(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
2007 年12 月13 日 (13.12.2007)

PCT

(10) 国際公開番号
WO 2007/142102 A1

(51) 国際特許分類:
G10L 15/06 (2006.01) *G10L 15/18* (2006.01)
G10L 15/02 (2006.01)

(21) 国際出願番号: PCT/JP2007/061023

(22) 国際出願日: 2007 年5 月30 日 (30.05.2007)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願2006-150962 2006 年5 月31 日 (31.05.2006) JP

(71) 出願人 (米国を除く全ての指定国について): 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目７番１号 Tokyo (JP).

(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 江森 正 (EMORI, Tadashi) [JP/JP]; 〒1088001 東京都港区芝五丁目７番１号 日本電気株式会社内 Tokyo (JP). 大西 祥史 (ONISHI, Yoshifumi) [JP/JP]; 〒1088001 東京都港区芝五丁目７番１号 日本電気株式会社内 Tokyo (JP).

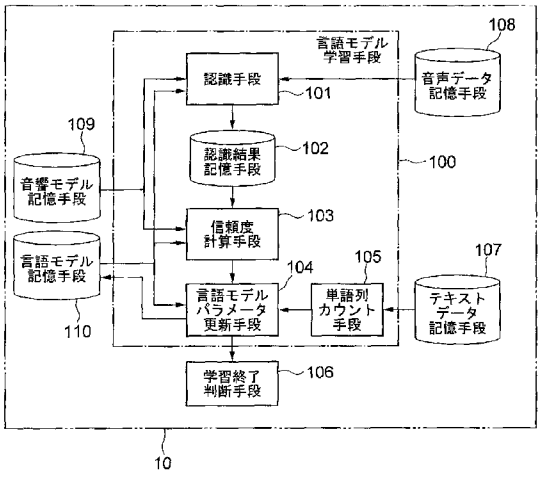
(74) 代理人: 高橋 勇 (TAKAHASHI, Isamu); 〒1010031 東京都千代田区東神田１丁目１０番７号 篠田ビル７階 Tokyo (JP).

(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, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[続葉有]

(54) **Title:** LANGUAGE MODEL LEARNING SYSTEM, LANGUAGE MODEL LEARNING METHOD, AND LANGUAGE MODEL LEARNING PROGRAM

(54) 発明の名称: 言語モデル学習システム、言語モデル学習方法、および言語モデル学習用プログラム



109... SOUND MODEL STORAGE MEANS
110... LANGUAGE MODEL STORAGE MEANS
100... LANGUAGE MODEL LEARNING MEANS
101... RECOGNIZING MEANS
102... RECOGNITION RESULT STORAGE MEANS
103... RELIABILITY DEGREE COMPUTING MEANS
104... LANGUAGE MODEL PARAMETER UPDATING MEANS
105... WORD SEQUENCE COUNTING MEANS
106... LEARNING END JUDGING MEANS
108... SPEECH DATA STORAGE MEANS
107... TEXT DATA STORAGE MEANS

(57) Abstract: A language model learning system for learning a language model on an identifiable basis relating to the word error rate used in speech recognition. The language model learning system (10) comprises recognizing means (101) for recognizing an input speech by using a sound model and a language model and outputting the recognized word sequence as the recognition result, reliability degree computing means (103) for computing the degree of reliability of the word sequence, and language model parameter updating means (104) for updating the parameters of the language model by using the degree of reliability. The language model parameter updating means updates the parameters of the language model to heighten the degree of reliability of the word sequence the computed degree of reliability of which is low when the recognizing means recognizes by using the updated language model and the reliability degree computing means computes the degree of reliability.

[続葉有]

WO 2007/142102 A1

(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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告書
— 補正書

２文字コード及び他の略語については、定期発行される各*PCT*ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

(57) 要約:

音声認識で用いられる単語誤り率などに関係がある識別的な基準で言語モデルを学習できる言語モデル学習システム等を提供する。言語モデル学習システム１０は、入力音声に対し音響モデルと言語モデルを用いて認識し、認識した単語列を認識結果として出力する認識手段１０１と、単語列の信頼度を計算する信頼度計算手段１０３と、信頼度を用いて言語モデルのパラメータを更新する言語モデルパラメータ更新手段１０４とを備え、言語モデルパラメータ更新手段は、信頼度が低く計算された単語列が、認識手段が更新後の言語モデルを用いて認識を行い、信頼度計算手段が信頼度を計算したときに、信頼度が高くなるように言語モデルのパラメータを更新する。

明 細 書

言語モデル学習システム、言語モデル学習方法、および言語モデル学習用プログラム

技術分野

[0001] 本発明は音声認識における言語モデル学習システム、言語モデル学習方法および言語モデル学習用プログラムに関し、識別的な基準を用いて言語モデルの学習を行うため、従来の方法よりも高精度な言語モデルを構築でき、これを音声認識システムに用いることで高精度な音声認識システムを構築できることができる、言語モデル学習システム、言語モデル学習方法、および言語モデル学習用プログラムに関する。

背景技術

[0002] 従来の技術を用いた言語モデルの学習方法を述べる。

[0003] 従来の言語モデルの学習方法においては、たとえば、非特許文献1の57ページから62ページに記載されているように、言語モデルをNグラムモデル(N-gram model)で表している。Nグラムモデル(N-gram model)は、N個の単語からなる単語列の出現確率を、履歴となる(N－1)個の単語の単語列の次にN個目の単語が出現する確率で近似したものである。ここで、単語列が、単数および複数の単語または単語以下の文字列で構成されているとすると、Nグラムモデル(N-gram model)は、大容量のテキストデータである学習コーパス(corpus)があれば、最尤推定によって算出できる。

[0004] このような従来技術で構成される言語モデル学習手段システムの構成を図6に示す。図6によると従来の言語モデル学習システムは、テキストデータ記憶手段107と単語列数カウント手段105と言語モデルパラメータ更新手段301と言語モデル記憶手段110で構成されている。

[0005] 単語列数カウント手段105は、テキストデータ記憶手段107に記憶された学習コーパス(corpus)であるテキストデータからN個の単語からなる単語列を全て抽出し、その単語列の種類ごとに出現回数を計算する。例えば、「of」と「the」との2単語が連鎖した単語列「of the」に対しテキストデータから「of the」が何回出現したのかを計算す

る。

[0006] 言語モデルパラメータ更新手段301は、対象となる単語列の出現回数を全単語列数で割りその単語列の出現確率を計算する。すなわち、単語列「of the」の出現確率は、単語列「of the」の出現回数を2単語連鎖の総数で割ったものに相当する。音声認識の場合、デコードのプロセスで条件付確率を用いる。例えば、「of」の後に「the」の出現する確率をP(the | of)とし、単語列「of the」が出現する同時確率であるP(of, the)とすると、ベイズの定理を用いてP(the | of) =P(of, the) /P(of)と計算することが出来る。ここで、P(of)は、単語「of」が出現する確率を表している。

[0007] 非特許文献1:「言語と計算4:確率的言語モデル」、東京大学出版会、1999年、北研二

発明の開示

発明が解決しようとする課題

[0008] 従来の言語モデル学習システムの第1の問題点は、従来の言語モデル学習方法である最尤推定が、音声認識の評価尺度として使われている単語誤り率等が考慮されていないため、この従来の方法で学習を行った言語モデルに基づいて音声認識を実行しても信頼性の高い認識結果が得られない可能性があるという点である。

[0009] 第2の問題点は、従来の言語モデルの学習方法である最尤学習法が、言語モデルの学習時に音響モデルの影響を考慮していないため、音響モデルと言語モデルとを双方の影響を鑑みながら同時又は逐次的に最適化することができない点である。

[0010] 本発明の目的は、音声認識の評価尺度として用いられる単語誤り率などに関係がある識別的な基準で言語モデルを学習させることにある。また、本発明の他の目的は、音響モデル及び言語モデルの学習を統一された識別的な基準で実行し、言語モデルの学習時に音響モデルの認識性能を考慮し音響モデルの学習時に言語モデルの性能を考慮して音響モデル及び言語モデルの学習を行うことにより、高精度な音声認識を実現するための音響モデル及び言語モデルを構築することにある。

課題を解決するための手段

[0011] 本発明の言語モデル学習システムは、音声認識用の言語モデルを記憶する言語モデル記憶手段と、予め記憶された学習用音声データを言語モデル記憶手段に記

憶された言語モデルを用いて音声認識し認識結果を出力する認識手段と、認識結果における各単語列の信頼度を計算する信頼度計算手段と、信頼度計算手段により算出された各単語列の信頼度に基づいて前記言語モデル記憶手段に記憶された言語モデルのパラメータを更新する前記言語モデルパラメータ更新手段とを備えたことを特徴とする。

[0012] このような言語モデル学習システムによれば、言語パラメータ更新手段が、音声認識の評価に用いられる基準に関係のある識別的な基準にしたがって言語モデルのパラメータを更新することで、言語モデルの学習を実行するので、信頼性の高い言語モデルを構築することができ、高精度な音声認識を実現することができる。

[0013] 上記言語モデル学習システムにおいて、信頼度計算手段により算出される各単語列の信頼度として、認識結果から算出される各単語列の事後確率、各単語列に対応する音声信号の信号対雑音比、各単語列の継続時間と期待継続時間の比のいずれかを用いるか、または、これらを組み合わせた値を用いるようにしても同様に信頼性の高い言語モデルを構築することができる。

[0014] 上記言語モデル学習システムにおいて、学習用音声データに対応する学習用テキストデータ内の全単語列数と各単語列の出現回数とを計算する単語列数カウント手段を備え、言語モデルパラメータ更新手段は、この単語列数カウント手段により算出された全単語列数と各単語列の出現回数とから各単語の出現頻度を計算し、その各単語列の出現頻度と信頼度計算手段により算出された各単語列の信頼度とに基づいて言語モデル記憶手段に記憶された言語モデルを更新してもよい。

[0015] また、言語モデルパラメータ更新手段は、信頼度計算手段により算出された信頼度が最大値でない場合に、これに対応する単語列の出現頻度を大きい値に補正して、その補正された出現頻度に基づいて言語モデル記憶手段に記憶された言語モデルのパラメータを更新してもよい。さらに、学習用テキストデータ内の単語列ω_jの出現回数をN_j、学習用テキストデータに含まれるω_jと同じ単語数の単語列の総数をR、認識結果において観測時系列O_rが観測された場合の単語列ω_jの信頼度を$p(\omega_j \mid O_r)$とし、定数をD、更新前の言語モデルの値をp_jとすると、[数1]式にしたがって単語列ω_jに対応する言語モデルのパラメータP_jを算出し当該パラメータを算出値に更新

してもよい。

[0016] [数1]

$$P_j = \frac{N_j - \sum_r p(\omega_j \mid O_r) + Dp_j}{R - \sum_j \sum_r p(\omega_j \mid O_r) + D}$$

$\sum_r$: 学習音声データ数分の和

$\sum_j$: 全単語列分の和

[0017] 上記言語モデル学習システムにおいて、学習用音声データと初期音響モデルと言語モデルを用いて音響モデルを更新する音響モデル学習手段をさらに備えるようにしても良い。このようにすれば、音響モデル学習手段と言語モデルパラメータ更新手段は、それぞれ音響モデルと言語モデルを統一された識別的な基準で学習する。そのため、音響モデルと言語モデルの両方を同時に最適化することができる。また、音響モデル学習手段は、相互情報量基準を用いて前記音響モデルの学習を行うようにしても良い。

[0018] 本発明の言語モデル学習方法は、予め記憶された言語モデルを用いて学習用音声データを音声認識し認識結果を出力する認識工程と、この認識結果における各単語列の信頼度を計算する信頼度計算工程と、この各単語列の信頼度に基づいて前記言語モデルのパラメータを更新する言語モデルパラメータ更新工程とを含むことを特徴とする。また、言語モデルパラメータ更新工程では、認識結果における各単語列の信頼度が最大になるように言語モデルのパラメータを更新してもよい。

[0019] 上記言語モデル学習方法によれば、音声認識の評価に用いられる基準に関係のある識別的な基準にしたがって言語モデルのパラメータの更新を行うことで、上述した言語モデル学習システムと同様に、信頼性の高い言語モデルを構築することができ、高精度な音声認識を実現することができる。

[0020] 本発明の言語モデル学習プログラムは、予め記憶された言語モデルを用いて学習用音声データを音声認識し認識結果を出力する認識処理と、この認識結果における

各単語列の信頼度を計算する信頼度計算処理と、この各単語列の信頼度を用いて前記言語モデルのパラメータを更新する言語モデルパラメータ更新処理とをコンピュータに実行させることを特徴とする。また、言語モデルパラメータ更新処理を、認識結果における各単語列の信頼度が最大になるように言語モデルのパラメータを更新するという内容に特定してもよい。

[0021] 上記言語モデル学習プログラムによれば、音声認識の評価に用いられる基準に関係のある識別的な基準にしたがって言語モデルパラメータ更新処理をコンピュータに実行させることで、上述した言語モデル学習システムと同様に、信頼性の高い言語モデルを構築することができ、高精度な音声認識を実現することができる。

発明の効果

[0022] 本発明によれば、音声認識の認識結果における各単語列の信頼度、すなわち音声認識の評価に用いられる基準に関係のある識別的な基準に従って言語モデルのパラメータを更新し、言語モデルの学習を実行するので、高精度な音声認識を実現できる信頼性の高い言語モデルを構築することができる。

発明を実施するための最良の形態

[0023] 以下、図を参照しながら本発明の一実施例である言語モデル学習システム10の構成と動作について説明する。

[0024] 図1は、言語モデル学習システム10の構成を示す機能ブロック図である。言語モデル学習システム10は、言語モデル学習手段100とテキストデータ記憶手段107と音響モデル記憶手段109と言語モデル記憶手段110と学習終了判断手段106とを含んでいる。

[0025] 言語モデル学習手段100は、認識手段101と認識結果記憶手段102と信頼度計算手段103と音響モデル記憶手段109と言語モデルパラメータ更新手段104と単語列数カウント手段105とを含んでいる。

[0026] テキストデータ記憶手段107は、言語モデルの学習のための学習用テキストデータを記憶しており、音声データ記憶部108は、言語モデルの学習のための学習用音声データを記憶している。このテキストデータ記憶手段107に記憶されたテキストデータは、音声データ記憶部108に記憶された音声データを書き起こしたものか、あるいは

逆に、音声データがテキストデータを音読したものである。

[0027] 音声データ記憶部108に記憶された学習用音声データは、例えば、アナログの音声信号をサンプリング周波数を44．1kHz、1サンプルあたり16ビットにA／D変換したデータである。

[0028] 音響モデル記憶手段109は、音響モデルを記憶している。この音響モデルは、音声の音響的特長を音素ごとに表現した確率モデルであり、例えば、ケンブリッジ大学で発行されている隠れマルコフモデル（HMM：Hidden Markov Model）のツールキットのマニュアルである「HTKBook for HTK Version.3.3 ヤング等著（以下、「参考文献2」と称する）」の35ページから40ページに記載されているHMMである。

[0029] 言語モデル記憶部110は、言語モデルを記憶している。この言語モデルは、単語が出現する順番を考慮した同時出現確率である。すなわち、単語と単語との言語的なつながり易さを数値化したものである。例えば、N個の単語からなる単語列の言語モデルは、P（w[1]，w[2]，・・・，w[N]）で表される。これは、単語w[1]の次に単語w[2]と続き単語w[N]まで連なる単語列の出現確率を示す。これをベイズのルールで展開すると、P（w[1]，w[2]，・・・，w[N]）＝P（w[1]）P（w[2]｜w[1]）・・P（w[N]｜w[1]，w[2]・・w[N－1]）となる。ただし、Nが大きくなるとP（w[N]｜w[1]，w[2]・・w[N－1]）の履歴となる単語列w[1]，w[2]・・w[N－1]の組み合わせが莫大になり学習できないため、通常の実装では履歴単語の数は3～4とされる。このようなモデルが、Nグラムモデル（N-gram model）である。本実施例では、言語モデルにNグラムモデル（N-gram model）を用いている。

[0030] 認識手段101は、音響モデル記憶手段109に記憶された音響モデルと言語モデル記憶手段110に記憶された言語モデルとを用いて、音声データ記憶手段108に記憶されている学習用音声データを音声認識し認識結果を出力する。

[0031] 認識手段101で実行される音声認識処理は、大きく分けると音響分析と探索に分けられ、音響分析は、音声データの特徴量を計算する処理であり、参考文献1の55ページから66ページに記載されているように、音声データに対しプリエンファシス、窓関数、FFT（Fast Fourier Transform）、フィルターバンク、対数化、コサイン変換の順に計算を行うことでメルケプストラムやパワー、それらの時間変化量を計算する。探索

は、音声データの特徴量と音響モデルとを用いて単語の音響尤度を計算し、音響尤度の高い単語を認識結果として出力する。また、探索において音響尤度のほかに言語モデルも考慮したスコア付けを行う場合も考えられる。

[0032] 認識結果の出力形態は、図3に表されるような単語グラフ形式である。図3(a)における単語グラフは、参考文献1の333ページから337ページに記載されているSLF(HTK Standard Lattice Format)と同様に、円で表されたノード(I1～I5)と棒線で表されたアークとから成り立つ。単語はアークに付随しており、図3(a)においてa～gで示している。実際に認識手段101から出力される単語グラフは、図3(b)のようなテキストで出力され、ノードの時刻と、それぞれのアークの始終端ノードと単語、音響尤度とが出力される。

[0033] 認識結果記憶手段102は、認識手段101から出力された認識結果である単語グラフを記憶する。信頼度計算手段103は、音声の観測時系列O_rに対し単語列ω_jが認識できたかどうかを表す値である信頼度を認識結果を基に計算する。信頼度は、音響モデルと言語モデルとがともに高精度に構築できた場合、正解単語列に対して1に近づき、不正解単語に対して0に近づく。

[0034] 単語列カウント手段105は、テキストデータ記憶手段107に記憶されているテキストデータから単語列を抽出し、単語列の種類ごとにその出現回数を計算する。例えば、「of」と「the」とが連鎖した単語列「of the」が、学習用テキストデータ内に何回出現したのかを計算する。

[0035] 言語モデルパラメータ更新手段104は、[数1]式を用いて言語モデルのパラメータを更新する。

[0036] [数1]

$$P_j = \frac{N_j - \sum_r p(\omega_j \mid O_r) + Dp_j}{R - \sum_j \sum_r p(\omega_j \mid O_r) + D}$$

$\sum_r$: 学習音声データ数分の和

$\sum_j$: 全単語列分の和

[0037] [数1]式において、N_jは学習用テキストデータ内に単語列ω_jが出現した数を示し、Rは学習用テキストデータに含まれるω_jと同じ単語数の単語列の総数を示し、Dは定数であり、p_jは更新前の言語モデルの値であり、$p(\omega_j \mid O_r)$は認識結果において観測時系列O_rが観測された場合の単語列ω_jの信頼度を示している。

[0038] [数1]式の$p(\omega_j \mid O_r)$には言語モデルパラメータの更新における寄与度を表すパラメータを指定することができ、その場合は、$p(\omega_j \mid O_r)$の前にパラメータをかけるか、べき乗のパラメータとすることができる。また、[数1]式の定数Dは、推定値の収束具合によって実験的に値を決めることができる。

[0039] ここで、この信頼度を統計的な観点で計算したものが単語事後確率である。単語事後確率は、「Frank Wessel,Ralf Schluter,Kalus Macherey,ans Herman Ney,''Confidence Measures for Large Vocabulary Continuos Speech Recognition,''IEEE Trans. on Speech and Audio Processing. Vol 9, No.3,March 2001(以下、「参考文献2」と称する)」に記載されている方法を用いて計算することができる。

[0040] ここで、参考文献2に従い、図3に示す認識結果に基づく単語cの事後確率の計算方法を説明する。認識結果に基づく単語cの事後確率を計算するためには、単語cの前向き確率αと後ろ向き確率βとを求める必要があり、言語モデルを3単語連鎖確率(tri-gram model)とした場合、前向き確率αは[数2]式で表される。

[0041] [数2]

$$\alpha(\mathrm{a};\mathrm{c}) = P_A(o_c \mid \mathrm{c}) \sum_{\mathrm{z}} \alpha(\mathrm{z};\mathrm{a}) P_L(\mathrm{c} \mid \mathrm{az})$$

$\sum_{\mathrm{z}}$: z ∈ a の始端の全単語

[0042] ここで、o_cは単語cの特徴量であり、全区間の特徴量を表す場合はOとする。$P_A(o_c \mid c)$は単語cの音響尤度、$P_L(c \mid az)$は単語z→a→cの順で構成される単語列の出現確率を表している。[数2]式に示すように、単語cの前向き確率αは、単語aの始端につながる全ての単語の前向き確率と言語確率との積を全て足し合わせたものになっている。単語c以外の単語の前向き確率を算出する場合、算出対象の単語より前の時刻に出現した単語の前向き確率を求めておくことで、対象の前向き確率を算出することができる。

[0043] 後ろ向き確率βは[数3]式で表される。

[0044] [数3]

$$\beta(c;e) = P_A(o_c \mid c)\sum_{z'} \beta(e;z')P_L(z' \mid ce)$$

z'∈eの終端に接続される全単語

[0045] [数3]式に示すように、単語cの後ろ向き確率βは、[数2]式で示す前向き確率αに比べて、cとeとz'等の関係が前後逆になっている。

[0046] 認識結果における単語cの事後確率$P(c \mid o_c)$は、[数2]及び[数3]を用いて[数4]で表される。

[0047] [数4]

$$P(c \mid o_c) = \frac{\sum_{z}\sum_{z'} \alpha(z;c)\beta(c;z')P_L(z' \mid zc)}{P_A(O)P_A(o_c \mid c)}$$

z∈aの始端の全単語
z'∈eの終端に接続される全単語

[0048] ここで、Σのzは、単語aの始端に接続された全単語の総和、z'は単語eの終端に接続された全単語の総和を表す。$P_A(O)$は、全ての観測時系列Oの音響尤度であり[数5]式で表わされる。

[0049] [数5]

$$P_A(O)=\sum_{z}\sum_{z'}\alpha(z;z')$$

z ∈ a の始端の全単語
z'∈ e の終端に接続される全単語

[0050] ここで、事後確率の計算方法の定義を見てみると、事後確率は単語ごとに求められることがわかる。認識結果における単語cの事後確率P(c | o_c)は、単語cが同じ区間の単語dまたはh等(図3参照)と比べて観測時系列O_cとどの程度マッチしたかを示す値で、0～1の値に正規化されている。単語cの事後確率は、単語cが2つの単語で構成されていても計算可能である。

[0051] [数2]、[数3]、[数4]においては、音響モデル及び言語モデルの寄与度を表すパラメータを設定することが可能で、そのときは、$P_A(o_c \mid c)^y$や$P_L(c \mid \alpha z)^x$のようにべき乗のパラメータを設定する。

[0052] $p(\omega_j \mid O_r)$を認識結果に基づく単語列ω_jの事後確率とした場合、[数1]は学習後の音声認識に対して単語列の事後確率を最大にするパラメータを推定する基準から得られたものであり、この基準は音響モデルの識別的な推定方法にも使われている。音響モデルの学習については第2実施例にて説明する。

[0053] [数1]式を用いて言語モデルのパラメータを更新する場合、$p(\omega_j \mid O_r)$は、観測時系列O_rに対する単語列ω_jの信頼度であり、[数1]式は、学習用テキストデータにおける単語列ω_jの出現頻度から認識結果における信頼度の総和を引く定式になっている。これは、総合的に信頼度が高い単語列の場合、出現頻度から引かれる数が大きくなるため、更新後の言語モデルのパラメータは小さくなる。また、信頼度が低い単語列の場合、出現頻度から引かれる数が小さくなるため、言語モデルのパラメータは大きくなる。ここで、「信頼度が高い」とは信頼度が1の場合であり、「信頼度が低い」とは信頼度が1以外の場合である。

[0054] 信頼度に事後確率を用いる場合、言語モデルのパラメータ更新は、認識手段101の認識性能に依存することになる。

[0055] また、本実施例においては、信頼度に事後確率を用いて言語モデルのパラメータを更新したが、前記の性質を満たす尺度であれば、信頼度にどのようなものを用いて

もよく、例えば、単語列ごとの音声信号の信号雑音比（SNR：signal-to-noise ratio）や、単語列の継続時間と期待継続時間との比などを信頼度としてもよい。

[0056] また、対象単語列の音声信号の信号対雑音比、対象単語列の継続時間と期待継続時間との比、認識結果に基づく対象単語列の事後確率とを組み合わせて信頼度として用いても良い。例えば、[数1]式の右辺の分母と分子の$p(\omega_j \mid O_r)$をそれぞれ次の[数6]式により算出される$p'(\omega_j \mid O_r)$に置き換えてもよい。

[0057] [数6]

$$p'(\omega_j \mid O_r) = Ap(\omega_j \mid O_r) + B(SNR) + C(\text{継続時間と期待継続時間の比})$$

A、B、Cは係数

[0058] 学習終了判断手段106は、言語モデルの更新後、全音声データの事後確率を計算し、その和SUM[t]をとる。その後、SUM[t]から言語モデルを更新する前の単語事後確率の総和SUM[t－1]を差し引いた値をSUM[t]で割ったものを学習進捗係数T_pとする。学習進捗係数T_pが、予め定められた閾値を超えている場合は、言語モデルの学習をやり直し、閾値を下回る場合は言語モデルの学習を終了する。

[0059] 図2は、言語モデル学習システム10の動作を示すフローチャートである。

[0060] Step1にて、認識手段101が音響モデル記憶手段109に記憶された音響モデルと言語モデル記憶手段110に記憶された言語モデルとを用いて、音声データ記憶手段108に記憶されている学習用音声データを音声認識し認識結果を認識結果記憶手段102へ出力する。ここで用いられる音響モデルや言語モデルは前述の形式であれば、そのパラメータ値がどのような学習方法で学習されたものでもよい、さらに全くの乱数でも良い。また、出力される認識結果は、単語グラフとする。

[0061] Step2にて、信頼度計算手段103が、認識結果記憶手段102に記憶された認識結果と言語モデル記憶手段110に記憶された言語モデルとを用いて各単語列の事後確率を計算する。この計算する動作は、認識手段101による認識結果全てに対して行われる。

[0062] Step3にて、単語列数カウント手段105が、テキストデータ記憶手段に記憶された

学習用テキストデータから対象となる単語列の数をカウントする。

[0063] Step4にて、言語モデルパラメータ更新手段104が、信頼度計算手段103により算出された単語列の事後確率と、単語列数カウント手段105によりカウントされた数値とを[数1]式に代入して言語モデルの確率値を算出し更新する。ここで更新された言語モデルは、それを用いて音声認識を行うことが可能なものである。

[0064] Step5にて、学習終了判断手段106が、言語モデルパラメータ更新手段104により更新された言語モデルのパラメータを用いて学習データ全てに対する単語事後確率を計算し、それを元に学習進捗係数T_pが閾値を下回っている場合は、言語モデル学習システム10の動作を終了し、学習進捗係数T_pが閾値を上回っている場合はStep1に戻る。

[0065] このような言語モデル学習システム10によれば、言語モデルパラメータ更新手段104が、認識結果における単語列の信頼度、すなわち音声認識の評価に用いられる基準に関係のある識別的な基準により言語モデルのパラメータの更新を行うことにより、言語モデルの学習を実行する。そのため、高精度な音声認識を実現するための言語モデルを構築することができる。

[0066] 次に、本発明の第2の実施例である言語モデル学習システム20について図面を参照して詳細に説明する。ここで、言語モデル学習システム20は、多くの構成が図1の言語モデル学習システム10と共通するので、共通する構成要素には図1と同一の符号を付して説明を省略する。

[0067] 図4は、言語モデル学習システム20の構成を示す機能ブロック図である。言語モデル学習システム20は、図1に開示した言語モデル学習システム10の構成に加えて、音響モデル学習手段200を含んでいる。音響モデル学習手段200は、音声データ記憶手段108に記憶された学習用音声データと、音響モデル記憶手段109に記憶された音響モデルと、言語モデル記憶手段110に記憶された言語モデルとを用いてこの音響モデルの学習を行う。

[0068] 音響モデル学習手段200が実行する音響モデルの学習方法としては、例えば、スピーチコミュニケーションの1997年のボリューム22の303ページから314ページに記載されている「大語彙認識のMMIE学習 V. Veltchev, J.J. Odell, P.C. Woodland

, S.J. Yang, "MMIE training of large vocabulary recognition systems,"Speech Communication vol.22, 303-314, 1997(以下、これを「参考文献3」と称する)」に記載されているような、相互情報量基準による推定を用いる。相互情報量基準による音響モデルの学習について、参考文献3の308ページから309ページを基に説明する。

[0069] 音響モデル学習手段200は、まず、音響モデルと言語モデルとを用いて音声データ記憶手段108に記憶された学習用音声データを音声認識する。この認識結果は、単語グラフで出力され、認識結果に基づく各単語列の事後確率を計算する。単語内の音素や状態のセグメンテーションを計算する必要があるが、その計算をビタービアルゴリズムで計算する。音素セグメンテーションの計算後、状態ごとの十分統計量を計算する。十分統計量の計算時には音素や状態ごとの事後確率を計算する必要があるが、参考文献3では単語の事後確率を用いている。十分統計量の計算は、認識結果に対してだけでなく、正解の文字列に対しても同様に実行される。認識結果と認識結果に対する十分統計量を用いて、参考文献3の305ページに記載されている式(4)と式(5)と306ページに記載されている式(8)に適用して音響モデルのパラメータを更新する。

[0070] 図5は、言語モデル学習システム20の動作を示すフローチャートである。

[0071] Step101では、音響モデル学習手段200が、音響モデル記憶手段109に記憶された音響モデルと言語モデル記憶手段110に記憶された言語モデルと音声データ記憶手段108に記憶されている音声データとを用いて音響モデルの学習を実行する。音響モデルの学習は前述の相互情報量を用いた学習のほかに参考文献2の6ページから8ページに記載されているBaum=Welchアルゴリズムによる最尤基準による方法も考えられる。音響モデルの学習後、音響モデル記憶手段109に記憶された音響モデルを更新し、Step102の処理へ移る。

[0072] Step102では、Step101で更新された音響モデルと言語モデル記憶手段110に記憶された言語モデルと、音声データ記憶手段108に記憶された学習用音声データと、テキストデータ記憶手段107に記憶された学習用テキストデータとを用いて、実施例1と同様に、言語モデルのパラメータの更新を行う。

[0073] Step103では、学習終了判断手段106が、実施例1と同様に、言語モデルパラメ

ータ更新後の認識結果に基づく各単語列の事後確率の総和SUM[t]から更新前の総和SUM[t－1]を差し引いた値を、SUM[t]で割ったものを学習進捗係数Tpとし、学習進捗係数Tpが予め定められた閾値を超えている場合はStep101から学習をやり直し、閾値を下回る場合は言語モデルの学習を終了する。

[0074] ここで、第1の実施例のSUM[t]と第2の実施例のSUM[t]との違いは、第1の実施例では音響モデルを更新していないが、第2の実施例では音響モデルを更新している点である。また、参考文献3の305ページに記載されている式(4)と式(5)と306ページに記載されている式(8)は、導出元になる式が上述した[数1]式と同じである。

[0075] このように本第2実施例の言語モデル学習システム20は、音響モデル学習手段200を含み、音響モデルと言語モデルとを統一された識別的な基準で学習する。そのため、音響モデルと言語モデルの両方を同時に最適化することができ、高精度な音声認識を実現するための音響モデル及び言語モデルを構築することができる。

図面の簡単な説明

[0076] [図1]本発明の第1の実施例である言語モデル学習システムの構成を示すブロック図である。

[図2]図1に開示した言語モデル学習システムの動作を示す流れ図である。

[図3]図1に開示した認識手段から出力される認識結果である単語グラフの一例を説明するための図である。

[図4]本発明の第2の実施例である言語モデル学習システムの構成を示すブロック図である。

[図5]図4に開示した言語モデル学習システムの動作を示す流れ図である。

[図6]従来の技術で構成される言語モデル学習システムを示すブロック図である。

符号の説明

[0077] 10、20 言語モデル学習システム

100 言語モデル学習手段

101 認識手段

102 認識結果記憶手段

103 信頼度計算手段

104 言語モデルパラメータ更新手段

105 単語列数カウント手段

106 学習終了判断手段

107 テキストデータ記憶手段

108 音声データ記憶手段

109 音響モデル記憶手段

110 言語モデル記憶手段

200 音響モデル学習手段

請求の範囲

[1] 音声認識用の言語モデルを記憶する言語モデル記憶手段と、予め記憶された学習用音声データを前記言語モデル記憶手段に記憶された言語モデルを用いて音声認識し認識結果を出力する認識手段と、前記認識結果における単語列それぞれの信頼度を計算する信頼度計算手段と、前記信頼度計算手段により算出された各単語列の信頼度に基づいて前記言語モデル記憶手段に記憶された言語モデルのパラメータを更新する前記言語モデルパラメータ更新手段とを備えたことを特徴とする言語モデル学習システム。

[2] 前記言語モデルパラメータ更新手段が、前記信頼度計算手段により算出される各単語列の信頼度が最大になるように前記言語モデル記憶手段に記憶された言語モデルのパラメータを更新することを特徴とする請求項1に記載の言語モデル学習システム。

[3] 前記信頼度計算手段は、前記認識結果に基づく各単語列の事後確率をその単語列の信頼度として計算することを特徴とする請求項1または2に記載の言語モデル学習システム。

[4] 前記信頼度計算手段は、前記認識結果に基づく各単語列の音声信号の信号対雑音比をその単語列の信頼度として計算することを特徴とする請求項1または2に記載の言語モデル学習システム。

[5] 前記信頼度計算手段は、前記認識結果に基づく各単語列の継続時間と期待継続時間との比をその単語列の信頼度として計算することを特徴とする請求項1または2に記載の言語モデル学習システム。

[6] 前記信頼度計算手段は、前記認識結果に基づく各単語列の事後確率と、その単語列の音声信号の信号対雑音比と、その単語列の継続時間と期待継続時間との比とを組み合わせた値を、この単語列の信頼度として計算することを特徴とする請求項1または2に記載の言語モデル学習システム。

[7] 前記学習用音声データに対応する学習用テキストデータ内の全単語列数と各単語列の出現回数とを計算する単語列数カウント手段を備え、前記言語モデルパラメータ更新手段が、前記単語列数カウント手段により算出された全単語列数と各単語列

の出現回数とから各単語の出現頻度を計算し、その各単語列の出現頻度と前記信頼度計算手段により算出された各単語列の信頼度とに基づいて前記言語モデル記憶手段に記憶された言語モデルのパラメータを更新することを特徴とする請求項1ないし請求項6のいずれかひとつに記載の言語モデル学習システム。

[8] 前記言語モデルパラメータ更新手段は、前記信頼度計算手段により算出された信頼度が最大値でない場合に、これに対応する単語列の前記出現頻度を大きい値に補正して、その補正された出現頻度に基づいて前記言語モデル記憶手段に記憶された言語モデルのパラメータを更新することを特徴とする請求項7に記載の言語モデル学習システム。

[9] 前記言語モデルパラメータ更新手段は、前記学習用テキストデータ内の単語列ω_jの出現回数をN_j、前記学習用テキストデータに含まれるω_jと同じ単語数の単語列の総数をR、前記認識結果において観測時系列O_rが観測された場合の単語列ω_jの信頼度を$p(\omega_j \mid O_r)$とし、定数をD、更新前の言語モデルの値をp_jとすると、[数1]式にしたがって単語列ω_jに対応する言語モデルのパラメータP_jを算出し当該パラメータを算出した値に更新することを特徴とする請求項7または8に記載の言語モデル学習システム。

[数1]

$$P_j = \frac{N_j - \sum_r p(\omega_j \mid O_r) + Dp_j}{R - \sum_j \sum_r p(\omega_j \mid O_r) + D}$$

$\sum_r$: 学習音声データ数分の和

$\sum_j$: 全単語列分の和

[10] 音声認識用の音響モデルを記憶する音響モデル記憶手段と、この音響モデルと前記学習用音声データと前記言語モデル記憶手段に記憶された言語モデルとに基づいて前記音響モデル記憶手段に記憶された音響モデルを更新する音響モデル学習手段とをさらに備えたことを特徴とする請求項1ないし請求項9のいずれかひとつに

記載の言語モデル学習システム。

[11] 前記音響モデル学習手段が、相互情報量基準を用いて前記音響モデル記憶手段に記憶された音響モデルを更新することを特徴とする請求項10に記載の言語モデル学習システム。

[12] 予め記憶された言語モデルを用いて学習用音声データを音声認識し認識結果を出力する認識工程と、

前記認識結果における各単語列の信頼度を計算する信頼度計算工程と、

前記認識結果における各単語列の信頼度に基づいて前記言語モデルのパラメータを更新する言語モデルパラメータ更新工程とを含むことを特徴とする言語モデル学習方法。

[13] 前記言語モデルパラメータ更新工程では、前記信頼度計算工程で算出される各単語列の信頼度が最大になるように前記言語モデルのパラメータを更新することを特徴とする請求項1に記載の言語モデル学習方法。

[14] 前記信頼度計算工程では、前記認識結果に基づく各単語列の事後確率をその単語列の信頼度として算出することを特徴とする請求項12または13に記載の言語モデル学習方法。

[15] 前記信頼度計算工程では、前記認識結果に基づく各単語列の音声信号の信号対雑音比をその単語列の信頼度として算出することを特徴とする請求項12または13に記載の言語モデル学習方法。

[16] 前記信頼度計算工程では、前記認識結果に基づく各単語列の継続時間と期待継続時間との比をその単語列の信頼度として算出することを特徴とする請求項12または13に記載の言語モデル学習方法。

[17] 前記信頼度計算工程では、前記認識結果に基づく各単語列の事後確率と、その単語列の音声信号の信号対雑音比と、その単語列の継続時間と期待継続時間との比とを組み合わせた値を、この単語列の信頼度として算出することを特徴とする請求項12または13に記載の言語モデル学習方法。

[18] 前記言語モデルのパラメータ更新工程では、前記学習用音声データに対応する学習用テキストデータ内の各単語列の出現頻度と前記信頼度計算工程で算出された

各単語列の信頼度とに基づいて前記言語モデルのパラメータを更新することを特徴とする請求項12ないし請求項17のいずれかひとつに記載の言語モデル学習方法。

[19] 前記言語モデルパラメータ更新工程では、前記信頼度計算工程で算出された信頼度が最大値でない場合に、これに対応する単語列の前記出現頻度を大きい値に補正して、その補正された出現頻度に基づいて前記言語モデルのパラメータを更新することを特徴とする請求項18に記載の言語モデル学習方法。

[20] 前記言語モデルパラメータ更新手段は、前記学習用テキストデータ内の単語列ω_jの出現回数をN_j、前記学習用テキストデータに含まれるω_jと同じ単語数の単語列の総数をR、前記認識結果において観測時系列O_rが観測された場合の単語列ω_jの信頼度を$p(\omega_j \mid O_r)$とし、定数をD、更新前の言語モデルの値をp_jとすると、[数1]式にしたがって単語列ω_jに対応する言語モデルのパラメータP_jを算出し当該パラメータを更新することを特徴とする請求項18または19に記載の言語モデル学習方法。

[数1]

$$P_j = \frac{N_j - \sum_r p(\omega_j \mid O_r) + Dp_j}{R - \sum_j \sum_r p(\omega_j \mid O_r) + D}$$

$\sum_r$: 学習音声データ数分の和

$\sum_j$: 全単語列分の和

[21] 予め記憶された音響モデルと前記言語モデルと前記学習用音声データとを用いて当該音響モデルを更新する音響モデル学習工程を含むことを特徴とする請求項12ないし請求項20のいずれかひとつに記載の言語モデル学習方法。

[22] 前記音響モデル学習工程では、相互情報量基準を用いて前記音響モデルを更新することを特徴とする請求項21に記載の言語モデル学習方法。

[23] 予め記憶された言語モデルを用いて学習用音声データを音声認識し認識結果を出力する認識処理と、

前記認識結果における各単語列の信頼度を計算する信頼度計算処理と、

前記認識結果における各単語列の信頼度を用いて前記言語モデルのパラメータを更新する言語モデルパラメータ更新処理とをコンピュータに実行させることを特徴とする言語モデル学習プログラム。

[24] 前記言語モデルパラメータ更新処理を、前記認識結果における各単語列の信頼度が最大になるように前記言語モデルのパラメータを更新するという内容に特定したことを特徴とする請求項23に記載の言語モデル学習プログラム。

[25] 前記信頼度計算処理を、前記認識結果に基づく各単語列の事後確率をその単語列の信頼度として算出するという内容に特定したことを特徴とする請求項23または24に記載の言語モデル学習プログラム。

[26] 前記信頼度計算処理を、前記認識結果に基づく各単語列の音声信号の信号対雑音比をその単語列の信頼度として算出するという内容に特定したことを特徴とする請求項23または24に記載の言語モデル学習プログラム。

[27] 前記信頼度計算処理を、前記認識結果に基づく各単語列の継続時間と期待継続時間との比をその単語列の信頼度として算出するという内容に特定したことを特徴とする請求項23または24に記載の言語モデル学習プログラム。

[28] 前記信頼度計算処理を、前記認識結果に基づく各単語列の事後確率と、その単語列の音声信号の信号対雑音比と、その単語列の継続時間と期待継続時間との比とを組み合わせた値を、この単語列の信頼度として算出するという内容に特定したことを特徴とする請求項23または24に記載の言語モデル学習プログラム。

[29] 前記言語モデルパラメータ更新処理を、前記学習用音声データに対応する学習用テキストデータ内の各単語列の出現頻度と前記信頼度計算工程で算出された各単語列の信頼度とに基づいて前記言語モデルのパラメータを更新するという内容に特定したことを特徴とする請求項23乃至28にいずれかひとつに記載の言語モデル学習プログラム。

[30] 前記言語モデルパラメータ更新処理を、前記信頼度計算処理で算出された信頼度が最大値でない場合に、これに対応する単語列の前記出現頻度を大きい値に補正して、その補正された出現頻度に基づいて前記言語モデルのパラメータを更新するという内容に特定したことを特徴とする請求項29に記載の言語モデル学習プログラ

ム。

[31] 前記言語モデルパラメータ更新処理を、前記学習用テキストデータ内の単語列ω_jの出現回数をN_j、前記学習用テキストデータに含まれるω_jと同じ単語数の単語列の総数をR、前記認識結果において観測時系列O_rが観測された場合の単語列ω_jの信頼度を$p(\omega_j \mid O_r)$とし、定数をD、更新前の言語モデルの値をp_jとすると、[数1]式にしたがって単語列ω_jに対応する言語モデルのパラメータP_jを算出し当該パラメータを更新するという内容に特定したことを特徴とする請求項29または30に記載の言語モデル学習プログラム。

[数1]

$$P_j = \frac{N_j - \sum_r p(\omega_j \mid O_r) + Dp_j}{R - \sum_j \sum_r p(\omega_j \mid O_r) + D}$$

$\sum_r$:学習音声データ数分の和

$\sum_j$:全単語列分の和

[32] 予め記憶された音響モデルと前記言語モデルと前記学習用音声データとを用いて当該音響モデルを更新する音響モデル学習処理を前記コンピュータに実行させることを請求項23乃至31のいずれかひとつに記載の言語モデル学習プログラム。

[33] 前記音響モデル学習処理を、相互情報量基準を用いて前記音響モデルを更新するという内容に特定したことを特徴とする請求項32に記載の言語モデル学習プログラム。

補正書の請求の範囲

[2007年10月12日（12.10.2007）国際事務局受理]

記載の言語モデル学習システム。

[11] 前記音響モデル学習手段が、相互情報量基準を用いて前記音響モデル記憶手段に記憶された音響モデルを更新することを特徴とする請求項10に記載の言語モデル学習システム。

[12] 予め記憶された言語モデルを用いて学習用音声データを音声認識し認識結果を出力する認識工程と、

前記認識結果における各単語列の信頼度を計算する信頼度計算工程と、

前記認識結果における各単語列の信頼度に基づいて前記言語モデルのパラメータを更新する言語モデルパラメータ更新工程とを含むことを特徴とする言語モデル学習方法。

[13] （補正後） 前記言語モデルパラメータ更新工程では、前記信頼度計算工程で算出される各単語列の信頼度が最大になるように前記言語モデルのパラメータを更新することを特徴とする請求項12に記載の言語モデル学習方法。

[14] 前記信頼度計算工程では、前記認識結果に基づく各単語列の事後確率をその単語列の信頼度として算出することを特徴とする請求項12または13に記載の言語モデル学習方法。

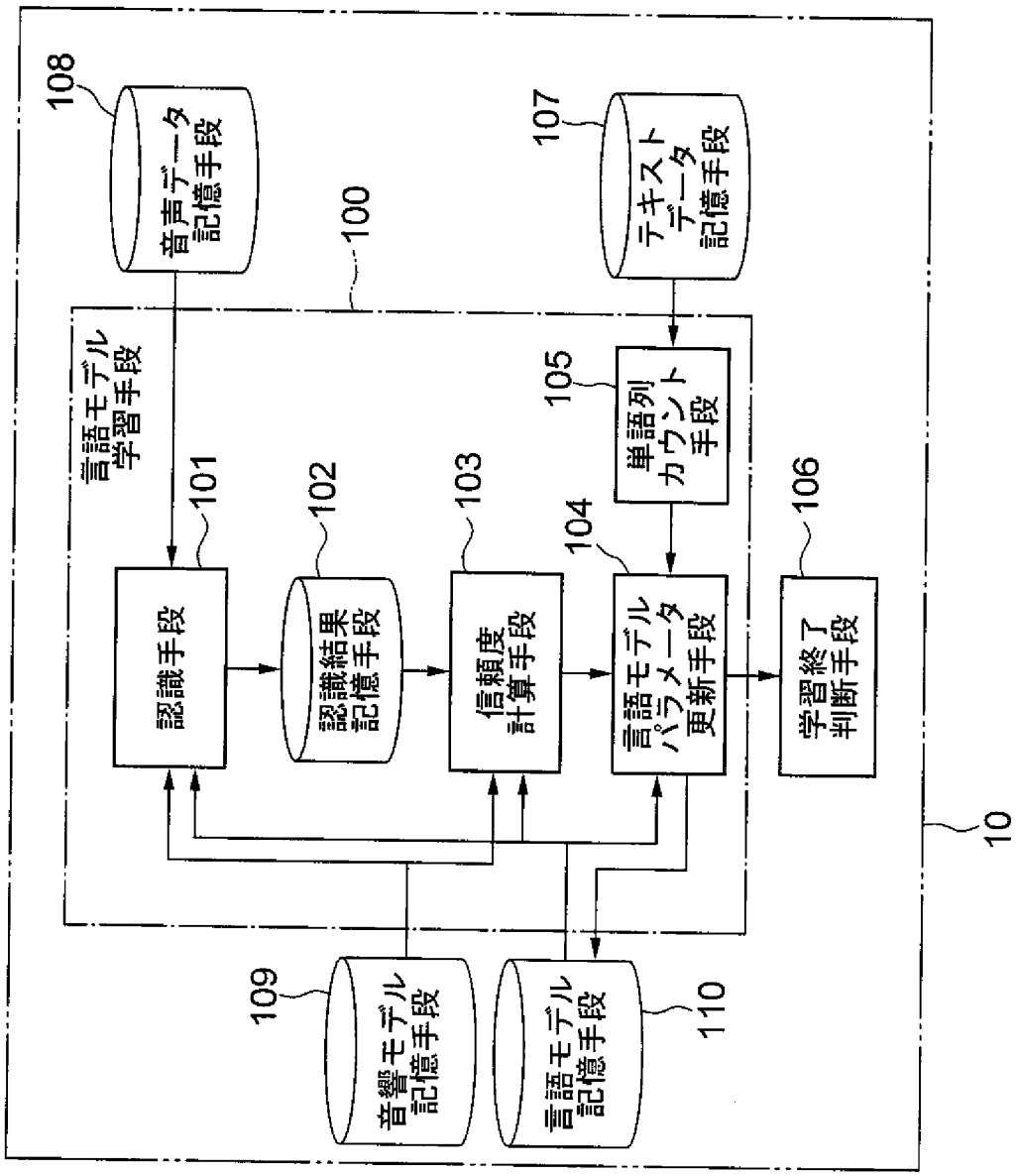
[15] 前記信頼度計算工程では、前記認識結果に基づく各単語列の音声信号の信号対雑音比をその単語列の信頼度として算出することを特徴とする請求項12または13に記載の言語モデル学習方法。

[16] 前記信頼度計算工程では、前記認識結果に基づく各単語列の継続時間と期待継続時間との比をその単語列の信頼度として算出することを特徴とする請求項12または13に記載の言語モデル学習方法。

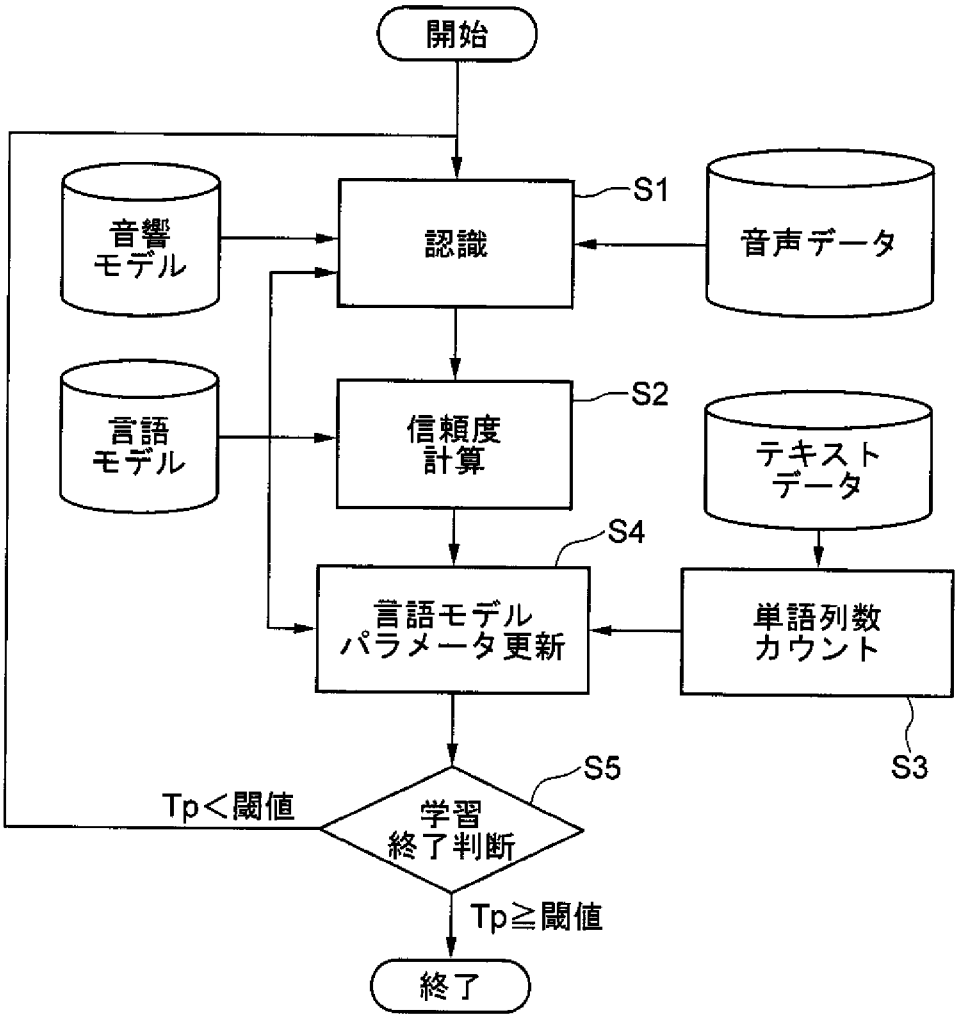
[17] 前記信頼度計算工程では、前記認識結果に基づく各単語列の事後確率と、その単語列の音声信号の信号対雑音比と、その単語列の継続時間と期待継続時間との比とを組み合わせた値を、この単語列の信頼度として算出することを特徴とする請求項12または13に記載の言語モデル学習方法。

[18] 前記言語モデルのパラメータ更新工程では、前記学習用音声データに対応する学習用テキストデータ内の各単語列の出現頻度と前記信頼度計算工程で算出された

[図1]

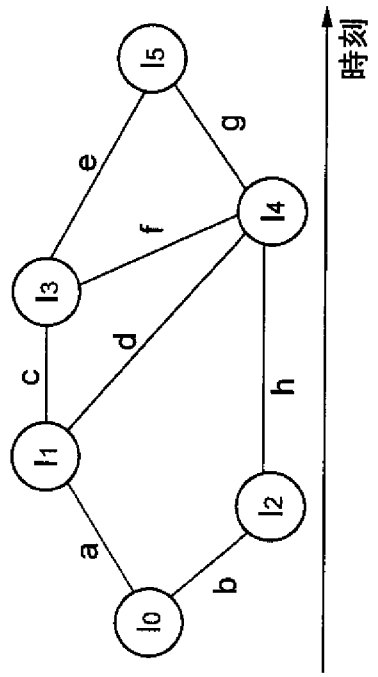

108
音声データ記憶手段
107
テキストデータ記憶手段
100
言語モデル学習手段
105
単語列カウント手段
101
認識手段
102
認識結果記憶手段
103
信頼度計算手段
104
言語モデルパラメータ更新手段
106
学習終了判断手段
10
109
音響モデル記憶手段
110
言語モデル記憶手段

[図2]


開始
音響
モデル
認識
S1
音声データ
言語
モデル
信頼度
計算
S2
テキスト
データ
S4
言語モデル
パラメータ更新
単語列数
カウント
S3
S5
Tp<閾値
学習
終了判断
Tp≧閾値
終了

[図3]

(a)

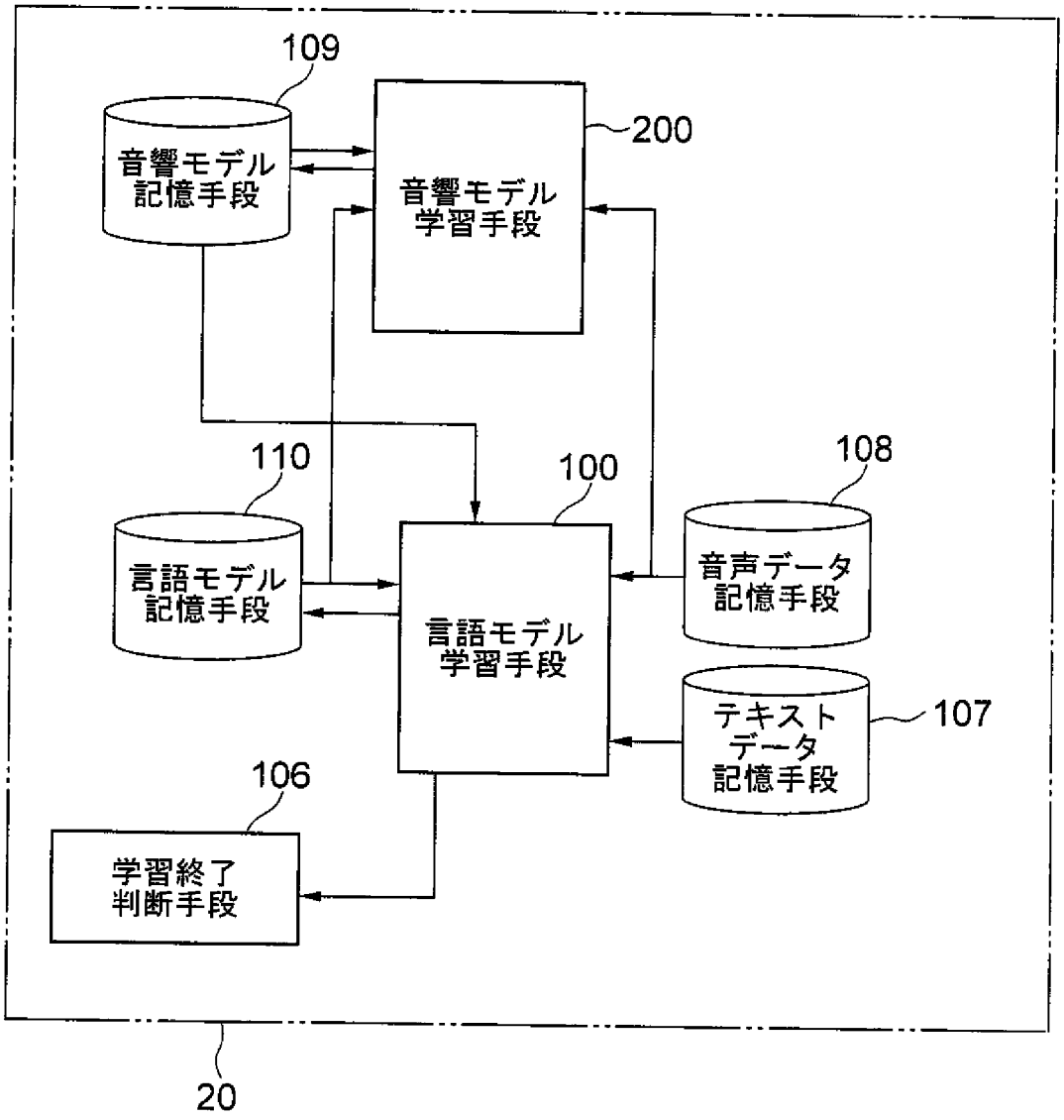


(b)

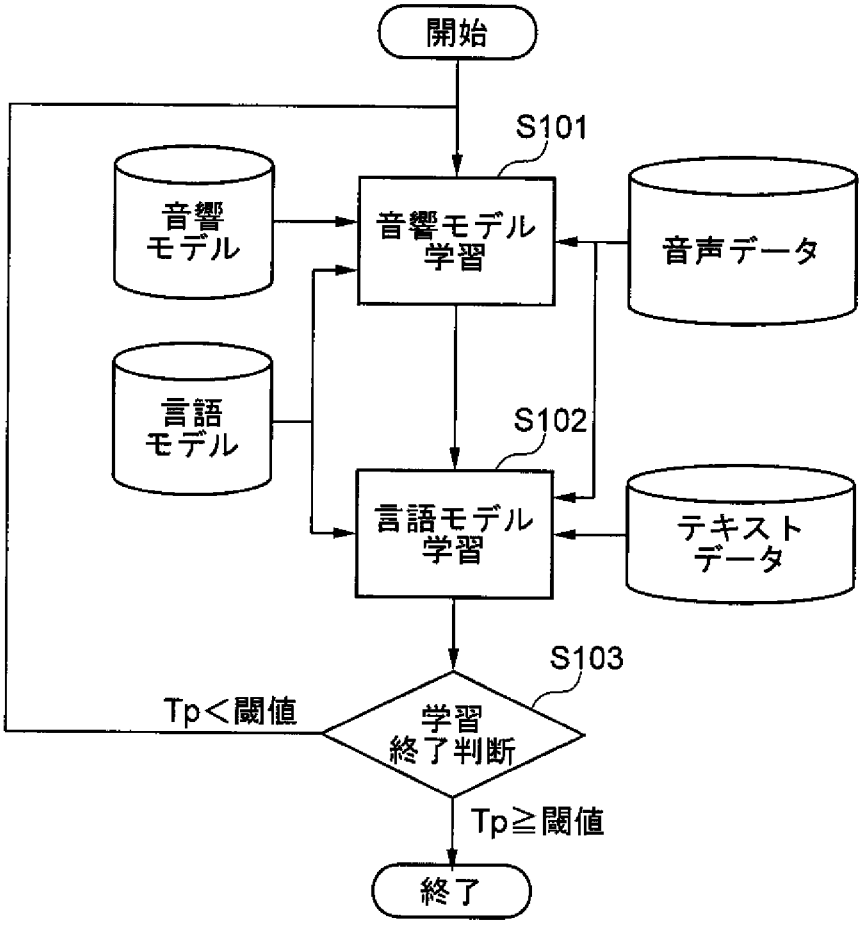
ノード	時刻
l_0	= 1.0秒
l_1	= 2.0秒
l_2	= 2.5秒
l_3	= 3.0秒
l_4	= 3.5秒
l_5	= 4.0秒

始端	終端	単語	音響尤度
l_0	l_1	a	-10.0
l_0	l_2	b	-12.7
l_1	l_3	c	-15.1
l_1	l_4	d	-20.4
l_2	l_4	h	-18.5
l_3	l_4	f	-13.6
l_3	l_5	e	-16.0
l_4	l_5	g	-10.2

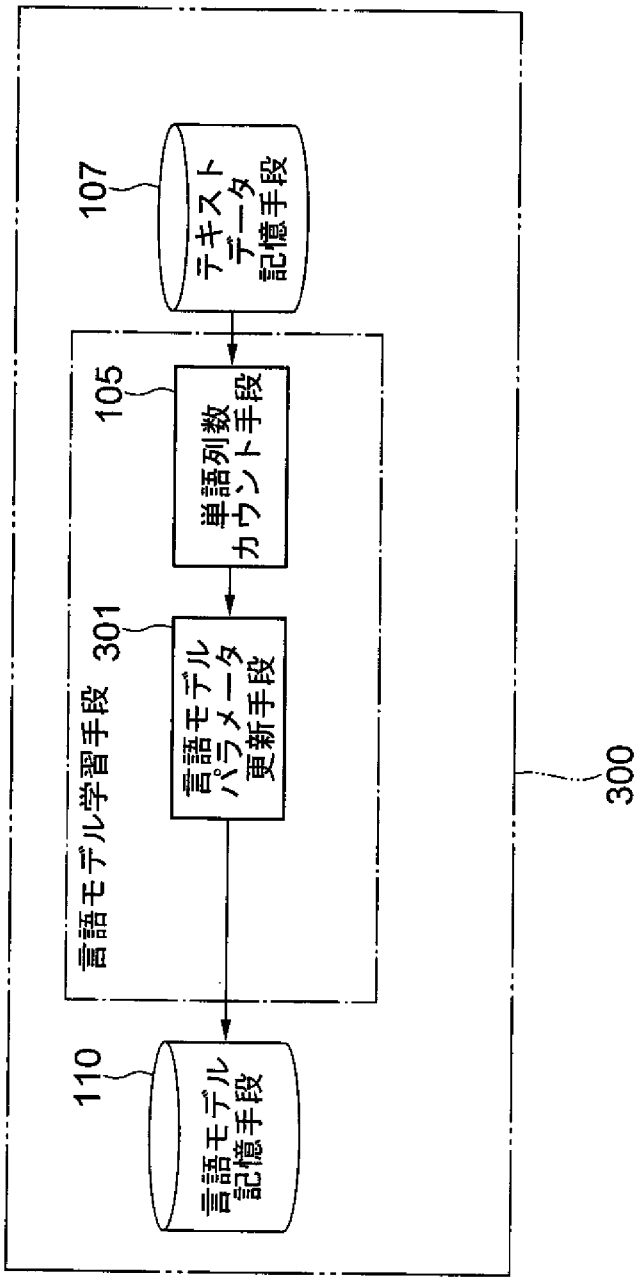
[図4]


109
音響モデル
記憶手段
200
音響モデル
学習手段
110
100
108
言語モデル
記憶手段
言語モデル
学習手段
音声データ
記憶手段
テキスト
データ
記憶手段
107
106
学習終了
判断手段
20

[図5]


開始
S101
音響
モデル
音響モデル
学習
音声データ
言語
モデル
S102
言語モデル
学習
テキスト
データ
S103
Tp＜閾値
学習
終了判断
Tp≧閾値
終了

[図6]


107
テキストデータ記憶手段
105
単語列数カウント手段
301
言語モデルパラメータ更新手段
言語モデル学習手段
110
言語モデル記憶手段
300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2007/061023

A. CLASSIFICATION OF SUBJECT MATTER
G10L15/06(2006.01)i, *G10L15/02*(2006.01)i, *G10L15/18*(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L15/00-17/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2007
Kokai Jitsuyo Shinan Koho 1971-2007 Toroku Jitsuyo Shinan Koho 1994-2007

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2001-109491 A (Waseda University), 20 April, 2001 (20.04.01), Full text; Figs. 1 to 3 (Family: none)	1-33
A	JP 2002-91477 A (Mitsubishi Electric Corp.), 27 March, 2002 (27.03.02), Full text; Figs. 1 to 17 (Family: none)	1-33
A	JP 2002-533771 A (Koninklijke Philips Electronics N.V.), 08 October, 2002 (08.10.02), Full text; Figs. 1 to 5 & WO 2000/038175 A1 & EP 1055227 A1 & US 6823307 B1	1-33

☒ 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
28 August, 2007 (28.08.07)	11 September, 2007 (11.09.07)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2007/061023

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2000-259173 A (Nippon Hoso Kyokai), 22 September, 2000 (22.09.00), Full text; Figs. 1 to 4 (Family: none)	1-33
A	JP 2000-75886 A (Kabushiki Kaisha ATR Onsei Honyaku Tsushin Kenkyusho), 14 March, 2000 (14.03.00), Full text; Figs. 1 to 6 (Family: none)	1-33
A	JP 2000-75892 A (Nippon Hoso Kyokai), 14 March, 2000 (14.03.00), Full text; Figs. 1 to 4 (Family: none)	1-33
A	JP 10-198395 A (Kabushiki Kaisha ATR Onsei Honyaku Tsushin Kenkyusho), 31 July, 1998 (31.07.98), Full text; Figs. 1 to 5 (Family: none)	1-33
A	JP 2000-356997 A (Kabushiki Kaisha ATR Onsei Honyaku Tsushin Kenkyusho), 26 December, 2000 (26.12.00), Full text; Figs. 1 to 11 (Family: none)	1-33
A	JP 10-240288 A (Philips Electronics N.V.), 11 September, 1998 (11.09.98), Full text & EP 862161 A2 & US 6157912 A	1-33
A	Tatsuya KAWAHARA, Akinobu LEE, "Renzoku Onsei Ninshiki Software Julius", Journal of Japanese Society for Artificial Intelligence, Vol.20, No.1, 2005.01, pages 41 to 49	1-33

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２００７／０６１０２３

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G10L15/06(2006.01)i, G10L15/02(2006.01)i, G10L15/18(2006.01)i

Ｂ．　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Int.Cl.　G10L15/00-17/00

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

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２００７年
日本国実用新案登録公報	１９９６－２００７年
日本国登録実用新案公報	１９９４－２００７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）

WPI, CiNii

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	関連する請求の範囲の番号
A	JP 2001-109491 A（学校法人早稲田大学） 2001.04.20, 全文, 図１－３（ファミリーなし）	１－33
A	JP 2002-91477 A（三菱電機株式会社） 2002.03.27, 全文, 図１－17（ファミリーなし）	１－33
A	JP 2002-533771 A（コーニンクレッカ フィリップス エレクトロニクス エヌ ヴィ）2002.10.08, 全文, 図１－５ & WO 2000/038175 A1 & EP 1055227 A1 & US 6823307 B1	１－33

☑　Ｃ欄の続きにも文献が列挙されている。　　☐　パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの
「Ｅ」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「Ｏ」口頭による開示、使用、展示等に言及する文献
「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「＆」同一パテントファミリー文献

国際調査を完了した日 ２８．０８．２００７	国際調査報告の発送日 １１．０９．２００７
国際調査機関の名称及びあて先 日本国特許庁（ＩＳＡ／ＪＰ） 郵便番号１００－８９１５ 東京都千代田区霞が関三丁目４番３号	特許庁審査官（権限のある職員）　５Ｚ　８９４６ 山下　剛史 電話番号　０３－３５８１－１１０１　内線　３５４１

Ｃ（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
A	JP 2000-259173 A（日本放送協会） 2000.09.22，全文，図１－４（ファミリーなし）	１－33
A	JP 2000-75886 A（株式会社エイ・ティ・アール音声翻訳通信研究所） 2000.03.14，全文，図１－６（ファミリーなし）	１－33
A	JP 2000-75892 A（日本放送協会） 2000.03.14，全文，図１－４（ファミリーなし）	１－33
A	JP 10-198395 A（株式会社エイ・ティ・アール音声翻訳通信研究所） 1998.07.31，全文，図１－５（ファミリーなし）	１－33
A	JP 2000-356997 A（株式会社エイ・ティ・アール音声翻訳通信研究所）2000.12.26，全文，図１－11（ファミリーなし）	１－33
A	JP 10-240288 A（フィリップス エレクトロニクス ネムローゼ フェンノートシャップ） 1998.09.11，全文 ＆ EP 862161 A2 ＆ US 6157912 A	１－33
A	河原達也，李晃伸，”連続音声認識ソフトウエア Julius”， 人工知能学会誌，Vol.20，No.1，2005.01，p.41-49	１－33