

(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
2022年8月25日(25.08.2022)



(10) 国際公開番号

WO 2022/176124 A1

(51) 国際特許分類:

G10L 15/06 (2013.01) G10L 15/16 (2006.01)
G10L 15/10 (2006.01)

(21) 国際出願番号: PCT/JP2021/006176

(22) 国際出願日: 2021年2月18日(18.02.2021)

(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: 佐藤 宏 (SATO, Hiroshi); 〒1808585 東京都武蔵野市緑町三丁目9番11号 N

T T 知的財産センタ内 Tokyo (JP). 福富隆朗(FUKUTOMI, Takaaki); 〒1808585 東京都武蔵野市緑町三丁目9番11号 N T T 知的財産センタ内 Tokyo (JP). 篠原 雄介 (SHINOHARA, Yusuke); 〒1808585 東京都武蔵野市緑町三丁目9番11号 N T T 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NS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,

(54) Title: LEARNING DEVICE, ESTIMATION DEVICE, METHODS THEREFOR, AND PROGRAM

(54) 発明の名称: 学習装置、推定装置、それらの方法、およびプログラム

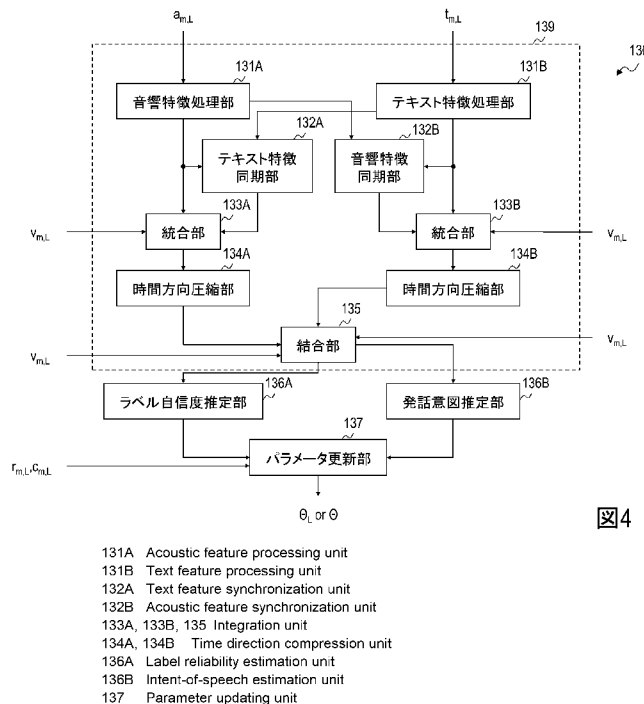


図4

(57) Abstract: The present invention estimates the intent of speech more accurately than conventional art. This learning device learns an estimation model on the basis of learning data which includes: an acoustic signal for learning; and a label indicating whether the acoustic signal was spoken to a specific object. The learning device includes: a feature synchronization unit that obtains a post-synchronization feature amount by synchronizing an acoustic feature amount obtained from the acoustic signal for learning, and a text feature amount corresponding to the acoustic signal; an intent-of-speech

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

estimation unit that uses the post-synchronization feature amount to estimate whether the acoustic signal was spoken to a specific object; and a parameter updating unit that updates a parameter of the estimation model on the basis of the label included in the learning data and the estimation results of the intent-of-speech estimation unit.

(57) 要約: 従来技術よりも精度よく発話意図を推定する。学習装置は、学習用の音響信号と、音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データに基づき、推定モデルを学習する。学習装置は、学習用の音響信号から得られる音響特徴量と、音響信号に対応するテキスト特徴量とを同期させることによって同期後特徴量を得る特徴同期部と、同期後特徴量を用いて、音響信号が所定の対象に対して発話されたものか否かを推定する発話意図推定部と、学習データに含まれるラベルと、発話意図推定部の推定結果とに基づいて推定モデルのパラメータを更新するパラメータ更新部とを含む。

明 細 書

発明の名称：

学習装置、推定装置、それらの方法、およびプログラム

技術分野

[0001] 本発明は、入力信号が所定の対象に向けて発話された音声かどうかを推定する推定モデルを学習する学習装置、学習済みの推定モデルを用いて推定する推定装置、それらの方法、およびプログラムに関する。

背景技術

[0002] 音声対話エージェントにおいて、入力される音声は必ずしもユーザがエージェントに向けて発した発話であるとは限らない。例えばユーザがその場にいる別の人間に向けて発した発話や、背景で再生されているテレビの音声などが入力される場合がある。こうしたユーザの意図しない発話が対話エージェントに入力されると、それによって対話シナリオが起動し、話しかけていないのにエージェントが喋り出す現象や、意図しない認識結果を検索してしまう現象が生じる。こうした誤作動はユーザエクスペリエンス低下の原因となる。

[0003] 入力信号が音声か否かを識別する技術として、発話区間検出技術が存在する。発話区間検出技術は入力信号から発話区間（音声区間）のみを検出し、発話していない区間（非音声区間）を除外する。しかし、発話区間検出技術では音声か否かを識別することはできるが、応答すべき音声か否かを識別することはできない。つまり、発話区間検出技術では、テレビの音声や別の話者に向けた発話などを、応答すべきでない音声として識別することはできない。

[0004] 応答すべきでない音声を識別する技術として、特に対話ロボットをはじめとする音声対話インターフェースを対象として、入力音声がユーザからロボットに向けられたものであるかどうかを判別し、音声に対して応答をするべきかどうかを識別する技術が存在する。例えば、非特許文献 1 が知られてい

る。

[0005] 非特許文献1ではスマートスピーカを想定し、音声から得られる音響的な特徴と、音声を認識した結果から得られる言語的な特徴とから発話意図の有無を識別する。なお、「発話意図」とは、ユーザが所定の対象に向けて音声を発する意図を意味し、発話意図の有無とは、所定の対象に入力された音声、ユーザが意図して対象に向けて発した音声であるか否かを意味する。ここで、所定の対象とは、入力された音声の対象に向けて発した音声か否かを識別することで、より適切に対象の目的を達することができるものであり、例えば、対話システムや電話等である。

先行技術文献

非特許文献

[0006] 非特許文献1: Mallidi, S. H., Maas, R., Goehner, K., Rastrow, A., Matsoukas, S., & Hoffmeister, B., "Device-directed utterance detection", arXiv preprint arXiv:1808.02504., 2018.

発明の概要

発明が解決しようとする課題

[0007] 非特許文献1では、音響的な特徴と認識結果の言語的な特徴とを用いて識別を行うが、音響特徴と言語特徴は別々にモデル化されており、両者の系列間に存在する時間的な対応関係は考慮できない。例えば、認識結果のある単語に対して、入力された音韻が、機械に話しかけているにだけは打ち解けすぎている、などのように言語的な特徴量のある一部分が、音響的な特徴量のどの部分に対応するかを考慮した精緻なモデル化ができない。

[0008] 本発明は、音響系列と言語系列の対応関係を捉えながら処理を行うことで、両者の系列間に存在する時間的な対応関係を考慮し、従来技術よりも精度よく発話意図を推定することができるモデルを学習する学習装置、モデルを用いた推定装置、それらの方法、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

[0009] 上記の課題を解決するために、本発明の一態様によれば、学習装置は、学習用の音響信号と、音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データに基づき、推定モデルを学習する。学習装置は、学習用の音響信号から得られる音響特徴量と、音響信号に対応するテキスト特徴量とを同期させることによって同期後特徴量を得る特徴同期部と、同期後特徴量を用いて、音響信号が所定の対象に対して発話されたものか否かを推定する発話意図推定部と、学習データに含まれるラベルと、発話意図推定部の推定結果とに基づいて推定モデルのパラメータを更新するパラメータ更新部とを含む。

[0010] 上記の課題を解決するために、本発明の他の態様によれば、推定装置は、学習用の音響信号と、学習用の音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データを用いて予め学習された推定モデルに基づき推定を行う。推定装置は、推定対象の音響信号から得られる音響特徴量と、推定対象の音響信号に対応するテキスト特徴量とを同期させることによって同期後特徴量を得る特徴同期部と、同期後特徴量を用いて、推定対象の音響信号が所定の対象に対して発話されたものか否かを推定する発話意図推定部とを含む。

発明の効果

[0011] 本発明によれば、音響系列と言語系列の対応関係を捉えながら処理を行うことで、両者の系列間に存在する時間的な対応関係が考慮し、従来技術よりも精度よく発話意図を推定することができるという効果を奏する。

図面の簡単な説明

[0012] [図1]第一実施形態に係る推定システムの構成例を示す図。

[図2]第一実施形態に係る学習装置の機能ブロック図。

[図3]第一実施形態に係る学習装置の処理フローを示す図。

[図4]第一実施形態に係るモデル学習部の機能ブロック図。

[図5]第一実施形態に係るモデル学習部の処理フローを示す図。

[図6]第一実施形態に係る推定装置の機能ブロック図。

[図7]第一実施形態に係る推定装置の処理フローを示す図。

[図8]第一実施形態に係る推定部の機能ブロック図。

[図9]第一実施形態に係る推定部の処理フローを示す図。

[図10]変形例2の構成、変形例1と変形例2とを組合せた構成の実験結果を示す図。

[図11]第二実施形態に係る学習装置の機能ブロック図。

[図12]第二実施形態に係る学習装置の処理フローを示す図。

[図13]第二実施形態に係るモデル学習部の機能ブロック図。

[図14]第二実施形態に係るモデル学習部の処理フローを示す図。

[図15]第二実施形態に係る推定装置の機能ブロック図。

[図16]第二実施形態に係る推定装置の処理フローを示す図。

[図17]第二実施形態に係る推定部の機能ブロック図。

[図18]第二実施形態に係る推定部の処理フローを示す図。

[図19]本手法を適用するコンピュータの構成例を示す図。

発明を実施するための形態

[0013] 以下、本発明の実施形態について、説明する。なお、以下の説明に用いる図面では、同じ機能を持つ構成部や同じ処理を行うステップには同一の符号を記し、重複説明を省略する。以下の説明において、ベクトルや行列の各要素単位で行われる処理は、特に断りが無い限り、そのベクトルやその行列の全ての要素に対して適用されるものとする。

[0014] <第一実施形態のポイント>

(1)音響的な特徴と認識結果の言語的な特徴の両者を時系列上で対応付けながらモデル化することで、より精緻なモデル化を行い、従来のモデルより精度の高い推定を行う。

[0015] (2)発話意図のラベリングをする際に、ラベリングの自信度も同時に記録し、自信度も学習時に利用することで、ラベルの信頼性を考慮したモデル学習を行う。これにより、不確実なラベルの影響を低減することができる。

[0016] (3)特徴量として、従来考慮されていなかった音源の放射方向や直間比などに

着目した新規の特徴量や、所定の対象に入力される発話としての妥当性に関連した新規の特徴量を導入することで、発話意図の有無をより明示的に捉えることができる。

[0017] <第一実施形態に係る推定システム>

図1は推定システムの構成例を示す。

[0018] 推定システムは、学習装置100と推定装置200とを含む。

[0019] 学習装置100は、学習データ S_L を入力とし、学習データ S_L に基づき、推定モデル Θ_L を学習し、学習済みの推定モデル Θ を出力する。なお、学習データ S_L には、学習用の M 個の音響信号 $s_{m,L}$ と、ラベル $r_{m,L}$ と、自信度 $c_{m,L}$ とが含まれる。

[0020] $S_L = ((s_{1,L}, r_{1,L}, c_{1,L}), (s_{2,L}, r_{2,L}, c_{2,L}), \dots, (s_{M,L}, r_{M,L}, c_{M,L}))$

ラベル $r_{m,L}$ は m 番目の学習用の音響信号 $s_{m,L}$ が所定の対象に対して発話されたものか否か（発話意図の有無）を表し、例えば $r_{m,L}=0$ は発話意図が無いことを意味し、 $r_{m,L}=1$ は発話意図が有ることを意味する。自信度 $c_{m,L}$ は、アノテータ（対象にラベル付けを行うもの）のラベル付与の自信度を表す。

[0021] 推定装置200は、推定処理に先立ち学習済みの推定モデル Θ を受け取る。推定装置200は、推定対象の音響信号 s_T を入力とし、推定モデル Θ に基づき、音響信号 s_T が所定の対象に向けられて発話された音声か否かを推定し、推定結果 R （発話意図の有無の推定値）を出力する。

[0022] 学習装置および推定装置は、例えば、中央演算処理装置（CPU:Central Processing Unit）、主記憶装置（RAM:Random Access Memory）などを有する公知又は専用のコンピュータに特別なプログラムが読み込まれて構成された特別な装置である。学習装置および推定装置は、例えば、中央演算処理装置の制御のもとで各処理を実行する。学習装置および推定装置に入力されたデータや各処理で得られたデータは、例えば、主記憶装置に格納され、主記憶装置に格納されたデータは必要に応じて中央演算処理装置へ読み出されて他の処理に利用される。学習装置および推定装置の各処理部は、少なくとも一部が集積回路等のハードウェアによって構成されていてもよい。学習装置およ

び推定装置が備える各記憶部は、例えば、RAM (Random Access Memory) などの主記憶装置、またはリレーショナルデータベースやキーバリューストアなどのミドルウェアにより構成することができる。ただし、各記憶部は、必ずしも学習装置および推定装置がその内部に備える必要はなく、ハードディスクや光ディスクもしくはフラッシュメモリ (Flash Memory) のような半導体メモリ素子により構成される補助記憶装置により構成し、学習装置および推定装置の外部に備える構成としてもよい。

[0023] まず、学習装置 100 について説明する。

[0024] <第一実施形態に係る学習装置 100>

図 2 は第一実施形態に係る学習装置 100 の機能ブロック図を、図 3 はその処理フローを示す。

[0025] 学習装置 100 は、音声認識部 110、特徴量算出部 120、モデル学習部 130 を含む。

[0026] 各部について説明する。

[0027] <音声認識部 110>

音声認識部 110 は、学習用の音響信号 $s_{m,L}$ を入力とし、音声認識を実行し (S 110)、音声認識に基づく情報 $y_{m,L}$ を得、出力する。音声認識に基づく情報は、音声認識結果および、音声認識を実行した際の認識結果の信頼度や音声認識の計算時間などのデータの少なくとも何れかを含む。こうした音声認識結果の言語的な情報や、認識時の信頼度などのデータは、発話意図の有無を推定するために用いられる。

[0028] <特徴量算出部 120>

特徴量算出部 120 は、音響信号 $s_{m,L}$ および音声認識に基づく情報 $y_{m,L}$ を入力とし、特徴量 $o_{m,L}$ を算出し (S 120)、出力する。特徴量 $o_{m,L}$ は、発話意図の有無を推定するために用いられる。例えば、 m 番目の発話の特徴量 $o_{m,L}$ は、 N_m 個の特徴量 $o_{m,L,n}$ を含むものとし、 $n=1, 2, \dots, N_m$ 、 $o_{m,L}=(o_{m,L,1}, \dots, o_{m,L,N_m})$ とする。ただし、下付き添え字 A_B は A_B を意味する。

[0029] 特徴量 $o_{m,L}$ は「音響特徴量 $a_{m,L}$ 」と「テキスト特徴量 $t_{m,L}$ 」と「その他の特徴

量 $v_{m,L}$ 」とのいずれかあるいは組み合わせを含むベクトルであり、「音響特徴量 $a_{m,L}$ 」と「テキスト特徴量 $t_{m,L}$ 」と「その他の特徴量 $v_{m,L}$ 」はそれぞれ1つ以上の要素(特徴量)を含むベクトルである。

[0030] 「音響特徴量」としては短時間のフレームに対して算出されたMFCC(Mel-Frequency Cepstrum Coefficients)やFBANK特徴量などの公知の音響特徴量の、時系列データ、あるいはそれらに対して時間方向に平均を取るなどの処理を施したものを利用することができる。音響特徴量は、音響信号 $s_{m,L}$ から直接求めてもよいし、音声認識部110の音声認識の処理過程で算出される公知の音響特徴量を利用してもよい。なお、音声認識部110の音声認識の処理過程で算出される公知の音響特徴量を利用する場合には、音響信号 $s_{m,L}$ を入力として受け付けなくともよい。

[0031] 「テキスト特徴量」としては音声認識に基づく情報 $y_{m,L}$ に含まれる音声認識結果や認識結果の候補に対して、単語系列や文字の系列をword 2 vecなどの公知の手法でベクトル系列化したものである。音声認識結果や認識結果の候補からは、当該発話が所定の対象に入力されやすいものであるかどうかを推定することができる。

[0032] 「その他の特徴量」は、音響信号 $s_{m,L}$ から得られる特徴量と、音声認識に基づく情報 $y_{m,L}$ から得られる特徴量とがある。

[0033] 音響信号 $s_{m,L}$ から得られる「その他の特徴量」として、以下の(i),(ii)が考えられる。

(i)音源の位置あるいは方向、音源までの距離に関する情報:入力音から公知の手法で算出される音源の位置あるいは方向、また入力音から公知の方法で算出される直間比のような遠近感の情報(音源までの距離)を利用できる。加えてこれらの時間変動を特徴量として利用できる。例えば音声から得られる直間比からは音源の遠さを捉えることができ発話意図の推定に有用である。複数チャネルの音声を得られる場合は、精密に音源の距離や方向、音源からの音の放射方向の情報を算出可能である。さらに音源の距離や方向の情報は時間変動を見ることで音源位置に揺らぎがある人間の発話か、固定された

テレビやスピーカー等の音源かを判別できる。

[0034] (ii)音響信号帯域幅または周波数特性に関する情報：入力音の帯域幅や周波数特性などの情報を利用できる。これらの情報は、公知の技術を利用して音響信号 $s_{m,L}$ を用いて求めることができる。音声の帯域幅からは入力音がラジオやテレビなどの再生音であることを捉えることができる。

[0035] 音声認識に基づく情報 $y_{m,L}$ から得られる「その他の特徴量」として、以下の(iii)～(v)が考えられる。

(iii)音声認識結果の信頼度または音声認識の計算時間に関する情報：音声認識に基づく情報 $y_{m,L}$ に含まれる音声認識結果の信頼度や音声認識の計算時間などの情報を利用することができる。発話意図のない発話は一般に音声認識が困難であることから音声認識の信頼度などの情報も特徴量として有用である。

[0036] (iv)音声認識結果から計算される発話のコマンドとしての妥当性に関する情報：音声認識結果から計算される発話のコマンドとしての妥当性などを利用することができる。発言のコマンドとしての妥当性とは、例えばデバイスの持つコマンドのリストの各要素と、認識結果との一致度合いの最大値とする。一致度合いとして例えばコマンドの単語数のうち認識結果に含まれた割合を採用することができる。又はコマンドと認識結果それぞれをTF-IDF(Term Frequency - Inverse Document Frequency)やbag of wordsなど公知の手法でベクトル化したものの距離とすることができる。

[0037] (v)音声認識結果から得られる入力発話の解釈難度に関する情報：音声認識結果から得られる入力発話の解釈難度などを利用することができる。発言の解釈難度は人間が機械にむけて発話する際にわかりやすい言葉を選ぶ特性も捉えるもので、例えば単語数に表れる発話の長さや、認識結果を構文解析した結果から得られる、指示代名詞の有無、助詞の省略の有無とすることができる。

[0038] これらの特徴の一部あるいは組み合わせを「その他の特徴量」として利用できる。

[0039] 以上の特徴量をモデル学習部130に入力することで、モデルの識別性能を高めることができる。

[0040] <モデル学習部130>

モデル学習部130は、学習データ S_L に含まれるラベル $r_{m,L}$ と自信度 $c_{m,L}$ と特徴量 $0_L=(o_{1,L}, o_{2,L}, \dots, o_{M,L})$ を入力とし、これらの情報を用いて推定モデル Θ_L を学習し(S130)、学習済みの推定モデル Θ を出力する。推定モデルは特徴量 0_L から発話意図の有無を推定する2値分類モデルであり、公知の深層学習技術を用いて学習することができる。前述の通り、特徴量 $o_{m,L}$ は、「音響特徴量 $a_{m,L}$ 」と「テキスト特徴量 $t_{m,L}$ 」と「その他の特徴量 $v_{m,L}$ 」とのいずれかあるいは組み合わせを含むベクトルである。

[0041] 本実施形態では、1発話分の音響信号に対応する特徴量 $o_{m,L}$ と、その発話の発話意図ラベル $r_{m,L}$ と、発話意図のラベリングを行う際にラベリングの自信度 $c_{m,L}$ とからなる学習データを用いて学習を行うことができる。この場合、発話意図の識別モデルは、1発話の音声から発話意図の予測ラベルだけでなく、同時にそのデータに対するアノータのラベリングの自信度も予測を行う。学習時には、発話意図の予測の正誤に関する損失関数と、アノータのラベリング自信度の予測誤差に関する損失関数の値を重み付け和したものを損失関数とするマルチタスク学習を実施する。これにより、モデルはアノータのラベリング自信度を考慮しつつ発話意図の学習を行うことができるようになる。

[0042] 発話意図の識別モデルは、特徴量算出部120で算出された、音響特徴量 $a_{m,L}$ の時系列データ、またテキスト特徴量 $t_{m,L}$ の時系列データ、またその他の特徴量 $v_{m,L}$ の時系列データ、また時系列ではないその他の特徴量 $v_{m,L}$ を入力に持ち、発話意図の有無を表すラベルの推定値と自信度の推定値を出力に持つDNN(Deep Neural Network)のモデルである。このうち時系列の特徴量については、CNN(Convolutional Neural Network)やLSTM(Long Short Term Memory)、Self-Attentionと呼ばれる技術を用いることで、長時間の系列の関連性を考慮しながら、時系列の長さによらない固定長のベクトルに変換することができる。

。時系列方向の次元を持たないもともと固定長の特徴量については、例えば時系列方向の情報を持つ特徴量の各時刻と結合したり、時系列方向の情報を持つベクトルを固定長化した後で結合したりすることで統合を行うことができる。各特徴量を統合したベクトルに対して公知のDNNによって発話意図を出力するモデルを構築する。

[0043] 特に、音響特徴量の時系列データとテキスト特徴量の時系列データについては両者の間に時間的な対応関係が存在することから、これを考慮しつつモデル化を行うことで、より精緻に発話意図の有無を捉えることができる。これらを固定長化する際にこれを実現するために、与えられた発話意図ラベルから発話意図の識別を学習するのと同時に、両時系列データの時間的な対応関係を捉えることのできるモデル化手法を採用することが有効である。これには、例えば参考文献1のSource Target Attentionと呼ばれるNW構造を採用してもよい。

[0044] (参考文献1) Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, "Neural machine translation by jointly learning to align and translate", in International Conference on Learning Representations (ICLR), 2015

具体的には、音響系列、言語系列それぞれをLSTMなどのDNNで処理した時系列 X_A 、 X_L に対して、

$$Y_{A+L} = \text{Source Target Attention}(X_A, X_L, X_L)$$

$$Y_{L+A} = \text{Source Target Attention}(X_L, X_A, X_A)$$

とあらわされる処理を行うことで、音響特徴量系列と同期した言語特徴量系列 Y_{A+L} および言語特徴量系列と同期した音響特徴量系列 Y_{L+A} を得ることができる。ただしここでSource Target Attention (Q, K, V)とはQをクエリ、Kをキー、VをバリューとするSource Target Attentionを示す。こうして得られた Y_{A+L} 、 Y_{L+A} を例えばそれぞれ X_A 、 X_L と特徴次元方向に結合したり、加算したりすることで統合し、それらに対してLSTMをはじめとするDNNで処理を行うことで、両者の系列の時間的な対応関係を捉える精緻なモデル化が可能である。例えば、

音響特徴量にテキスト特徴量を同期させる場合、音響特徴量の各フレームに対し、その時刻に対応するテキスト特徴量のフレームを重み付けて取得する。なお、重みは、ニューラルネットワークにより与えるか、または、音声認識によって得られた音響特徴量系列と言語系列のアラインメント情報をattentionの重みとして利用してもよい。逆に、テキスト特徴量に音響特徴量を同期させる場合、テキスト特徴量の各フレームに対し、その時刻に対応する音響特徴量のフレームを重み付けて取得する。なお、重みは、ニューラルネットワークにより与えるか、または、音声認識によって得られた音響特徴量系列と言語系列のアラインメント情報をattentionの重みとして利用してもよい。

[0045] 上述の処理を実現するモデル学習部130の構成について説明する。

[0046] 図4はモデル学習部130の機能ブロック図を、図5はその処理フローの例を示す。

[0047] モデル学習部130は、特徴同期部139、ラベル自信度推定部136A、発話意図推定部136Bおよびパラメータ更新部137を含む。

[0048] <特徴同期部139>

特徴同期部139は、特徴量 $0_L = (o_{1,L}, o_{2,L}, \dots, o_{M,L})$ を入力とし、音響特徴量 $a_{m,L}$ と、音響信号 $a_{m,L}$ に対応するテキスト特徴量 $t_{m,L}$ とを同期させることによって同期後特徴量を得（S139）、出力する。例えば、特徴同期部139は、音響特徴処理部131A、テキスト特徴処理部131B、テキスト特徴同期部132A、音響特徴同期部132B、統合部133A、133B、時間方向圧縮部134A、134Bおよび結合部135を含み（図4参照）、以下の処理を行う。

[0049] <音響特徴処理部131A>

音響特徴処理部131Aは、音響特徴量 $a_{m,L}$ を入力とし、テキスト特徴同期部132Aおよび統合部133Aにおいて、処理しやすいデータに変換し（S131A）、変換後の音響特徴量を出力する。なお、変換後の音響特徴量を単に音響特徴量ともいう。例えば、時系列モデリングを行うDNNを用いて、

音響特徴量を変換する。ただし、テキスト特徴同期部 1 3 2 A および統合部 1 3 3 A が変換前の音響特徴量 $a_{m,L}$ をそのまま用いる場合には、音響特徴処理部 1 3 1 A を設けなくともよい。

[0050] <テキスト特徴処理部 1 3 1 B>

テキスト特徴処理部 1 3 1 B は、テキスト特徴量 $t_{m,L}$ を入力とし、テキスト特徴同期部 1 3 2 B および統合部 1 3 3 B において、処理しやすいデータに変換し (S 1 3 1 B)。変換後のテキスト特徴量を出力する。なお、変換後のテキスト特徴量を単にテキスト特徴量ともいう。例えば、時系列モデリングを行う DNN を用いて、テキスト特徴量を変換する。ただし、テキスト特徴同期部 1 3 2 B および統合部 1 3 3 B が変換前のテキスト特徴量 $t_{m,L}$ をそのまま用いる場合には、テキスト特徴処理部 1 3 1 B を設けなくともよい。

[0051] <テキスト特徴同期部 1 3 2 A>

テキスト特徴同期部 1 3 2 A は、音響特徴量とテキスト特徴量とを入力とし、音響特徴量にテキスト特徴量を同期させ (S 1 3 2 A)、音響特徴量の各フレームに対応付けたテキスト特徴量 (以下、同期後テキスト特徴量ともいう) を出力する。例えば、音響特徴量の時系列、テキスト特徴量の時系列それぞれを LSTM などの DNN で処理した時系列 X_A 、 X_L に対して、

$$Y_{A+L} = \text{Source Target Attention} (X_A, X_L, X_L)$$

とあらわされる処理を行うことで、音響特徴量の時系列 X_A と同期したテキスト特徴量の時系列 Y_{A+L} を得る。

[0052] <音響特徴同期部 1 3 2 B>

音響特徴同期部 1 3 2 B は、テキスト特徴量と音響特徴量とを入力とし、テキスト特徴量に音響特徴量を同期させ (S 1 3 2 B)、テキスト特徴量の各フレーム (各文字や単語) に対応付けた音響特徴量 (以下、同期後音響特徴量ともいう) を出力する。例えば、音響特徴量の時系列、テキスト特徴量の時系列それぞれを LSTM などの DNN で処理した時系列 X_A 、 X_L に対して、

$$Y_{L+A} = \text{Source Target Attention} (X_L, X_A, X_A)$$

とあらわされる処理を行うことで、テキスト特徴量の時系列 X_L と同期した音響

特徴量の時系列 Y_{L+A} を得る。

[0053] <統合部 1 3 3 A, 1 3 3 B>

統合部 1 3 3 A は、同期後テキスト特徴量と音響特徴量とを入力とし、これらの特徴量を結合し (S 1 3 3 A)、出力する。

[0054] 統合部 1 3 3 B は、同期後音響特徴量とテキスト特徴量とを入力とし、これらの特徴量を結合し (S 1 3 3 B)、出力する。

[0055] 統合部 1 3 3 A は、その他の特徴量 $v_{m,L}$ を入力とし、同期後テキスト特徴量と音響特徴量と結合し、出力してもよい。同様に、統合部 1 3 3 B は、その他の特徴量 $v_{m,L}$ を入力とし、同期後音響特徴量とテキスト特徴量と結合し、出力してもよい。

[0056] なお、その他の特徴量 $v_{m,L}$ が時間方向に長さを持つ場合には、統合部 1 3 3 A は「音響特徴量」と「同期後テキスト特徴量」と「その他の特徴量 $v_{m,L}$ 」とを時系列を考慮して結合し、統合部 1 3 3 B は「テキスト特徴量」と「同期後音響特徴量」と「その他の特徴量 $v_{m,L}$ 」とを時系列を考慮して結合する。

[0057] その他の特徴量 $v_{m,L}$ が時間方向に長さを持たない場合には、統合部 1 3 3 A は音響特徴量のフレーム数分、その他の特徴量を複製し、音響特徴量 $a_{m,L}$ のフレームごとに「音響特徴量 $a_{m,L}$ 」と「同期後テキスト特徴量」と「その他の特徴量」を結合し、統合部 1 3 3 B はテキスト特徴量のフレーム (各文字や単語) 数分、その他の特徴量を複製し、テキスト特徴量 $t_{m,L}$ のフレームごとに「テキスト特徴量」と「同期後音響特徴量」と「その他の特徴量」を結合する。

[0058] <時間方向圧縮部 1 3 4 A, 1 3 4 B>

時間方向圧縮部 1 3 4 A, 1 3 4 B は、それぞれ統合部 1 3 3 A, 1 3 3 B の出力した特徴量を、言い換えると、時間方向に長さを持つ、音響由来の特徴量、テキスト由来の特徴量、その他の特徴量を入力とし、時間方向に圧縮して (S 1 3 4 A, S 1 3 4 B)、次元の固定長のベクトルにし、出力する。時間方向の圧縮処理には、様々な公知の技術を利用することができ、例えばself-attention poolingなどを利用してもよい。時間方向集約部 1 3

4 A, 1 3 4 B には、一次元の固定長のベクトルにする前に、LSTMなどの時系列モデルを含める構成としてもよい。

[0059] <結合部 1 3 5>

結合部 1 3 5 は、時間方向圧縮部 1 3 4 A, 1 3 4 B がそれぞれ出力した一次元の固定長のベクトルを入力とし、これらのベクトルを結合し (S 1 3 5)、結合後のベクトルである同期後特徴量を出力する。

[0060] なお、結合部 1 3 5 は、その他の特徴量 $v_{m,L}$ のうち時系列方向の次元を持たない特徴量を入力とし、時間方向圧縮部 1 3 4 A, 1 3 4 B がそれぞれ出力した一次元の固定長のベクトルとその他の特徴量 $v_{m,L}$ のうち時系列方向の次元を持たない特徴量とを結合し、結合後のベクトルである同期後特徴量を出力してもよい。

[0061] <ラベル自信度推定部 1 3 6 A>

ラベル自信度推定部 1 3 6 A は、同期後特徴量を入力とし、同期後特徴量を用いて、ラベル自信度推定モデルに基づき、ラベルを付与する際の自信度を推定し (S 1 3 6 A)、推定結果 (ラベル自信度の推定値) を出力する。なお、ラベル自信度推定モデルは、同期後特徴量を入力とし、ラベル自信度の推定値を出力とするモデルであり、例えば、DNN からなる。

[0062] <発話意図推定部 1 3 6 B>

発話意図推定部 1 3 6 B は、同期後特徴量を入力とし、同期後特徴量を用いて、発話意図推定モデルに基づき、学習用の音響信号が所定の対象に対して発話されたものか否かを推定し (S 1 3 6 B)、推定結果 (発話意図ラベルの推定値) を出力する。なお、発話意図推定モデルは、同期後特徴量を入力とし、発話意図ラベルの推定値を出力とするモデルであり、例えば、DNN からなる。

[0063] <パラメータ更新部 1 3 7>

パラメータ更新部 1 3 7 は、学習データ S_L に含まれるラベル $r_{m,L}$ と自信度 $c_{m,L}$ と、ラベル自信度の推定値と、発話意図ラベルの推定値とを入力とし、これらの値に基づいて、推定モデルのパラメータを更新する (S 1 3 7)。なお

、推定モデルは、音響信号から得られる音響特徴量とその音響信号に対応するテキスト特徴量とを入力として、その音響信号の発話意図ラベルの推定値を出力とするモデルである。例えば、パラメータ更新部137は、ラベル $r_{m,L}$ と発話意図ラベルの推定値とが一致し、自信度 $c_{m,L}$ とラベル自信度の推定値とが一致するように、音響特徴処理部131A、テキスト特徴処理部131B、テキスト特徴同期部132A、音響特徴同期部132B、時間方向圧縮部134A、134B、ラベル自信度推定部136A、発話意図推定部136Bにおいて用いるパラメータを更新する。

[0064] 収束条件を満たさない場合には（S137-2のnoの場合）、パラメータ更新部137は更新後のパラメータを各部に出力し、更新されたパラメータを用いて、上述の処理S131A～S136Bを繰り返す。

[0065] 収束条件を満たす場合には（S137-2のyesの場合）、パラメータ更新部137は更新後のパラメータを学習済みのパラメータを含む推定モデル Θ として出力する。

[0066] 収束条件は、パラメータの更新が収束したか否かを判別するための条件である。例えば、収束条件は、更新回数が所定の回数を超えることや、更新前後のパラメータの差分が所定の閾値未満であることなどである。

[0067] 次に、推定装置200について説明する。

[0068] <第一実施形態に係る推定装置200>

図6は第一実施形態に係る推定装置200の機能ブロック図を、図7はその処理フローを示す。

[0069] 学習装置100は、音声認識部210、特徴量算出部220、推定部230を含む。

[0070] 各部について説明する。

[0071] <音声認識部210>

音声認識部210は、推定対象の音響信号 s_T を入力とし、音声認識を実行し（S210）、音声認識に基づく情報 y_T を得、出力する。例えば、音声認識部210は音声認識部110と同様の音声認識処理を行う。

[0072] <特徴量算出部 2 2 0>

特徴量算出部 2 2 0 は、音響信号 s_T および音声認識に基づく情報 y_T を入力とし、特徴量 o_T を算出し (S 2 2 0)、出力する。例えば、特徴量算出部 2 2 0 は特徴量算出部 1 2 0 と同様の特徴量算出処理を行う。

[0073] <推定部 2 3 0>

推定部 2 3 0 は、推定処理に先立ち学習済みの推定モデル Θ を受け取る。

[0074] 推定部 2 3 0 は、特徴量 o_T を入力とし、学習済みの推定モデル Θ を用いて、発話意図の有無を推定し (S 2 3 0)、推定結果 R を出力する。学習済みの推定モデル Θ に対して入力として特徴量 o_T を与え、出力として発話意図の有無の推定結果 R を得る。推定結果 R は、例えば、発話意図の有無を示す 2 値のラベルである。

[0075] 図 8 は推定部 2 3 0 の機能ブロック図を、図 9 はその処理フローの例を示す。

[0076] 推定部 2 3 0 は、特徴同期部 2 3 9 および発話意図推定部 2 3 6 を含む。さらに、特徴同期部 2 3 9 は、音響特徴処理部 2 3 1 A、テキスト特徴処理部 2 3 1 B、テキスト特徴同期部 2 3 2 A、音響特徴同期部 2 3 2 B、統合部 2 3 3 A、2 3 3 B、時間方向圧縮部 2 3 4 A、2 3 4 B および結合部 2 3 5 を含む。

[0077] 特徴同期部 2 3 9 および発話意図推定部 2 3 6 は、それぞれ特徴同期部 1 3 9 および発話意図推定部 1 3 6 と同様の処理 S 2 3 9、S 2 3 6 を行う。よって、特徴同期部 2 3 9 内の音響特徴処理部 2 3 1 A、テキスト特徴処理部 2 3 1 B、テキスト特徴同期部 2 3 2 A、音響特徴同期部 2 3 2 B、統合部 2 3 3 A、2 3 3 B、時間方向圧縮部 2 3 4 A、2 3 4 B および結合部 2 3 5 は、それぞれ特徴同期部 1 3 9 内の音響特徴処理部 1 3 1 A、テキスト特徴処理部 1 3 1 B、テキスト特徴同期部 1 3 2 A、音響特徴同期部 1 3 2 B、統合部 1 3 3 A、1 3 3 B、時間方向圧縮部 1 3 4 A、1 3 4 B および結合部 1 3 5 と同様の処理 S 2 3 1 S ~ S 2 3 5 を行う。ただし、特徴量 $o_{m,L}$ に基づく値に代えて特徴量 o_T に基づく値に対して各処理を行う。

[0078] <効果>

このような構成により、音響系列と言語系列の対応関係を捉えながら処理を行うことで、両者の系列間に存在する時間的な対応関係が考慮し、従来技術よりも精度よく発話意図を推定することができる。

[0079] また、非特許文献1では、モデルを学習する際、人手でアノテーションした発話意図の正解ラベルを必要とする。しかし、音声対話インターフェースの音声ログに対してアノテーションを実施する場合、一部のデータについては人間にも発話意図の識別が困難であることから、不正確なラベルが付与される場合があり、そうした不正確なラベルで学習を行うことで、識別精度が低下する問題がある。ラベルの不正確さに対処する方法として一般に、複数人のアノテータでラベリングを行い、多数決をとる方法がとられるが、アノテータ数に比例したコストがかかる上に、いずれのアノテータも判別が困難な難しいデータについては対処できない。

[0080] 本実施形態によれば、人間でもアノテーションが困難なデータを、「正解するのが困難である」という事を認識したうえでモデルを学習することができる。

[0081] また、発話意図を精度よく推定することで、こうしたユーザエクスペリエンス低下の原因となる誤作動の防止に寄与する。

[0082] さらに、音声認識以外のタスクとして電話等のテレコミュニケーションにおいて、本実施形態の推定システムを用いることにより、対話しているユーザの音声のみを精度よく相手に提示できる。

[0083] 例えば、本実施形態の推定処理S230を音声認識装置の後段の処理として行い、認識仮説とともにアプリケーションに渡す形での利用などが考えられる。

[0084] <変形例1：自信度を用いない構成>

第一実施形態と異なる部分を中心に説明する。

[0085] 第一実施形態では、アノテータ(対象にラベル付けを行うもの)のラベル付与の自信度 $c_{m,L}$ を学習に利用したが、本変形例では自信度 $c_{m,L}$ を利用しない。

[0086] この場合、学習データ S_L には、学習用の M 個の音響信号 $s_{m,L}$ と、ラベル $r_{m,L}$ とが含まれる。

[0087] $S_L = ((s_{1,L}, r_{1,L}), (s_{2,L}, r_{2,L}), \dots, (s_{M,L}, r_{M,L}))$

モデル学習部 130 は、ラベル自信度推定部 136 A を含まない。

[0088] パラメータ更新部 137 は、学習データ S_L に含まれるラベル $r_{m,L}$ と、発話意図ラベルの推定値とを入力とし、これらの値に基づいて、推定モデルのパラメータを更新する (S 137)。例えば、パラメータ更新部 137 は、ラベル $r_{m,L}$ と発話意図ラベルの推定値とが一致するように、音響特徴処理部 131 A、テキスト特徴処理部 131 B、テキスト特徴同期部 132 A、音響特徴同期部 132 B、時間方向圧縮部 134 A、134 B、発話意図推定部 136 B において用いるパラメータを更新する (S 137)。

[0089] <変形例 2 : その他の特徴量を用いない構成>

第一実施形態と異なる部分を中心に説明する。

[0090] 第一実施形態では、その他の特徴量を学習および推定に利用したが、本変形例ではその他の特徴量を利用しない。

[0091] 特徴量算出部 120、220 で算出する特徴量には、その他の特徴量を含まない。そのため、統合部 133 A、133 B、233 A、233 B、統合部 135、235 でその他の特徴量を統合、結合する処理は行わない。

[0092] 図 10 は、変形例 2 の構成、変形例 1 と変形例 2 とを組合せた構成の実験結果を示す。何れの構成についても従来技術よりも精度よく発話意図を推定することができる。

[0093] <変形例 3>

第一実施形態と異なる部分を中心に説明する。

[0094] ラベル自信度推定部 136 A および発話意図推定部 136 B は、必ずしも結合後のベクトルを入力としなくともよく、時間方向圧縮部 134 A の出力するベクトルと時間方向圧縮部 134 B の出力するベクトルとの少なくとも何れか一方を入力とし、ラベル自信度および発話意図ラベルの推定値を求める構成としてもよい。この場合、同期後特徴量は、時間方向圧縮部 134 A

の出力するベクトルと時間方向圧縮部 1 3 4 B の出力するベクトルとの少なくとも何れか一方を含む。モデル学習部 1 3 0 は、時間方向圧縮部 1 3 4 A の出力するベクトルを同期後特徴量とする場合、音響特徴同期部 1 3 2 B、統合部 1 3 3 B、時間方向圧縮部 1 3 4 B、結合部 1 3 5 を含まなくともよく、時間方向圧縮部 1 3 4 B の出力するベクトルを同期後特徴量とする場合、音響特徴同期部 1 3 2 A、統合部 1 3 3 A、時間方向圧縮部 1 3 4 A、結合部 1 3 5 を含まなくともよい。この場合、ラベル自信度推定モデルは、時間方向圧縮部 1 3 4 A の出力するベクトルと時間方向圧縮部 1 3 4 B の出力するベクトルとの少なくとも何れか一方を含む同期後特徴量を入力とし、ラベル自信度の推定値を出力とするモデルである。同様に、発話意図推定モデルは、時間方向圧縮部 1 3 4 A の出力するベクトルと時間方向圧縮部 1 3 4 B の出力するベクトルとの少なくとも何れか一方を含む同期後特徴量を入力とし、発話意図ラベルの推定値を出力とするモデルである。

[0095] 発話意図推定部 2 3 6 も同様に、必ずしも結合後のベクトルを入力としなくともよく、時間方向圧縮部 2 3 4 A の出力するベクトルと時間方向圧縮部 2 3 4 B の出力するベクトルとの少なくとも何れか一方を含む同期後特徴量を入力とし、発話意図ラベルの推定値を求める構成としてもよい。推定部 2 3 0 は、時間方向圧縮部 2 3 4 A の出力するベクトルを同期後特徴量とする場合、音響特徴同期部 2 3 2 B、統合部 2 3 3 B、時間方向圧縮部 2 3 4 B、結合部 2 3 5 を含まなくともよく、時間方向圧縮部 2 3 4 B の出力するベクトルを同期後特徴量とする場合、音響特徴同期部 2 3 2 A、統合部 2 3 3 A、時間方向圧縮部 2 3 4 A、結合部 2 3 5 を含まなくともよい。

[0096] このような構成により、音響特徴量とテキスト特徴量との何れか一方を他方に同期させ、音響系列と言語系列の対応関係を捉えながら処理を行っており、第一実施形態と同様の効果を得ることができる。なお、時間方向圧縮部 1 3 4 A、2 3 4 A の出力するベクトル、つまり、音響特徴量にテキスト特徴量を同期させて得られる同期後テキスト特徴量と、音響特徴量とに基づき得られる固定長のベクトルを用いたほうが、推定精度が高い。

[0097] <変形例 4>

第一実施形態と異なる部分を中心に説明する。

[0098] 本実施形態では、学習データに学習用の音響信号が含まれるが、S 1 1 0、S 1 2 0を外部の装置で行い、学習用の音響信号に代えて、対応する音響特徴量及びテキスト特徴量が学習データに含まれる構成としてもよい。なお、音響特徴処理 1 3 1 Aおよびテキスト特徴処理 1 3 1 Bは、必要に応じて、学習装置 1 0 0内で行ってもよいし、外部の装置で行ってもよい。推定装置 2 0 0についても同様であり、S 2 1 0、S 2 2 0を外部の装置で行い、推定対象の音響信号に代えて、対応する音響特徴量及びテキスト特徴量を入力としてもよい。

[0099] <第二実施形態>

第一実施形態と異なる部分を中心に説明する。

[0100] <第二実施形態に係る学習装置 1 0 0>

図 1 1 は第二実施形態に係る学習装置 1 0 0の機能ブロック図を、図 1 2 はその処理フローを示す。

[0101] 学習装置 1 0 0は、特徴量算出部 1 2 0、モデル学習部 1 3 0を含む。つまり、第二実施形態に係る学習装置 1 0 0は、音声認識部 1 1 0を含まない。

[0102] <特徴量算出部 1 2 0>

特徴量算出部 1 2 0は、音響信号 $s_{m,L}$ を入力とし、特徴量 $o_{m,L}$ を算出し（S 1 2 0）、出力する。特徴量 $o_{m,L}$ は、発話意図の有無を推定するために用いられる。

[0103] 特徴量 $o_{m,L}$ は「音響特徴量 $a_{m,L}$ 」と「その他の特徴量 $v_{m,L}$ 」とのいずれかあるいは組み合わせを含むベクトルであり、「音響特徴量 $a_{m,L}$ 」と「その他の特徴量 $v_{m,L}$ 」はそれぞれ1つ以上の要素(特徴量)を含むベクトルである。

[0104] 「音響特徴量」は第一実施形態で説明した通りである。

[0105] 「その他の特徴量」は、第一実施形態で説明した音響信号 $s_{m,L}$ から得られる特徴量のみからなる。また、本実施形態では、その他の特徴量は、固定長の

ベクトルとする。

[0106] <モデル学習部 130>

モデル学習部 130 は、学習データ S_L に含まれるラベル $r_{m,L}$ と自信度 $c_{m,L}$ と、特徴量 $0_L = (o_{1,L}, o_{2,L}, \dots, o_{M,L})$ を入力とし、これらの情報を用いて推定モデル Θ_L を学習し (S 130)、学習済みの推定モデル Θ を出力する。

[0107] 第二実施形態において音響特徴量と出力ラベルの関係は、公知の LSTM や self-attention 機構つき LSTM といった技術を用いることで学習可能である。MFCC や FBANK 特徴量のように時系列の情報を保持した特徴量については、公知の LSTM や self-attention 機構つき LSTM といったモデルに入力して固定長のベクトル化し、出力されたベクトルと、時系列の情報を保持しない特徴量を結合したベクトルを DNN などのモデルに入力して、対象に向けた発話かどうかを 0 から 1 で出力するモデルを学習する。

[0108] 上述の処理を実現するモデル学習部 130 の構成について説明する。

[0109] 図 13 はモデル学習部 130 の機能ブロック図を、図 14 はその処理フローの例を示す。

[0110] モデル学習部 130 は、音響特徴処理部 131 A、時間方向圧縮部 134 A、ラベル自信度推定部 136 A、発話意図推定部 136 B およびパラメータ更新部 137 を含む。

[0111] また、モデル学習部 130 は、結合部 133 C, 133 D, 133 E の少なくとも 1 つを含む。さらに、モデル学習部 130 は、必要に応じて付加特徴量処理部 138 を含む。

[0112] <付加特徴量処理部 138>

付加特徴量処理部 138 は、その他の特徴量 $v_{m,L}$ を入力とし、結合部 133 C, 133 D, 133 E において、処理しやすいデータに変換し (S 138)、変換後のその他の特徴量を出力する。なお、変換後のその他の特徴量を単にその他の特徴量ともいう。例えば、時系列モデリングを行う DNN を用いて、その他の特徴量を変換する。ただし、結合部 133 C, 133 D, 133 E が変換前のその他の特徴量 $v_{m,L}$ をそのまま用いる場合には、付加特徴量処理

部 1 3 8 を設けなくともよい。

[0113] <結合部 1 3 3 C>

結合部 1 3 3 C は、音響特徴量 $a_{m,L}$ とその他の特徴量とを入力とし、音響特徴量 $a_{m,L}$ のフレーム数分、その他の特徴量を複製し、音響特徴量 $a_{m,L}$ の各フレームにその他の特徴量を結合し (S 1 3 3 C)、出力する。なお、その他の特徴量を結合した音響特徴量 $a_{m,L}$ を、単に音響特徴量 $a_{m,L}$ ともいう。

[0114] <音響特徴処理部 1 3 1 A>

音響特徴処理部 1 3 1 A は、音響特徴量 $a_{m,L}$ を入力とし、結合部 1 3 3 D または時間方向圧縮部 1 3 4 A において、処理しやすいデータに変換し (S 1 3 1 A)、変換後の音響特徴量を出力する。なお、変換後の音響特徴量を単に音響特徴量ともいう。例えば、時系列モデリングを行う DNN を用いて、音響特徴量を変換する。ただし、結合部 1 3 3 D または時間方向圧縮部 1 3 4 A が変換前の音響特徴量 $a_{m,L}$ をそのまま用いる場合には、音響特徴処理部 1 3 1 A を設けなくともよい。

[0115] <結合部 1 3 3 D>

結合部 1 3 3 D は、音響特徴量 $a_{m,L}$ とその他の特徴量とを入力とし、音響特徴量 $a_{m,L}$ のフレーム数分、その他の特徴量を複製し、音響特徴量 $a_{m,L}$ の各フレームにその他の特徴量を結合し (S 1 3 3 D)、出力する。なお、その他の特徴量を結合した音響特徴量 $a_{m,L}$ を、単に音響特徴量 $a_{m,L}$ ともいう。

[0116] <時間方向圧縮部 1 3 4 A>

時間方向圧縮部 1 3 4 A は、音響特徴量処理部 1 3 1 A または結合部 1 3 3 D の出力した特徴量を、言い換えると、時間方向に長さを持つ音響由来の特徴量を、時間方向に圧縮して (S 1 3 4 A)、一次元の固定長のベクトルにし、出力する。

[0117] <結合部 1 3 3 E>

結合部 1 3 5 は、時間方向圧縮部 1 3 4 A が出力した一次元の固定長のベクトルと、その他の特徴量 $v_{m,L}$ とを結合し (S 1 3 3 E)、結合後のベクトルを出力する。

[0118] ラベル自信度推定部 1 3 6 A および発話意図推定部 1 3 6 B における処理は第一実施形態と同様である。

[0119] <パラメータ更新部 1 3 7>

パラメータ更新部 1 3 7 は、学習データ S_L に含まれるラベル $r_{m,L}$ と自信度 $c_{m,L}$ と、ラベル自信度の推定値と、発話意図ラベルの推定値とを入力とし、これらの値に基づいて、推定モデルのパラメータを更新する (S 1 3 7)。例えば、パラメータ更新部 1 3 7 は、ラベル $r_{m,L}$ と発話意図ラベルの推定値とが一致し、自信度 $c_{m,L}$ とラベル自信度の推定値とが一致するように、音響特徴処理部 1 3 1 A、時間方向圧縮部 1 3 4 A、ラベル自信度推定部 1 3 6 A、発話意図推定部 1 3 6 B において用いるパラメータを更新する (S 1 3 7)。

[0120] 収束条件を満たさない場合には (S 1 3 7 - 2 のnoの場合)、パラメータ更新部 1 3 7 は更新後のパラメータを各部に出力し、更新されたパラメータを用いて、上述の処理 S 1 3 8 ~ S 1 3 6 B を繰り返す。

[0121] 収束条件を満たす場合には (S 1 3 7 - 2 のyesの場合)、パラメータ更新部 1 3 7 は更新後のパラメータを学習済みのパラメータとして出力する。

[0122] <第二実施形態に係る推定装置 2 0 0>

図 1 5 は第二実施形態に係る推定装置 2 0 0 の機能ブロック図を、図 1 6 はその処理フローを示す。

[0123] 学習装置 1 0 0 は、特徴量算出部 2 2 0、推定部 2 3 0 を含む。

[0124] 各部について説明する。

[0125] <特徴量算出部 2 2 0>

特徴量算出部 2 2 0 は、音響信号 s_T を入力とし、特徴量 o_T を算出し (S 2 2 0)、出力する。例えば、特徴量算出部 2 2 0 は第二実施形態に係る特徴量算出部 1 2 0 と同様の特徴量算出処理を行う。

[0126] <推定部 2 3 0>

推定部 2 3 0 は、推定処理に先立ち学習済みの推定モデル Θ を受け取る。

[0127] 推定部 2 3 0 は、特徴量 o_T を入力とし、学習済みのモデルを用いて、発話意図の有無を推定し (S 2 3 0)、推定結果 R を出力する。学習済みのモデルに

対して入力として特徴量 o_T を与え、出力として発話意図の有無の推定結果 R を得る。推定結果 R は、例えば、発話意図の有無を示す2値のラベルである。

[0128] 図17は推定部230の機能ブロック図を、図18はその処理フローの例を示す。

[0129] 推定部230は、音響特徴処理部231A、時間方向圧縮部234Aおよび発話意図推定部236を含む。

[0130] また、モデル学習部130の結合部133C、133D、133E、付加特徴量処理部138に対応する結合部233C、233D、233E、付加特徴量処理部238を含む。

[0131] 音響特徴処理部231A、時間方向圧縮部234A、発話意図推定部236、結合部233C、233D、233E、付加特徴量処理部238は、それぞれ第二実施形態に係る音響特徴処理部131A、時間方向圧縮部134A、発話意図推定部136B、結合部133C、133D、133E、付加特徴量処理部138と同様の処理S238～S236を行う。ただし、特徴量 $o_{m,L}$ に基づく値に代えて特徴量 o_T に基づく値に対して各処理を行う。

[0132] <効果>

このような構成により、従来考慮されていなかった音源の放射方向や直間比などに着目した新規の特徴量を導入することで、発話意図の有無をより明示的に捉えることができる。例えば音声認識のポストフィルタとして利用する場合など、音声認識の仮説出力を待たずに発話者の意図を捉えたい場合に利用すると良い。

[0133] <その他の変形例>

本発明は上記の実施形態及び変形例に限定されるものではない。例えば、上述の各種の処理は、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。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。

[0134] <プログラム及び記録媒体>

上述の各種の処理は、図19に示すコンピュータの記憶部2020に、上記方法の各ステップを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。

[0135] 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。

[0136] 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。さらに、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。

[0137] 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、自己の記録媒体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、さらに、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。

[0138] 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。

[0139] 上記実施形態でCPUがソフトウェア（プログラム）を読み込んで実行したプログラムを、CPU以外の各種のプロセッサが実行してもよい。この場合のプロセッサとしては、GPU（Graphics Processing Unit）、FPGA（Field-Programmable Gate Array）等の製造後に回路構成を変更可能なPLD（Programmable Logic Device）、及びASIC（Application Specific Integrated Circuit）等の特定の処理を実行させるために専用に設計された回路構成を有するプロセッサである専用電気回路等が例示される。また、プログラムを、これらの各種のプロセッサのうちの1つで実行してもよいし、同種又は異種の2つ以上のプロセッサの組み合わせ（例えば、複数のFPGA、及びCPUとFPGAとの組み合わせ等）で実行してもよい。また、これらの各種のプロセッサのハードウェア的な構造は、より具体的には、半導体素子等の回路素子を組み合わせた電気回路である。

[0140] 以上の実施形態に関し、更に以下の付記を開示する。

[0141] （付記項1）

学習用の音響信号と、前記音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データに基づき、推定モデルを学習する学習装置であって、

メモリと、

前記メモリに接続された少なくとも1つのプロセッサと、

を含み、

前記プロセッサは、

学習用の音響信号から得られる音響特徴量と、前記音響信号に対応するテ

キスト特徴量とを同期させることによって同期後特徴量を得、

前記同期後特徴量を用いて、前記音響信号が所定の対象に対して発話されたものか否かを推定し、

学習データに含まれるラベルと、前記発話意図推定部の推定結果とに基づいて前記推定モデルのパラメータを更新する。

[0142] (付記項 2)

学習用の音響信号と、前記学習用の音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データを用いて予め学習された推定モデルに基づき推定を行う推定装置であって、

メモリと、

前記メモリに接続された少なくとも 1 つのプロセッサと、

を含み、

前記プロセッサは、

推定対象の音響信号から得られる音響特徴量と、前記推定対象の音響信号に対応するテキスト特徴量とを同期させることによって同期後特徴量を得、

前記同期後特徴量を用いて、前記推定対象の音響信号が所定の対象に対して発話されたものか否かを推定する、

推定装置。

請求の範囲

- [請求項1] 学習用の音響信号と、前記音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データに基づき、推定モデルを学習する学習装置であって、
- 学習用の音響信号から得られる音響特徴量と、前記音響信号に対応するテキスト特徴量とを同期させることによって同期後特徴量を得る特徴同期部と、
- 前記同期後特徴量を用いて、前記音響信号が所定の対象に対して発話されたものか否かを推定する発話意図推定部と、
- 学習データに含まれるラベルと、前記発話意図推定部の推定結果とに基づいて前記推定モデルのパラメータを更新するパラメータ更新部とを含む、
- 学習装置。
- [請求項2] 請求項1の学習装置であって、
- 前記同期後特徴量は、
- 前記音響特徴量に、前記テキスト特徴量を同期させて得られる同期後テキスト特徴量と、前記音響特徴量とに基づき得られる固定長のベクトルと、
- 前記テキスト特徴量に、前記音響特徴量を同期させて得られる同期後音響特徴量と、前記テキスト特徴量とに基づき得られる固定長のベクトルとの、少なくとも一方を含む、
- 学習装置。
- [請求項3] 請求項1または請求項2の学習装置であって、
- 学習データには、学習用の音響信号と、前記学習用の音響信号が所定の対象に対して発話されたものか否かを表すラベルと、ラベルを付与する際の自信度とが含まれるものとし、
- 前記同期後特徴量を用いて、ラベルを付与する際の自信度を推定するラベル自信度推定部を含み、

前記パラメータ更新部は、前記ラベルと前記発話意図推定部の推定結果、および、学習データに含まれる自信度と前記ラベル自信度推定部の推定結果に基づいて、前記推定モデルのパラメータを更新する、学習装置。

[請求項4]

請求項1から請求項3の何れかの学習装置であって、
その他の特徴量は、

(i)音源の位置あるいは方向、音源までの距離に関する情報、

(ii)音響信号帯域幅または周波数特性に関する情報、

(iii)音声認識結果の信頼度または音声認識の計算時間に関する情報、

(iv)音声認識結果から計算される発話のコマンドとしての妥当性に関する情報、

(v)音声認識結果から得られる入力発話の解釈難度に関する情報、
の少なくとも何れかを含むものとし、

学習データに含まれるラベルと、前記音響特徴量と前記テキスト特徴量とその他の特徴量とを用いて、前記推定モデルを学習する、
学習装置。

[請求項5]

学習用の音響信号と、前記学習用の音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データを用いて予め学習された推定モデルに基づき推定を行う推定装置であって、

推定対象の音響信号から得られる音響特徴量と、前記推定対象の音響信号に対応するテキスト特徴量とを同期させることによって同期後特徴量を得る特徴同期部と、

前記同期後特徴量を用いて、前記推定対象の音響信号が所定の対象に対して発話されたものか否かを推定する発話意図推定部とを含む、
推定装置。

[請求項6]

学習用の音響信号と、前記音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データに基づき、推定モ

デルを学習する学習方法であって、

学習用の音響信号から得られる音響特徴量と、前記音響信号に対応するテキスト特徴量とを同期させることによって同期後特徴量を得る特徴同期ステップと、

前記同期後特徴量を用いて、前記音響信号が所定の対象に対して発話されたものか否かを推定する発話意図推定ステップと、

学習データに含まれるラベルと、前記発話意図推定ステップの推定結果とに基づいて前記推定モデルのパラメータを更新するパラメータ更新ステップとを含む、

学習方法。

[請求項7]

学習用の音響信号と、前記学習用の音響信号が所定の対象に対して発話されたものか否かを表すラベルとが含まれる学習データを用いて予め学習された推定モデルに基づき推定を行う推定方法であって、

推定対象の音響信号から得られる音響特徴量と、前記推定対象の音響信号に対応するテキスト特徴量とを同期させることによって同期後特徴量を得る特徴同期ステップと、

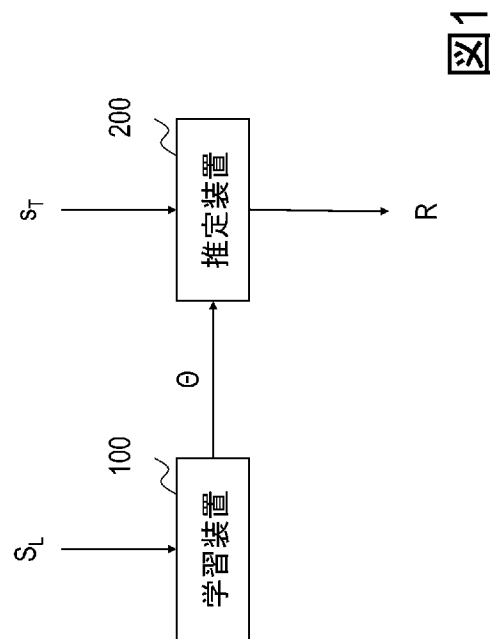
前記同期後特徴量を用いて、前記推定対象の音響信号が所定の対象に対して発話されたものか否かを推定する発話意図推定ステップとを含む、

推定方法。

[請求項8]

請求項1から請求項4の何れかの学習装置、または、請求項5の推定装置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]

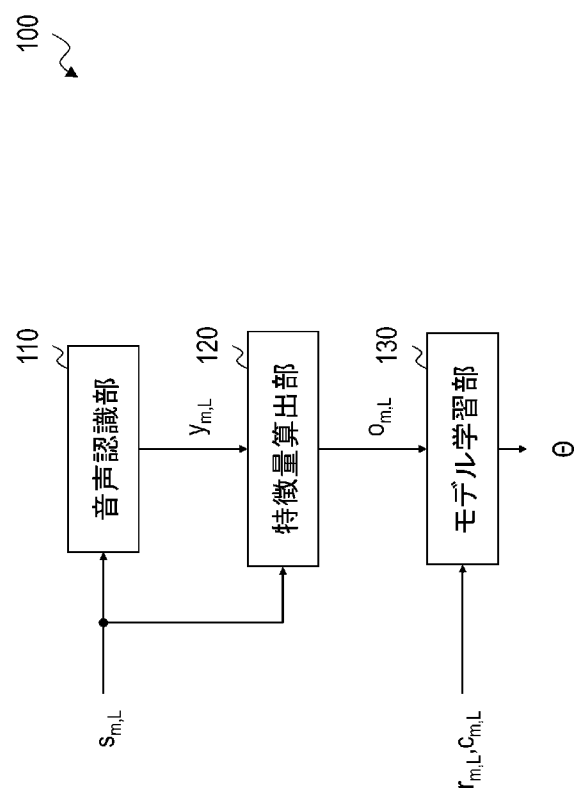
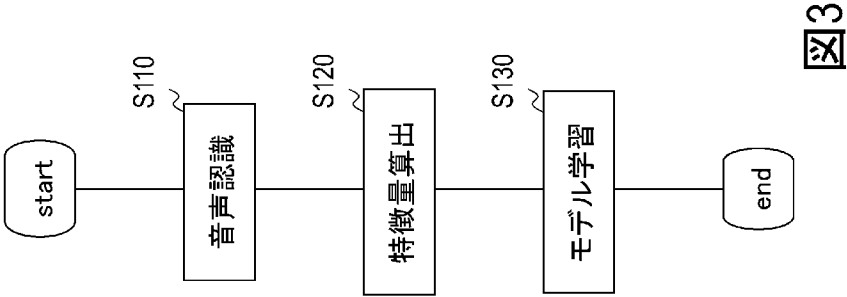


図2

[図3]



[図3]

[図4]

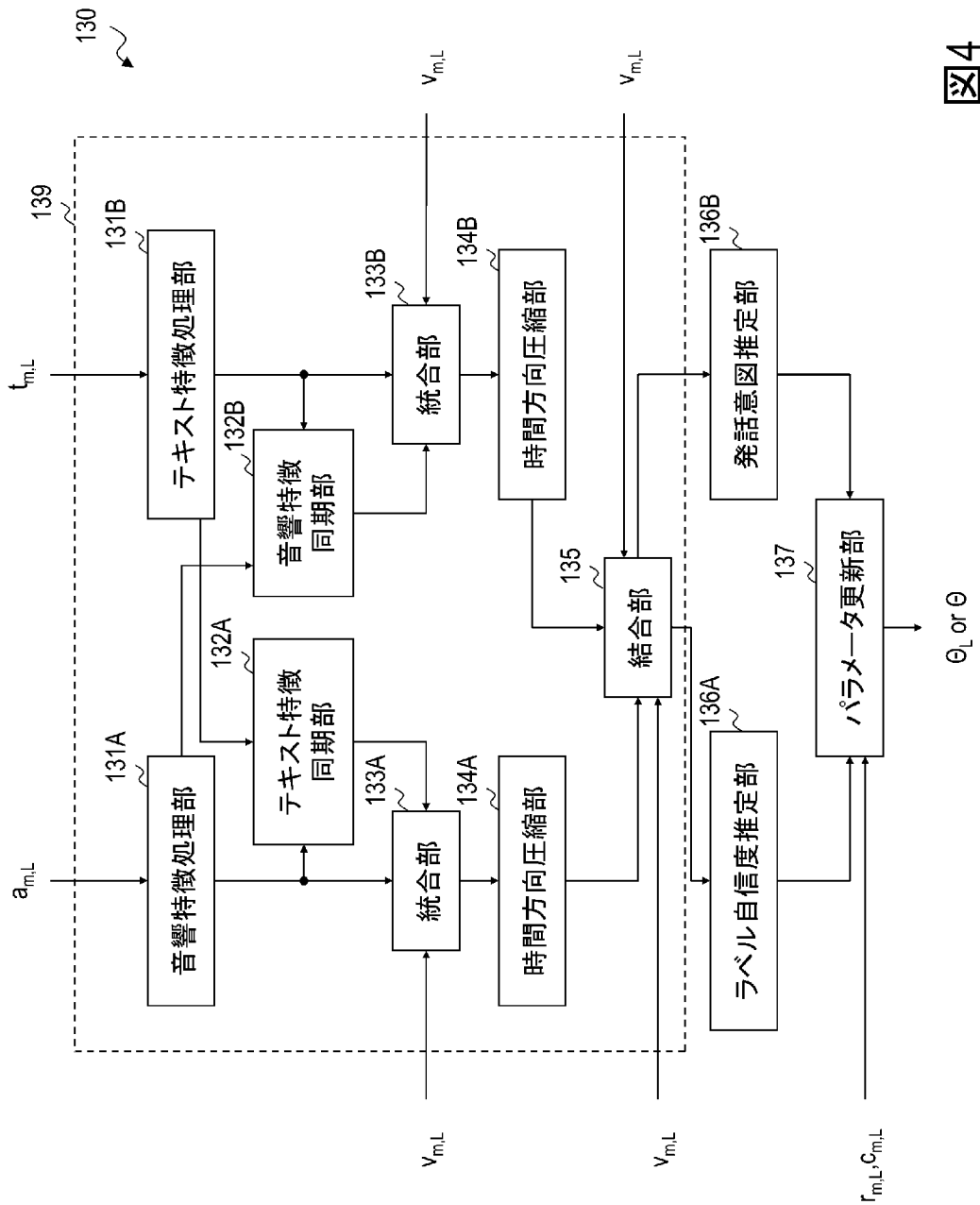


図4

[図5]

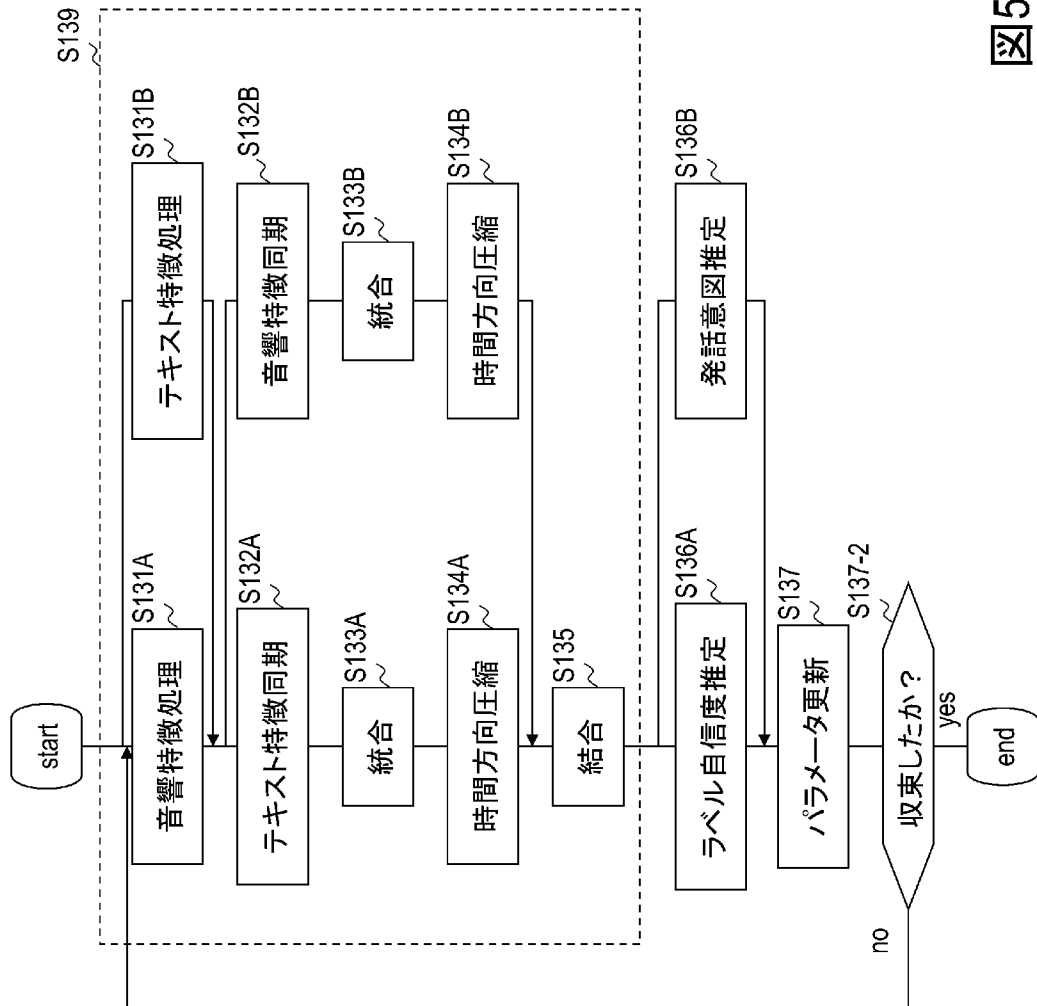
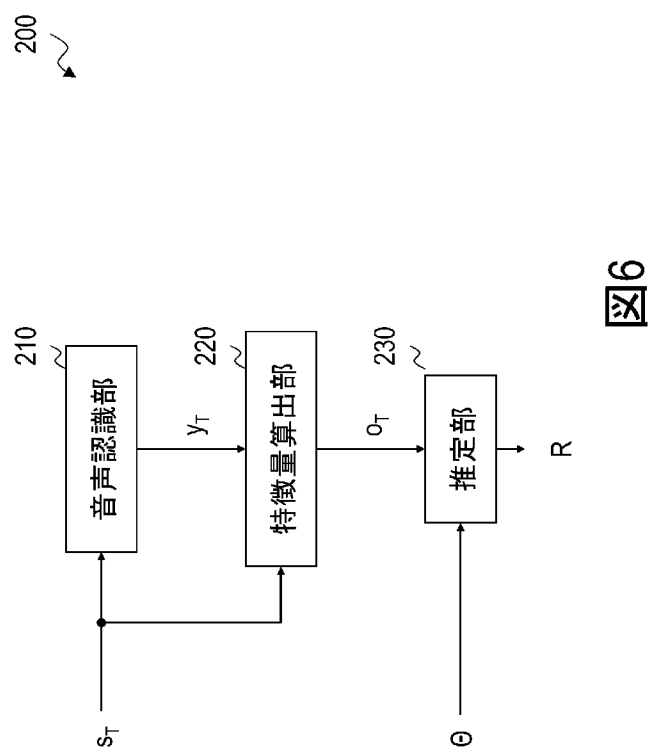
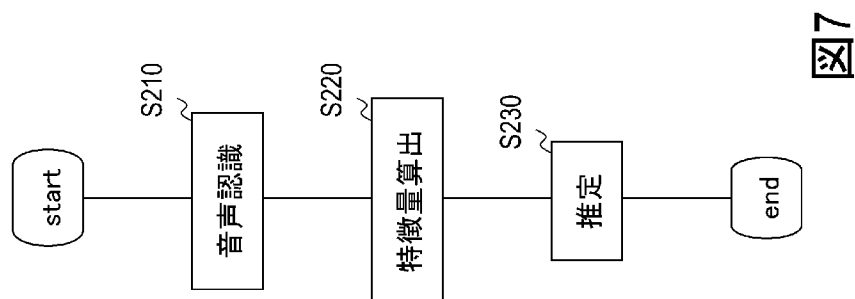


図5

[図6]

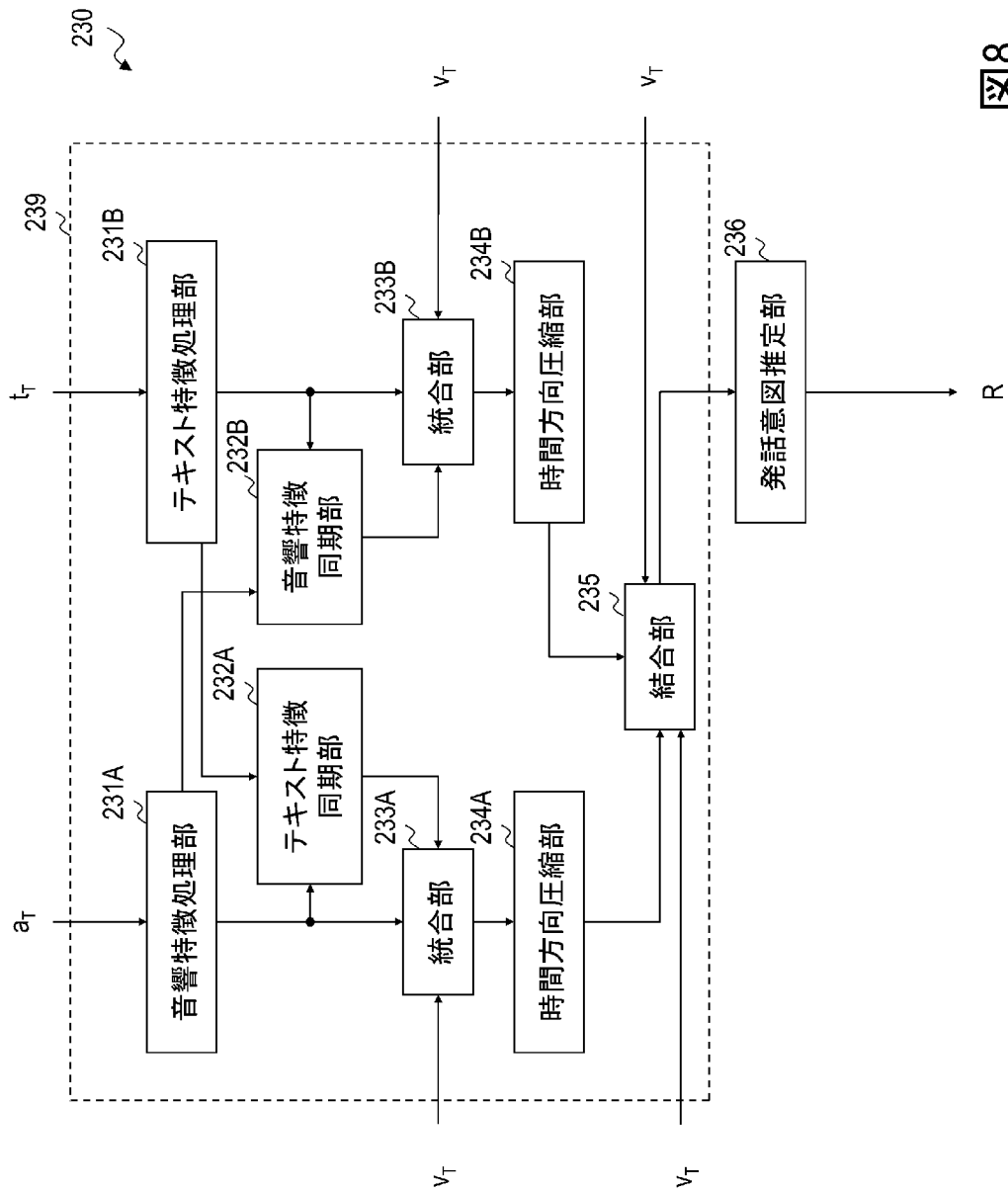


[図7]



[図7]

[図8]



[図8]

[図9]

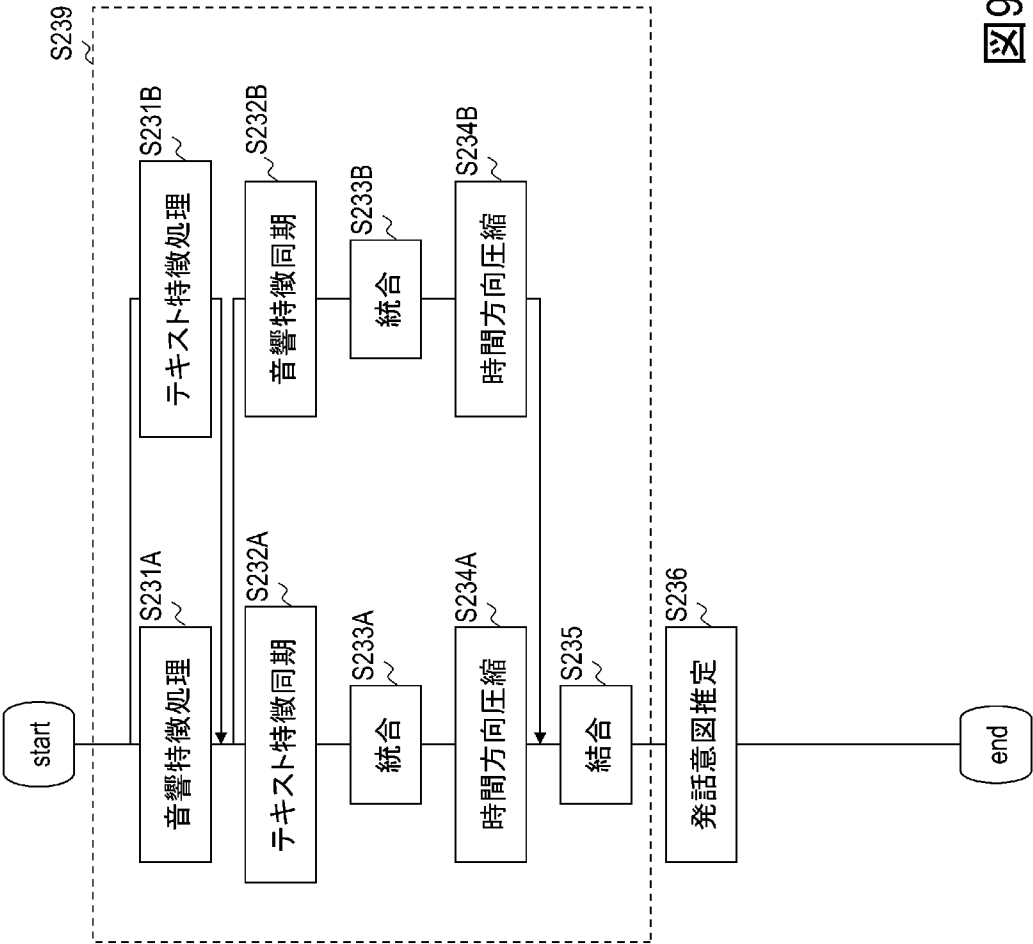


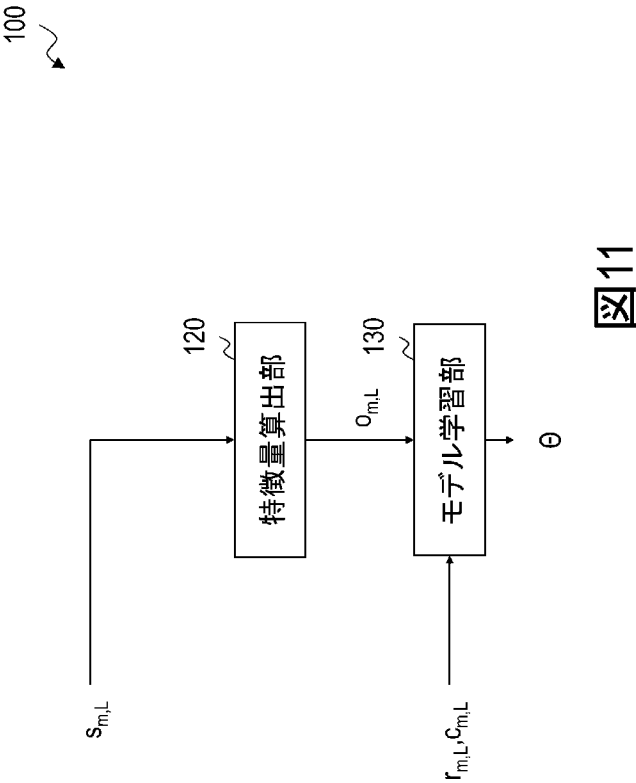
図9

[図10]

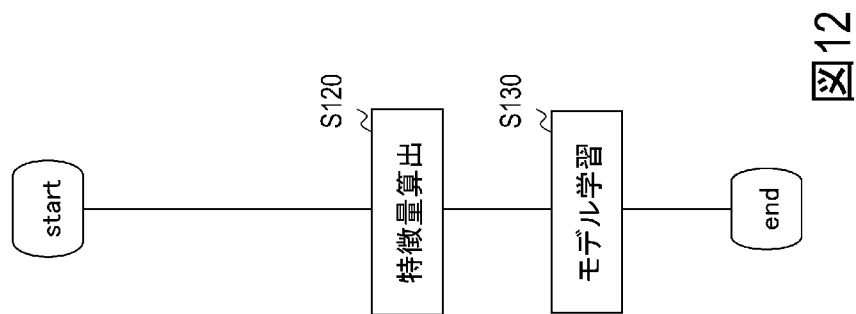
手法	EER [%]
音響特徴量系列を利用	10.8
言語特徴量系列を利用	23.8
言語特徴量系列と言語特徴量系列を併用(非特許文献1)	9.7
第一実施形態の推定システム(自信度なし(変形例1+2))	9.3
第一実施形態の推定システム(自信度あり(変形例2))	9.3

図10

[図11]



[図12]



[図13]

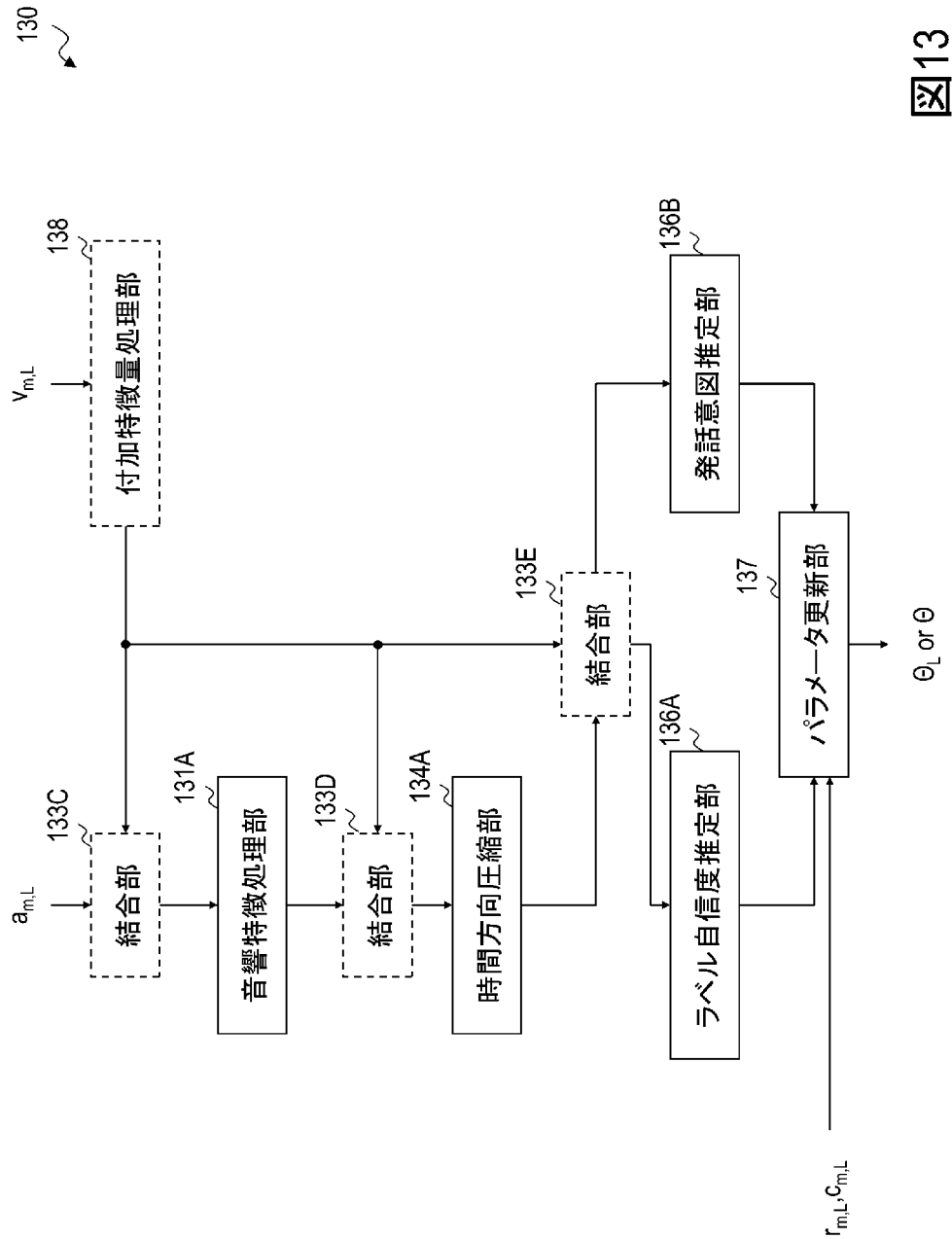


図13

[図14]

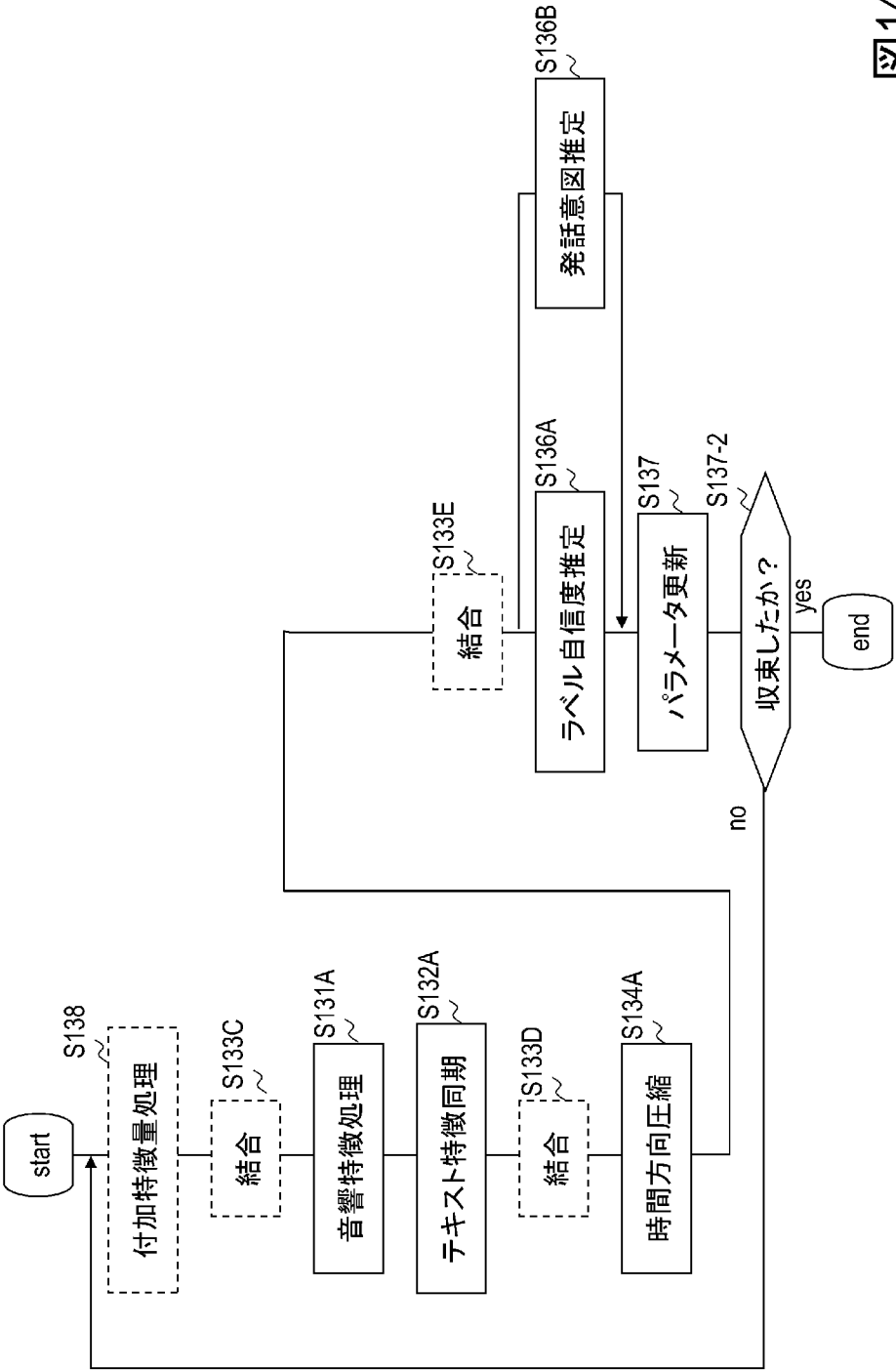


図14

[図15]

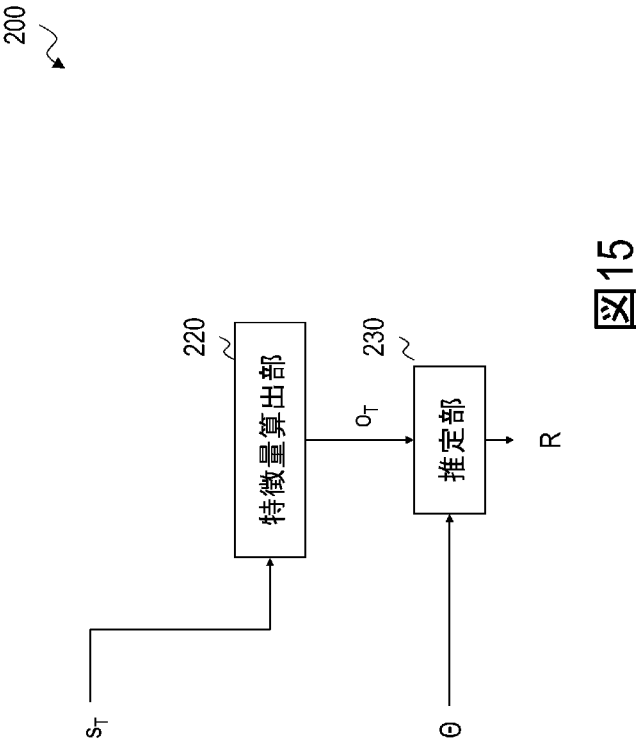
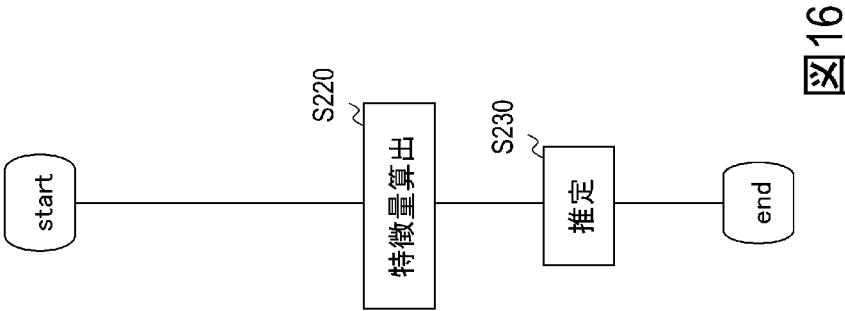


図15

[図16]



[図17]

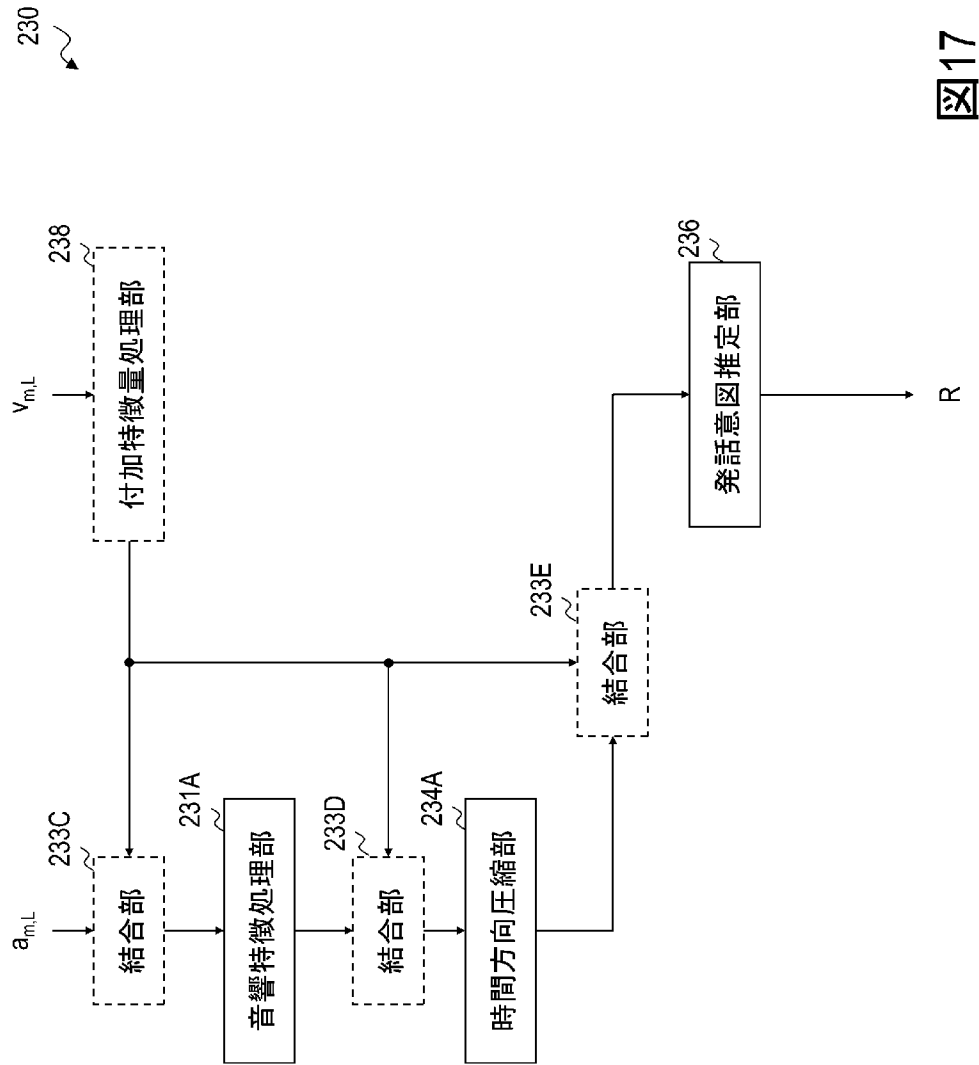


図17

[図18]

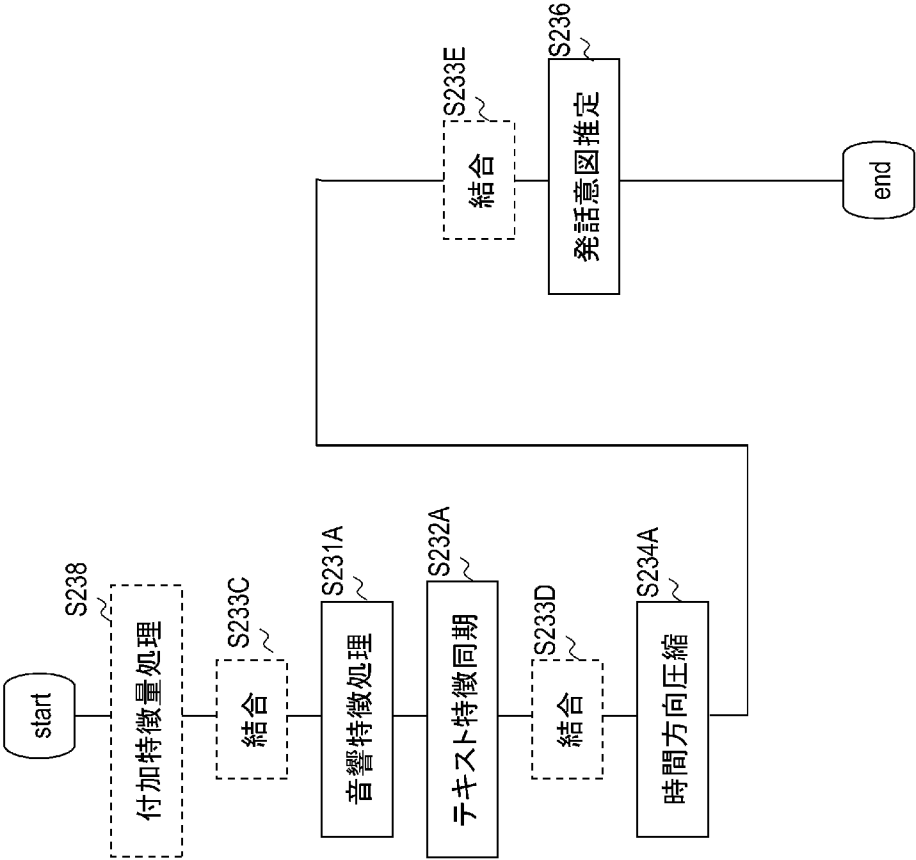


図18

[図19]

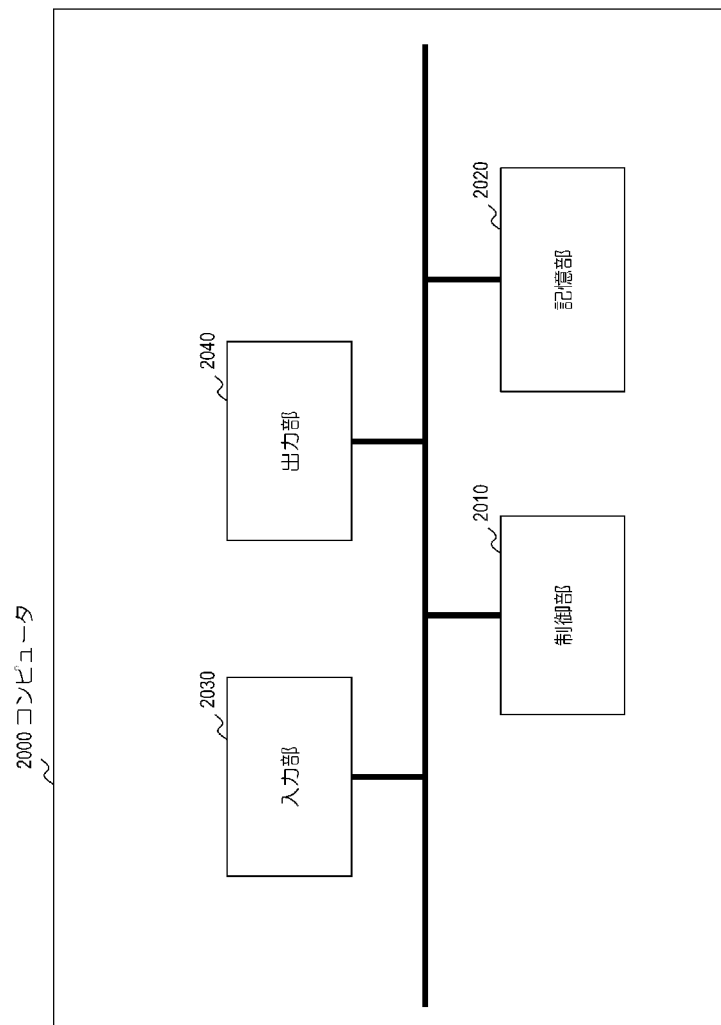


図19

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/006176

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/06(2013.01)i; G10L 15/10(2006.01)i; G10L 15/16(2006.01)i
FI: G10L15/10 500Z; G10L15/06 300C; G10L15/16

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)

G10L15/00-15/34; G06N3/00

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	1971-2021
Registered utility model specifications of Japan	1996-2021
Published registered utility model applications of Japan	1994-2021

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.
A	JP 2018-528458 A (MITSUBISHI ELECTRIC RESEARCH LABORATORIES, INC) 27 September 2018 (2018-09-27) paragraphs [0008]-[0035]	1-8
A	JP 2019-79034 A (HITACHI LTD) 23 May 2019 (2019-05-23) entire text	1-8
A	JP 2017-535823 A (XBRAIN, INC) 30 November 2017 (2017-11-30) entire text	1-8
A	JP 2020-67500 A (NIPPON TELEGR & TELEPH CORP <NTT>) 30 April 2020 (2020-04-30) entire text	1-8
A	JP 2018-146898 A (PANASONIC IP MAN CORP) 20 September 2018 (2018-09-20) entire text	1-8
A	CN 112182191 A (NORTHWEST NORMAL UNIVERSITY) 05 January 2021 (2021-01-05) entire text	1-8



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
20 May 2021 (20.05.2021)

Date of mailing of the international search report
01 June 2021 (01.06.2021)

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/JP2021/006176

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	高津弘明他，システム発話の文脈を考慮した発話意図理解，言語処理学会第25回年次大会発表論文集[online], March 2019, pp. 320-323, (TAKATSU, Hiroaki et al.), non-official translation (Utterance intention understanding in consideration of the context of system utterance. Proceedings of the 25th annual meeting of the Association for Natural Language Processing [online])	1-8

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2021/006176

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2018-528458 A	27 Sep. 2018	US 2017/0161256 A1 paragraphs [0014]- [0043] WO 2017/094911 A1 CN 108292305 A	
JP 2019-79034 A	23 May 2019	US 2019/0130904 A1 entire text EP 3477638 A2 KR 10-2019-0046623 A CN 110019745 A	
JP 2017-535823 A	30 Nov. 2017	US 2016/0098992 A1 entire text WO 2016/054230 A1 EP 3201913 A1 KR 10-2017-0070094 A CN 107004410 A	
JP 2020-67500 A	30 Apr. 2020	WO 2020/085070 A1 entire text	
JP 2018-146898 A	20 Sep. 2018	US 2018/0257236 A1 entire text EP 3373301 A1 CN 108573695 A	
CN 112182191 A	05 Jan. 2021	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G10L 15/06(2013.01)i; G10L 15/10(2006.01)i; G10L 15/16(2006.01)i FI: G10L15/10 500Z; G10L15/06 300C; G10L15/16														
B. 調査を行った分野														
調査を行った最小限資料（国際特許分類（IPC）） G10L15/00-15/34; G06N3/00														
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 <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2021年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2021年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2021年</td> </tr> </table>			日本国実用新案公報	1922-1996年	日本国公開実用新案公報	1971-2021年	日本国実用新案登録公報	1996-2021年	日本国登録実用新案公報	1994-2021年				
日本国実用新案公報	1922-1996年													
日本国公開実用新案公報	1971-2021年													
日本国実用新案登録公報	1996-2021年													
日本国登録実用新案公報	1994-2021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												
A	JP 2018-528458 A（三菱電機株式会社） 27.09.2018（2018-09-27） 段落[0008]-[0035]	1-8												
A	JP 2019-79034 A（株式会社日立製作所） 23.05.2019（2019-05-23） 全文	1-8												
A	JP 2017-535823 A（エクスプレイン・インコーポレーテッド） 30.11.2017（2017-11-30） 全文	1-8												
A	JP 2020-67500 A（日本電信電話株式会社） 30.04.2020（2020-04-30） 全文	1-8												
A	JP 2018-146898 A（パナソニックIPマネジメント株式会社） 20.09.2018（2018-09-20） 全文	1-8												
A	CN 112182191 A（NORTHWEST NORMAL UNIVERSITY） 05.01.2021（2021-01-05） 全文	1-8												
☒ 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。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>“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</td> </tr> <tr> <td>“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</td> <td>“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</td> </tr> <tr> <td>“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</td> <td>“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</td> </tr> <tr> <td>“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）</td> <td>“&” 同一パテントファミリー文献</td> </tr> <tr> <td>“O” 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>“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</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	“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	“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	“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	“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	“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	“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）	“&” 同一パテントファミリー文献	“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	
* 引用文献のカテゴリー	“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													
“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	“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													
“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	“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													
“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）	“&” 同一パテントファミリー文献													
“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														
国際調査を完了した日 20.05.2021		国際調査報告の発送日 01.06.2021												
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） 山下 剛史 5Z 8946 電話番号 03-3581-1101 内線 3591												

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
A	高津弘明他，システム発話の文脈を考慮した発話意図理解，言語処理学会第25回年次大会発表論文集 [online]，2019.03 pp. 320-323	1-8

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2021/006176

引用文献			公表日	パテントファミリー文献	公表日
JP	2018-528458	A	27.09.2018	US 2017/0161256 A1 段落[0014]-[0043] WO 2017/094911 A1 CN 108292305 A	
JP	2019-79034	A	23.05.2019	US 2019/0130904 A1 全文 EP 3477638 A2 KR 10-2019-0046623 A CN 110019745 A	
JP	2017-535823	A	30.11.2017	US 2016/0098992 A1 全文 WO 2016/054230 A1 EP 3201913 A1 KR 10-2017-0070094 A CN 107004410 A	
JP	2020-67500	A	30.04.2020	WO 2020/085070 A1 全文	
JP	2018-146898	A	20.09.2018	US 2018/0257236 A1 全文 EP 3373301 A1 CN 108573695 A	
CN	112182191	A	05.01.2021	(ファミリーなし)	