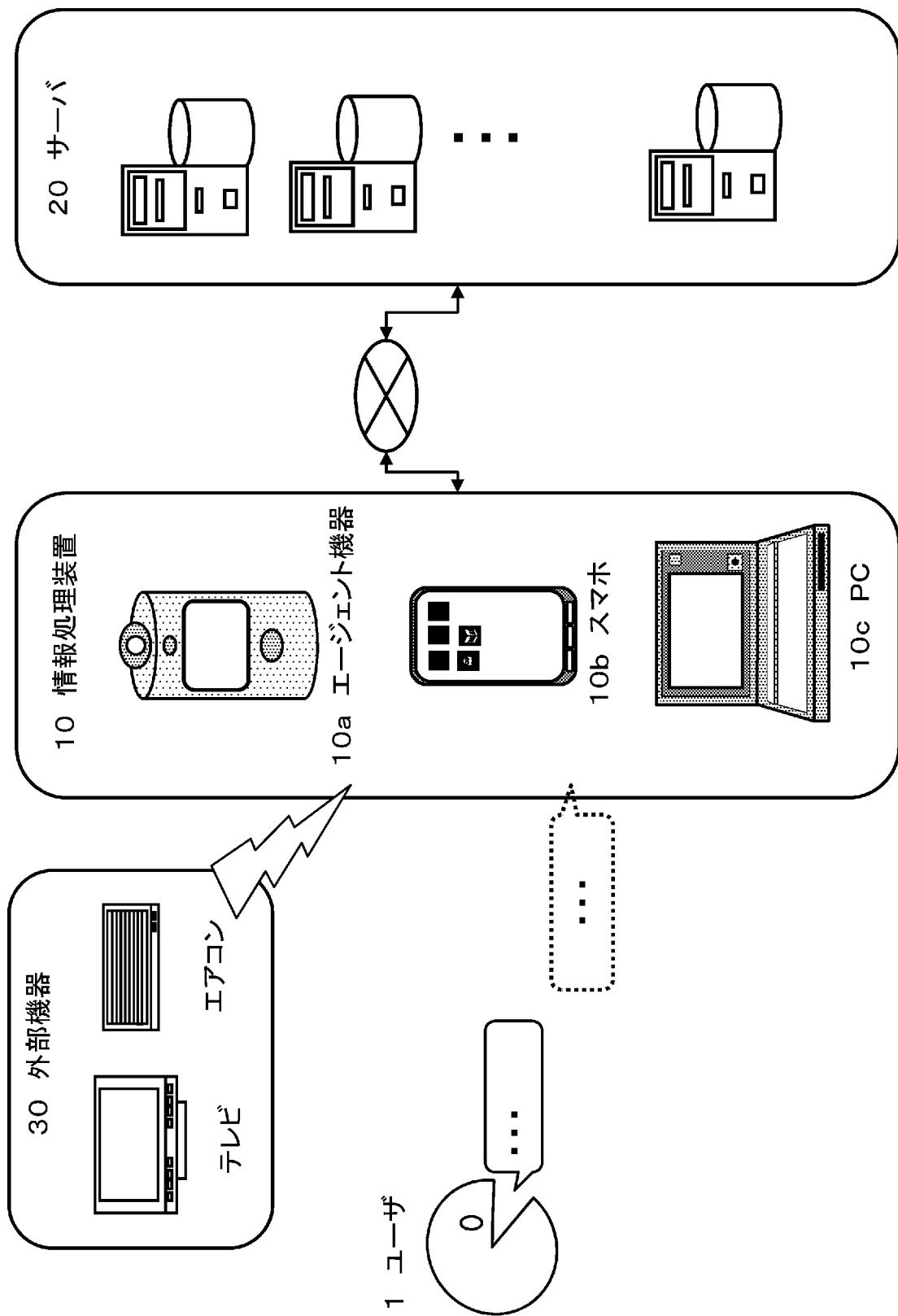
前記発話学習適応処理において、

ユーザから入力する複数のユーザ発話を解析して、意図の不明確なユーザ発話に含まれる実体情報（エンティティ）と、正しい意図を対応付けた学習データを生成して記憶部に格納する処理を実行させるプログラム。

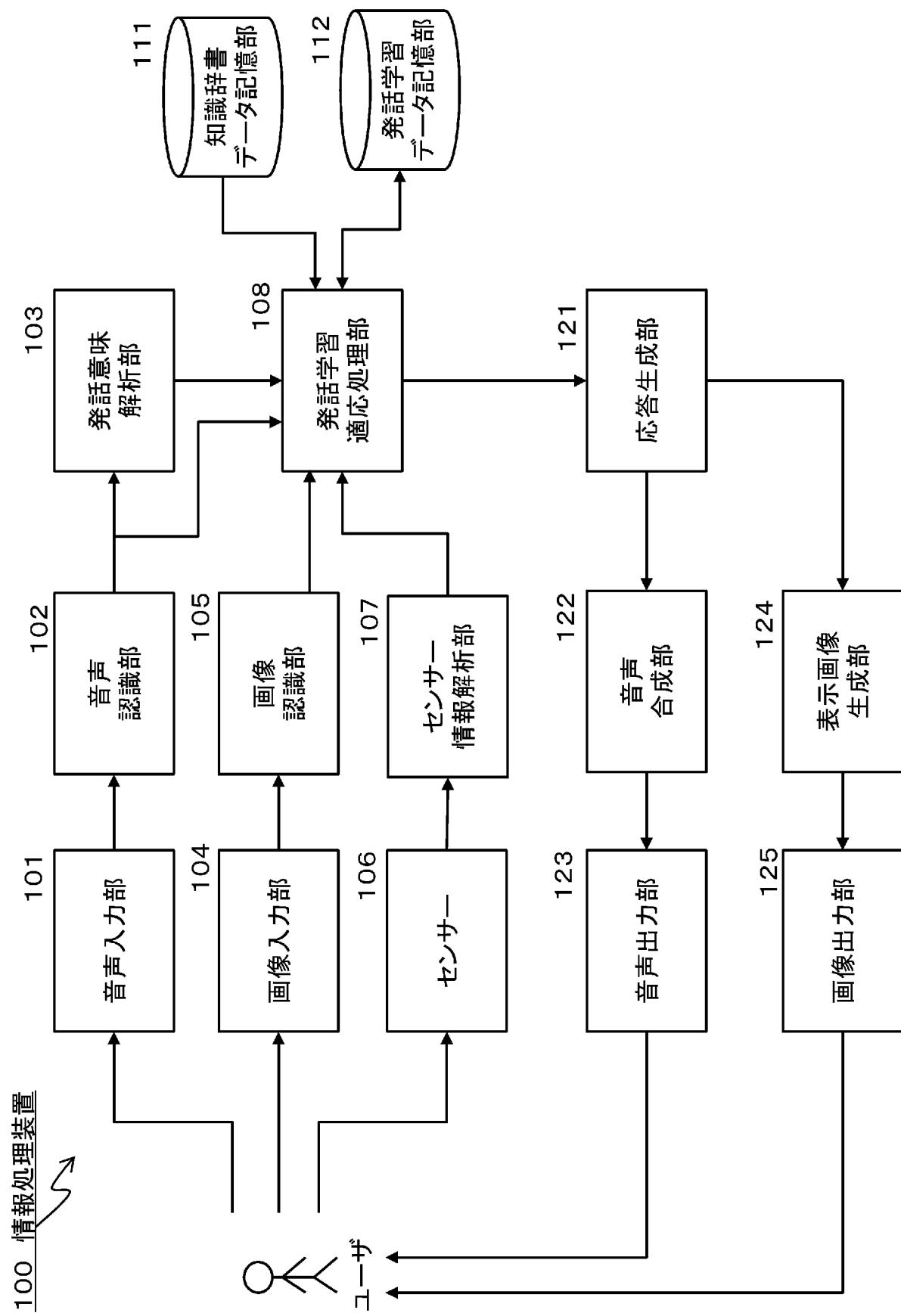
[図1]



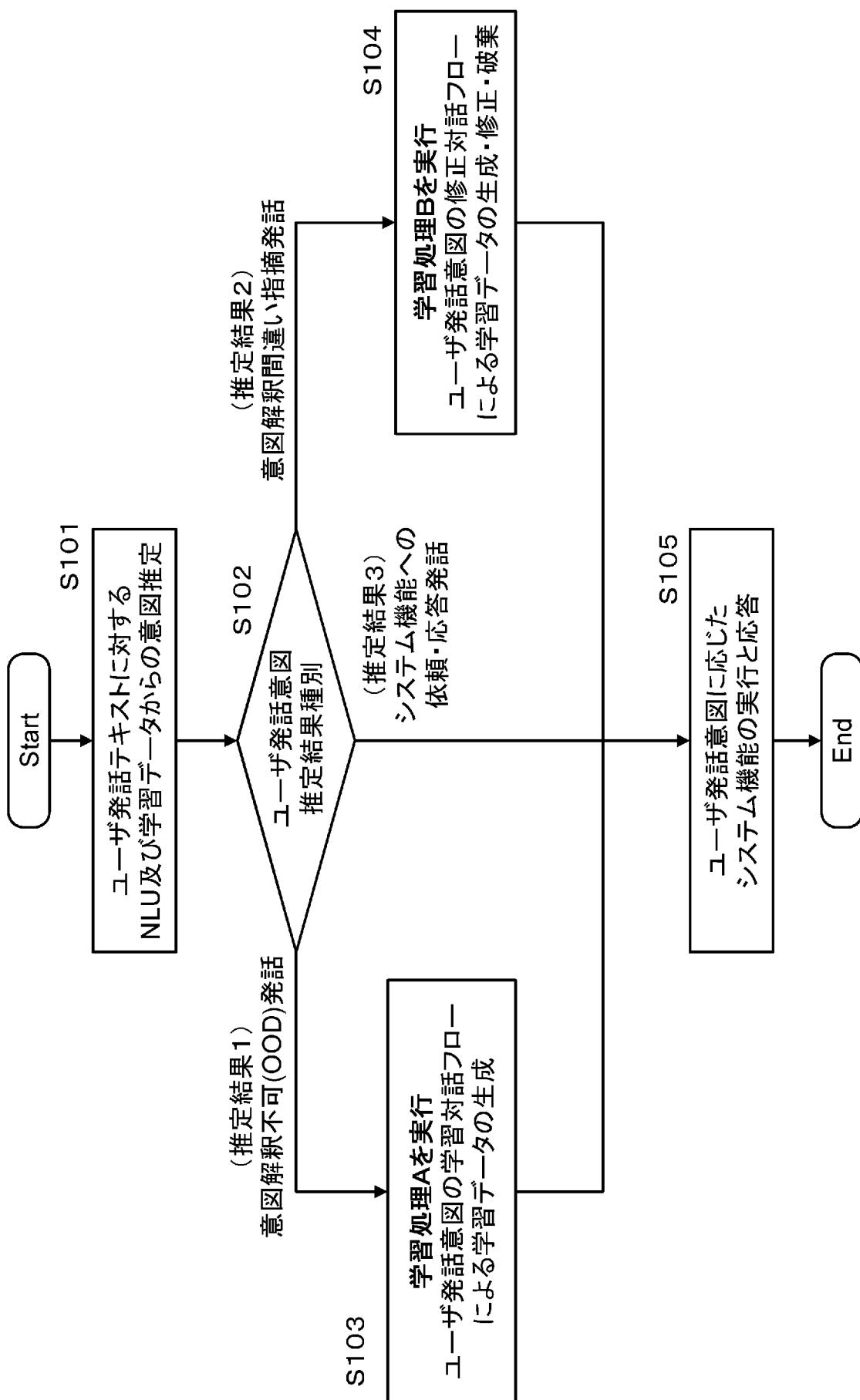
[図2]



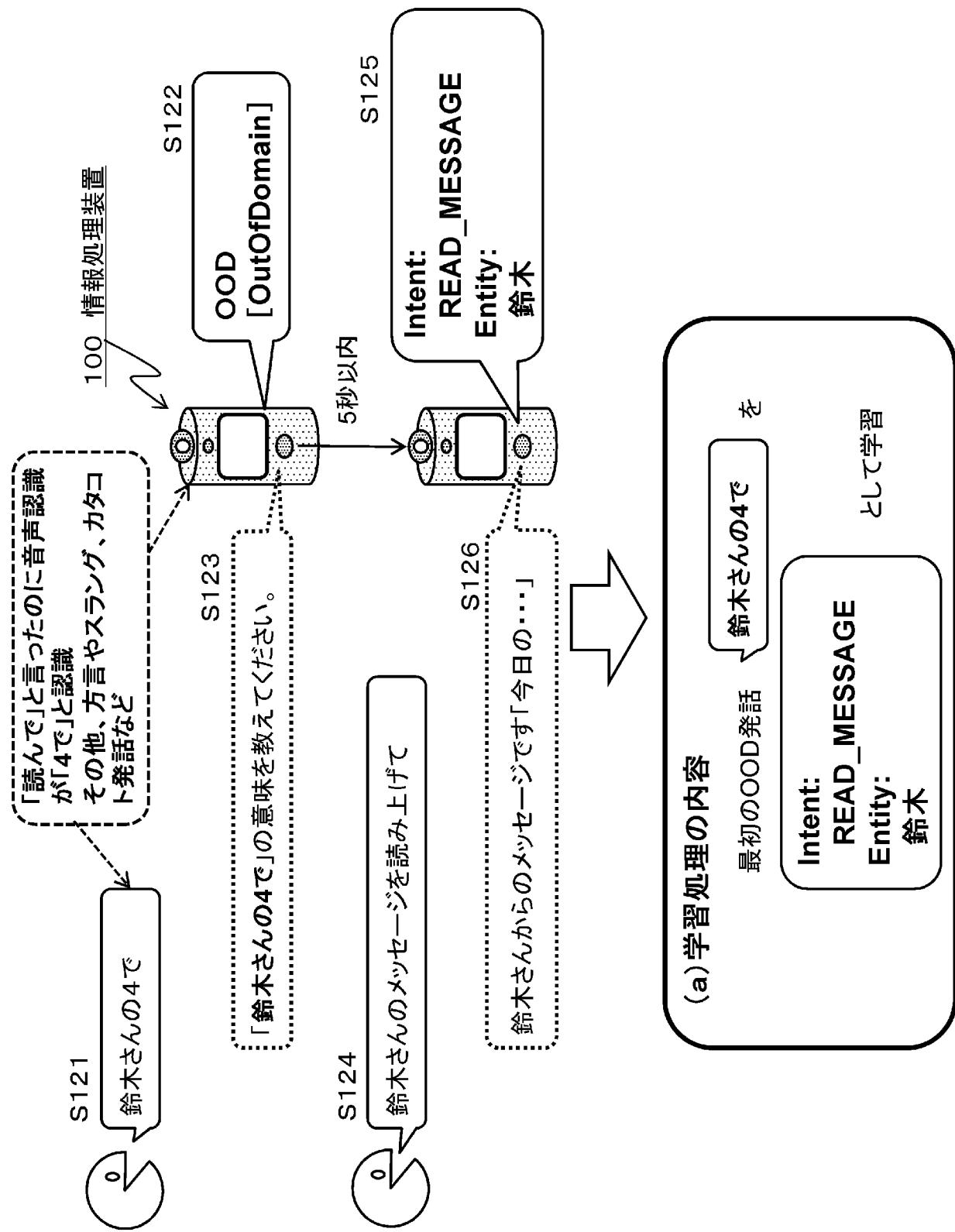
[図3]



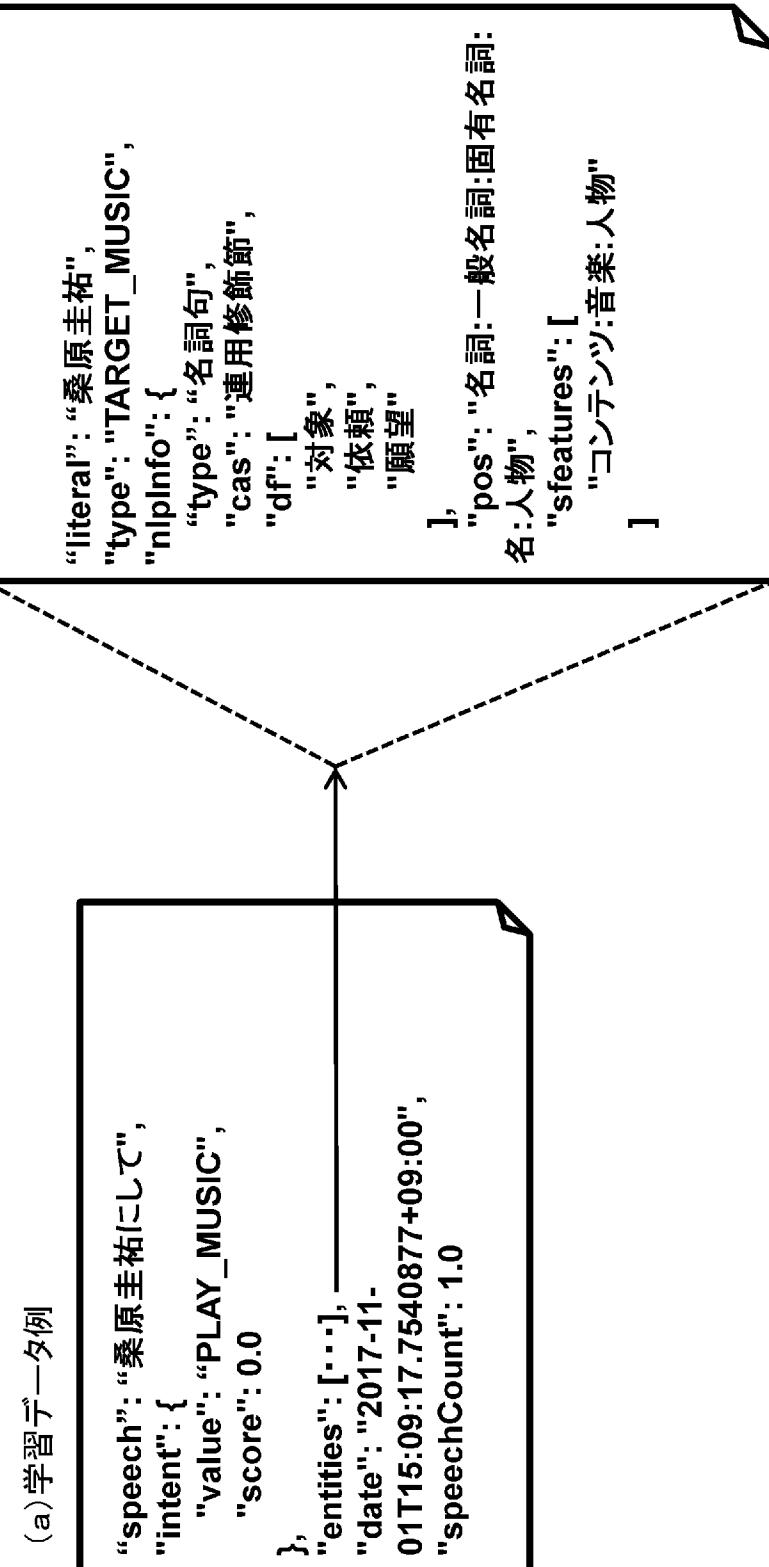
[図4]



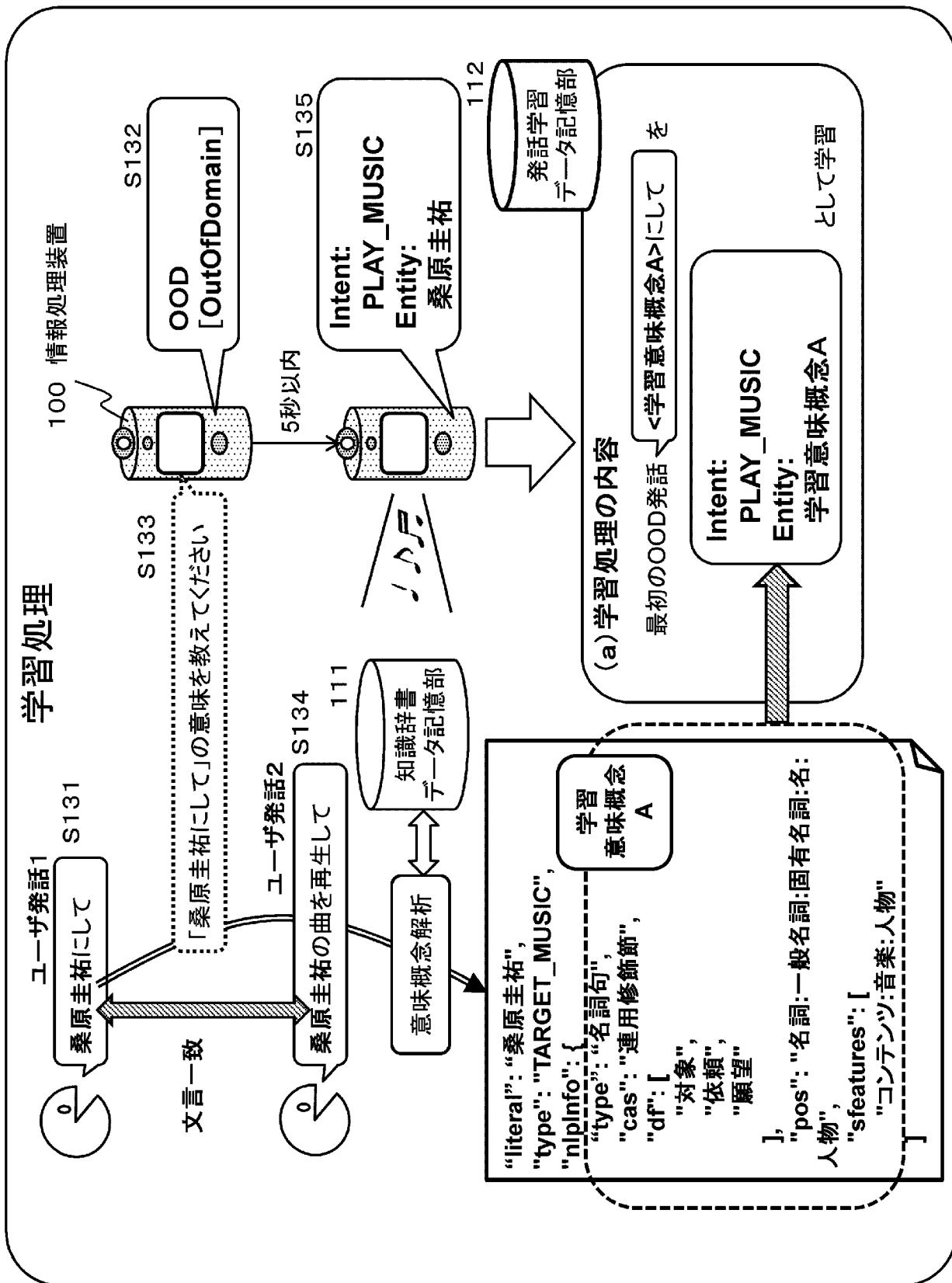
[図5]



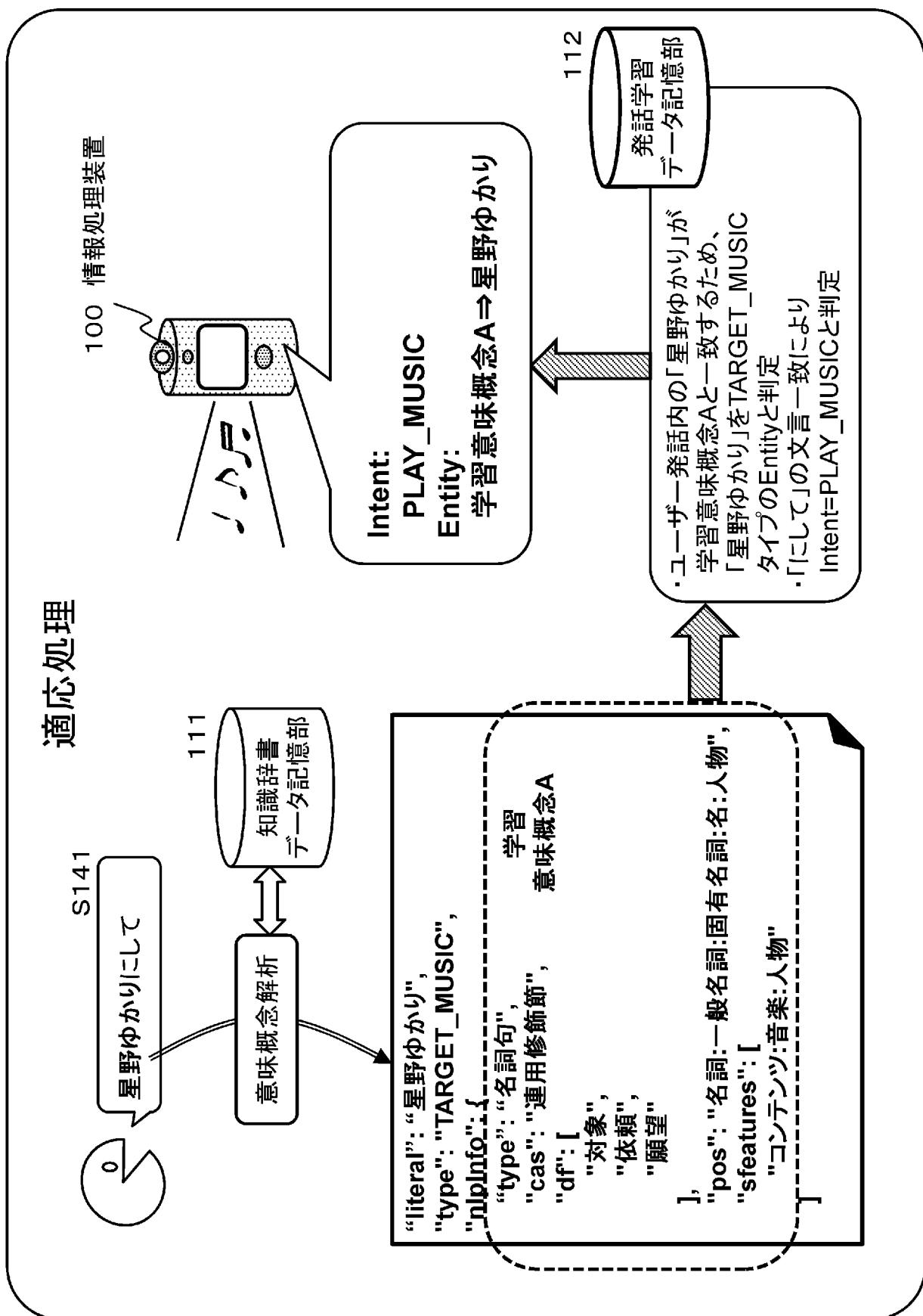
[図6]



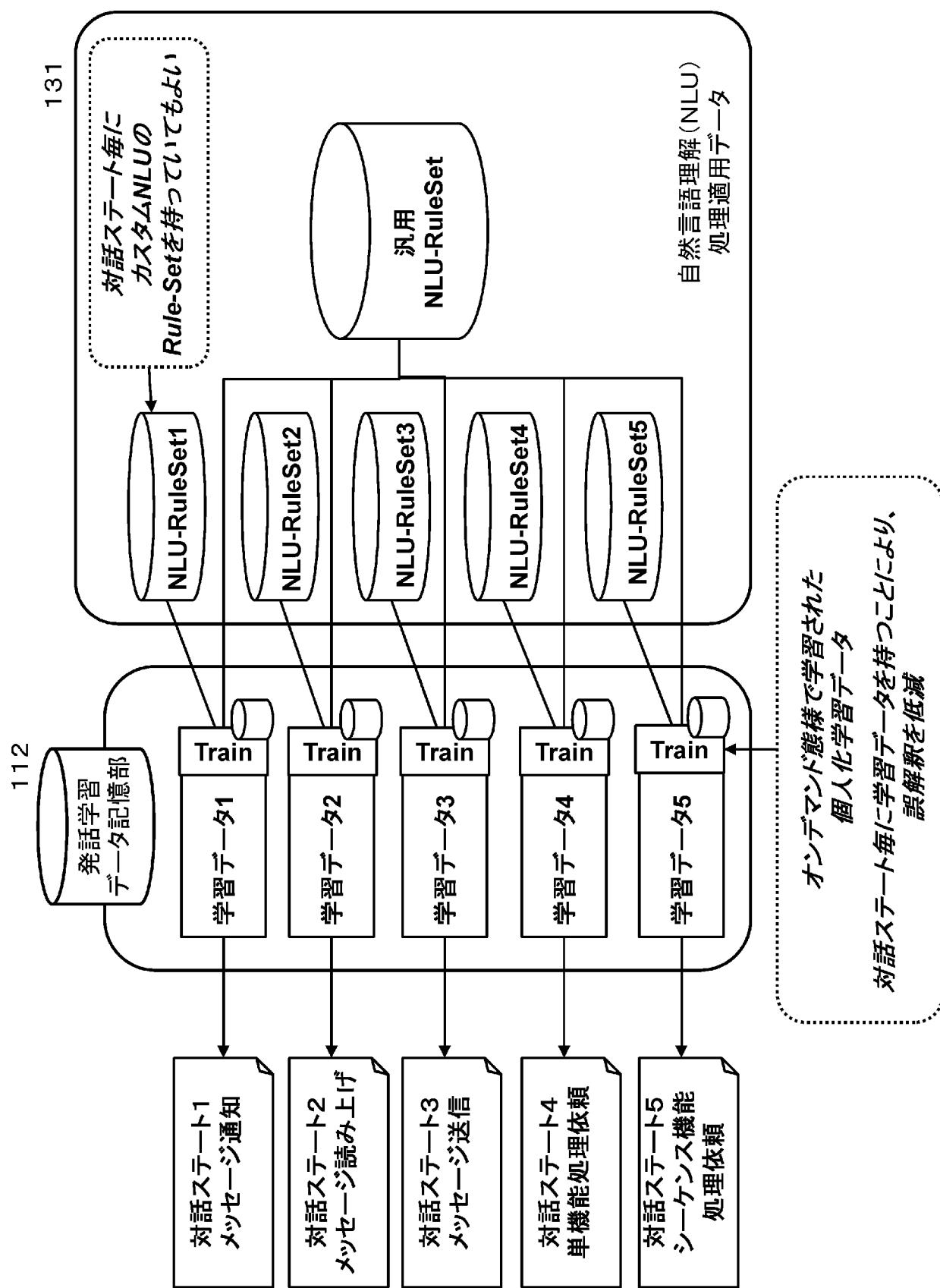
[図7]



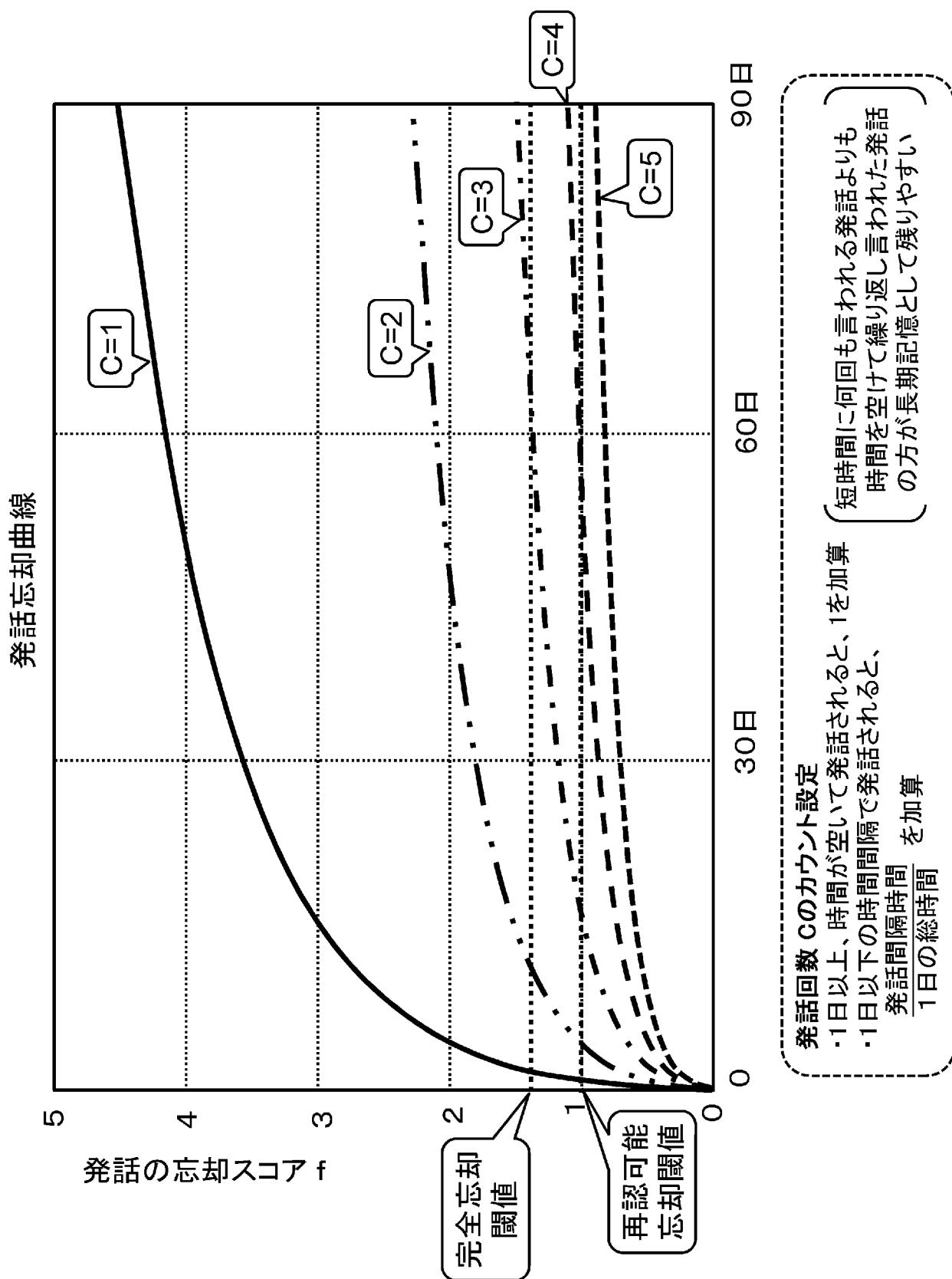
[図8]



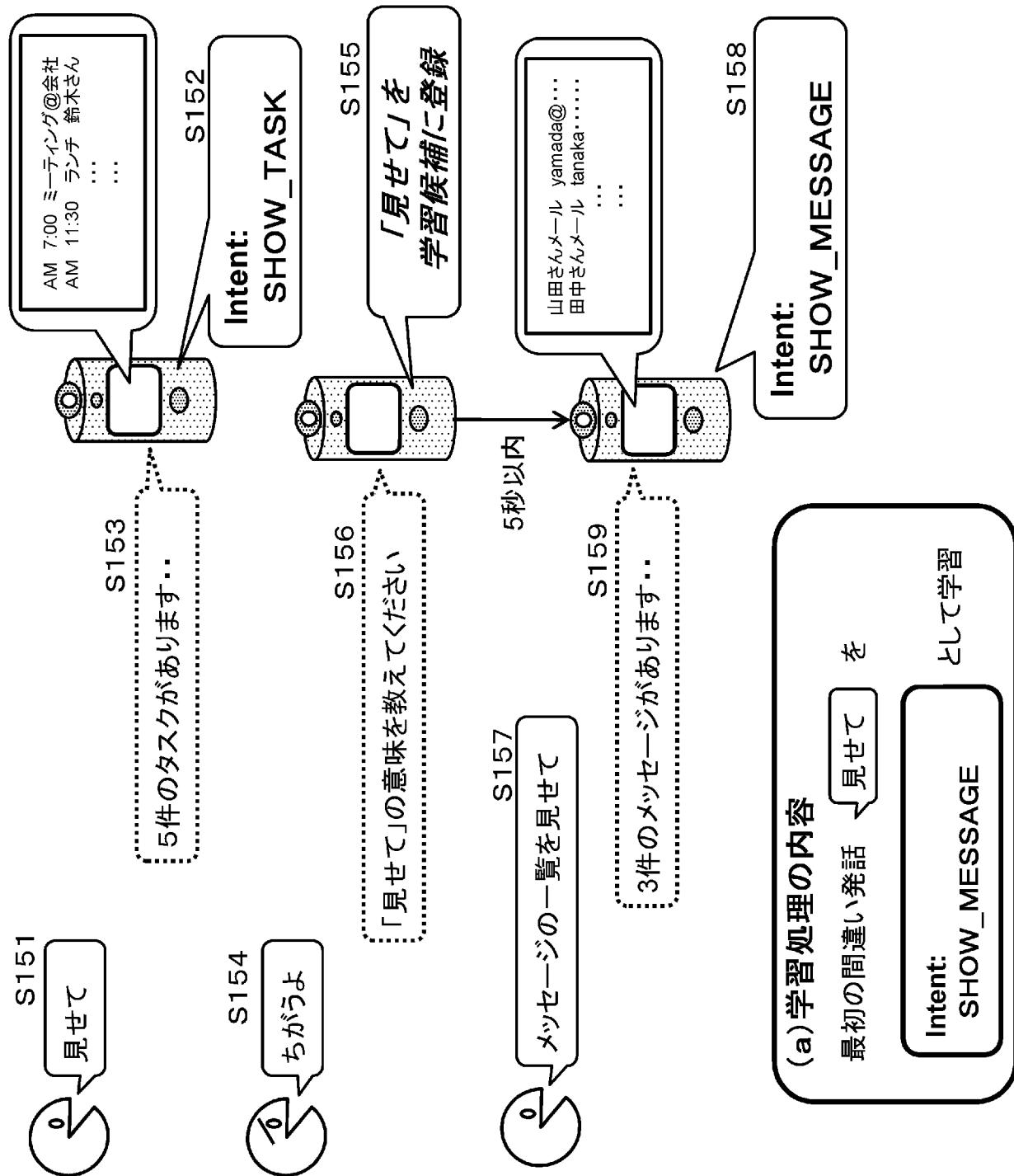
[図9]



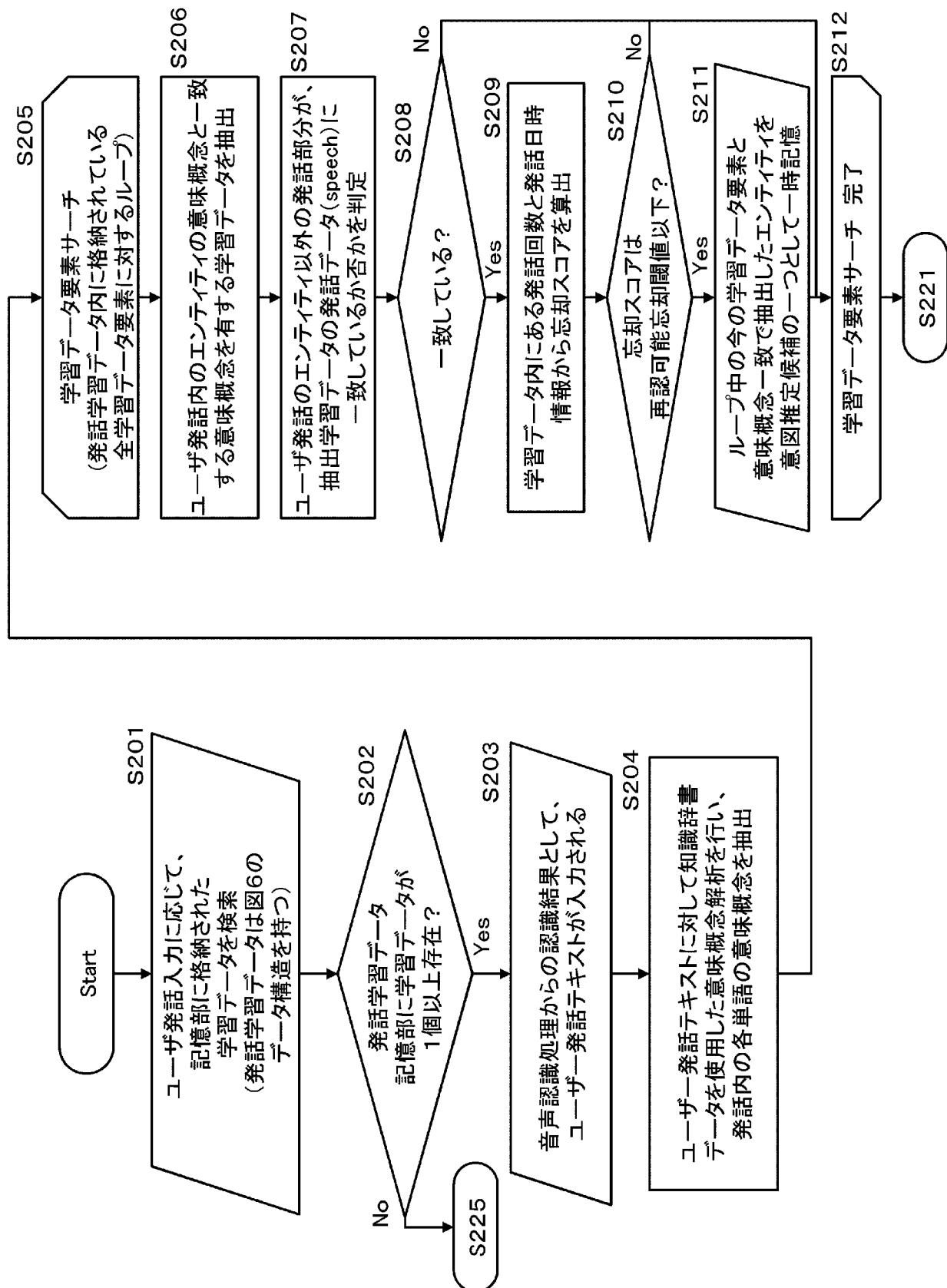
[図10]



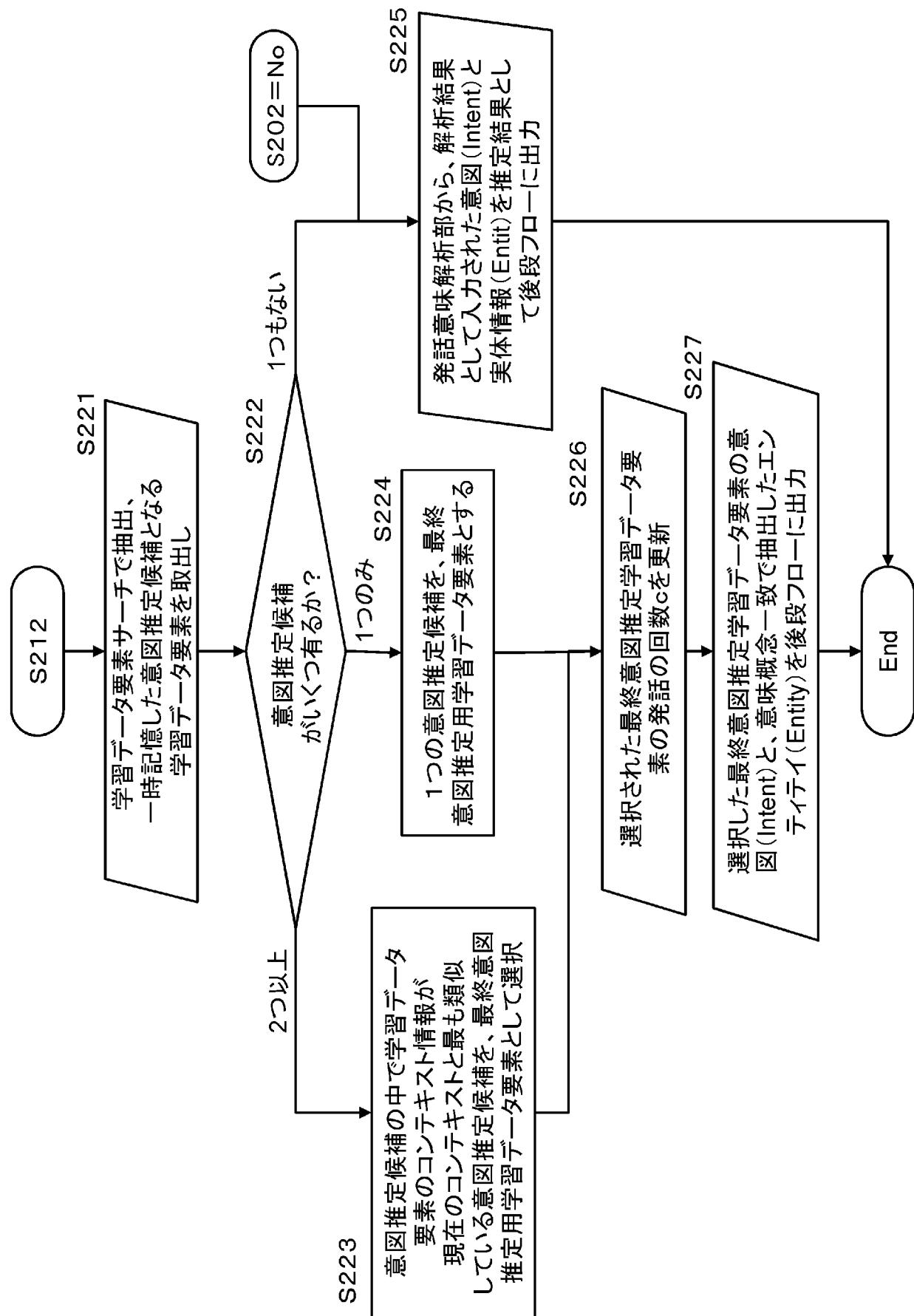
[図11]



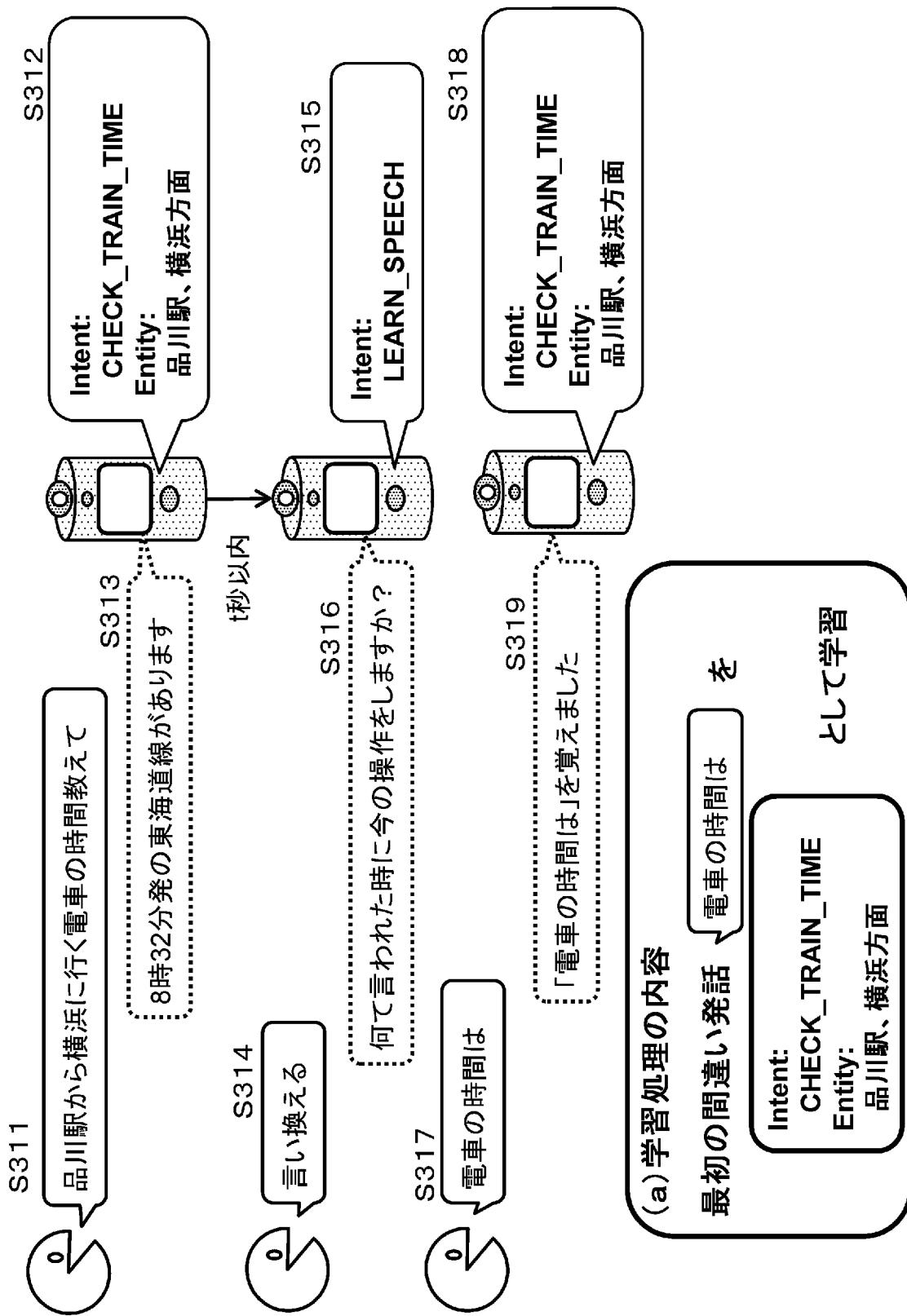
[図12]



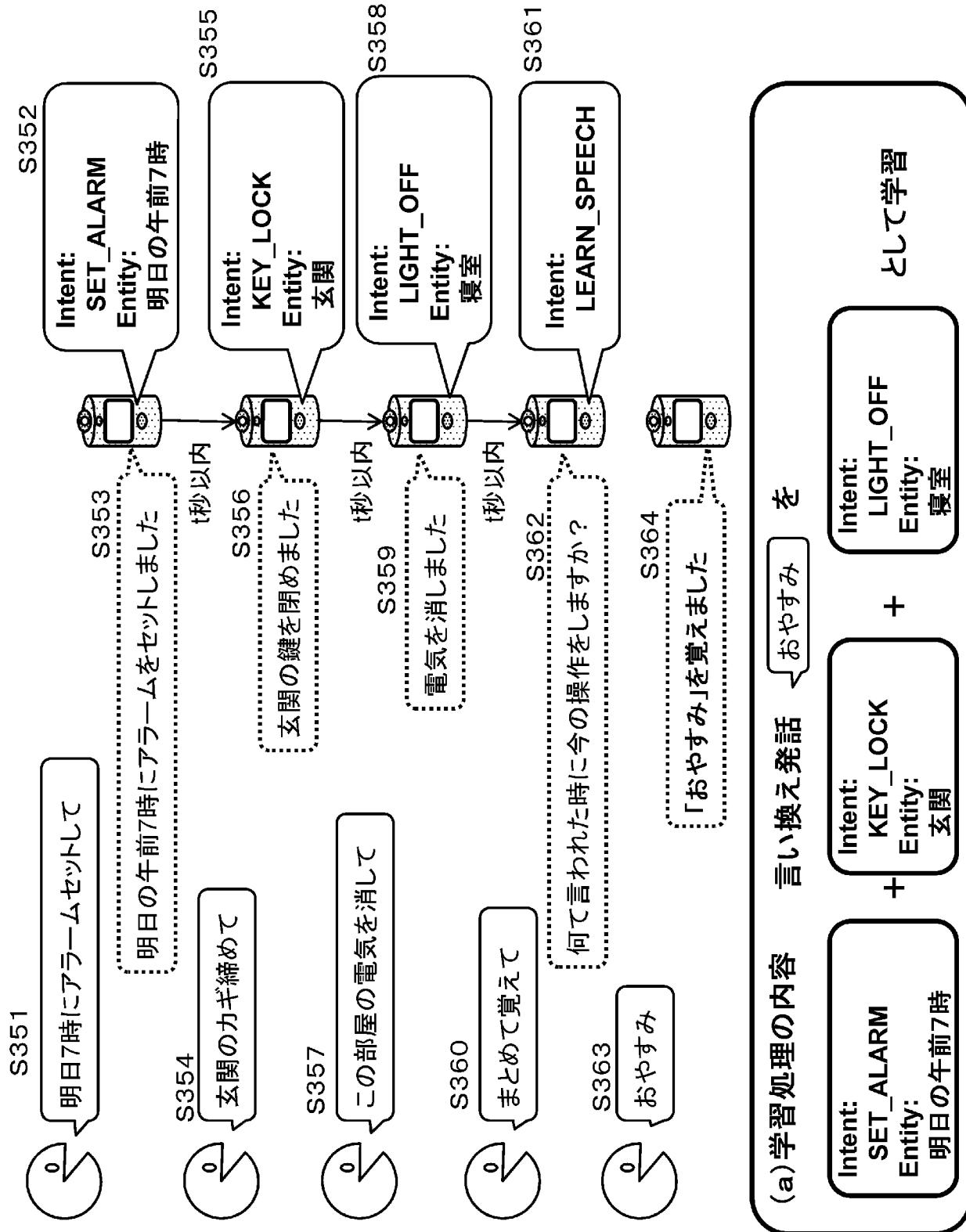
[図13]



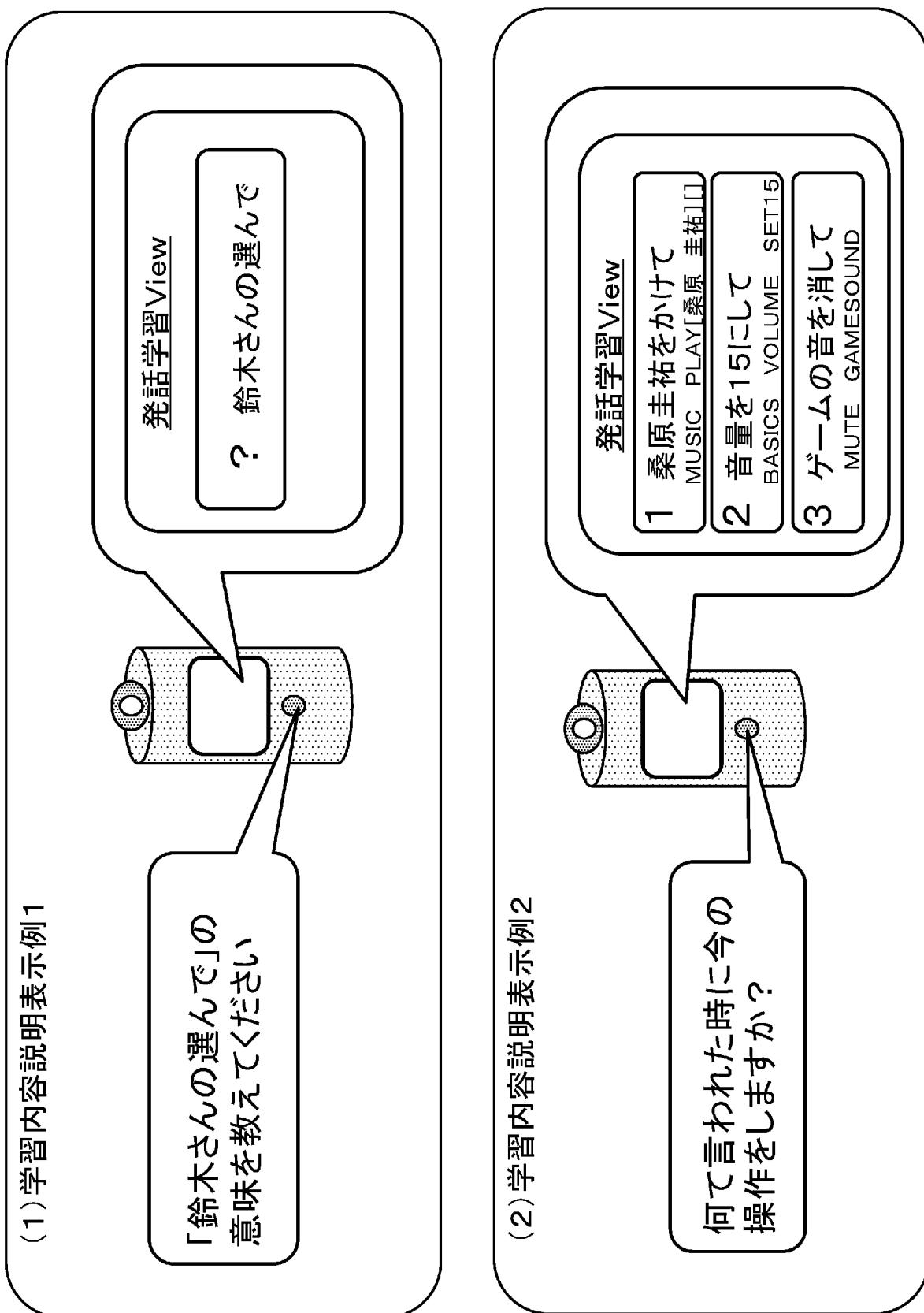
[図14]



[図15]

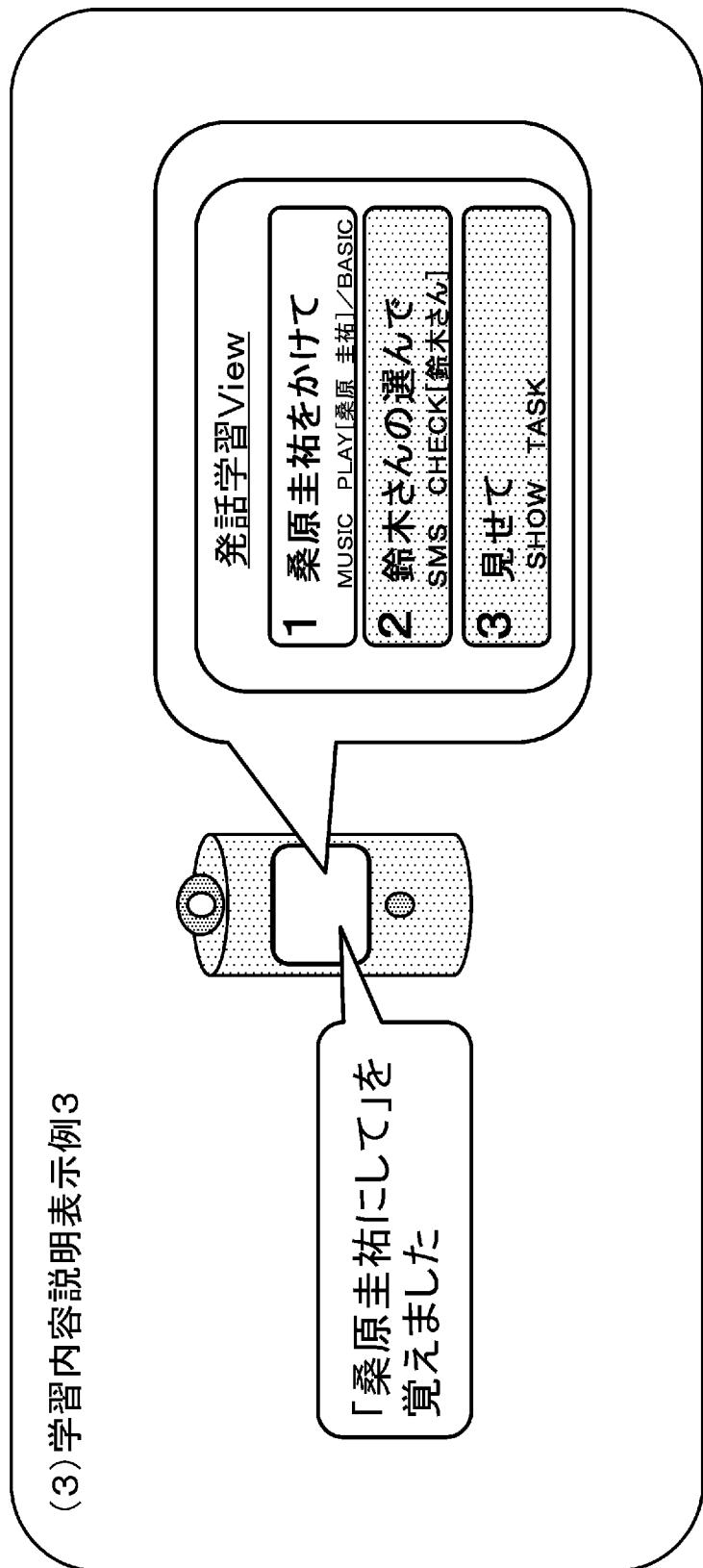


[図16]

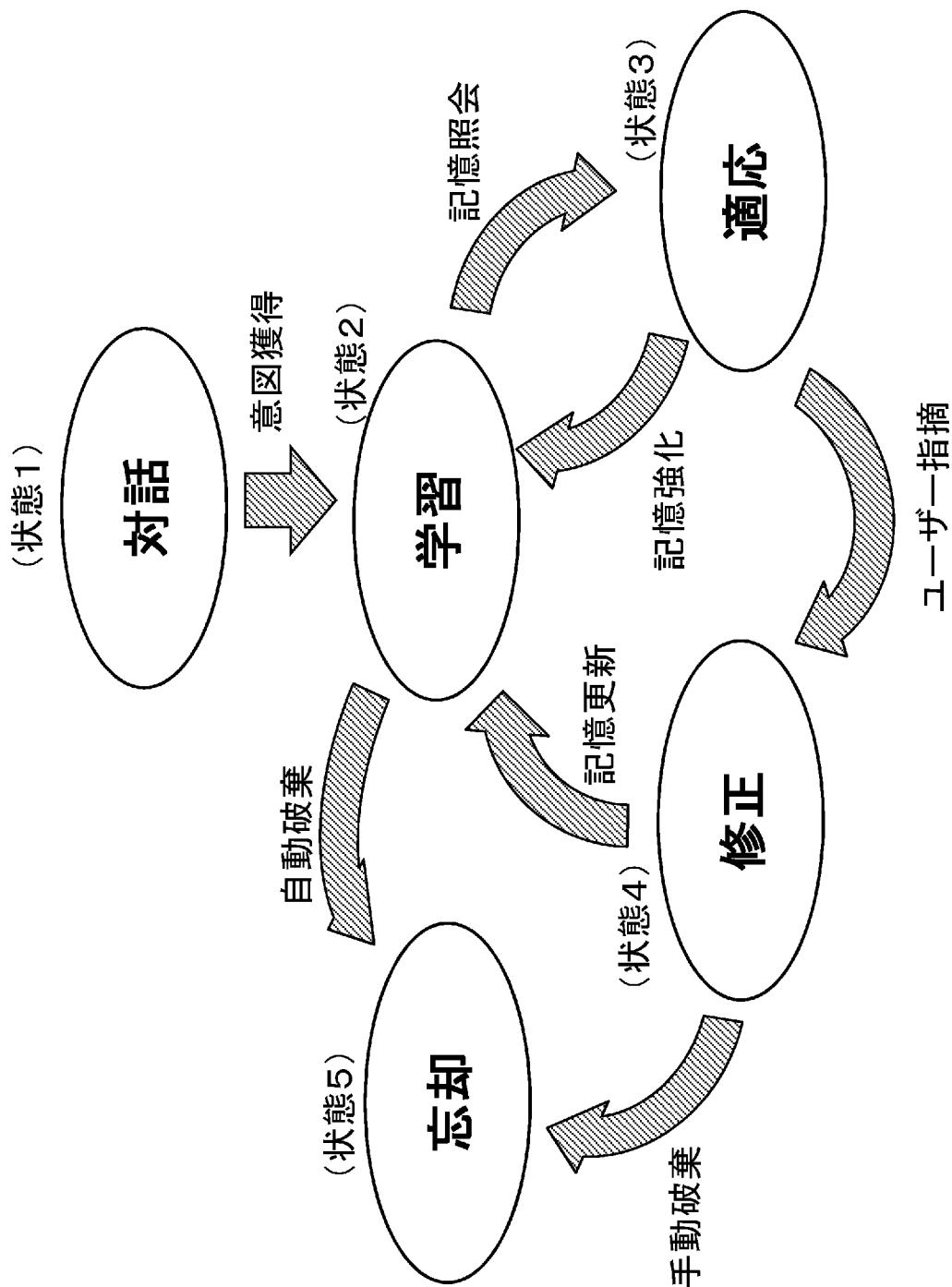


[図17]

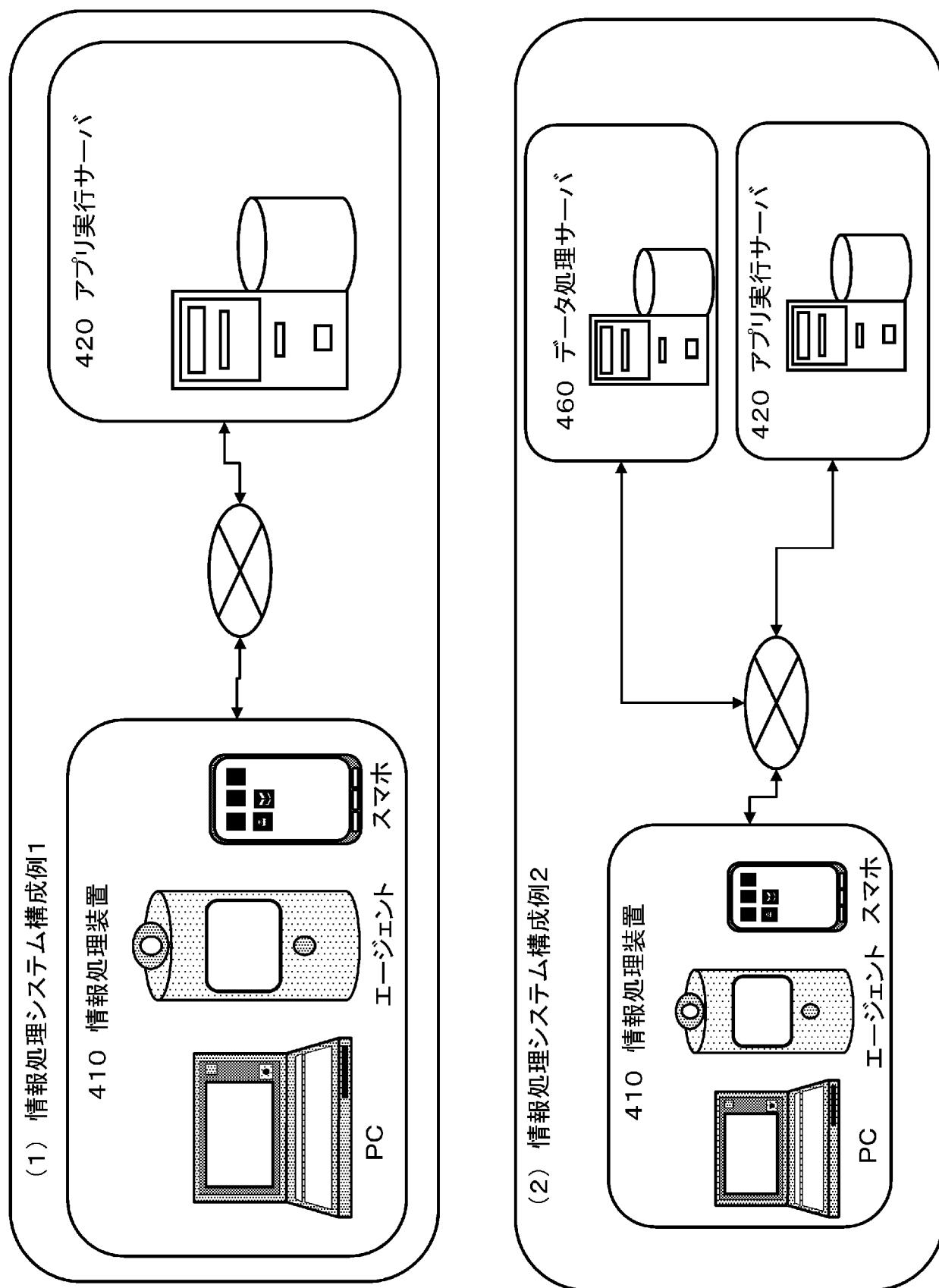
## (3) 学習内容説明表示例3



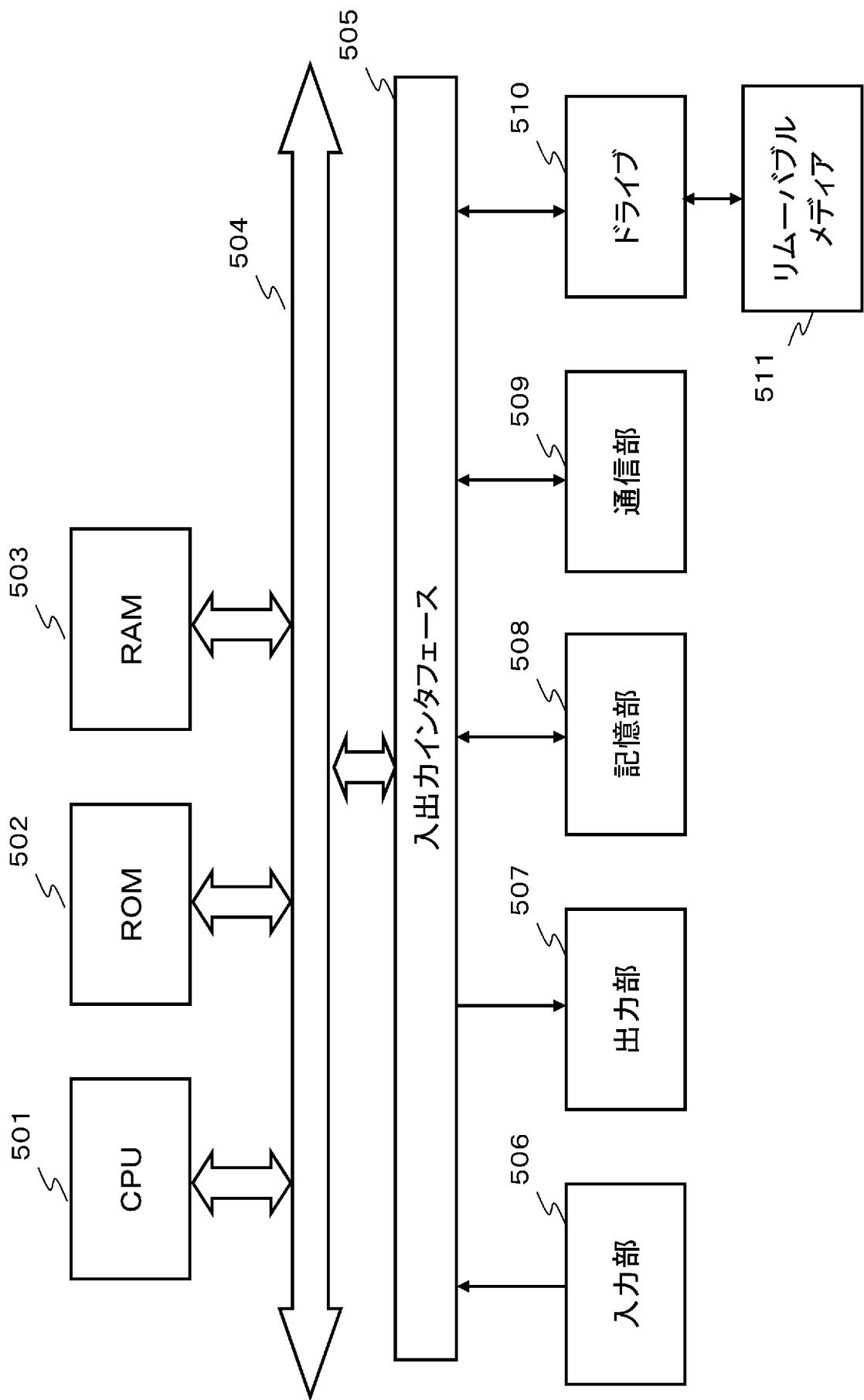
[図18]



[図19]



[図20]



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/039826

### A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G10L15/10 (2006.01)i, G06F3/16 (2006.01)i, G06F17/27 (2006.01)i, G10L15/22 (2006.01)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)

Int.Cl. G10L15/10, G06F3/16, G06F17/27, G10L15/22

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

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2019
Registered utility model specifications of Japan	1996-2019
Published registered utility model applications of Japan	1994-2019

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 A	WO 2016/151698 A1 (TOSHIBA CORPORATION) 29 September 2016, paragraphs [0047]-[0056] & US 2017/0140754 A1, paragraphs [0066]-[0085]	1-2, 10, 17-20 3-9, 11-16
A	WO 2016/067418 A1 (MITSUBISHI ELECTRIC CORPORATION) 06 May 2016, entire text, all drawings & US 2017/0199867 A1, entire text, all drawings & CN 107077843 A	1-20
A	JP 2008-58465 A (TOSHIBA CORPORATION) 13 March 2008, entire text, all drawings & US 2008/0059178 A1, entire text, all drawings	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  
09.01.2019

Date of mailing of the international search report  
22.01.2019

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/JP2018/039826

**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 2006-215317 A (HITACHI, LTD.; ZANAVY INFORMATICS KK; NISSAN MOTOR CO., LTD.) 17 August 2006, entire text, all drawings (Family: none)	1-20
A	US 9299346 B1 (AMAZON TECHNOLOGIES, INC.) 29 March 2016, entire text, all drawings & US 10026394 B1	1-20

## A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G10L15/10(2006.01)i, G06F3/16(2006.01)i, G06F17/27(2006.01)i, G10L15/22(2006.01)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（I P C））

Int.Cl. G10L15/10, G06F3/16, G06F17/27, G10L15/22

## 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	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019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	WO 2016/151698 A1 (株式会社東芝) 2016.09.29, 段落 [0047] – [0056] & US 2017/0140754 A1, 段落 [0066] – [0085]	1-2, 10, 17-20
A	WO 2016/067418 A1 (三菱電機株式会社) 2016.05.06, 全文, 全図 & US 2017/0199867 A1, 全文, 全図 & CN 107077843 A	3-9, 11-16
A	JP 2008-58465 A (株式会社東芝) 2008.03.13, 全文, 全図 & US 2008/0059178 A1, 全文, 全図	1-20

※ 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 09.01.2019	国際調査報告の発送日 22.01.2019
国際調査機関の名称及びあて先 日本国特許庁 (I S A / 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） 上田 雄 電話番号 03-3581-1101 内線 3591 5Z 5095

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 2006-215317 A (株式会社日立製作所, 株式会社ザナヴィ・インフォマティクス, 日産自動車株式会社) 2006. 08. 17, 全文, 全図 (ファミリーなし)	1-20
A	US 9299346 B1 (AMAZON TECHNOLOGIES, INC.) 2016. 03. 29, 全文, 全図 & US 10026394 B1	1-20