

(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
2021年6月3日(03.06.2021)



(10) 国際公開番号

WO 2021/106080 A1

(51) 国際特許分類:

G10L 15/22 (2006.01)

(21) 国際出願番号:

PCT/JP2019/046184

(22) 国際出願日:

2019年11月26日(26.11.2019)

(25) 国際出願の言語:

日本語

(26) 国際公開の言語:

日本語

(71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 山田 和徳 (YAMADA, Kazunori); 〒1808585 東京都武蔵野市緑町三丁目9番1号 NTT 知的財産センタ内 Tokyo (JP).

中村 孝(NAKAMURA, Takashi); 〒1808585 東京都武蔵野市緑町三丁目9番1号 NTT 知的財産センタ内 Tokyo (JP).

(74) 代理人: 中尾 直樹, 外 (NAKAO, Naoki et al.); 〒1600022 東京都新宿区新宿三丁目1番22号 新宿NSOビル6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, 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,

(54) Title: DIALOG DEVICE, METHOD, AND PROGRAM

(54) 発明の名称: 対話装置、方法及びプログラム

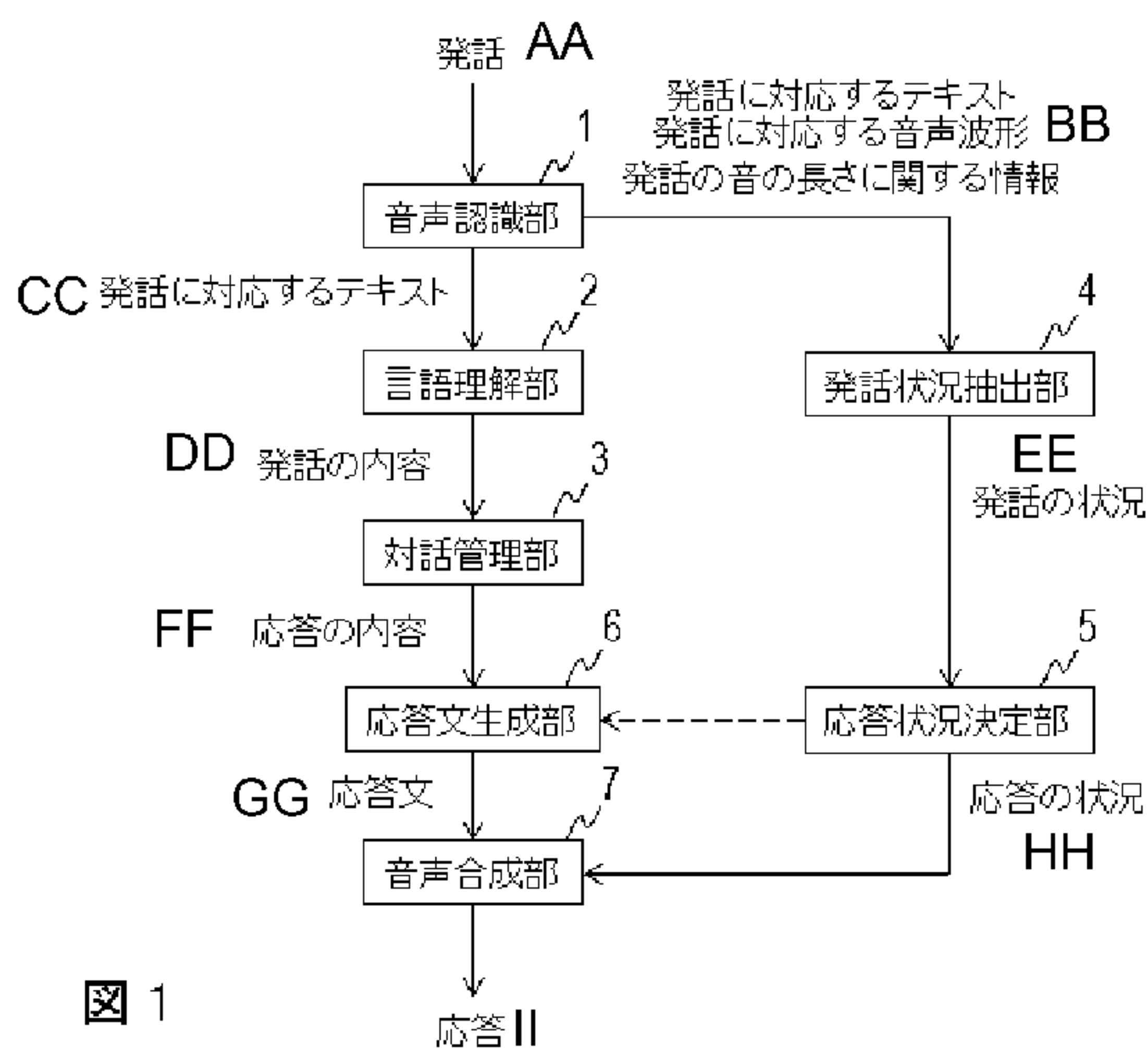


図 1

- 1 Speech recognition unit
 - 2 Language comprehension unit
 - 3 Dialog management unit
 - 4 Utterance status extraction unit
 - 5 Response status determination unit
 - 6 Response text generation unit
 - 7 Speech synthesis unit
- AA Utterance
BB Text corresponding to utterance, Speech waveform corresponding to utterance, information about length of utterance
CC Text corresponding to utterance
DD Content of utterance
EE Status of utterance
FF Content of response
GG Response text
HH Status of response
II Response

(57) Abstract: The dialog device is provided with: a speech recognition unit 1 that performs speech recognition on an input utterance and generates text corresponding to the utterance, a speech waveform corresponding to the utterance, and information about the length of the utterance; a language comprehension unit 2 that understands the content of the utterance using the text corresponding to the utterance; a dialog management unit 3 that determines the content of a response to the utterance using the content of the utterance; an utterance status extraction unit 4 that extracts the status of the utterance using the text corresponding to the utterance, the speech waveform corresponding to the utterance, and the information about the length of the utterance; a response status determination unit 5 that determines the status of the

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)）

response according to the status of the utterance; a response text generation unit 6 that generates response text using the content of the response; and a speech synthesis unit 7 that synthesizes a speech that corresponds to the response text taking the status of the response into consideration.

(57) 要約：対話装置は、入力された発話に対して音声認識を行い、発話に対応するテキストと、発話に対応する音声波形と、発話の音の長さに関する情報とを生成する音声認識部1と、発話に対応するテキストを用いて、発話の内容を把握する言語理解部2と、発話の内容を用いて、発話に対応する応答の内容を決定する対話管理部3と、発話に対応するテキストと、発話に対応する音声波形と、発話の音の長さに関する情報とを用いて、発話の状況を抽出する発話状況抽出部4と、発話の状況に応じて、応答の状況を決定する応答状況決定部5と、応答の内容を用いて、応答文を生成する応答文生成部6と、応答文に対応する音声であって、応答の状況を考慮した音声を合成する音声合成部7と、を備えている。

明 細 書

発明の名称：対話装置、方法及びプログラム

技術分野

[0001] 本発明は、合成音声を用いた音声対話において、より自然な応答発話を生成する技術に関する。

背景技術

[0002] 従来の一般的な音声合成においては、音声合成部に入力されたテキスト情報に応じて音声合成が行われている（例えば、特許文献1参照。）。

[0003] また、従来の一般的な音声対話システムにおいては、対話相手の発話を音声認識してテキスト化して言語理解し、対話の状態を管理しつつ、応答文を生成して音声合成を行うことで発話応答が行われている（例えば、特許文献2参照。）。

先行技術文献

特許文献

[0004] 特許文献1：特開平01-284898号公報

特許文献2：特開2018-133070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、対話システムにおいて、システムがどのように発話するかは、音声合成部に入力されたテキストに依存する。また、対話相手である人がシステムと自然に対話できるかどうかは、応答生成部より生成、出力されるテキストに依存する。

[0006] このように、応答発話される音声が発話生成部で生成されるテキスト情報のみに依存するため、テキスト上では適切に応答できていても、実際の対話相手の発話音声自体の状況と応答発話の音声の状況にギャップが発生する可能性がある。

[0007] 本発明は、より自然な対話を実現する対話装置、方法及びプログラムを提

供することを目的とする。

課題を解決するための手段

[0008] この発明の一態様による対話装置は、入力された発話に対して音声認識を行い、発話に対応するテキストと、発話に対応する音声波形と、発話の音の長さに関する情報とを生成する音声認識部と、発話に対応するテキストを用いて、発話の内容を把握する言語理解部と、発話の内容を用いて、発話に対応する応答の内容を決定する対話管理部と、発話に対応するテキストと、発話に対応する音声波形と、発話の音の長さに関する情報とを用いて、発話の状況を抽出する発話状況抽出部と、発話の状況に応じて、応答の状況を決定する応答状況決定部と、応答の内容を用いて、応答文を生成する応答文生成部と、応答文に対応する音声であって、応答の状況を考慮した音声を合成する音声合成部と、を備えている。

発明の効果

[0009] より自然な対話を実現することができる。

図面の簡単な説明

[0010] [図1]図1は、対話装置の機能構成の例を示す図である。

[図2]図2は、対話方法の処理手続きの例を示す図である。

[図3]図3は、応答状況決定部5の処理の例を説明するための図である。

[図4]図4は、応答状況決定部5の処理の他の例を説明するための図である。

[図5]図5は、コンピュータの機能構成例を示す図である。

発明を実施するための形態

[0011] 以下、本発明の実施の形態について詳細に説明する。なお、図面中において同じ機能を有する構成部には同じ番号を付し、重複説明を省略する。

[0012] [第一実施形態]

対話装置は、図1に示すように、音声認識部1、言語理解部2、対話管理部3、発話状況抽出部4、応答状況決定部5、応答文生成部6及び音声合成部7を例えば備えている。

[0013] 対話方法は、対話装置の各構成部が、以下に説明し、図 1 に示すステップ S 1 からステップ S 7 の処理を行うことにより例えば実現される。

[0014] 以下、対話装置の各構成部について説明する。

[0015] <音声認識部 1>

音声認識部 1 には、発話が入力される。

[0016] 音声認識部 1 は、入力された発話に対して音声認識を行い、発話に対応するテキストと、発話に対応する音声波形と、発話の音の長さに関する情報とを生成する（ステップ S 1）。

[0017] 発話に対応するテキストのことを「発話文」と呼ぶこともある。

[0018] 生成された発話に対応するテキストは、言語理解部 2 及び発話状況抽出部 4 に出力される。

[0019] 発話に対応する音声波形と、発話の音の長さに関する情報とは、発話状況抽出部 4 に出力される。

[0020] 発話の音の長さに関する情報は、発話自体の長さであってもよいし、発話を構成する各音素の長さであってもよい。

[0021] 音声認識部 1 に入力される発話の例は、「明日の天気は何ですか？」である。

[0022] <言語理解部 2>

言語理解部 2 には、音声認識部 1 で生成された、発話に対応するテキストが入力される。

[0023] 言語理解部 2 は、発話に対応するテキストを用いて、発話の内容を把握する（ステップ S 2）。把握された内容は、対話管理部 3 に出力される。

[0024] 発話の内容とは、例えばいわゆる対話行為についての情報である。対話行為は、例えば行為タイプ、属性の情報を少なくとも有している（例えば、参考文献 1 参照。）。

[0025] [参考文献 1] Hironan, “機械学習を使って作る対話システム”、[online]、[令和 1 年 11 月 13 日検索]、インターネット < URL : <https://qiita.com/Hironan/items/6425787ccbee75dfae36>>

発話の対話タイプの例は、質問、あいさつ、主張である。

[0026] 音声認識部 1 に入力される発話が「明日の天気は何ですか？」である場合の発話の内容の例は、（行為タイプ＝質問、時間の属性＝明日）である。

[0027] <対話管理部 3>

対話管理部 3 には、言語理解部 2 で把握された、発話の内容が入力される。

[0028] 対話管理部 3 は、発話の内容を用いて、発話に対応する応答の内容を決定する（ステップ S 3）。

[0029] 決定された応答の内容は、応答文生成部 6 に出力される。

[0030] 応答の内容とは、例えば対話タイプについての情報である。応答の対話タイプの例は、回答、回答（嘘）、質問、あいさつ、謝罪、確認である。

[0031] 対話管理部 3 は、例えば参考文献 1 に記載された方法により応答の内容を決定する。すなわち、対話管理部 3 は、入力された発話の内容に基づいて、内部状態を更新し、更新された内部状態に基づいて、発話の内容である対話タイプを決定する。その際、対話管理部 3 は、外部APIを用いて、発話の内容を決定してもよい。

[0032] 発話の内容が（行為タイプ＝質問、時間の属性＝明日）である場合の応答の内容の例は、（行為タイプ＝回答、天気の属性＝晴れ）である。

[0033] <発話状況抽出部 4>

発話状況抽出部 4 には、音声認識部 1 で生成された、発話に対応するテキストと、発話に対応する音声波形と、発話の音の長さに関する情報とが入力される。

[0034] 発話状況抽出部 4 は、発話に対応するテキストと、発話に対応する音声波形と、発話の音の長さに関する情報とを用いて、発話の状況を抽出する（ステップ S 4）。

[0035] 抽出された発話の状況は、応答状況決定部 5 に出力される。

[0036] 発話の状況とは、少なくとも、発話の話速、発話を行った者の感情等の発話の状況に関連する情報である。発話の状況には、発話を行った者の口調が

含まれてもよい。

[0037] 発話の話速とは、発話の速さに関する情報である。発話の話速は、例えば、単位時間あたりに含まれる文字数又は音素数である。

[0038] 発話を行った者の感情の例は、普通、喜び、悲しみ、怒り、穏やか、興奮、落ち着き、落ち込み、不安、恐縮、明るい、暗い等の感情である。例えば、発話状況抽出部4は、発話を行った者の感情を、普通、喜び、悲しみ、怒り、穏やか、興奮、落ち着き、落ち込み、不安、恐縮、明るい、暗い等の何れかの感情に分類することで決定する。発話状況抽出部4は、発話を行った者の感情を、普通、喜び、悲しみ、怒りの何れかに分類することで決定してもよい。発話状況抽出部4は、発話を行った者の感情を、穏やか、興奮、落ち着き、落ち込み、不安、恐縮の何れかに分類することで決定してもよい。発話状況抽出部4は、発話を行った者の感情を、明るい、暗いに分類することで決定してもよい。

[0039] 発話状況抽出部4は、例えば参考文献2に記載された方法により、発話を行った者の感情を決定することができる。発話を行った者の感情は、例えば、発話に対応するテキストと、発話に対応する音声波形とに基づいて決定される。

[0040] 〔参考文献2〕天沼沙織、樽松理樹、羽倉淳、藤田ハミド、“音声からの感情推定におけるクラス分け指標の提案”、情報処理学会第73回全国大会、2011

発話を行った者の口調とは、例えば、丁寧、口語である。ここで、口語とは、丁寧ではないことである。

[0041] 発話状況抽出部4は、例えば参考文献3に記載された方法により、発話を行った者の口調を決定することができる。発話を行った者の口調は、例えば、発話に対応するテキストと、発話に対応する音声波形とに基づいて決定される。

[0042] 〔参考文献3〕馬場朗、関根剛宏、日比谷新平、大林史明、寺澤章、西山高史、仲島了治、“ヒューマノイドエージェントへの口調識別の応用”、情

報処理学会第66回全国大会、2004

[0043] <応答状況決定部5>

応答状況決定部5は、発話状況抽出部4で抽出された発話の状況が入力される。

[0044] 応答状況決定部5は、発話の状況に応じて、応答の状況を決定する（ステップS5）。

[0045] 決定された応答の状況は、音声合成部7に出力される。

[0046] （応答状況決定部5の処理の例1）

応答状況決定部5は、例えば、所定のルールに基づいて、入力された発話の状況に応じた、応答の状況を決定することができる。所定のルールの例は、図3に示す変換表である。

[0047] 図3に示す変換表を用いると、入力される発話の状況が、例えば、（話速＝普通，発話を行った者の感情＝普通，発話を行った者の口調＝丁寧）である場合には、応答の状況として、（話速＝普通，応答の感情＝普通，応答の口調＝丁寧）という応答の状況が決定される。

[0048] また、図3に示す変換表を用いると、入力される発話の状況が、例えば、（話速＝普通，発話を行った者の感情＝喜び，発話を行った者の口調＝口語）である場合には、応答の状況として、（話速＝普通，応答の感情＝喜び，応答の口調＝口語）という応答の状況が決定される。このように、発話を行った者の口調が口語である場合に、応答の口調を口語とすることで、雑談でのフランクな質問に対するフランクな応答を実現することができる。

[0049] また、図3に示す変換表を用いると、入力される発話の状況が、例えば、（話速＝早い，発話を行った者の感情＝怒り，発話を行った者の口調＝口語）である場合には、応答の状況として、（話速＝遅い，応答の感情＝普通，応答の口調＝丁寧）という応答の状況が決定される。このように、発話を行った者の感情が怒りである場合に、応答の話速を遅くし、応答の感情を普通にし、応答の口調を丁寧にすることで、発話を行った者を落ち着かせることができる。

[0050] なお、図3の変換表では、3個の発話の状況のそれぞれに対応する応答の状況のみを示している。例えば、応答状況決定部5が実際に用いる変換表では、全ての発話の状況のそれぞれに対応する応答の状況が定められているとする。

[0051] なお、応答状況決定部5は、変換表に記載した特定の発話の状況については変換表を用いて応答の状況を決し、その他の発話の状況については、予め定められた応答の状況を、応答状況決定部5が出力する応答の状況として決定してもよい。

[0052] (応答状況決定部5の処理の例2)

応答状況決定部5は、ニューラルネットワークなどを用いた非線形変換を用いることで、応答の状況を決してもよい。

[0053] 例えば、ニューラルネットワークの入力層の次元は、発話の話速の種類と感情の種類と口調の種類を足した数とし、ニューラルネットワークの出力層の次元は、応答の話速の種類と感情の種類と口調の種類を足した数とする。ニューラルネットワークの中間層(隠れ層)の数は、任意である。中間層(隠れ層)のそれぞれの次元数も任意である。

[0054] ある入力された発話について、該当する話速・感情・口調の種類に1を入力し、そのほかの該当しない種類には0を入力する。例えば、話速が普通、感情が普通、口調が丁寧の発話については、話速が普通に該当する入力ノードには1を入力し(感情・口調についても同様)、その他の、話速が速い等に該当する入力ノードには0を入力する。

[0055] その入力によりニューラルネットから出力される出力値が、対応する応答の出力に近づくように、ニューラルネットのパラメータを調整することで、入力である発話の状況と応答の状況の変換のパターンを学習したモデルを生成する。上記の例では、応答の話速が普通、応答の感情が普通、応答の口調が丁寧に該当する出力ノードが1を、他の出力ノードは0を出力するようにパラメータ調整する。

[0056] ニューラルネットを活用することで、現状のパターンにない入力パターン

の発話が来ても、既存のパターンに似た形式で対応した応答を行うことができる可能性がある。

[0057] なお、上記の活用の仕方は0、1の入力に限っていたが、連続値を許容するように拡張すると、話速や感情など、中間くらいの微妙な発話についてもそれに応じた微妙なニュアンスで応対ができる可能性がある。

[0058] <応答文生成部6>

応答文生成部6には、対話管理部3で決定された応答の内容が入力される。

[0059] 応答文生成部6は、応答の内容を用いて、応答文を生成する（ステップS6）。

[0060] 生成された応答文は、音声合成部7に出力される。

[0061] 応答の内容の例が（行為タイプ＝回答、天気の種類＝晴れ）である場合の応答文の例は、「晴れです。」である。

[0062] <音声合成部7>

音声合成部7には、応答文生成部6で生成された応答文と、応答状況決定部5で決定された応答の状況とが入力される。

[0063] 音声合成部7は、応答文に対応する音声であって、応答の状況を考慮した音声を合成する（ステップS7）。

[0064] 合成された音声は、対話装置から出力される。

[0065] このように、テキストだけでなく、対話相手の発話音声から得られる相手の発話の状況の情報も入力とし、その状況も加味した音声合成を行う。これにより、より自然な対話を実現することができる。

[0066] [変形例1]

応答状況決定部5が決定した応答の状況は、応答の口調を含んでもよい。

[0067] この場合、応答文生成部6は、応答状況決定部5が決定した応答の状況に含まれる、応答の口調を考慮して、応答文を生成してもよい。

[0068] 発話を行った者の口調を考慮して応答文を生成することで、更に自然な対

話を実現することができる。

[0069] 例えば、応答の内容の例が（行為タイプ＝回答、天気の種類＝晴れ）であり、応答の口調＝丁寧である場合には、応答文生成部6は、「晴れです。」という応答文を生成する。また、応答の内容の例が（行為タイプ＝回答、天気の種類＝晴れ）であり、応答の口調＝口語である場合には、応答文生成部6は、「晴れだよ。」という応答文を生成する。

[0070] [変形例2]

なお、応答状況決定部5は、発話に対応するテキスト、発話の内容、応答の内容及び対話管理部3が応答の内容を決定するまでに得られた情報の少なくとも1つに更に応じて、応答の状況を決定してもよい。

[0071] 対話管理部3が応答の内容を決定するまでに得られた情報は、例えば対話管理部3における内部情報である。

[0072] 応答状況決定部5が、発話の内容である発話の対話タイプと、応答の内容である応答の対話タイプとに更に基づいて、応答の状況を決定する場合に用いる所定のルールである変換表の例を図4に示す。

[0073] 図4に示す変換表を用いると、発話についての入力が、例えば、（話速＝普通、発話を行った者の感情＝普通、発話を行った者の口調＝丁寧、発話の対話タイプ＝質問、応答の対話タイプ＝回答）である場合には、応答の状況として、（話速＝普通、応答の感情＝普通、応答の口調＝丁寧）という応答の状況が決定される。これにより、通常の問い合わせに対応することができる。

[0074] また、図4に示す変換表を用いると、発話についての入力が、例えば、（話速＝遅い、発話を行った者の感情＝不安、発話を行った者の口調＝丁寧、発話の対話タイプ＝質問、応答の対話タイプ＝回答）である場合には、応答の状況として、（話速＝普通、応答の感情＝穏やか、応答の口調＝丁寧）という応答の状況が決定される。これにより、不安感のある、遠慮がちな問い合わせに対応することができる。

[0075] また、図4に示す変換表を用いると、発話についての入力が、例えば、（

話速＝遅い，発話を行った者の感情＝不安，発話を行った者の口調＝丁寧，発話の対話タイプ＝質問，応答の対話タイプ＝質問）である場合には、応答の状況として、（話速＝遅い，応答の感情＝恐縮，応答の口調＝丁寧）という応答の状況が決定される。これにより、不安感のある、遠慮がちな問い合わせに対応しつつ、質問をすることができる。

[0076] また、図4に示す変換表を用いると、発話についての入力が、例えば、（話速＝普通，発話を行った者の感情＝喜び，発話を行った者の口調＝口語，発話の対話タイプ＝あいさつ，応答の対話タイプ＝あいさつ）である場合には、応答の状況として、（話速＝普通，応答の感情＝喜び，応答の口調＝口語）という応答の状況が決定される。これにより、あいさつのやり取りを実現することができる。

[0077] また、図4に示す変換表を用いると、発話についての入力が、例えば、（話速＝遅い，発話を行った者の感情＝落ち込み，発話を行った者の口調＝口語，発話の対話タイプ＝あいさつ，応答の対話タイプ＝質問）である場合には、応答の状況として、（話速＝遅い，応答の感情＝穏やか，応答の口調＝丁寧）という応答の状況が決定される。これにより、落ち込んだ発話に対応する、丁寧な応答（例えば、「どうしましたか？」）を実現することができる。

[0078] また、図4に示す変換表を用いると、発話についての入力が、例えば、（話速＝普通，発話を行った者の感情＝明るい，発話を行った者の口調＝口語，発話の対話タイプ＝質問，応答の対話タイプ＝回答）である場合には、応答の状況として、（話速＝普通，応答の感情＝明るい，応答の口調＝口語）という応答の状況が決定される。これにより、雑談でのフランクな質問に対する通常の回答を行うことができる。

[0079] また、図4に示す変換表を用いると、発話についての入力が、例えば、（話速＝普通，発話を行った者の感情＝明るい，発話を行った者の口調＝口語，発話の対話タイプ＝質問，応答の対話タイプ＝回答（嘘））である場合には、応答の状況として、（話速＝普通，応答の感情＝悲しみ，応答の口調＝

口語)という応答の状況が決定される。これにより、雑談でのフランクな質問に対するワザとずらした回答を行うことができる。

[0080] また、図4に示す変換表を用いると、発話についての入力、例えば、(話速＝速い、発話を行った者の感情＝怒り、発話を行った者の口調＝口語、発話の対話タイプ＝主張、応答の対話タイプ＝謝罪)である場合には、応答の状況として、(話速＝遅い、応答の感情＝落ち込み、応答の口調＝丁寧)という応答の状況が決定される。これにより、コールセンター等におけるクレームの対応を実現することができる。

[0081] また、図4に示す変換表を用いると、発話についての入力、例えば、(話速＝速い、発話を行った者の感情＝興奮、発話を行った者の口調＝丁寧、発話の対話タイプ＝質問、応答の対話タイプ＝確認)である場合には、応答の状況として、(話速＝普通、応答の感情＝落ち着き、応答の口調＝丁寧)という応答の状況が決定される。これにより、緊急の問い合わせに対する復唱等を行うことができる。

[0082] [他の変形例]

以上、本発明の実施の形態及び変形例について説明したが、具体的な構成は、これらの実施の形態に限られるものではなく、本発明の趣旨を逸脱しない範囲で適宜設計の変更等があっても、本発明に含まれることはいうまでもない。

[0083] 実施の形態において説明した各種の処理は、記載の順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。

[0084] 例えば、対話装置の構成部間のデータのやり取りは直接行われてもよいし、図示していない記憶部を介して行われてもよい。

[0085] [プログラム、記録媒体]

上記説明した装置における各種の処理機能をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記各装

置における各種の処理機能がコンピュータ上で実現される。例えば、上述の各種の処理は、図5に示すコンピュータの記録部2020に、実行させるプログラムを読み込ませ、制御部2010、入力部2030、出力部2040などに動作させることで実施できる。

[0086] この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。

[0087] また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。更に、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

[0088] このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記憶装置に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、更に、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コ

ンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

[0089] また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

符号の説明

- [0090]
- | | |
|---|---------|
| 1 | 音声認識部 |
| 2 | 言語理解部 |
| 3 | 対話管理部 |
| 4 | 発話状況抽出部 |
| 5 | 応答状況決定部 |
| 6 | 応答文生成部 |
| 7 | 音声合成部 |

請求の範囲

- [請求項1] 入力された発話に対して音声認識を行い、前記発話に対応するテキストと、前記発話に対応する音声波形と、前記発話の音の長さに関する情報とを生成する音声認識部と、
- 前記発話に対応するテキストを用いて、前記発話の内容を把握する言語理解部と、
- 前記発話の内容を用いて、前記発話に対応する応答の内容を決定する対話管理部と、
- 前記発話に対応するテキストと、前記発話に対応する音声波形と、前記発話の音の長さに関する情報とを用いて、前記発話の状況を抽出する発話状況抽出部と、
- 前記発話の状況に応じて、応答の状況を決定する応答状況決定部と、
- 前記応答の内容を用いて、応答文を生成する応答文生成部と、
- 前記応答文に対応する音声であって、前記応答の状況を考慮した音声を合成する音声合成部と、
- を含む対話装置。
- [請求項2] 請求項1の対話装置であって、
- 前記発話の状況は、少なくとも、発話の話速、発話を行った者の感情である、
- 対話装置。
- [請求項3] 請求項1又は2の対話装置であって、
- 前記応答の状況は、応答の口調を含み、
- 前記応答文生成部は、前記応答の状況に含まれる、応答の口調を考慮して、応答文を生成する、
- 対話装置。
- [請求項4] 請求項1から3の何れかの対話装置であって、
- 前記応答状況決定部は、前記発話に対応するテキスト、前記発話の

内容、前記応答の内容及び前記対話管理部が前記応答の内容を決定するまでに得られた情報の少なくとも1つに更に応じて、応答の状況を決定する、

対話装置。

[請求項5]

音声認識部が、入力された発話に対して音声認識を行い、前記発話に対応するテキストと、前記発話に対応する音声波形と、前記発話の音の長さに関する情報とを生成する音声認識ステップと、

言語理解部が、前記発話に対応するテキストを用いて、前記発話の内容を把握する言語理解ステップと、

対話管理部が、前記発話の内容を用いて、前記発話に対応する応答の内容を決定する対話管理ステップと、

発話状況抽出部が、前記発話に対応するテキストと、前記発話に対応する音声波形と、前記発話の音の長さに関する情報とを用いて、前記発話の状況を抽出する発話状況抽出ステップと、

応答状況決定部が、前記発話の状況に応じて、応答の状況を決定する応答状況決定ステップと、

応答文生成部が、前記応答の内容を用いて、応答文を生成する応答文生成ステップと、

音声合成部が、前記応答文に対応する音声であって、前記応答の状況を考慮した音声を合成する音声合成ステップと、

を含む対話方法。

[請求項6]

請求項1から4の何れかの対話装置の各部としてコンピュータを機能させるためのプログラム。

[図1]

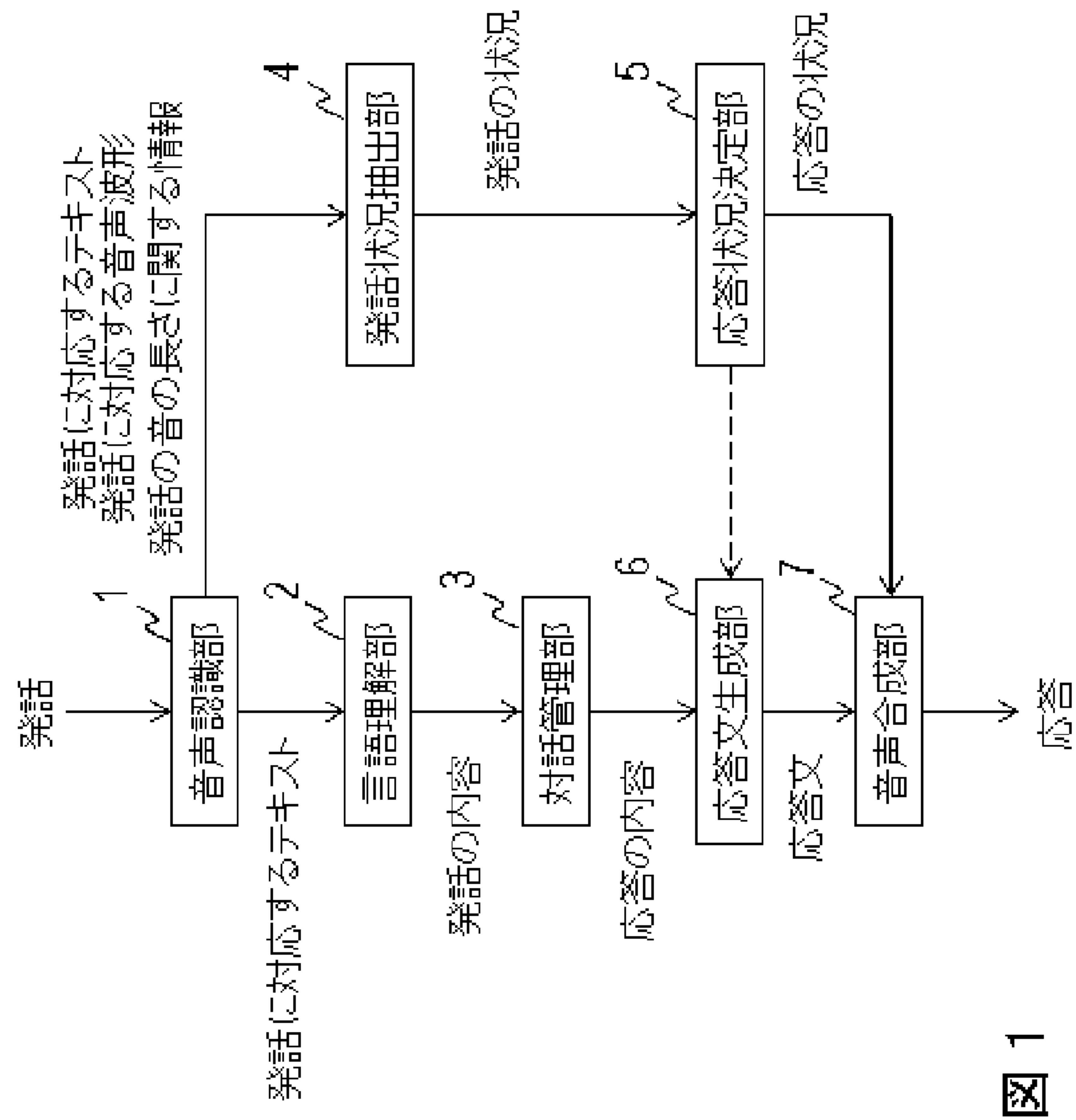


図 1

[図2]

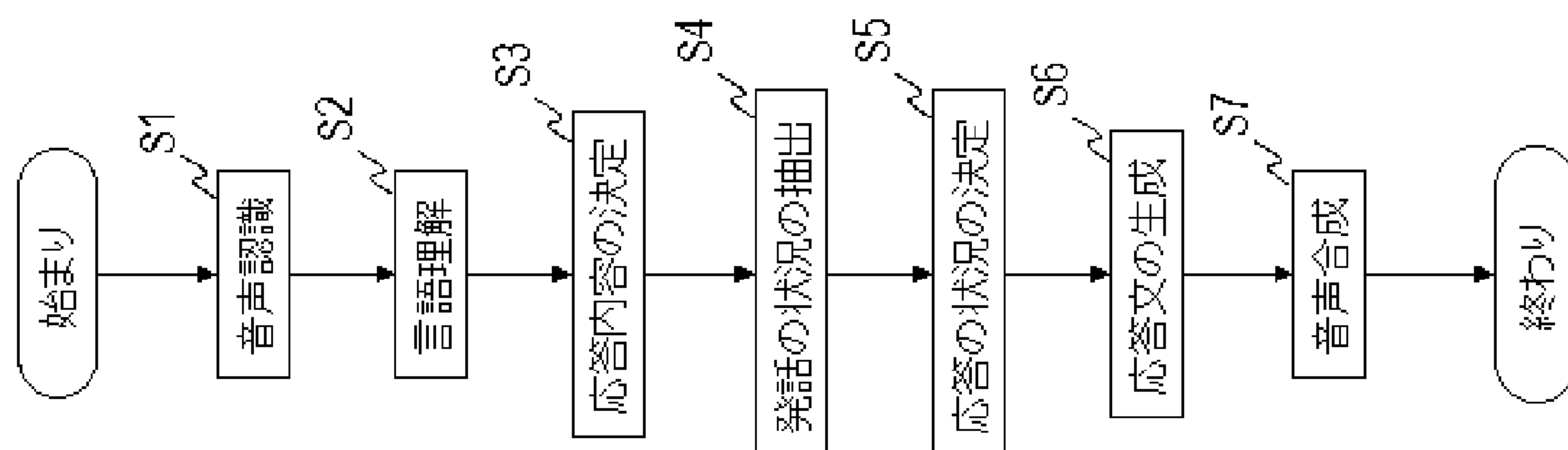


図2

[図3]

入力				出力(応答)			
話速	感情	口調	話速	感情	口調	話速	感情
普通	普通	丁寧	普通	普通	丁寧	普通	丁寧
普通	喜び	口語	普通	喜び	口語	普通	丁寧
速い	怒り	口語	遅い	普通	丁寧	普通	丁寧

図3

[図4]

入力				出力			
話速	感情	口調	発話の対話タイプ	応答の対話タイプ	話速	感情	口調
普通	普通	丁寧	質問	回答	普通	普通	丁寧
遅い	不安	丁寧	質問	回答	普通	穏やか	丁寧
遅い	不安	丁寧	質問	質問	遅い	恐縮	丁寧
普通	喜び	口語	あいさつ	あいさつ	普通	喜び	口調
遅い	落ち込み	口語	あいさつ	質問	遅い	穏やか	丁寧
普通	明るい	口語	質問	回答	普通	明るい	口調
普通	明るい	口語	質問	回答(嘘)	普通	悲しみ	口語
速い	怒り	口語	主張	謝罪	遅い	落ち込み	丁寧
速い	興奮	丁寧	質問	確認	普通	落ち着き	丁寧

図4

[図5]

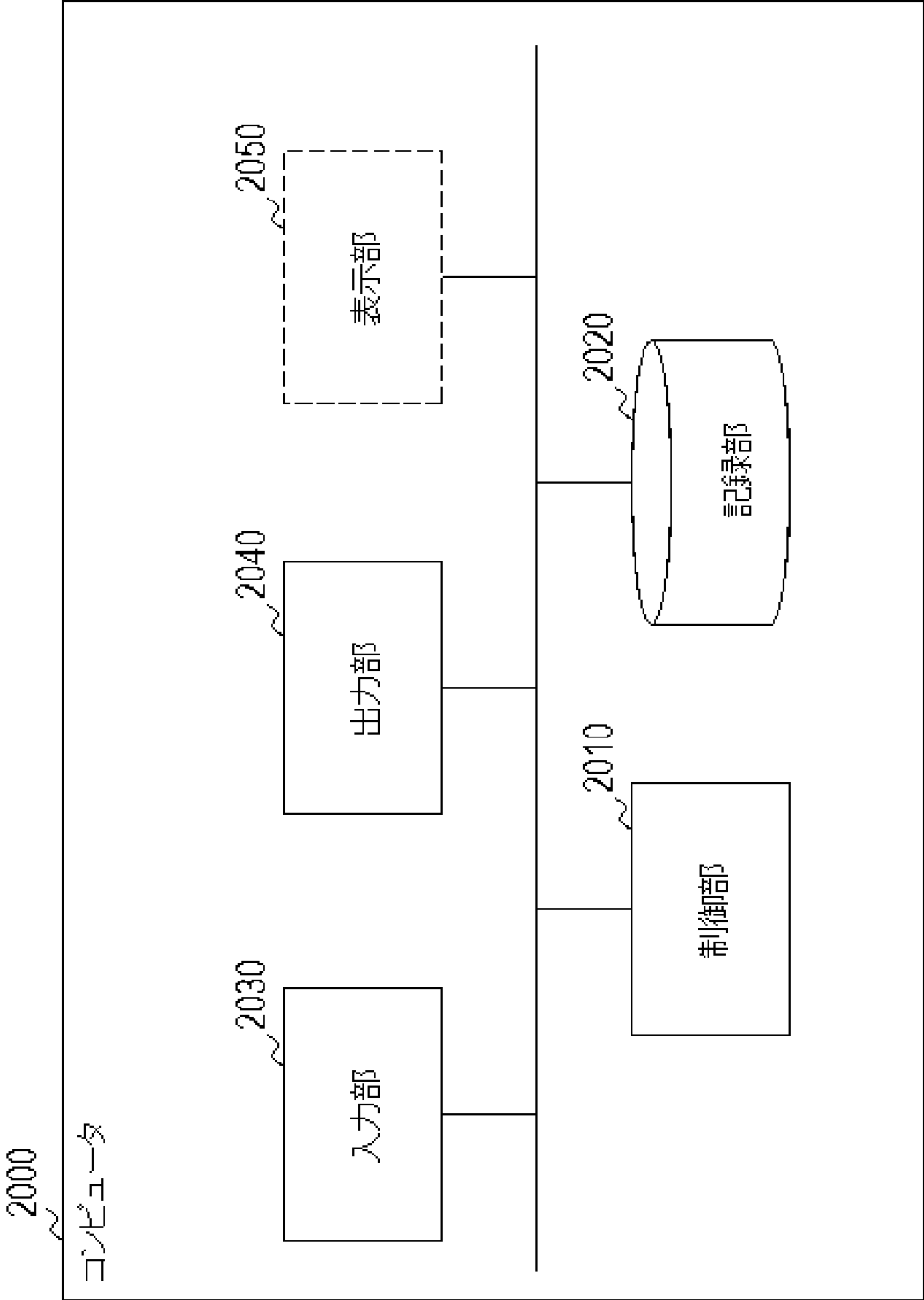


図5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2019/046184

A. CLASSIFICATION OF SUBJECT MATTER		
Int. Cl. G10L15/22(2006.01) i		
FI: G10L15/22 300U		
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. G10L15/22		
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 1971-2020		
Registered utility model specifications of Japan 1996-2020		
Published registered utility model applications of Japan 1994-2020		
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.
Y	JP 2001-272991 A (SANYO ELECTRIC CO., LTD.) 05 October 2001, in particular, paragraphs [0007], [0025]-[0046], fig. 5	1-6
Y	JP 2004-90109 A (SONY CORP.) 25 March 2004, in particular, paragraphs [0068]-[0076]	1-6
A	JP 2012-128440 A (DENSO CORP.) 05 July 2012, in particular, paragraph [0019]	1-6
☐ 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		Date of mailing of the international search report
03.06.2020		16.06.2020
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
Information on patent family members

International application No.
PCT/JP2019/046184

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
JP 2001-272991 A	05.10.2001	(Family: none)	
JP 2004-90109 A	25.03.2004	(Family: none)	
JP 2012-128440 A	05.07.2012	(Family: none)	

国際調査報告		国際出願番号 PCT/JP2019/046184
A. 発明の属する分野の分類（国際特許分類（I P C）） G10L 15/22(2006.01)i FI: G10L15/22 300U		
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）） 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 1 9 2 2 - 1 9 9 6 年 日本国公開実用新案公報 1 9 7 1 - 2 0 2 0 年 日本国実用新案登録公報 1 9 9 6 - 2 0 2 0 年 日本国登録実用新案公報 1 9 9 4 - 2 0 2 0 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
Y	JP 2001-272991 A（三洋電機株式会社）05.10.2001（2001 - 10 - 05） 特に、段落 [0 0 0 7] 、 [0 0 2 5] - [0 0 4 6] 、図 5	1-6
Y	JP 2004-90109 A（ソニー株式会社）25.03.2004（2004 - 03 - 25） 特に、段落 [0 0 6 8] - [0 0 7 6] 。	1-6
A	JP 2012-128440 A（株式会社デンソー）05.07.2012（2012 - 07 - 05） 特に、段落 [0 0 1 9] 。	1-6
☐ 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抵触するものではなく、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 03.06.2020	国際調査報告の発送日 16.06.2020	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 渡部 幸和 5Z 1162 電話番号 03-3581-1101 内線 3589	

PCT/JP2019/046184

引用文献			公表日	パテントファミリー文献	公表日
JP	2001-272991	A	05.10.2001	(ファミリーなし)	
JP	2004-90109	A	25.03.2004	(ファミリーなし)	
JP	2012-128440	A	05.07.2012	(ファミリーなし)	