

(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2月11日(11.02.2021)



(10) 国際公開番号

WO 2021/024491 A1

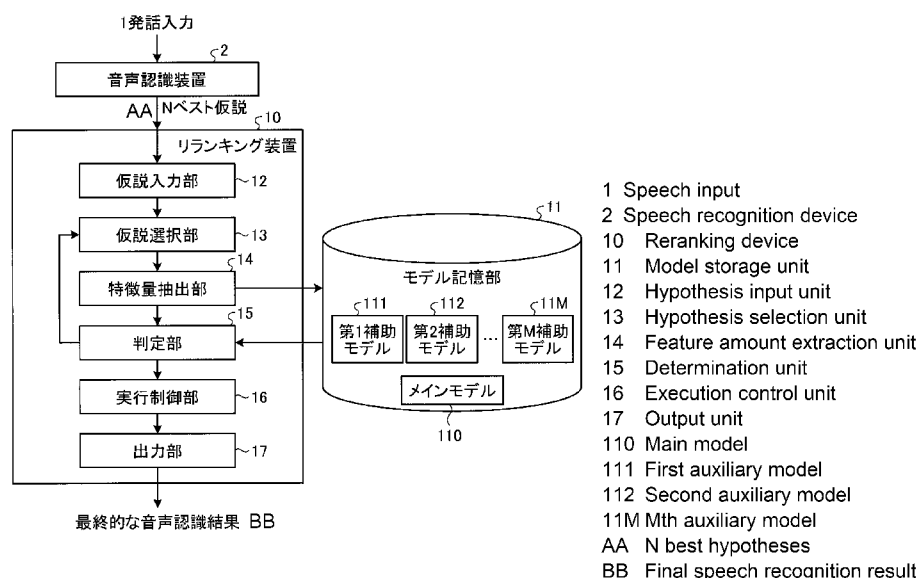
- (51) 国際特許分類:
G10L 15/08 (2006.01) *G10L 15/16* (2006.01)
- (21) 国際出願番号: PCT/JP2019/031517
- (22) 国際出願日: 2019年8月8日(08.08.2019)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 小川 厚徳 (OGAWA, Atsunori); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財産センタ内 Tokyo (JP). デル

クロア マーク (DELCROIX, Marc); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財産センタ内 Tokyo (JP). 苅田 成樹 (KARITA, Shigeki); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財産センタ内 Tokyo (JP). 中谷 智広 (NAKATANI, Tomohiro); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財産センタ内 Tokyo (JP).

(74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).

(54) Title: DETERMINATION DEVICE, LEARNING DEVICE, DETERMINATION METHOD, AND DETERMINATION PROGRAM

(54) 発明の名称: 判定装置、学習装置、判定方法及び判定プログラム



(57) Abstract: A reranking device (10) has: a hypothesis input unit (12) that receives an input of N best hypotheses with which the scores of speech recognition accuracy are associated; a hypothesis selection unit (13) that selects two hypotheses to be determined from among the N best hypotheses the input of which has been received; and a determination unit (15) that determines the levels of accuracy of the two hypotheses by using a plurality of first to Mth auxiliary models (111-11M) represented by neural networks capable of converting the two hypotheses into hidden vectors when the selected

(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, 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))

two hypotheses are provided, and determining the levels of accuracy of the two hypotheses on the basis of the hidden vectors of the two hypotheses, and by using a main model (110) represented by a neural network capable of determining the levels of accuracy of the two hypotheses on the basis of the hidden vectors of the two hypotheses respectively converted by the auxiliary models.

(57) 要約: リランキング装置(10)は、音声認識精度のスコアが対応付けられたNベスト仮説の入力を受け付ける仮説入力部(12)と、入力を受け付けたNベスト仮説のうち、判定対象である二つの仮説を選択する仮説選択部(13)と、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の第1補助モデル(111)~第M補助モデル(11M)と、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデル(110)とを用いて、二つの仮説の精度の高低を判定する判定部(15)と、を有する。

明 細 書

発明の名称： 判定装置、学習装置、判定方法及び判定プログラム
技術分野

[0001] 本発明は、判定装置、学習装置、判定方法及び判定プログラムに関する。

背景技術

[0002] 音声認識は、人間が発した音声（発話）を計算機により単語列（テキスト）に変換する技術である。通常、音声認識システムは、入力された一つの発話に対して、音声認識スコアの最も高い仮説（音声認識結果）である一つの単語列（1 ベスト仮説）を出力する。ただし、音声認識装置による音声認識の精度は、100%ではない。このため、一つの入力発話に対して、1 ベスト仮説のみを出力するのではなく、 N (≥ 2) 個の仮説を出力して、 N ベストリスコアリング装置を用いて、その N 個仮説の中から音声認識精度が最も高いと推定される仮説を最終的な音声認識結果として出力する、 N ベストリスコアリングと呼ばれる手法がある。なお、 N ベストリスコアリングと N ベストリランキングとは同義として扱われている。

[0003] N ベストリスコアリング方法では、音声認識結果である仮説の中からスコアの高い所定数（ N 個）の仮説を出力する。そして、 N ベストリスコアリング方法では、この中から尤もらしい仮説を音声認識結果として出力する。ここで、スコアが最大となる仮説が必ずしもベストな仮説とは限らない。このため、二つの仮説のうち尤もらしい仮説（正解に近い仮説）を選択する二択問題をトーナメント方式で繰り返し適用することで、尤もらしい仮説を選択するリランキング装置が提案されている（例えば、非特許文献1 参照）。

先行技術文献

非特許文献

[0004] 非特許文献1: Atsunori Ogawa, Marc Delcroix, Shigeki Karita, Tomohiro Nakatani, "RESCORING N-BEST SPEECH RECOGNITION LIST BASED ON ONE-ON-ONE HYPOTHESIS COMPARISON USING ENCODER-CLASSIFIER

MODEL”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6099-6103, 2018.

発明の概要

発明が解決しようとする課題

[0005] 非特許文献1に記載のリランキング方法では、N仮説をスコアの降順に並べ、先頭の仮説（スコアが最も高い仮説）から順に二つの仮説を選択し、学習済みの二択問題を解くニューラルネットワーク（NN）にこれらの仮説を入力することによって、いずれかの仮説を選択する処理を繰り返し行い、最終的に選択された仮説を音声認識結果として出力することが記載されている。非特許文献1に記載のリランキング方法では、一定の精度で音声認識結果を出力するが、さらに、近年では、音声認識結果の出力に対して、精度の安定化が要求されている。

[0006] 本発明は、上記に鑑みてなされたものであって、ある音声信号に対する解の候補として挙げられた複数の仮説に対し、最も精度が高い仮説を安定した精度で判定することができる判定装置、学習装置、判定方法及び判定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

[0007] 上述した課題を解決し、目的を達成するために、本発明に係る判定装置は、音声認識精度のスコアが対応付けられたNベスト仮説の入力を受け付ける入力部と、入力を受け付けたNベスト仮説のうち、判定対象である二つの仮説を選択する選択部と、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の補助モデルと、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデルとを用いて、二つの仮説の精度の高低を判定する判定部と、を有することを特徴とする。

[0008] また、本発明に係る学習装置は、音声認識精度が既知である学習用の二つ

の仮説の入力を受け付ける入力部と、二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の補助モデルと、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデルとに対し、各ニューラルネットワークが二つの仮説の精度の高低を判定するタスクを個別に行うとみなしたマルチタスク学習を行わせる学習部と、を有することを特徴とする。

[0009] また、本発明に係る判定方法は、判定装置が実行する判定方法であって、音声認識精度のスコアが対応付けられたNベスト仮説の入力を受け付ける工程と、入力を受け付けたNベスト仮説のうち、判定対象である二つの仮説を選択する工程と、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の補助モデルと、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデルとを用いて、二つの仮説の精度の高低を判定する工程と、を含んだことを特徴とする。

[0010] また、本発明に係る判定プログラムは、音声認識精度のスコアが対応付けられたNベスト仮説の入力を受け付けるステップと、入力を受け付けたNベスト仮説のうち、判定対象である二つの仮説を選択するステップと、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の補助モデルと、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデルとを用いて、二つの仮説の精度の高低を判定するステップと、をコンピュータに実行させる。

発明の効果

[0011] 本発明によれば、ある音声信号に対する解の候補として挙げられた複数の仮説に対し、最も精度が高い仮説を、安定した精度で判定することができる。

図面の簡単な説明

[0012] [図1]図1は、実施の形態1に係るリランキング装置の機能構成の一例を示す図である。

[図2]図2は、第1補助モデル～第M補助モデル及びメインモデルの構成を説明する図である。

[図3]図3は、第1補助モデルの構築例を示す図である。

[図4]図4は、実施の形態1に係るリランキング処理の処理手順を示すフローチャートである。

[図5]図5は、図1に示すリランキング装置が、Nベスト仮説に対して実行するリランキング処理を説明する図である。

[図6]図6は、実施の形態2に係る学習装置の機能構成の一例を示す図である。

[図7]図7は、図6に示す入替部の処理を説明する図である。

[図8]図8は、実施の形態2に係る学習処理の処理手順を示すフローチャートである。

[図9]図9は、実施の形態3に係るリランキング装置の要部構成を示す図である。

[図10]図10は、実施の形態3に係るリランキング処理の処理手順を示すフローチャートである。

[図11]図11は、プログラムが実行されることにより、リランキング装置及び学習装置が実現されるコンピュータの一例を示す図である。

発明を実施するための形態

[0013] 以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施の形態により本発明が限定されるものではない。また、図面の記載に

において、同一部分には同一の符号を付して示している。

[0014] 本実施の形態では、音声認識結果である N ($N \geq 2$) ベスト仮説のうち、最終的な音声認識結果である最も音声認識精度が高い仮説（単語列）を得るためのモデルを用いたリランキング装置、及び、 N ベストのリランキング処理に用いるモデルを実現する学習装置について説明する。なお、本実施の形態については、 N ベストリスコアリングではなく、 N ベストリランキングと表現を統一して説明する。

[0015] まず、本実施の形態に係るリランキング装置が N ベスト仮説のリランキングを行う上で、本実施の形態におけるモデルが有すべき必要最低限な機能について述べる。本実施の形態では、 N ベスト仮説から最も音声認識精度が高い仮説（オラクル仮説）を、最終的な音声認識結果として見つけ出すことが目的である。

[0016] すなわち、本実施の形態では、 N ベスト仮説の中からオラクル仮説をリランキングにより見つけ出すためにモデルに必要最低限な機能は、 N ベスト仮説中の二つの仮説に着目したときに、どちらの仮説の方がより高い音声認識精度を有しているかを判定できることである点に着目した。言い換えると、本実施の形態におけるモデルに必要最低限な機能は、 N ベスト仮説中の二つの仮説を対象に、一対一の仮説比較を行うことができることである。

[0017] そこで、本実施の形態に係るリランキング装置は、一対一の二つの仮説の比較を行う機能を持つモデルを用いることによって、二つの仮説のうち音声認識精度がより高い仮説を判定する機能を持たせた。さらに、本実施の形態では、モデルとして、ニューラルネットワーク（ NN ）で表されるメインモデルと、 NN で表される複数の補助モデルとを用いる。各補助モデルは、二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定するモデルである。メインモデルは、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定するモデルである。

[0018] そして、本実施の形態に係るリランキング装置は、Nベスト仮説のスコアの昇順に二つの仮説を選択し、選択した二つの仮説のうち、音声認識精度がより高い仮説を次の判定対象の一方の仮説として残し、未判定の仮説から昇順に他方の仮説を選択して、複数の補助モデル及びメインモデルを用いた比較を行う。本実施の形態に係るリランキング装置は、前回の判定で音声認識精度がより高いと判定された仮説を判定対象の一方の仮説として選択し、未判定の仮説のうち最も順位の低い仮説を他方の仮説として選択し、複数の補助モデル及びメインモデルによる二つの仮説に対する比較処理を繰り返す。これによって、本実施の形態では、安定した精度で、Nベスト仮説の中からオラクル仮説を見つけ出すことを可能にした。

[0019] [実施の形態1]

[リランキング装置]

まず、実施の形態1に係るリランキング装置について説明する。このリランキング装置は、音声認識結果であるNベスト仮説のうちの二つの仮説に対して音声認識精度の高低の判定を繰り返し実行して、最も音声認識精度の高い仮説を最終的な音声認識結果として出力する。

[0020] 図1は、実施の形態1に係るリランキング装置の機能構成の一例を示す図である。実施の形態1に係るリランキング装置10は、例えば、ROM (Read Only Memory)、RAM (Random Access Memory)、CPU (Central Processing Unit) 等を含むコンピュータ等に所定のプログラムが読み込まれて、CPUが所定のプログラムを実行することで実現される。

[0021] リランキング装置10は、音声認識装置2から出力されたNベスト仮説の入力を受け付ける。そして、リランキング装置10は、このNベスト仮説のうち、二つの仮説に対する音声認識精度の高低についての判定を、全Nベスト仮説について実行し、音声認識精度が高い仮説として残った仮説を、スコアと対応付けて、最終的な音声認識結果として出力する。なお、音声認識装置2は、1発話が入力されると、例えば、音声認識用のモデルを用いて音声認識を行い、音声認識結果としてNベスト仮説を出力する。音声認識用のモ

デルは、学習用の複数の発話と、各発話に対応する書き起こし（正解単語列）を学習データとして用いて学習（モデルパラメータが最適化）されている。

[0022] リランキング装置 10 は、モデル記憶部 11、仮説入力部 12、仮説選択部 13（選択部）、特徴量抽出部 14、判定部 15、実行制御部 16 及び出力部 17 を有する。

[0023] モデル記憶部 11 は、補助モデル及びメインモデル 110 を記憶する。図 1 の例では、モデル記憶部 11 は、補助モデルとして、第 1 補助モデル 111 ~ 第 M 補助モデル 11M を記憶する。第 1 補助モデル 111 ~ 第 M 補助モデル 11M 及びメインモデル 110 は、NN で表されるモデルである。第 1 補助モデル 111 ~ 第 M 補助モデル 11M 及びメインモデル 110 は、音声認識精度が既知である学習用の N ベスト仮説を用いて予め学習される。

[0024] 第 1 補助モデル補助 111 ~ 第 M 補助モデル 11M は、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、NN で表される。第 1 補助モデル 111 ~ 第 M 補助モデル 11M は、学習用の N ベスト仮説のうち二つの仮説が与えられたときに、二つの仮説について、その二つの仮説の音声認識精度の高低を判定できるように学習される。第 1 補助モデル 111 ~ 第 M 補助モデル 11M は、二つの仮説を、それぞれ RNN を用いて隠れ状態ベクトルに変換する。そして、第 1 補助モデル 111 ~ 第 M 補助モデル 11M は、NN を用いて、隠れ状態ベクトルを基に二つの仮説の精度の高低の並びが正しいことを示す事後確率をそれぞれ生成する。

[0025] メインモデル 110 は、第 1 補助モデル 111 ~ 第 M 補助モデル 11M においてそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、NN で表される。メインモデル 110 は、第 1 補助モデル 111 ~ 第 M 補助モデル 11M においてそれぞれ変換された学習用の二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるように学習される。メインモデル 110 は、NN を用いて、

第1補助モデル111～第M補助モデル11Mにおいてそれぞれ変換された学習用の二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低の並びが正しいことを示す事後確率を生成する。

[0026] 第1補助モデル111～第M補助モデル11M及びメインモデル110は、各ニューラルネットワークが二つの仮説の精度の高低を判定するタスクを個別に行うとみなしたマルチタスク学習によって学習が実行される。

[0027] 仮説入力部12は、Nベスト仮説の入力を受け付ける。Nベスト仮説は、音声認識装置2が出力する。或いは、他の装置が、ネットワーク等を介して、Nベスト仮説をリランキング装置10に入力してもよい。

[0028] 仮説選択部13は、入力を受け付けたNベスト仮説のうち、一対一の比較対象である二つの仮説を、Nベスト仮説のスコアの昇順に選択する。仮説選択部13は、最初の判定においては、Nベスト仮説のうち、スコアが最下位である仮説と、最下位の仮説より1つ順位が高い仮説とを判定対象として選択する。仮説選択部13は、以降の判定においては、二つの仮説の一方の仮説として、前回の判定で音声認識精度がより高いと判定された仮説を選択する。そして、仮説選択部13は、二つの仮説の他方の仮説として、未判定の仮説のうち、最もスコアの順位が低い仮説を選択する。このように、仮説選択部13は、全Nベスト仮説について一対一の比較が実行されるように、Nベスト仮説から、昇順に、比較対象の二つの仮説を選択する。

[0029] 特徴量抽出部14は、一対一の比較対象である二つの仮説について、それぞれの特徴量を抽出する。特徴量抽出部14は、一対一の比較対象であるNベスト仮説中の v 位の仮説とNベスト仮説中の u ($u < v \leq N$) 位の仮説（単語列）と、について、それぞれの特徴量を抽出する。特徴量抽出部14は、仮説中の各単語単位で特徴量ベクトルを抽出する。各単語の特徴量ベクトルは、例えば、離散値である単語IDをNNによる単語の埋め込み処理により連続値のベクトルとして表現した単語ベクトルに、音声認識処理により得られる単語単位の音響スコア（対数尤度）や言語スコア（対数確率）などを補助特徴量として、単語ベクトルに連結したものである。

[0030] 判定部 15 は、一対一の比較対象の二つの仮説に対し、第 1 補助モデル 111 ~ 第 M 補助モデル 11M 及びメインモデル 110 を用いて、二つの仮説の精度の高低を判定する。判定部 15 は、一対一の比較対象である v 位の仮説と u 位の仮説とを第 1 補助モデル 111 ~ 第 M 補助モデル 11M にそれぞれ入力し、メインモデル 110 による出力結果を用いて、どちらの仮説が高い音声認識精度を有しているかを判定する。u 位及び v 位で表す仮説の順位は、N ベスト仮説において既に付与されているものである。リランキング装置 10 では、順位の再設定を行わない。

[0031] ここで、第 1 補助モデル 111 ~ 第 M 補助モデル 11M は、u 位の仮説の特徴量及び v 位の仮説の特徴量が入力されると、u 位の仮説が v 位の仮説よりも音声認識精度が高いことを示す事後確率を出力する。メインモデル 110 は、第 1 補助モデル 111 ~ 第 M 補助モデル 11M においてそれぞれ変換された二つの仮説の隠れ状態ベクトルが入力されると、u 位の仮説が v 位の仮説よりも音声認識精度が高いことを示す事後確率を出力する。判定部 15 は、メインモデル 110 による事後確率が 0.5 以上である場合には、u 位の仮説が v 位の仮説よりも音声認識精度が高いと判定する。また、判定部 15 は、メインモデル 110 による事後確率が 0.5 未満である場合には、v 位の仮説が u 位の仮説よりも音声認識精度が高いと判定する。

[0032] なお、リランキング装置 10 では、特徴量抽出部 14 の機能を、第 1 補助モデル 111 ~ 第 M 補助モデル 11M が有してもよい。この場合、判定部 15 は、比較対象である二つの仮説を第 1 補助モデル 111 ~ 第 M 補助モデル 11M にそれぞれ入力する。

[0033] そして、判定部 15 は、比較対象の二つの系列のうち、より精度が高いと判定した仮説を次の判定時における比較対象として残し、他方の仮説を以降の比較対象から外す。仮説選択部 13 は、判定部 15 によって精度が高いと判定された仮説を二つの系列の一方の仮説として選択し、判定部 15 による判定が行われていない仮説のうち最もスコアの順位が低い仮説を他方の仮説として選択する。具体的には、前述したように、仮説選択部 13 は、判定部

15が残した仮説を二つの仮説の一方の仮説として選択し、Nベスト仮説のうち、前回比較対象となった仮説の順位の1つ上の順位の仮説を二つの仮説の他方の仮説として選択する。

[0034] 実行制御部16は、判定部15による判定処理と仮説選択部13による選択処理とを、所定条件に達するまで繰り返す制御を行う。この場合、実行制御部16は、全Nベスト仮説について一対一の比較が実行されるように、仮説選択部13における比較対象の二つの仮説の選択処理、特徴量抽出部14における特徴量抽出処理、及び、判定部15における判定処理を繰り返す制御を行う。具体的に、実行制御部16は、1位の仮説に対して比較処理が行われるまで、仮説の選択処理、特徴量抽出処理及び判定処理を繰り返す制御を行う。

[0035] 出力部17は、仮説の選択処理、特徴量抽出処理、判定処理及び順位の設定処理が繰り返された結果、所定条件に達した場合、Nベスト仮説のうち、比較対象として残っている仮説を、最も音声認識精度が高い仮説、すなわち、最終的な音声認識結果として出力する。出力部17は、最後の判定処理で精度が高いと判定された仮説を最終的な音声認識結果として出力する

[0036] [定義]

まず、リランキング装置10に必要最低限な機能要件を数式で定義する。

$W^{(u)} = w_1^{(u)}, w_2^{(u)}, \dots, w_{L(W^{(u)})}^{(u)}$ を、Nベスト仮説中のu位の仮説（単語列）と定義する。また、 $L(W^{(u)})$ を、 $W^{(u)}$ の長さ（単語数）と定義する。

[0037] また、 $A^{(u)} = a_1^{(u)}, a_2^{(u)}, \dots, a_{L(W^{(u)})}^{(u)}$ を $W^{(u)}$ に対応する補助特徴量ベクトル列と定義する。 $W^{(u)}$ 中のi番目の単語 $w_i^{(u)}$ の補助特徴量ベクトル $a_i^{(u)}$ は、例えば、音声認識装置による音声認識処理の結果として得られる音響スコア（対数尤度）や言語スコア（対数確率）などである（詳細は、例えば、A. Ogawa and T. Hori, “Error detection and accuracy estimation in automatic speech recognition using deep bidirectional recurrent neural networks”, Speech Commun

ication, vol. 89, pp.70–83, May 2017. (以降、参考文献 1 とする。)
) を参照)。

[0038] 補助特徴量ベクトル $a_i^{(u)}$ は、17次元の基本の補助特徴量ベクトルを含む。さらに、補助特徴量ベクトル $a_i^{(u)}$ では、前向き (forward) LSTMLMの単語予測スコアを18次元の補助特徴量として用いてもよい。LSTMLMは、長短期記憶メモリ (long short-term memory: LSTM) ユニットを用いた再帰的ニューラルネットワーク (Recurrent Neural Network: RNN) 言語モデルであり、後述するように、本実施の形態 1～3における各補助モデルを構成するモデルである。そして、補助特徴量ベクトル $a_i^{(u)}$ では、後向き (backward) LSTMLMの単語予測スコアを19次元の補助特徴量として用いてもよい。後向きLSTMLMは、未来の単語列から現在の単語の生起確率を予測するものであるであり、前向きLSTMLMと相補的な単語予測能力を持つことから各補助モデルが出力する判定情報の精度向上が期待できる。

[0039] また、 $X^{(u)} = x_1^{(u)}, x_2^{(u)}, \dots, x_{L(W^{(u)})}^{(u)}$ を $W^{(u)}$ に対応する特徴量ベクトル列と定義する。 $W^{(u)}$ 中の i 番目の単語 $w_i^{(u)}$ の特徴量ベクトル $x_i^{(u)}$ は、 $x_i^{(u)} = \text{concat}(\text{embed}(w_i^{(u)}), a_i^{(u)})$ で得られる。ここで、 $\text{concat}(\cdot)$ は、ベクトルの連結処理を表す。また、 $\text{embed}(\cdot)$ は、NNによる単語の埋め込み処理 (離散値の単語IDを連続値のベクトルで表現する処理) (詳細は、例えば、坪井祐太，海野裕也，鈴木潤，深層学習による自然言語処理，MLP機械学習プロフェッショナルシリーズ，講談社，2017。(以降、参考文献 2 とする。) を参照) を表す。なお、 $\text{embed}(\cdot)$ を行うNNも第1補助モデル111～第M補助モデル11Mの一部であり、そのパラメータは、後述のエンコーダRNN及び2クラス分類FFNNのパラメータと同時に学習 (最適化) される。

[0040] そして、 $P(O | X^{(u)}, X^{(v)})$ を、二つの仮説 $W^{(u)}, W^{(v)}$ の精度の高低の並びが正しいことを示す事後確率と定義する。 $P(O | X^{(u)}, X^{(v)})$ は、第1補助モデル111～第M補助モデル11M及びメインモデル11

0のそれぞれにおいて、生成される。

[0041] [補助モデル及びメインモデルの構成]

第1補助モデル111～第M補助モデル11M、メインモデル及び構成について説明する。図2は、第1補助モデル111～第M補助モデル11M及びメインモデル110の構成を説明する図である。図2では、処理の流れを説明するため、判定部15も記載される。

[0042] 図2に示すように、各補助モデルは、それぞれ、二つの第1変換部、第1結合部及び判定情報生成部を有する。具体的に、第1補助モデル111を例に説明する。第1補助モデル111は、二つの第1変換部111-1u、111-1v、第1結合部111-2及び判定情報生成部111-3を有する。

[0043] 第1変換部111-1uは、比較対象の二つの仮説 $W^{(u)}$ 、 $W^{(v)}$ のうち、仮説 $W^{(u)}$ の特徴量 $X^{(u)}$ の入力を受け付け、隠れ状態ベクトルに変換する。第1変換部111-1vは、比較対象の二つの仮説 $W^{(u)}$ 、 $W^{(v)}$ のうち、仮説 $W^{(v)}$ の特徴量 $X^{(v)}$ の入力を受け付け、隠れ状態ベクトルに変換する。

[0044] 第1結合部111-2は、第1変換部111-1u、第1変換部111-1vが変換した二つの隠れ状態ベクトルを結合する。判定情報生成部111-3は、二つの仮説 $W^{(u)}$ 、 $W^{(v)}$ の精度の高低の並びが正しいことを示す事後確率 $P(0 | X^{(u)}, X^{(v)})$ を、判定情報として生成する。他の補助モデルも、第1補助モデル111と同じ構成であり、与えられた二つの仮説 $W^{(u)}$ 、 $W^{(v)}$ に対し、それぞれ、隠れ状態ベクトルの変換、隠れ状態ベクトルの結合、及び、判定情報の生成を含むタスクをそれぞれ実行できるようにしている。なお、各補助モデルは、学習時におけるランダム初期化時における初期値がそれぞれ異なる。

[0045] メインモデル110は、メイン結合部110-1と、判定情報生成部110-2とを有する。メイン結合部110-1は、第1補助モデル111～第M補助モデル11Mでそれぞれ変換された二つの仮説の隠れ状態ベクトルを結合する。判定情報生成部110-3は、二つの仮説 $W^{(u)}$ 、 $W^{(v)}$ の精度の

高低の並びが正しいことを示す事後確率 $P(O | X^{(u)}, X^{(v)})$ を、判定情報として生成する。

[0046] N ベスト仮説中の u 番目の仮説 $W^{(u)}$ と v 番目の仮説 $W^{(v)}$ ($u < v \leq N$) の特徴量ベクトル列 $X^{(u)}$, $X^{(v)}$ が各補助モデルに与えられたとき、メインモデル 110 は、記号 $y = \{0\}$ の事後確率 $P(O | X^{(u)}, X^{(v)})$ を出力する。

[0047] 判定部 15 は、メインモデル 110 が出力した事後確率 $P(O | X^{(u)}, X^{(v)})$ を受け取り、判定を行う。 $P(O | X^{(u)}, X^{(v)})$ は、 u 位の仮説と v 位の仮説との順位の上下関係が正しさを確率的に表現する事後確率である。判定部 15 は、N ベストリランキングモデルから出力された事後確率 $P(O | X^{(u)}, X^{(v)})$ を取得し、取得した事後確率を所定の閾値と比較して、 u 位の仮説及び v 位の仮説のいずれがより音声認識精度が高いかを判定する。

[0048] 具体的には、判定部 15 は、事後確率 $P(O | X^{(u)}, X^{(v)})$ が 0.5 以上である場合には、 u 位の仮説が v 位の仮説よりも音声認識精度が高いと判定し、 $y = 0$ を出力する。また、判定部 15 は、事後確率 $P(O | X^{(u)}, X^{(v)})$ が 0.5 未満である場合には、 v 位の仮説が u 位の仮説よりも音声認識精度が高いと判定し、 $y = 1$ を出力する。

[0049] すなわち、判定部 15 は、以下の (1-1) 式及び (1-2) 式に示すように、 u 位の仮説及び v 位の仮説のいずれがより音声認識精度が高いかを判定する。

[0050] $P(O | X^{(u)}, X^{(v)}) \geq 0.5$
 if WER (Word error rate) of $W^{(u)} \leq$ WER of $W^{(v)} \dots (1-1)$
 $P(O | X^{(u)}, X^{(v)}) < 0.5$
 otherwise $\dots (1-2)$

[0051] ここで、与えられた仮説 (単語列) の音声認識精度を返す関数 $\sum_y P(y | X^{(u)}, X^{(v)}) = 1$ であるため、(1-1) 式の 1 段目に示す不等式が満足される場合、判定部 15 は、仮説 $W^{(u)}$ は仮説 $W^{(v)}$ 以上の音声認識精度を持つと判定する。また、(1-2) 式の不等式が満足される場合、判定部 15

は、 $W^{(u)}$ は $W^{(v)}$ よりも低い音声認識精度を持つと判定する。

[0052] したがって、(1-1) 式の 1 段目に示す不等式が満足される場合、 $W^{(u)}$ 及び $W^{(v)}$ のランキングの上下関係 ($u < v$) が正しいと推定される。このため、判定部 15 は、 $W^{(u)}$ を、 $W^{(v)}$ との一对一の仮説比較において $W^{(v)}$ よりも音声認識精度が高い仮説として残し、次の一对一の仮説比較では、 $W^{(v)}$ として使用する。なお、判定部 15 は、 $W^{(v)}$ を、 $W^{(u)}$ よりも音声認識精度が低い仮説として扱い、最も音声認識精度が高い仮説の候補、すなわち、最終的な音声認識結果の候補から除外する。

[0053] そして、(1-2) 式の 1 段目不等式が満足される場合は、 $W^{(u)}$ 及び $W^{(v)}$ のランキングの上下関係は、誤りであると推定される。すなわち、 $W^{(u)}$ 及び $W^{(v)}$ のランキングの上下関係は逆であると推定される。このため、判定部 15 は、 $W^{(v)}$ を、 $W^{(u)}$ との一对一の仮説比較において $W^{(u)}$ よりも音声認識精度が高い仮説として残し、次の一对一の仮説比較では、 $W^{(v)}$ として引き続き使用する。なお、判定部 15 は、元の $W^{(u)}$ を、元の $W^{(v)}$ よりも音声認識精度が低い仮説として扱い、最も音声認識精度が高い仮説の候補、すなわち、最終的な音声認識結果の候補から除外する。

[0054] [補助モデルの構築例]

第 1 補助モデル 111 ~ 第 M 補助モデル 11M の構築例について説明する。第 1 補助モデル 111 ~ 第 M 補助モデル 11M は、同じ構成であるため、図 3 を参照し、第 1 補助モデル 111 の構築例を説明する。図 3 は、第 1 補助モデル 111 の構築例を示す図である。なお、図 3 では、簡単のため、単語の埋め込み処理 $embed(\cdot)$ を行う NN は省略されている。以下、その詳細について説明する。

[0055] 比較対象の仮説 $W^{(u)}$ の長さ (単語数) $L(W^{(u)})$ と仮説 $W^{(v)}$ ($u < v \leq N$) の長さ $L(W^{(v)})$ とが異なる可能性がある。この長さの違いを吸収するため、第 1 補助モデル 111 は、二つの仮説の特徴量を、RNN を用いて隠れ状態ベクトルに変換する。具体的には、第 1 補助モデル 111 は、この処理を行うために、エンコーダデコーダモデル (詳細は、例えば、参考文

献2参照)のエンコーダRNN111-1aを第1変換部111-1u, 111-1vとして有する。

[0056] 第1補助モデル111は、エンコーダRNN111-1aを用いて $W^{(u)}$ と $W^{(v)}$ を固定長の隠れ状態ベクトルで表現する。そして、第1補助モデル111～第M補助モデル11Mは、これらの隠れ状態ベクトルを用いることによって、 $W^{(u)}$ と $W^{(v)}$ とを公平に比較することが可能になる。

[0057] エンコーダRNN111-1aの処理について説明する。エンコーダRNN111-1aは、RNNの一種である長短期記憶メモリ(long short-term memory: LSTM)ユニット(詳細は、例えば、参考文献2参照)を有する。LSTMユニットは、 $W^{(u)}$ のi番目の単語 $w_i^{(u)}$ の特徴量ベクトル $x_i^{(u)}$ と、i-1番目の隠れ状態ベクトル $h_{\{i-1\}}^{(u)}$ が与えられたとき、i番目の隠れ状態ベクトル $h_i^{(u)}$ を以下の(2)式のように与える。

[0058] $h_i^{(u)} = \text{lstm}(x_i^{(u)}, h_{\{i-1\}}^{(u)}) \cdots (2)$

[0059] ここで、 $\text{lstm}(\cdot)$ は、1層単方向(unidirectional)のLSTMユニットの処理を示す。また、 $h_i^{(u)} = 0$ (ゼロベクトル)である。 $h_i^{(u)}$ は、単語列 $w_1^{(u)}, w_2^{(u)}, \cdots, w_i^{(u)}$ の特徴量ベクトル列 $x_1^{(u)}, x_2^{(u)}, \cdots, x_i^{(u)}$ をエンコード(符号化)したものである。エンコーダRNN111-1aは、この処理を、特徴量ベクトル列 $X^{(u)}$ 中の各特徴量ベクトル $x_i^{(u)}$ に対して繰り返すことで、 $X^{(u)}$ をエンコードした隠れ状態ベクトル $h_{L(W^{(u)})}^{(u)}$ を得ることができる。

[0060] エンコーダRNN111-1aは、同様の処理を特徴量ベクトル列 $X^{(v)}$ に対しても行い、 $X^{(v)}$ をエンコードした隠れ状態ベクトル $h_{L(W^{(v)})}^{(v)}$ を得る。なお、 $X^{(u)}$ に対して処理を行うLSTMユニットと、 $X^{(v)}$ に対して処理を行うLSTMユニットは同じもの、すなわち、パラメータが共有されていてもよいし、別のLSTMユニットであってもよい。また、図3では、 $x_{L(W^{(u)})}^{(u)}, x_{L(W^{(v)})}^{(v)}, h_{L(W^{(u)})}^{(u)}, h_{L(W^{(v)})}^{(v)}$ の下付き部分 $L(W^{(u)})$ は、 $L(W^{(u)})$ と示している。

[0061] 第1補助モデル111は、以上で得た二つの隠れ状態ベクトル $h_{L(W^{(u)})}^{(u)}$

$u)$, $h_{L(W(v))}^{(v)}$ を、第1結合部111-2で連結した隠れ状態ベクトル $h_{\{(u,v)\}}$ をエンコーダRNN111-aの出力として以下の(3)式のよう
に得る。

$$[0062] \quad h_{\{(u,v)\}} = \text{concat} (h_{L(W(u))}^{(u)}, h_{L(W(v))}^{(v)}) \cdots (3)$$

[0063] そして、第1補助モデル111は、エンコーダRNN111-1aの後段に、クラス分類($y=0$ or 1)を行うためのNNを連結する。例えば、第1補助モデル111は、1クラス分類のためのNNとして、1層のフィードフォワード型NN(FFNN)111-3a(詳細は、例えば、参考文献2を参照)を、判定情報生成部113として用いる。エンコーダRNN111-1aの出力として得た隠れ状態ベクトル $h_{\{(u,v)\}}$ が、1層の1クラス分類FFNN111-3aに入力され、最終的に、1クラスの $y = \{0\}$ の事後確率 $P(y | X^{(u)}, X^{(v)})$ を以下の(4), (5)式のように得ることができる。

$$[0064] \quad z_{\{(u,v)\}} = \text{linear} (h_{\{(u,v)\}}) \cdots (4)$$

$$P(y | X^{(u)}, X^{(v)}) = \text{sigmoid} (z_{\{(u,v)\}})_y \cdots (5)$$

[0065] ここで、 $\text{linear}(\cdot)$ は、線形変換処理(詳細は、例えば、参考文献2を参照)を表す。 $\text{sigmoid}(\cdot)$ は、シグモイド処理を表す。

[0066] また、メインモデル110では、メイン結合部110-1は、第1結合部111-2と同様のベクトル連結処理を行う。また、メインモデル110では、判定情報生成部110-3は、判定情報生成部111-3の1層の1クラス分類FFNN111-3aと同様の構成の1クラス分類FFNNによって構成される。

[0067] [補助モデル及びメインモデルの他の構築例1]

なお、第1補助モデル111及びメインモデル110は、1クラス分類FFNNにおけるシグモイド処理に代えて、ソフトマックス処理を行ってもよい。この場合、エンコーダRNNの出力として得た隠れ状態ベクトル $h_{\{(u,v)\}}$ が、1層の2クラス分類FFNNに入力され、最終的に、2クラスの記号

$y = \{0, 1\}$ の事後確率 $P(y | X^{(u)}, X^{(v)})$ を以下 (6), (7) 式のように得ることができる。なお、 $y = 0$ は、 $W^{(u)}$ 及び仮説 $W^{(v)}$ の順位の上下関係が正しいことを示す。また、 $y = 1$ は、 $W^{(u)}$ 及び仮説 $W^{(v)}$ の順位の上下関係が誤りであることを示す。 $P(0 | X^{(u)}, X^{(v)})$ は、 u 位の仮説と v 位の仮説との順位の上下関係が正しさを確率的に表現する第 1 の事後確率である。 $P(1 | X^{(u)}, X^{(v)})$ は、 u 位の仮説と v 位の仮説との順位の上下関係が誤りであることを確率的に表現する第 2 の事後確率である。

[0068] $z^{\{(u, v)\}} = \text{linear}(h^{\{(u, v)\}}) \quad \dots (6)$

$P(y | X^{(u)}, X^{(v)}) = \text{softmax}(z^{\{(u, v)\}})_y \quad \dots (7)$

[0069] ここで、 $\text{softmax}(\cdot)$ は、ソフトマックス処理を表す。また、 $\text{softmax}(\cdot)_y$ は、ソフトマックス処理の結果として得られる事後確率ベクトルの y 番目の要素（確率値）を表す。

[0070] この場合、判定部 15 は、メインモデル 110 から出力された第 1 の事後確率 $P(0 | X^{(u)}, X^{(v)})$ 及び第 2 の事後確率 $P(1 | X^{(u)}, X^{(v)})$ を取得し、取得した二つの事後確率の大小を比較して、 u 位の仮説及び v 位の仮説のいずれがより音声認識精度が高いかを判定する。判定部 15 は、第 1 の事後確率 $P(0 | X^{(u)}, X^{(v)})$ が第 2 の事後確率 $P(1 | X^{(u)}, X^{(v)})$ よりも高い場合には、 u 位の仮説が v 位の仮説よりも音声認識精度が高いと判定する。また、判定部 15 は、第 1 の事後確率 $P(0 | X^{(u)}, X^{(v)})$ が第 2 の事後確率 $P(1 | X^{(u)}, X^{(v)})$ よりも低い場合には、 v 位の仮説が u 位の仮説よりも音声認識精度が高いと判定する。

[0071] [補助モデルの他の構築例 2]

なお、図 3 に示すエンコーダ RNN 111-1a の LSTM ユニットは、1 層単方向の LSTM ユニットとしたが、複数層または双方向 (bidirectional) の LSTM ユニットであってもよい。

[0072] [補助モデルの他の構築例 3]

また、LSTM ユニットの代わりに、単純な (sigmoid 関数等を活

活性化関数として持つ。) RNNや、Gated Recurrent Unit (GRU) を用いてもよい。

[0073] [補助モデル及びメインモデルの他の構築例4]

さらに、補助モデル及びメインモデル110は、図3の構築例では、1クラス分類NNとして、1層のフィードフォワード型NNを用いたが、複数層のフィードフォワード型NNを用いてもよい。Nベストリランキングモデルは、複数層のフィードフォワード型NNを用いる場合、活性化関数として、sigmoid関数、tanh関数、Rectified Linear Unit (ReLU) 関数、Parametric ReLU (PReLU) 関数などを用いることができる。なお、補助モデル及びメインモデル110の他の構築例1～4の用語の詳細については、例えば、参考文献2を参照いただきたい。

[0074] [補助モデルの他の構築例5]

また、補助モデルは、従来のNベストリスコアリングモデル（例えばRNN言語モデル）により計算されたスコアを、特徴量ベクトルにおける新たな次元として追加して利用することも可能である。

[0075] [リランキング処理の処理手順]

次に、図1に示すリランキング装置10が実行するリランキング処理の処理手順について説明する。図4は、実施の形態1に係るリランキング処理の処理手順を示すフローチャートである。

[0076] まず、仮説入力部12が、リランキング対象のNベスト仮説の入力を受け付けると（ステップS1）、仮説選択部13は、入力を受け付けたNベスト仮説のうち、スコアの昇順に、一対一の比較対象であるu位及びv位の二つの仮説を選択する ($u < v \leq N$)。まず、仮説選択部13は、 $u = N - 1$ 、 $v = N$ に設定する（ステップS2）。そして、仮説選択部13は、入力を受け付けたNベスト仮説から、u位及びv位の二つの仮説 $W^{(u)}$ 、 $W^{(v)}$ をNベスト仮説から選択する（ステップS3）。続いて、特徴量抽出部14は、仮説 $W^{(u)}$ 、 $W^{(v)}$ の特徴量を抽出する（ステップS4）。判定部15は、仮説

$W^{(u)}$, $W^{(v)}$ の特徴量 ($X^{(u)}$, $X^{(v)}$) を各補助モデル (第 1 補助モデル 111 ~ 第 M 補助モデル 11M) に入力する (ステップ S5)。

[0077] 判定部 15 は、N ベストリランキングモデルからの出力結果を取得する (ステップ S6)。具体的には、判定部 15 は、事後確率 $P(0 | X^{(u)}, X^{(v)})$ を取得する。

[0078] そして、(1-1) 式及び (1-2) 式において説明したように、判定部 15 は、 $P(0 | X^{(u)}, X^{(v)}) \geq 0.5$ であるか否かを判定する (ステップ S7)。 $P(0 | X^{(u)}, X^{(v)}) \geq 0.5$ である場合 (ステップ S7: Yes)、判定部 15 は、 u 位の仮説が v 位の仮説よりも音声認識精度が高いと判定し、実行制御部 16 は、 k について $k = u$ と設定する (ステップ S8)。 k は、比較処理後の仮説のうち、最も音声認識精度が高い仮説の N ベスト仮説における順位 (ランキング) である。一方、 $P(0 | X^{(u)}, X^{(v)}) \geq 0.5$ でない場合 (ステップ S7: No)、判定部 15 は、 v 位の仮説が u 位の仮説よりも音声認識精度が高いと判定し、実行制御部 16 は、 $k = v$ と設定する (ステップ S9)。

[0079] 続いて、実行制御部 16 は、 $u = 1$ であるか否かを判定する (ステップ S10)。 $u = 1$ でない場合 (ステップ S10: No)、必要な一対一の仮説比較処理がまだ全ては終了していないため、実行制御部 16 は、仮説選択部 13 に対し、比較対象の次の仮説の選択を行わせる。具体的には、仮説選択部 13 は、 $u = u - 1$ 、 $v = k$ に設定し (ステップ S11)、ステップ S3 に戻り、次の判定対象の N ベスト仮説 $W^{(u)}$, $W^{(v)}$ を選択する。そして、リランキング装置 10 は、この N ベスト仮説 $W^{(u)}$, $W^{(v)}$ に対して、ステップ S4 ~ ステップ S10 の処理を実行する。

[0080] また、 $u = 1$ である場合 (ステップ S10: Yes)、必要な一対一の比較処理が全て終了したため、実行制御部 16 は、 k 位の $W^{(k)}$ を最も音声認識精度が高いと推定される仮説、すなわち、最終的な音声認識結果として出力し (ステップ S12)、処理を終了する。このように、リランキング装置 10 では、任意の二つの仮説を 1 組とし、複数の組についてそれぞれ音声認識

精度の高低の判定を繰り返すことで、最も音声認識精度が高いと推定される仮説を、最終的な音声認識結果として出力することができる。

[0081] このように、実施の形態１に係るリランキング装置１０は、一対一の二つの仮説の比較を行う機能を持つモデルを用いることによって、二つの仮説のうち音声認識精度がより高い仮説を判定する機能を持たせた。さらに、リランキング装置１０では、モデルとして、ニューラルネットワーク（ＮＮ）で表されるメインモデル１１０と、ＮＮで表される複数の補助モデルとを用いる。

[0082] すなわち、リランキング装置１０では、複数の補助モデルを設け、入力された二つの仮説に対して、各補助モデルにタスクを実行させている。各補助モデルの構造は同じであっても、学習時においてパラメータのランダム初期化を行うので、同じ入力仮説に対しても異なる隠れ状態ベクトルを出力する。これにより、ある二つの入力仮説に対して、ある補助モデルが出力する隠れ状態ベクトルが適切なものでなかったとしても、別の補助モデルが適切な隠れ状態ベクトルを出力できる可能性が高まる。つまり、正確な仮説の判定結果を生成するのに適した隠れ状態ベクトルが、いずれかの補助ネットワークから出力される可能性が高くなる。この結果、リランキング装置１０のメインモデル１１０には、適切な二つの仮説に対応する隠れ状態ベクトルが安定して入力されるため、メインモデル１１０の出力値の精度も安定する。このように、実施の形態１に係るリランキングモデルは、安定した精度で、Ｎベスト仮説の中からオラクル仮説を見つけ出すことができる。

[0083] また、リランキング装置１０は、Ｎベスト仮説のスコアの昇順に二つの仮説を選択する。言い換えると、リランキング装置１０は、Ｎベスト仮説のうち、スコアが最も低い仮説から順に仮説ペアを選択する。図５は、図１に示すリランキング装置１０が、Ｎベスト仮説に対して実行するリランキング処理を説明する図である。

[0084] 一般には、スコアが高い仮説の方が、尤もらしい仮説である可能性が高い。スコアの高い順に仮説を選択していくと、最もスコアの高い仮説は、Ｎ－

1回の判定処理に勝ち抜かなければ、最終的な出力仮説として選択されず、尤もらしい仮説として選ばれにくくなってしまう。

[0085] そこで、図5に示すように、リランキング装置10は、最終的に出力仮説として選ばれる可能性の高い仮説について、少ない判定回数で済むように、Nベスト仮説のうち、スコアが最も低い仮説から順に仮説ペアを選択する。言い換えると、リランキング装置10は、図5に示すように、スコアの最も高い仮説については、シード権を与え、Nベスト仮説全体に対する比較処理の後の方の処理で比較処理が行われるようにし、尤もらしい仮説として選ばれやすくしている。このように、リランキング装置10は、最終的に出力仮説として選ばれる可能性の高い仮説が、尤もらしい仮説として選ばれやすいため、安定した精度で、Nベスト仮説の中からオラクル仮説を見つけ出すことができる。

[0086] [実施の形態2]

[学習装置]

次に、実施の形態2として、リランキング装置10が用いるNベストリランキングモデルを学習する学習装置について説明する。図6は、実施の形態2に係る学習装置の機能構成の一例を示す図である。実施の形態2に係る学習装置20は、例えば、ROM、RAM、CPU等を含むコンピュータ等に所定のプログラムが読み込まれて、CPUが所定のプログラムを実行することで実現される。図6に示すように、モデル記憶部21、学習装置20は、仮説入力部22及び学習部23を有する。

[0087] モデル記憶部21は、学習対象の第1補助モデル111～第M補助モデル11M及びメインモデル110を記憶する。第1補助モデル111～第M補助モデル11M及びメインモデル110は、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、NNで表される。第1補助モデル111～第M補助モデル11M及びは、Nベスト仮説のうちの二つの仮説を、RNNを用いて隠れ状態ベクトルに変換する。そして

、第1補助モデル111～第M補助モデル11Mは、NNを用いて、隠れ状態ベクトルを基に二つの仮説の精度の高低の並びが正しいことを示す事後確率を判定情報として生成する。

[0088] メインモデル110は、第1補助モデル111～第M補助モデル11Mにおいてそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、NNで表される。メインモデル110は、NNを用いて、第1補助モデル111～第M補助モデル11Mにおいてそれぞれ変換された学習用の二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低の並びが正しいことを示す事後確率を生成する。

[0089] 仮説入力部22は、音声認識精度が既知である学習用のNベスト仮説の入力を受け付ける。学習用のNベスト仮説として、学習データ中の各発話に対して音声認識が行われ、各発話のNベスト仮説が得られているものとする。また学習データであるので、全ての仮説の音声認識精度は、既知である。また、Nベスト仮説中の全ての仮説に対して、前述のように、特徴量ベクトル列が抽出されているものとする。

[0090] 学習部23は、学習用のNベスト仮説のうちの二つの仮説の特徴量がそれぞれ与えられたときに、第1補助モデル111～第M補助モデル11M及びメインモデル110に対し、各NNが二つの仮説の精度の高低を判定するタスクを個別に行うとみなしたマルチタスク学習を行わせる。学習部23は、各NNによって実行された各タスクについて所定の損失をそれぞれ計算し、各損失の重み付け和を全体の損失関数とする。そして、学習部23は、この全体の損失関数に基づいて、各NNのパラメータの値を更新する。

[0091] なお、学習部23は、各損失に対し、等重みで重み付けをしてもよい。また、メインモデル110が出力する判定情報が判定部15における判定に使用されるため、学習部23は、メインモデル110に、他の補助モデルよりも多めの重みを付けてもよい。

[0092] 学習部23では、学習用のNベスト仮説のうちの二つの仮説の特徴量ベクトル列と、これらに対応する教師ラベル（後述）とを、第1補助モデル11

1～第M補助モデル11M及びメインモデル110に与える。これによって、学習部23は、第1補助モデル111～第M補助モデル11M及びメインモデル110がこれら二つの仮説の音声認識精度の高低を正しく判定できるように、第1補助モデル111～第M補助モデル11M及びメインモデル110の学習（パラメータの最適化）を行う。

[0093] 具体的には、学習部23は、特徴量ベクトル列と、対応する教師ラベルとを第1補助モデル111～第M補助モデル11Mに入力し、第1補助モデル111～第M補助モデル11M及びメインモデル110がこれらの特徴量ベクトルを与えられたときに対応する教師ラベルを正しく出力できるように、第1補助モデル111～第M補助モデル11M及びメインモデル110の学習を行う。学習部23は、教師ラベル付与部231及び入替部232を有する。

[0094] 教師ラベル付与部231は、二つの仮説のうち音声認識精度がより高い仮説に他方の仮説よりも高い順位が付与されている場合に正解を表す教師ラベル（ $y = 0$ ）を付与して、第1補助モデル111～第M補助モデル11M及びメインモデル110に学習させる。また、教師ラベル付与部231は、二つの仮説のうち音声認識精度がより高い仮説に他方の仮説よりも低い順位が付与されている場合に誤りを表す教師ラベル（ $y = 1$ ）を付与し、第1補助モデル111～第M補助モデル11M及びメインモデル110に学習させる。

[0095] 入替部232は、学習用のNベスト仮説のうちの二つの仮説の順位を入れ換え、対応する教師ラベルも入れ換えて、第1補助モデル111～第M補助モデル11M及びメインモデル110の学習を行う。図7は、図6に示す入替部232の処理を説明する図である。例えば、教師ラベルとして $y = 0$ が付与されている二つの仮説については（図7の（1）参照）、二つの仮説の順位を入れ換え、教師ラベル y を1に変える（図7の（2）参照）。一方、教師ラベルとして $y = 1$ が付与されている二つの仮説については、二つの仮説の順位を入れ換え、教師ラベル y を0に変える。

[0096] [学習処理の処理手順]

次に、図6に示す学習装置20が実行する学習処理の処理手順について説明する。図8は、実施の形態2に係る学習処理の処理手順を示すフローチャートである。図8では、Nベスト仮説から二つの仮説として $W^{(u)}$ 、 $W^{(v)}$ ($u < v \leq N$) が与えられ、かつ、 $W^{(u)}$ の精度は、 $W^{(v)}$ の精度よりも高いときの学習処理の処理手順を示す。

[0097] 図8に示すように、教師ラベル付与部231が、教師ラベル $y=0$ を付与し（ステップS21）、 $W^{(u)}$ 、 $W^{(v)}$ の特徴量 $X^{(u)}$ 、 $X^{(v)}$ を第1補助モデル111～第M補助モデル11Mに入力する（ステップS22）。そして、学習部23は、第1補助モデル111～第M補助モデル11M及びメインモデル110にマルチタスク学習を行わせて、第1補助モデル111～第M補助モデル11M及びメインモデル110のモデルパラメータを更新させる（ステップS23）。

[0098] すなわち、この二つの仮説の $W^{(u)}$ 、 $W^{(v)}$ の特徴量ベクトル $X^{(u)}$ 、 $X^{(v)}$ を第1補助モデル111～第M補助モデル11Mに入力した場合、第1補助モデル111～第M補助モデル11M及びメインモデル110は、理想的には、 $P(0 | X^{(u)}, X^{(v)}) = 1$ の事後確率を出力すべきである。このため、教師ラベル付与部231は、教師ラベルとして、 $y=0$ を与える。以上の入力を基に、学習部23は、第1補助モデル111～第M補助モデル11M及びメインモデル110のモデルパラメータ（エンコーダRNN（LSTMユニット）、1クラス分類FFNN及び単語の埋め込み処理 $embed(\cdot)$ を行うNNのパラメータを同時に）を更新させる。

[0099] そして、入替部232は、仮説 $W^{(u)}$ 、 $W^{(v)}$ の順位を入れ替える（ステップS24）。すなわち、入替部232は、元々、 $W^{(v)}$ であった仮説を $W^{(u)}$ とし、元々、 $W^{(u)}$ であった仮説を $W^{(v)}$ とする。この場合には、 $W^{(u)}$ の精度は、 $W^{(v)}$ の精度よりも低い。よって、この二つの仮説 $W^{(u)}$ 、 $W^{(v)}$ の特徴量ベクトル $X^{(u)}$ 、 $X^{(v)}$ を第1補助モデル111～第M補助モデル11M及びメインモデル110に入力した場合、第1補助モデル111～第M補助

モデル11M及びメインモデル110は、理想的には、 $P(O | X^{(u)}, X^{(v)}) = 0$ の事後確率を出力すべきである。

[0100] このため、教師ラベル付与部231は、教師ラベルとして、 $y = 1$ を付与し（ステップS25）、 $W^{(u)}$ 、 $W^{(v)}$ の特徴量 $X^{(u)}$ 、 $X^{(v)}$ を第1補助モデル111～第M補助モデル11M及びメインモデル110に入力する（ステップS26）。学習部23は、以上の入力を基に、第1補助モデル111～第M補助モデル11M及びメインモデル110にマルチタスク学習を行わせて、第1補助モデル111～第M補助モデル11M及びメインモデル110を更新させて（ステップS27）、二つの仮説 $W^{(u)}$ 、 $W^{(v)}$ に対する学習処理を終了する。

[0101] 学習装置20は、上記の手順を、学習データ中の各発話のNベスト仮説について繰り返し、更にはその繰り返し自体を何度か（何エポックか）繰り返す。学習部23は、学習の更なる具体的な手順については、従来のNNの学習（詳細は、例えば、参考文献2参照）と同様に行うことができる。

[0102] [実施の形態2の効果]

このように、実施の形態2に係る学習装置20は、第1補助モデル111～第M補助モデル11M及びメインモデル110に、音声認識精度が既知である学習用のNベスト仮説のうちの二つの仮説を1組として、複数の組についてそれぞれ音声認識精度の高低を判定できるように予めマルチタスク学習を行わせている。したがって、学習装置20は、Nベストリランキングを行う上で最適な第1補助モデル111～第M補助モデル11M及びメインモデル110を、最新のNNに基づき実現することができる。そして、リランキング装置10は、学習装置20において学習された第1補助モデル111～第M補助モデル11M及びメインモデル110を使用することによって、一対一の二つの仮説の比較を精度よく行うことができ、安定した制度でオラクル仮説を抽出することができる。

[0103] [学習処理の効率化例1]

図8に示す学習処理の処理手順は、計算コストが高い。例えば、Eをエポ

ック数、 M を学習データ中の発話数とすると、上記の学習手順におけるモデルパラメータの更新回数は、最大で、 $E \times M \times N \times 2 \times {}_N C_2$ になる。通常、 E は数十程度、 M は少なくとも数万、 N は上記の通り100～1000程度であるので、モデルパラメータの更新回数は、膨大な数に達する。このため、本実施の形態では、学習の効率化を図ることが好ましい。そこで、以下に、学習の効率化例1について述べる。

[0104] 上述したように、 N ベストリスコアリングの主な目的は、 N ベスト仮説からオラクル仮説を最終的な音声認識結果として見つけ出すことである。言い換えれば、オラクル仮説をその他の $N-1$ 個の仮説から精度よく区別できればよい。これを実現するために、学習の際に、 N ベストリランキングモデルに入力する二つの仮説のうち的一方をオラクル仮説とする。これにより、モデルパラメータの更新回数を、 $E \times M \times N \times 2 \times (N-1)$ に削減することができる。

[0105] [学習処理の効率化例2]

次に、学習の効率化例2について説明する。学習の効率化例1では、 N ベスト仮説が与えられたとき、その中に含まれるオラクル仮説とその他の $N-1$ 個の仮説とを比較していた。学習処理の効率化例2では、オラクル仮説と比較するその他の仮説の個数を絞り込む。

[0106] 例えば、まず、下の典型的な四つの仮説を選択する。

仮説1は、オラクル仮説の次に高い音声認識精度を持つ仮説である。

仮説2は、音声認識スコアが最も高い仮説である。

仮説3は、最も低い音声認識精度を持つ仮説である。

仮説4は、音声認識スコアが最も低い仮説である。

[0107] 仮説1と仮説2とは、音声認識精度が高い（または高いと推定される）仮説で、オラクル仮説との区別が難しい仮説である。一方、仮説3と仮説4とは、音声認識精度が低い（または低いと推定される）仮説で、オラクル仮説との区別が容易な（確実に区別しないといけない）仮説である。その他の仮説をこの四つのみに絞り込む場合は、モデルパラメータの更新回数は、 $E \times$

$M \times N \times 2 \times 4$ にまで削減することができる。

[0108] ただし、上記の四つの仮説のみではオラクル仮説の対立仮説としての多様性が十分に確保できないと考えられる場合、 N ベスト仮説から、オラクル仮説とこれらの四つの仮説を除いた、残りの $N - 5$ 個の仮説から、所定のルールにしたがって抽出した所定数の仮説を選択して前記四つの仮説と共に対立仮説として用いてもよい。例えば、二つの仮説のうちの他方の仮説として、オラクル仮説とこれらの四つの仮説を除いた、残りの $N - 5$ 個の仮説から、等間隔に、或いは、はランダムに、 Q 個の仮説を選択して四つの仮説と共に他方の仮説として用いる。このとき、モデルパラメータの更新回数は、 $E \times M \times N \times 2 \times (4 + Q)$ となる。例えば、 Q は、 $5 \sim 50$ である。

[0109] [評価]

実際に、実施の形態 1 における N ベストリランキングと、非特許文献 1 記載の N ベストリランキングとの比較評価を行った。表 1 は、CSJ 音声コーパスを用いて、非特許文献 1 記載の N ベストリランキングとの比較評価する $100 (=N)$ ベストリランキング評価を行った結果を示す表である。表の数値は、WER (Word error rate) 率 [%] であり、Dev (Development)、Eval (Evaluation) を示す。

[0110] [表1]

(表1)

No.	Model	Dev	Eval
1	Single-encoder DDM	16.4	13.7
2	Eight-encoder DDM	16.1	13.4
3	2 with fwd & bwd LSTMLM scores	15.2	12.6
4	Oracle	11.6	9.7

[0111] 表 1 の通番「1」は、非特許文献 1 記載の N ベストリランキング結果である。表 1 の通番「2」は、実施の形態 1 に係るリランキング装置 10 であって 8 個の補助モデルを有する場合の結果である。通番「3」は、通番「2」

の条件に加え、前向き及び後ろ向き L S T M L M の単語予測スコアを 18 時
限目及び 19 次元目の補助特徴として用いている。通番「4」は、参考のため
に示されたオラクルである。

[0112] 表 1 に示すように、通番「1」の非特許文献 1 記載のリランキング方法で
も、十分に W E R を削減できるが、通番「2」の 8 個の補助モデルを有する
リランキング装置 10 では、さらに W E R 削減が実現できる。また、通番「
3」の評価結果に示すように、両方向の L S T M L M の単語予測スコアと 8
個の補助モデルを用いることで、相補的な W E R 削減効果が得られることが
確認できた。また、通番「2」以外にも、補助モデルの個数を、2 または 4
とした構成でも評価を行っており、補助モデルの個数を増やすほど W E R を
削減できる傾向が確認できた。

[0113] この評価結果から、本実施の形態 1 に係るリランキング装置 10 は、非特
許文献 1 記載のリランキング方法と比して、安定した W E R 削減が実現でき
る。

[0114] [実施の形態 3]

なお、実施の形態 1 のリランキング装置 10 は、メインモデル 110 の出
力を用いて判定を行ったが、メインモデル 110 の出力の他に各補助モデル
の出力を用いて判定を行ってもよい。図 9 は、実施の形態 3 に係るリランキ
ング装置の要部構成を示す図である。

[0115] 図 3 に示すように、リランキング装置は、判定部 15 の前段に、重み付け
部 18 を有する。重み付け部 18 は、第 1 補助モデル 111 ~ 第 M 補助モデ
ル 11M 及びメインモデル 110 から出力された全ての判定情報を取得し、
各判定情報に対して重み付け和を計算する。

[0116] なお、各判定情報に対応する重みは予め設定されている。重み付け部 18
は、各判定情報に対し、全補助モデル及びメインモデル 110 に対して等重
みで重み付けをしてもよい。また、重み付け部 18 は、メインモデル 110
に、他の補助モデルよりも多めの重みを付けてもよい。また、重み付け部 1
8 は、予め各判定情報に対する重みを学習した 1 層の線形 N N を有し、各判

定情報が入力されると各判定対象に対する重みを求めてもよい。

[0117] 判定部 15 は、重み付け部 18 が計算した重み付け和の値に基づいて二つの仮説の精度の高低を判定する。例えば、判定部 15 は、判定情報のそれぞれが、仮説 $W^{(u)}$ が選択される確率を示すものとして、判定情報の重みづけ和を 0 ～ 1 の範囲に収まるように正規化した値が 0.5 以上であれば仮説 $W^{(u)}$ を選択し、そうでなければ仮説 $W^{(v)}$ を選択する。

[0118] [判定処理の処理手順]

図 10 は、実施の形態 3 に係るリランキング処理の処理手順を示すフローチャートである。

[0119] 図 10 に示すステップ S31 ～ステップ S36 は、図 4 に示すステップ S1 ～ステップ S6 と同じ処理である。重み付け部 18 は、第 1 補助モデル 111 ～第 M 補助モデル 11M 及びメインモデル 110 から出力された全ての判定情報を取得し、各判定情報に対して重み付け和を計算する重み付け処理を行う（ステップ S37）。そして、重み付け部 18 が計算した重み付け和の値に基づいて二つの仮説の精度の高低を判定する。ステップ S38 ～ステップ S43 は、図 4 に示すステップ S7 ～ステップ S12 と同じ処理である。

[0120] [実施の形態 3 の効果]

この実施の形態 3 に示すように、メインモデル 110 による判定情報に加え、全補助モデルによる判定情報を用いて、判定を行うことも可能である。この際、実施の形態 3 では、各補助モデル或いはメインモデル 110 に応じて、各判定情報に対する重み付けを行い、重み付け和の値に基づいて二つの仮説の精度の高低を判定するため、オラクル仮説を抽出精度を保持することができる。

[0121] なお、本実施の形態では、全ての仮説に対して、比較処理を行うため、N ベスト仮説のソートも可能である。

[0122] また、本実施の形態 1 ～ 3 では、音声認識の N ベスト仮説をリランキングするためのモデルとして、図 1 に例示する第 1 補助モデル 111 ～第 M 補助

モデル 1 1 M 及びメインモデル 1 1 0 について説明した。ただし、本実施の形態 1 ～ 3 の第 1 補助モデル 1 1 1 ～第 M 補助モデル 1 1 M 及びメインモデル 1 1 0 は、音声認識の N ベスト仮説への適用にとどまらず、N ベスト仮説を採用しているあらゆるタスクに適用可能である。例えば、機械翻訳や文章要約などにも本実施の形態を適用することが可能である。また、文字列に限らず、数字やアルファベットを含む複数の系列にも本実施の系列を適用することが可能である。

[0123] このため、本実施の形態 1 ～ 3 は、ある一つの入力に対する解の候補として挙げられた複数の系列であれば、このうちの二つの系列に対し、NN で表されるモデルを用いて、二つの系列のうちより精度が高い（誤りが少ない）系列を判定できる。そして、本実施の形態 1 ～ 3 では、二つの系列のうち、より精度が高いと判定した系列を比較対象として残し、他方の系列を比較対象から外し、精度が高いと判定した系列を二つの系列の一方の仮説として選択し、複数の系列のうち、判定が行われていない系列のいずれかを他方の仮説として選択する。そして、本実施の形態 1 ～ 3 では、判定処理と選択処理とを、所定条件に達するまで順次実行させせる。これによって、本実施の形態 1 ～ 3 によれば、所定条件に達した場合に比較対象として残っている系列を、最も精度が高い系列、すなわち、最終的な出力として出力することができる。

[0124] また、この場合には、本実施の形態 1 ～ 3 では、精度が既知である学習用の複数の系列のうちの二つの系列の特徴量が与えられたとき、それら二つの系列の精度の高低が判定できるような、NN で表される第 1 補助モデル 1 1 1 ～第 M 補助モデル 1 1 M 及びメインモデル 1 1 0 にマルチタスク学習を行わせる。そして、本実施の形態 1 ～ 3 では、二つの系列のうち精度がより高い（誤りがより少ない）系列に他方の系列よりも高い順位が付与されている場合に正解を示す教師ラベルを付与して第 1 補助モデル 1 1 1 ～第 M 補助モデル 1 1 M 及びメインモデル 1 1 0 に学習させる。そして、本実施の形態 1 ～ 3 では、二つの系列のうち精度がより高い（誤りがより少ない）系列に他

方の系列よりも低い順位が付与されている場に誤りを示す教師ラベルを付与して第1補助モデル111～第M補助モデル11M及びメインモデル110に学習させる。本実施の形態1～3では、この第1補助モデル111～第M補助モデル11M及びメインモデル110によって、一対一の二つの系列の比較が高精度で行うことができ、この結果、最も精度の高い系列を精度よく得ることができる。

[0125] [システム構成等]

図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部又は一部を、各種の負荷や使用状況等に応じて、任意の単位で機能的又は物理的に分散・統合して構成することができる。例えば、リランキング装置10及び学習装置20は、一体の装置であってもよい。さらに、各装置にて行なわれる各処理機能は、その全部又は任意の一部が、CPU及び当該CPUにて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

[0126] また、本実施形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部又は一部を手動的におこなうこともでき、あるいは、手動的におこなわれるものとして説明した処理の全部又は一部を公知の方法で自動的におこなうこともできる。また、本実施形態において説明した各処理は、記載の順にしたが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。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

[0127] [プログラム]

図11は、プログラムが実行されることにより、リランキング装置10或いは学習装置20が実現されるコンピュータの一例を示す図である。コンピ

ユータ１０００は、例えば、メモリ１０１０、CPU１０２０を有する。また、コンピュータ１０００は、ハードディスクドライバインタフェース１０３０、ディスクドライバインタフェース１０４０、シリアルポートインタフェース１０５０、ビデオアダプタ１０６０、ネットワークインタフェース１０７０を有する。これらの各部は、バス１０８０によって接続される。

[0128] メモリ１０１０は、ROM１０１１及びRAM１０１２を含む。ROM１０１１は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライバインタフェース１０３０は、ハードディスクドライブ１０３１に接続される。ディスクドライバインタフェース１０４０は、ディスクドライブ１０４１に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ１０４１に挿入される。シリアルポートインタフェース１０５０は、例えばマウス１１１０、キーボード１１２０に接続される。ビデオアダプタ１０６０は、例えばディスプレイ１１３０に接続される。

[0129] ハードディスクドライブ１０３１は、例えば、OS１０９１、アプリケーションプログラム１０９２、プログラムモジュール１０９３、プログラムデータ１０９４を記憶する。すなわち、リランキング装置１０或いは学習装置２０の各処理を規定するプログラムは、コンピュータ１０００により実行可能なコードが記述されたプログラムモジュール１０９３として実装される。プログラムモジュール１０９３は、例えばハードディスクドライブ１０３１に記憶される。例えば、リランキング装置１０或いは学習装置２０における機能構成と同様の処理を実行するためのプログラムモジュール１０９３が、ハードディスクドライブ１０３１に記憶される。なお、ハードディスクドライブ１０３１は、SSD (Solid State Drive) により代替されてもよい。

[0130] また、上述した実施形態の処理で用いられる設定データは、プログラムデータ１０９４として、例えばメモリ１０１０やハードディスクドライブ１０３１に記憶される。そして、CPU１０２０が、メモリ１０１０やハードディスクドライブ１０３１に記憶されたプログラムモジュール１０９３やプロ

グラムデータ 1094 を必要に応じて RAM 1012 に読み出して実行する。

[0131] なお、プログラムモジュール 1093 やプログラムデータ 1094 は、ハードディスクドライブ 1031 に記憶される場合に限らず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ 1041 等を介して CPU 1020 によって読み出されてもよい。あるいは、プログラムモジュール 1093 及びプログラムデータ 1094 は、ネットワーク（LAN (Local Area Network)、WAN (Wide Area Network) 等）を介して接続された他のコンピュータに記憶されてもよい。そして、プログラムモジュール 1093 及びプログラムデータ 1094 は、他のコンピュータから、ネットワークインタフェース 1070 を介して CPU 1020 によって読み出されてもよい。

[0132] 以上、本発明者によってなされた発明を適用した実施形態について説明したが、本実施形態による本発明の開示の一部をなす記述及び図面により本発明は限定されることはない。すなわち、本実施形態に基づいて当業者等によりなされる他の実施形態、実施例及び運用技術等は全て本発明の範疇に含まれる。

符号の説明

- [0133] 2 音声認識装置
- 10 リランキング装置
 - 11, 21 モデル記憶部
 - 12 仮説入力部
 - 13 仮説選択部
 - 14 特徴量抽出部
 - 15 判定部
 - 16 実行制御部
 - 17 出力部
 - 18 重み付け部
 - 20 学習装置

2 2 仮説入力部

2 3 学習部

1 1 0 メインモデル

1 1 1 ～ 1 1 M 第 1 補助モデル～第 M 補助モデル

2 3 1 教師ラベル付与部

2 3 2 入替部

請求の範囲

- [請求項1] 音声認識精度のスコアが対応付けられたNベスト仮説の入力を受け付ける入力部と、
- 入力を受け付けた前記Nベスト仮説のうち、判定対象である二つの仮説を選択する選択部と、
- 選択された二つの仮説が与えられたとき、前記二つの仮説を隠れ状態ベクトルに変換し、前記二つの仮説の隠れ状態ベクトルを基に前記二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の補助モデルと、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に、前記二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデルとを用いて、前記二つの仮説の精度の高低を判定する判定部と、
- を有することを特徴とする判定装置。
- [請求項2] 前記選択部は、前記Nベスト仮説のスコアの昇順に前記二つの仮説を選択することを特徴とする請求項1に記載の判定装置。
- [請求項3] 前記判定部は、前記メインモデルから出力された判定情報、または、各補助モデルから出力された判定情報と前記メインモデルから出力された判定情報とに対して計算した重み付け和の値、に基づいて前記二つの仮説の精度の高低を判定することを特徴とする請求項1または2に記載の判定装置。
- [請求項4] 各補助モデルは、前記二つの仮説を、再帰的ニューラルネットワークを用いて隠れ状態ベクトルに変換し、ニューラルネットワークを用いて、前記隠れ状態ベクトルを基に二つの系列の精度の高低の並びが正しいことを示す事後確率を出力し、
- 前記メインモデルは、ニューラルネットワークを用いて、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に二つの系列の精度の高低の並びが正しいことを示す事後確率

を出力する

ことを特徴とする請求項 1 ～ 3 のいずれか一つに記載の判定装置。

[請求項5] 音声認識精度が既知である学習用の二つの仮説の入力を受け付ける入力部と、

前記二つの仮説が与えられたとき、前記二つの仮説を隠れ状態ベクトルに変換し、前記二つの仮説の隠れ状態ベクトルを基に前記二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の補助モデルと、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に、前記二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデルとに対し、各ニューラルネットワークが前記二つの仮説の精度の高低を判定するタスクを個別に行うとみなしたマルチタスク学習を行わせる学習部と、

を有することを特徴とする学習装置。

[請求項6] 前記学習部は、前記二つの仮説のうち音声認識精度がより高い仮説に他方の仮説よりも高い順位が付与されている場合に正解ラベルを付与して前記複数の補助モデル及び前記メインモデルに学習させ、前記二つの仮説のうち音声認識精度がより高い仮説に他方の仮説よりも低い順位が付与されている場合に誤りラベルを付与して前記複数の補助モデル及び前記メインモデルに学習させることを特徴とする請求項 5 に記載の学習装置。

[請求項7] 前記学習部は、各ニューラルネットワークによって実行された各タスクについて所定の損失をそれぞれ計算し、各損失の重み付け和に基づいて、各ニューラルネットワークのパラメータの値を更新することを特徴とする請求項 5 または 6 に記載の学習装置。

[請求項8] 各補助モデルは、前記二つの仮説を、再帰的ニューラルネットワークを用いて隠れ状態ベクトルに変換し、ニューラルネットワークを用いて、前記隠れ状態ベクトルを基に二つの系列の精度の高低の並びが

正しいことを示す事後確率を出力し、

前記メインモデルは、ニューラルネットワークを用いて、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に二つの系列の精度の高低の並びが正しいことを示す事後確率を出力する

ことを特徴とする請求項5～7のいずれか一つに記載の学習装置。

[請求項9]

判定装置が実行する判定方法であって、

音声認識精度のスコアが対応付けられたNベスト仮説の入力を受け付ける工程と、

入力を受け付けた前記Nベスト仮説のうち、判定対象である二つの仮説を選択する工程と、

選択された二つの仮説が与えられたとき、前記二つの仮説を隠れ状態ベクトルに変換し、前記二つの仮説の隠れ状態ベクトルを基に前記二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の補助モデルと、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に、前記二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデルとを用いて、前記二つの仮説の精度の高低を判定する工程と、

を含んだことを特徴とする判定方法。

[請求項10]

音声認識精度のスコアが対応付けられたNベスト仮説の入力を受け付けるステップと、

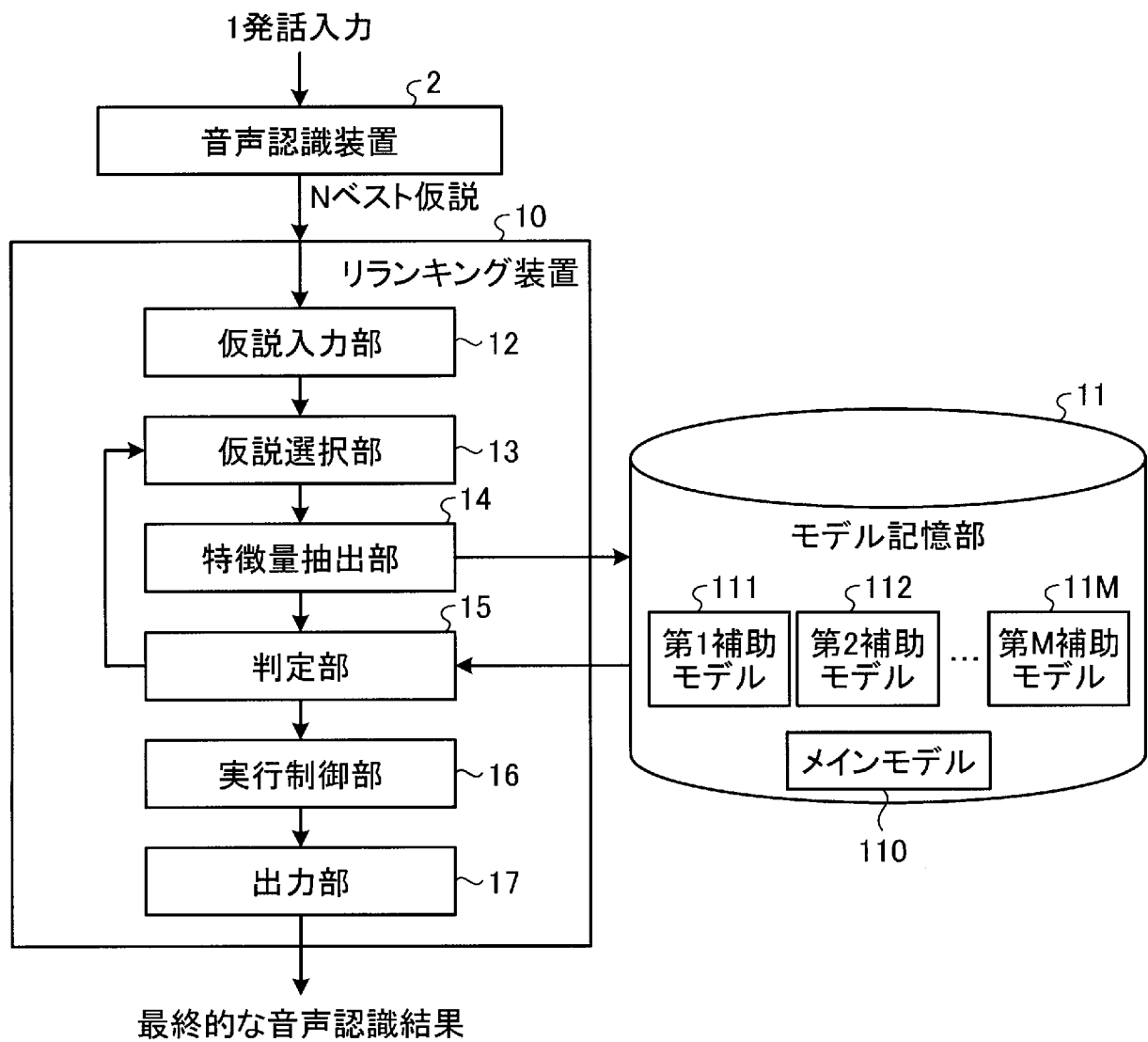
入力を受け付けた前記Nベスト仮説のうち、判定対象である二つの仮説を選択するステップと、

選択された二つの仮説が与えられたとき、前記二つの仮説を隠れ状態ベクトルに変換し、前記二つの仮説の隠れ状態ベクトルを基に前記二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表される複数の補助モデルと、前記複数の補助モデルでそれぞれ

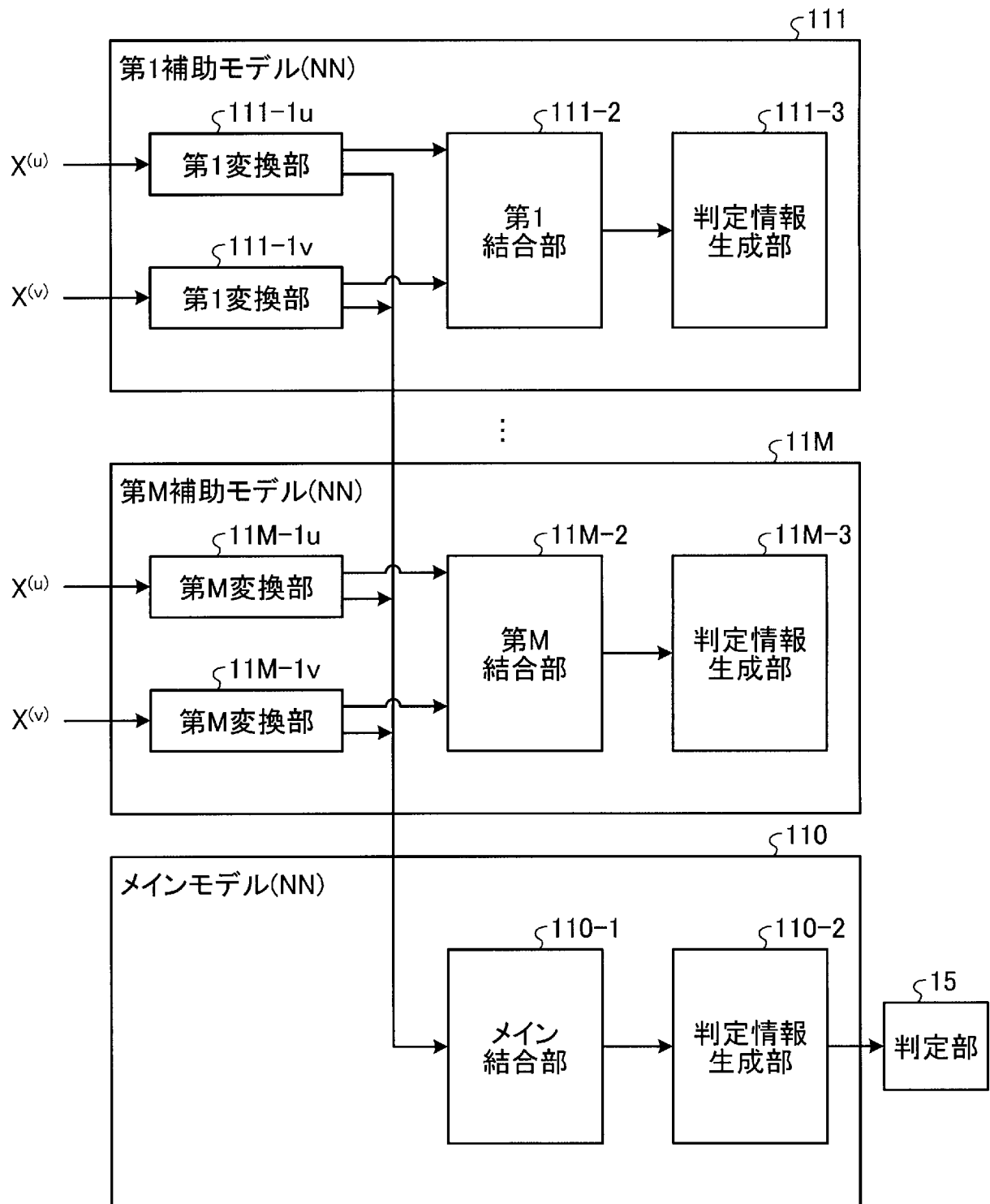
変換された前記二つの仮説の隠れ状態ベクトルを基に、前記二つの仮説の精度の高低を判定できるような、ニューラルネットワークで表されるメインモデルとを用いて、前記二つの仮説の精度の高低を判定するステップと、

をコンピュータに実行させるための判定プログラム。

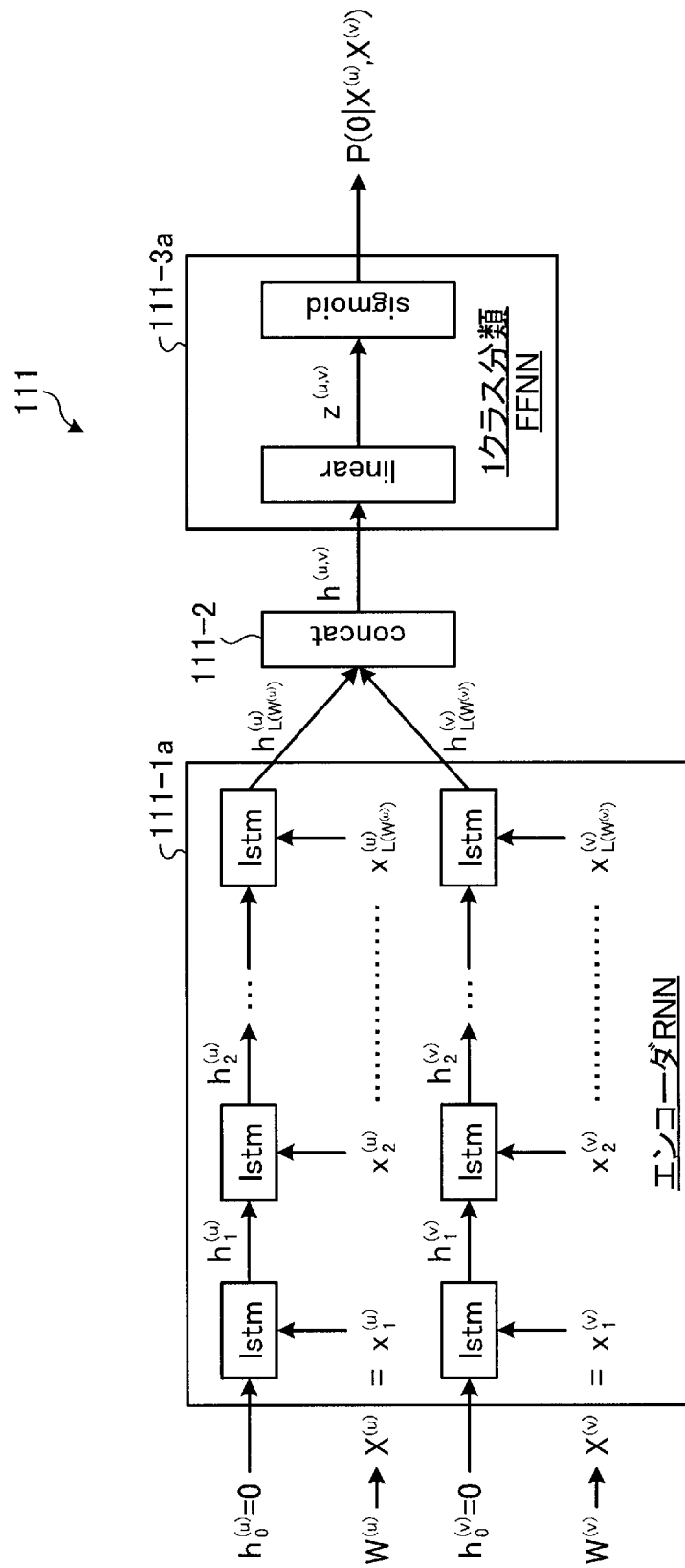
[図1]



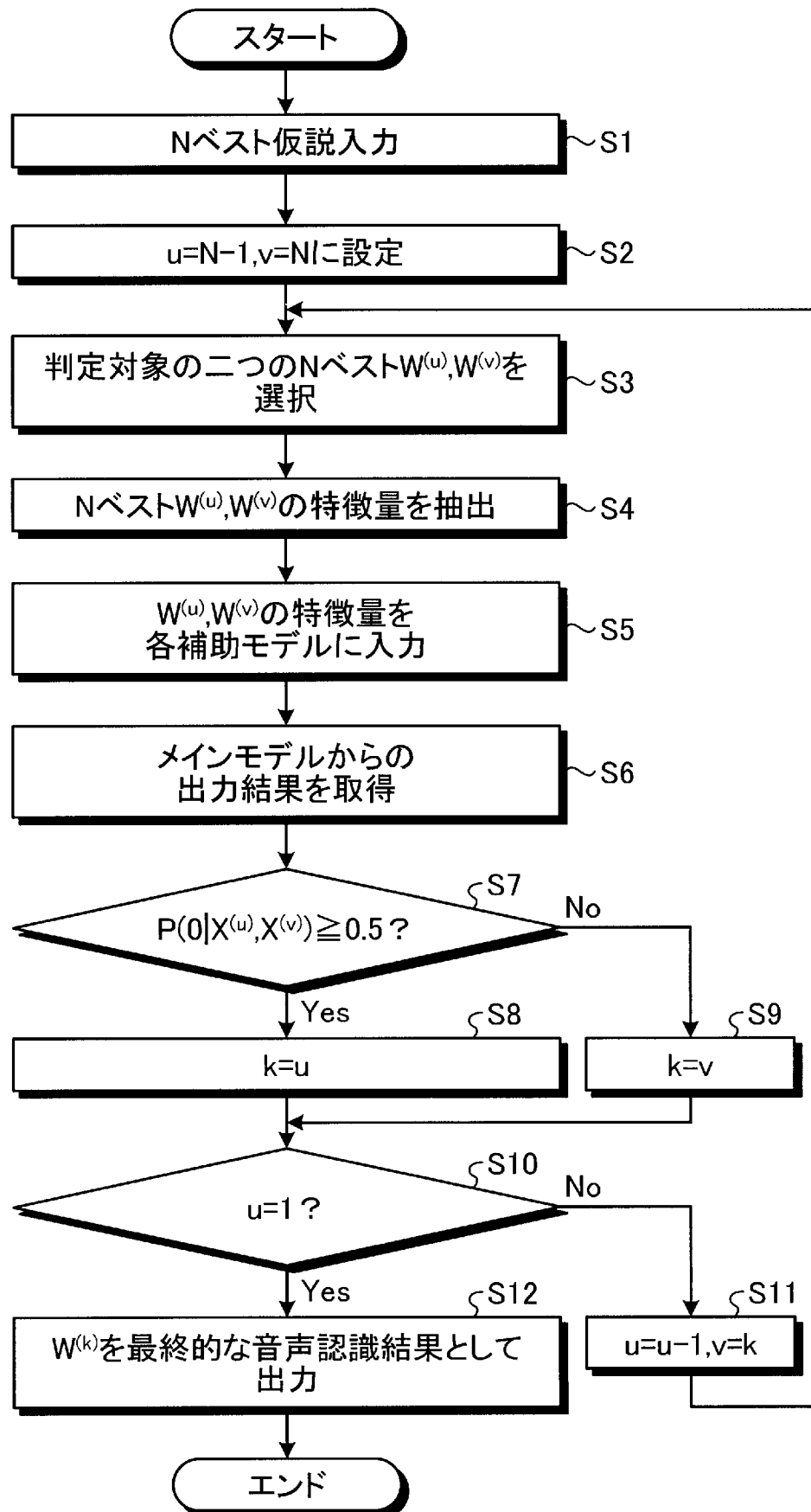
[図2]



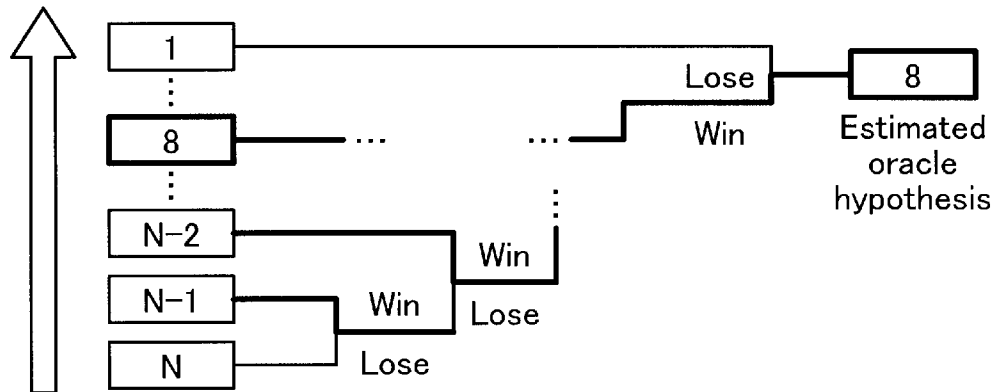
[図3]



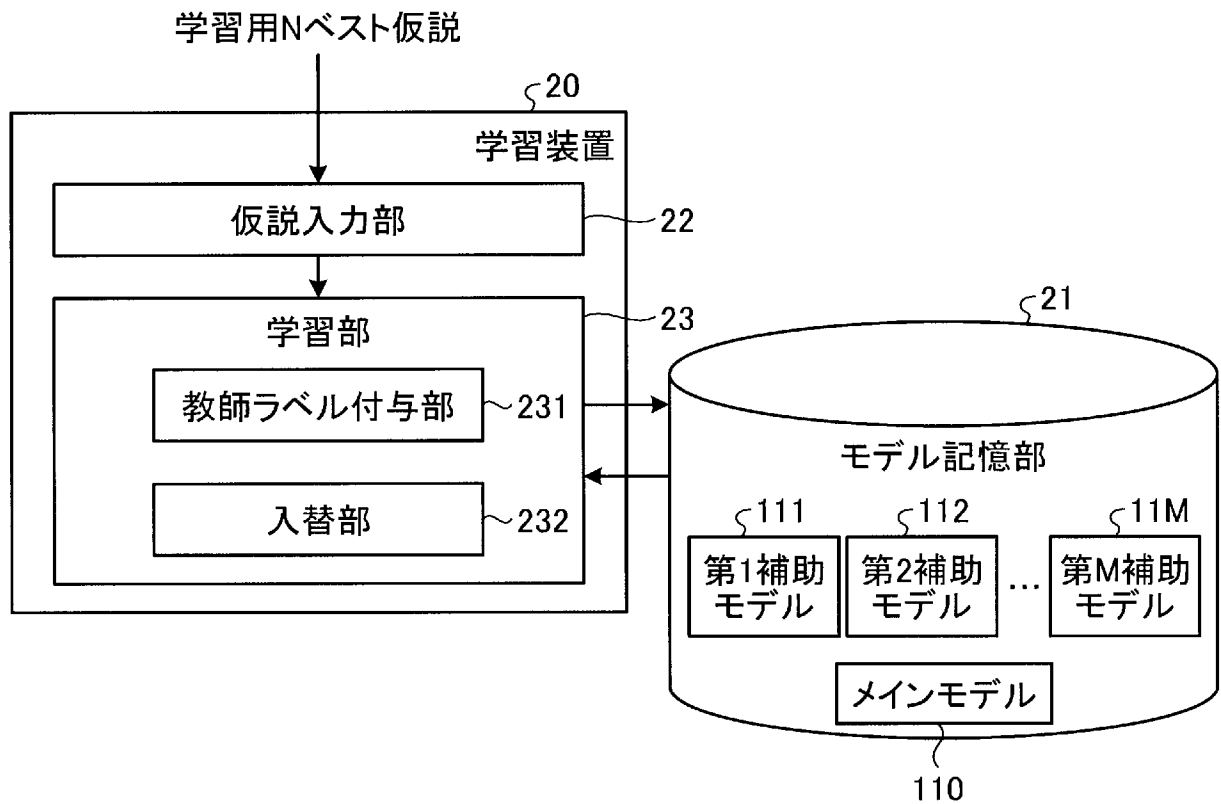
[図4]



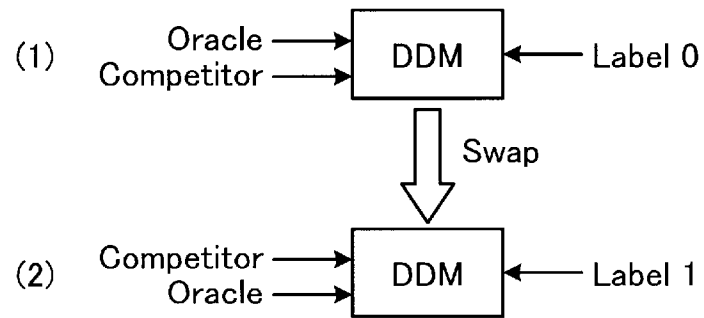
[図5]



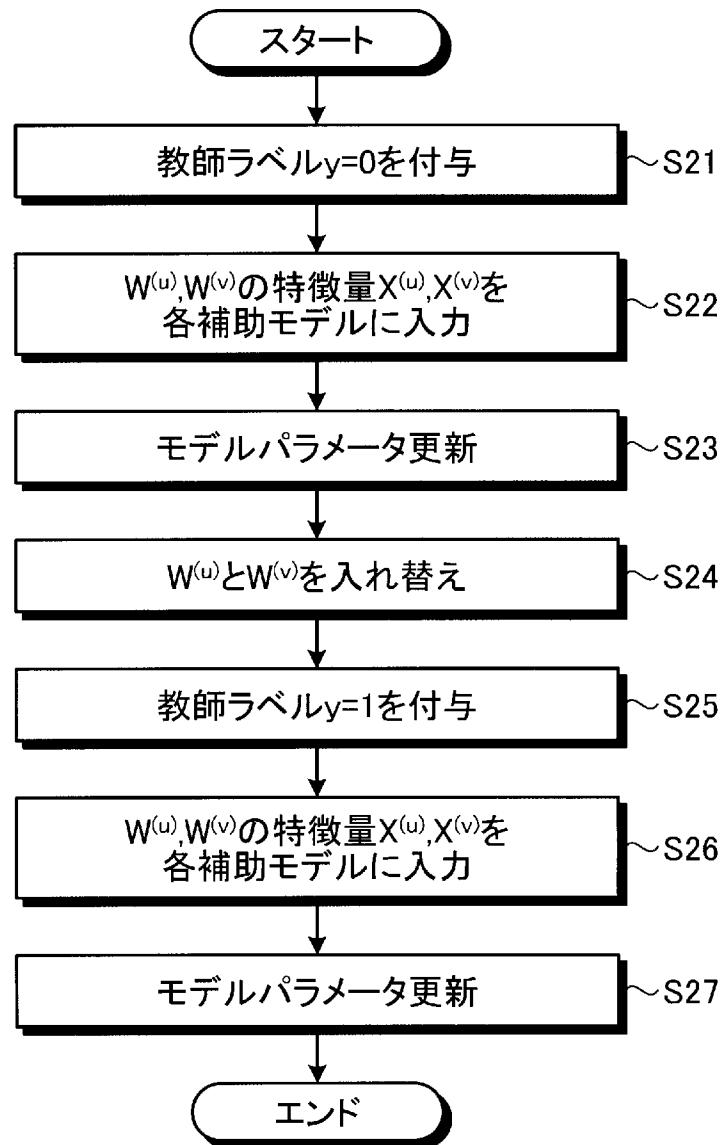
[図6]



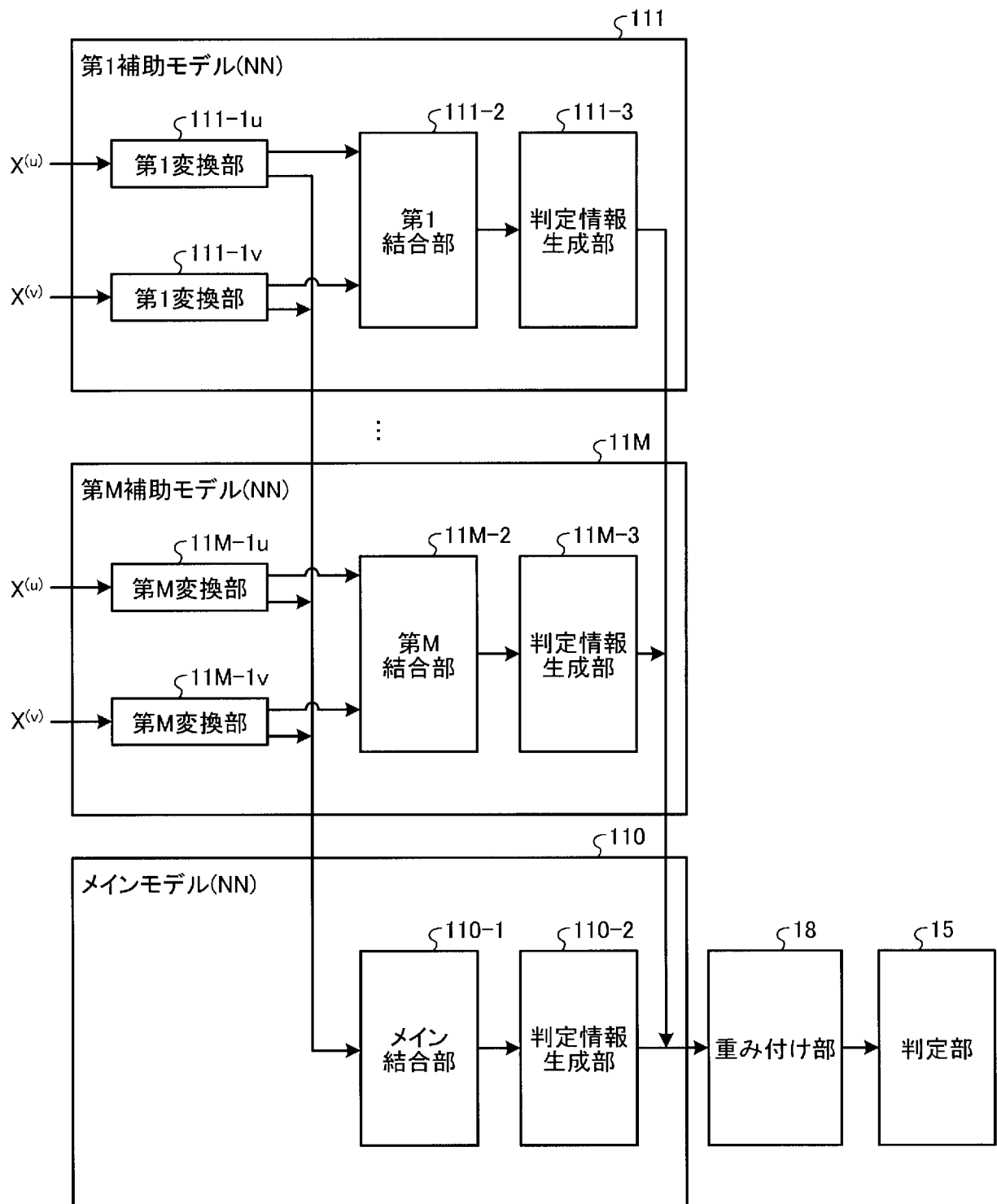
[図7]



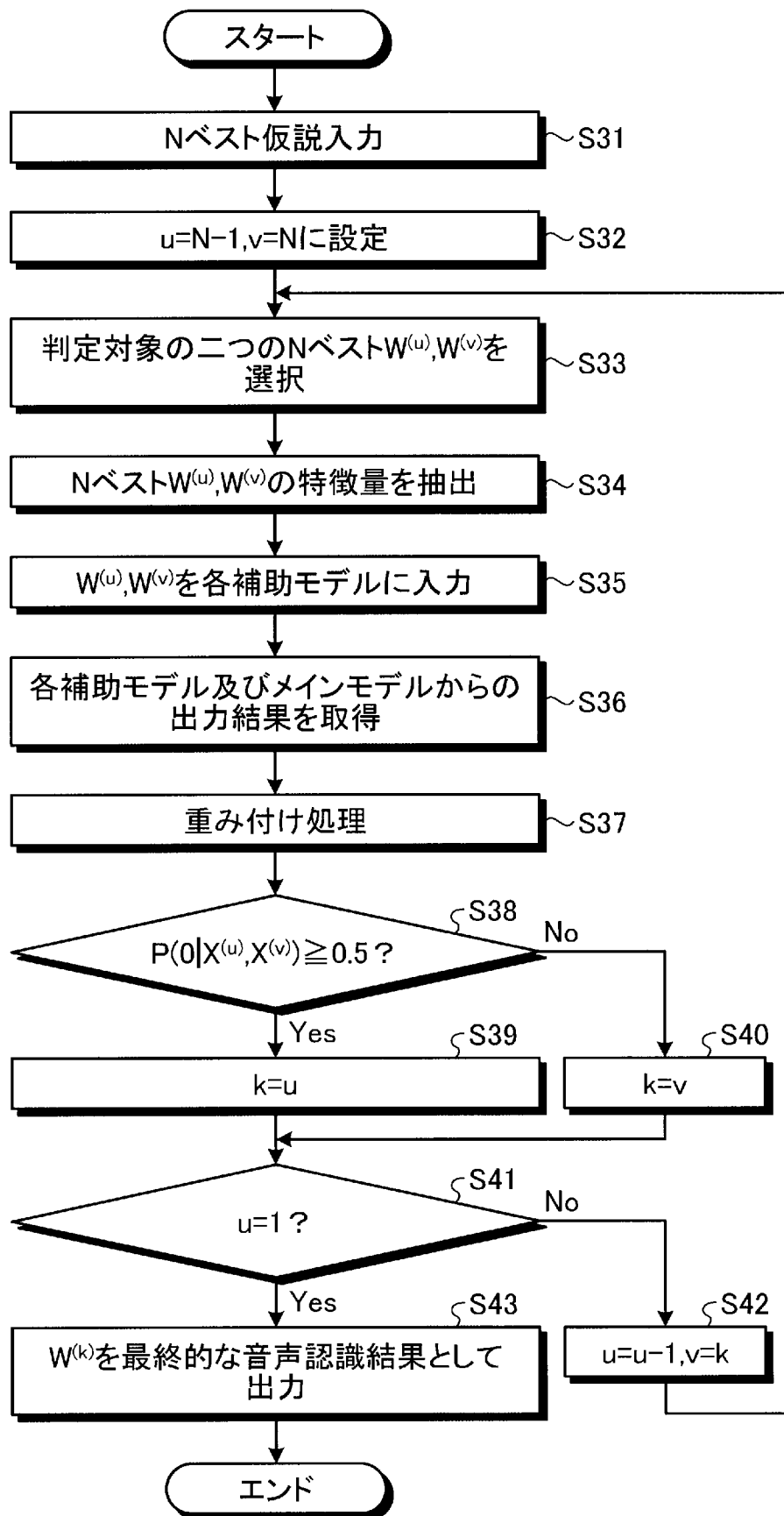
[図8]



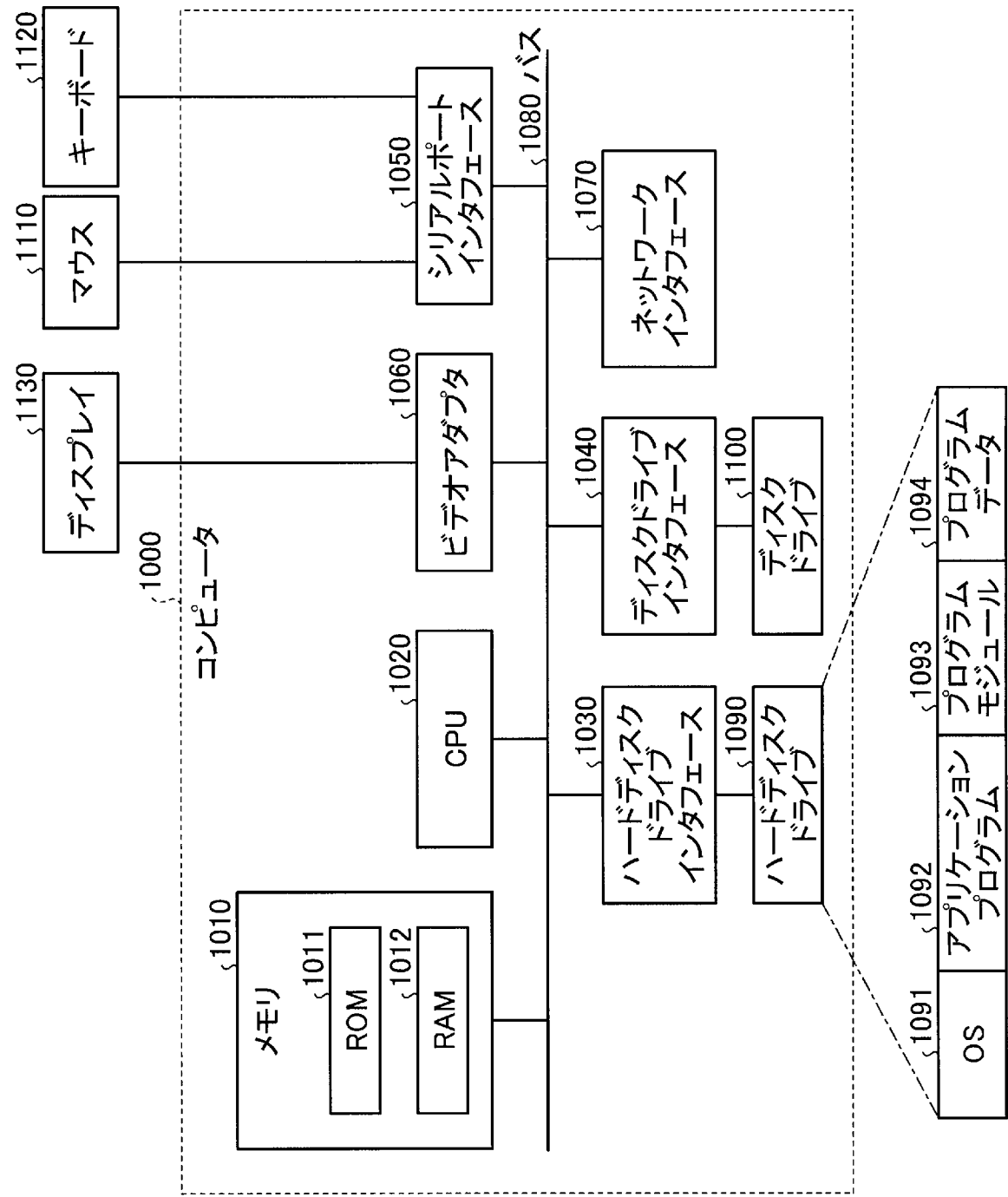
[図9]



[図10]



[図11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/031517

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/16 (2006.01) i

FI: G10L15/08 200Z; 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

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.
A	小川厚德他, 一対一の仮説比較を行う encoder-classifier モデルを用いた N ベスト音声認識仮説のリスクアリング, 日本音響学会 2018 年春季研究発表会講演論文集[CD-ROM], March 2018, pp. 23-24, non-official translation (OGAWA, Atsunori et al., "Rescoring of N-Best speech recognition hypotheses using an encoder-classifier model that performs one-to-one hypothesis comparison", Lecture proceedings of 2018 spring research conference of the Acoustical Society of Japan[CD-ROM])	1-10



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

"I"

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

21 January 2020 (21.01.2020)

Date of mailing of the international search report

28 January 2020 (28.01.2020)

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

International application No.

PCT/JP2019/031517

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	田中智大他, 複数仮説を考慮したニューラル誤り訂正言語モデルの検討, 電子情報通信学会技術研究報告[online], August 2018, vol. 118, no. 198, pp. 31-36, (TANAKA, Tomohiro et al., "Neural Error Corrective Language Models with Multiple Hypotheses", IEICE technical report[online])	1-10
A	JP 2011-243147 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 01.12.2011 (2011-12-01) paragraphs [0022]-[0033]	1-10
A	JP 2018-60047 A (NATIONAL INSTITUTE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY) 12.04.2018 (2018-04-12) paragraph [0033]	1-10
A	US 2017/0221474 A1 (MITSUBISHI ELECTRIC RESEARCH LABORATORIES INC.) 03.08.2017 (2017-08-03) paragraphs [0025]-[0049]	1-10
A	US 10032463 B1 (AMAZON TECHNOLOGIES, INC.) 24.07.2018 (2018-07-24) column 2, line 20 to column 12, line 44	1-10
E, A	JP 2019-133084 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 08.08.2019 (2019-08-08) paragraphs [0013]-[0084]	1-10
E, A	JP 2019-133085 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 08.08.2019 (2019-08-08) paragraphs [0018]-[0084]	1-10

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2019/031517

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2011-243147 A	01 Dec. 2011	(Family: none)	
JP 2018-60047 A	12 Apr. 2018	WO 2018/066436 A1	
US 2017/0221474 A1	03 Aug. 2017	WO 2017/135334 A1	
US 10032463 B1	24 Jul. 2018	(Family: none)	
JP 2019-133084 A	08 Aug. 2019	WO 2019/151506 A1	
JP 2019-133085 A	08 Aug. 2019	WO 2019/151503 A1	

A. 発明の属する分野の分類（国際特許分類（IPC）） G10L 15/08(2006.01)i; G10L 15/16(2006.01)i FI: G10L15/08 200Z; 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		
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 1971-2020年 日本国実用新案登録公報 1996-2020年 日本国登録実用新案公報 1994-2020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	小川厚德他，一対一の仮説比較を行うencoder-classifierモデルを用いたNベスト音声認識仮説のリスコアリング，日本音響学会2018年春季研究発表会講演論文集 [CD-ROM]，2018.03 pp.23-24	1-10
A	田中智大他，複数仮説を考慮したニューラル誤り訂正言語モデルの検討，電子情報通信学会技術研究報告 [online]，2018.08, Vol.118, No.198 pp.31-36	1-10
A	JP 2011-243147 A（日本電信電話株式会社）01.12.2011（2011-12-01） 段落[0022]-[0033]	1-10
A	JP 2018-60047 A（国立研究開発法人情報通信研究機構）12.04.2018（2018-04-12） 段落[0033]	1-10
A	US 2017/0221474 A1（MITSUBISHI ELECTRIC RESEARCH LABORATORIES INC.） 03.08.2017（2017-08-03） 段落[0025]-[0049]	1-10
A	US 10032463 B1（AMAZON TECHNOLOGIES, INC.）24.07.2018（2018-07-24） 第2欄第20行～第12欄第44行	1-10
☒ 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 21.01.2020		国際調査報告の発送日 28.01.2020
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） 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
E, A	JP 2019-133084 A (日本電信電話株式会社) 08.08.2019 (2019 - 08 - 08) 段落[0013]-[0084]	1-10
E, A	JP 2019-133085 A (日本電信電話株式会社) 08.08.2019 (2019 - 08 - 08) 段落[0018]-[0084]	1-10

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/JP2019/031517

引用文献			公表日	パテントファミリー文献	公表日
JP	2011-243147	A	01.12.2011	(ファミリーなし)	
JP	2018-60047	A	12.04.2018	W0	2018/066436 A1
US	2017/0221474	A1	03.08.2017	W0	2017/135334 A1
US	10032463	B1	24.07.2018	(ファミリーなし)	
JP	2019-133084	A	08.08.2019	W0	2019/151506 A1
JP	2019-133085	A	08.08.2019	W0	2019/151503 A1