

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4478925号
(P4478925)

(45) 発行日 平成22年6月9日(2010.6.9)

(24) 登録日 平成22年3月26日(2010.3.26)

(51) Int.Cl. F I
G 1 0 L 15/08 (2006.01) G 1 0 L 15/08 3 0 0 B
 G 1 0 L 15/08 2 0 0 Z

請求項の数 6 (全 16 頁)

<p>(21) 出願番号 特願2003-401724 (P2003-401724)</p> <p>(22) 出願日 平成15年12月1日(2003.12.1)</p> <p>(65) 公開番号 特開2005-164837 (P2005-164837A)</p> <p>(43) 公開日 平成17年6月23日(2005.6.23)</p> <p>審査請求日 平成17年6月10日(2005.6.10)</p> <p>(出願人による申告)平成15年度通信・放送機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願</p> <p>特許権者において、実施許諾の用意がある。</p>	<p>(73) 特許権者 393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2</p> <p>(74) 代理人 100099933 弁理士 清水 敏</p> <p>(72) 発明者 フランク・スーン 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 ロー・ウェイキット 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 中村 哲 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p style="text-align: right;">最終頁に続く</p>
--	---

(54) 【発明の名称】 音声認識結果の信頼度検証装置、コンピュータプログラム、及びコンピュータ

(57) 【特許請求の範囲】

【請求項1】

音声認識デコーダから出力される、各々単語事後確率が付与された単語からなる、複数の仮説単語列をあらわす音声認識結果を受け、前記単語事後確率に基づいて当該音声認識結果の信頼度を検証するための、音声認識結果の信頼度検証装置であって、

前記音声認識結果に含まれる各単語について、前記音声認識結果に含まれる単語の単語事後確率に基づいて一般化単語事後確率を算出するための一般化単語事後確率算出手段と

前記音声認識結果に含まれる各単語の単語事後確率を、前記一般化単語事後確率算出手段により算出された一般化単語事後確率で更新するための更新手段と、

前記更新手段により単語事後確率が更新された前記音声認識結果に基づき、前記複数の仮説単語列の中で、当該仮説単語列に含まれる単語の単語事後確率の和が最大となるものを探索するための探索手段と、

前記探索手段により探索された仮説単語列の単語事後確率の和が所定の条件を充足するか否かを判定することにより、前記音声認識結果の信頼度を検証するための判定手段とを含み、

前記音声認識結果に含まれる各単語には、さらに前記音声認識デコーダへの入力発話中における時間期間を定める情報が付されており、

前記一般化単語事後確率算出手段は、

前記複数の仮説単語列により構成される単語ラティス中に含まれる各単語について、当

該単語の時間期間と重なる時間期間であって、かつ当該単語と一致する単語を前記音声認識結果中から検索するための単語検索手段と、

前記複数の仮説単語列により構成される単語ラティス中に含まれる各単語について、前記単語ラティスに含まれるパスのうち、前記単語検索手段により検索された当該単語を通るパスの尤度の総和を、前記単語ラティスに含まれる全てのパスの尤度の総和で除算することにより、前記単語の一般化単語事後確率を算出するための手段とを含む、音声認識結果の信頼度検証装置。

【請求項 2】

前記一般化単語事後確率算出手段による一般化単語確率の算出に先立って、前記音声認識結果のうち、所定の基準により定められるしきい値よりも尤度が高いものからなる単語列のみを選択して前記一般化単語事後確率算出手段に与えるための手段をさらに含む、請求項 1 に記載の音声認識結果の信頼度検証装置。

10

【請求項 3】

前記一般化単語事後確率を算出するための手段は、前記単語検索手段により検索された単語の単語事後確率の総和と、前記音声認識結果に含まれる全ての単語の単語事後確率の総和との比率によって、前記各単語の一般化単語事後確率を算出するための手段を含む、請求項 1 または請求項 2 に記載の音声認識結果の信頼度検証装置。

【請求項 4】

前記仮説単語列中の単語 w の一般化単語事後確率 $p([w; s, t] | x_1^T)$ (ただし s 及び t はそれぞれ単語 w の時間期間の開始時刻及び終了時刻) は次の式

20

【数 1】

$$p([w; s, t] | x_1^T) = \sum_{\substack{M, [w; s, t]_1^M \\ \exists n, 1 \leq n \leq M \\ w = w_n \\ [s, t] \cap [s_n, t_n] \neq \emptyset}} \frac{\prod_{m=1}^M p^\alpha(x_{s_m}^{t_m} | w_m) p^\beta(w_m | w_1^M)}{p(x_1^T)}$$

30

で与えられ、ただし $x_1^T = x_1, \dots, x_T$ は観測された音声シーケンスであり、 M は音声認識結果の仮説に含まれる単語数であり、 s_n 及び t_n はそれぞれ、単語 w と一致する n 番目の単語 w_n の開始時刻及び終了時刻であり、 $p(x_{s_m}^{t_m} | w_m)$ は音響尤度であり、 $p(w_m | w_1^M)$ は言語尤度であり、 $p(x_1^T)$ は音響観測尤度であり、 α 及び β はそれぞれ所定の定数である、請求項 1 ~ 請求項 3 のいずれかに記載の音声認識結果の信頼度検証装置。

【請求項 5】

コンピュータにより実行されると、請求項 1 ~ 請求項 4 のいずれかに記載の音声認識結果の信頼度検証装置の各手段を実現するよう、当該コンピュータを動作させる、コンピュータプログラム。

40

【請求項 6】

請求項 5 に記載のコンピュータプログラムによりプログラムされたコンピュータ。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は音声認識の結果の受入 / 拒否を判定するための技術に関し、特に、音声認識結果の単語の信頼度尺度を用いて判定を行なう技術に関する。

【背景技術】

50

【0002】

最先端の音声認識技術を適用できる可能性のある分野は非常に広い。それら分野のうちには、好結果が得られているものも存在する。しかし、音声認識技術はまだ確実なものではない。そのため、音声認識結果を監視して、認識結果を受入れるか拒否するかを容易に判定できるような何らかの尺度が求められている。そうした技術の必要性は、音声認識結果がさらに新しく困難な分野に適用されて行くに伴い、今後増大していくと考えられる。

【0003】

そのような尺度は、容易に計算が可能で、かつ統計的に意味のあるものでなければならない。これまで提案されよく研究されてきたものに、単語事後確率 (Word Posterior Probability: wpp) がある。単語事後確率を音声認識結果の単語ラティス/グラフ又はN ベストリストに適用した研究として、非特許文献1, 2, 3に記載のものがある。

【0004】

【非特許文献1】ステファン・オルトマンズ、ヘルマン・ネイ、サビアー・オーバート、「大語彙連続音声認識のための単語グラフアルゴリズム」、コンピュータ・スピーチ及言語、第11巻1号、pp. 43 - 72、1997年1月 (Stefan Ortman ns, Hermann Ney, and Xavier Aubert, "A Word Graph Algorithm for Large Vocabulary Continuous Speech Recognition," Computer Speech and Language, vol. 11, no. 1, pp. 43 72, January 1997.)

【非特許文献2】リヒャルト・シュヴァルツ及びイェン・ルー・チョウ、「N ベストア ルゴリズム: N個の最尤文仮説を発見するための効率的かつ正確な手続き」、ICASS P1990予稿集、1990年、第1巻、pp. 81 94 (Richard Schw artz and Yen Lu Chow, "The N best Algorithm: An Efficient and Exact Procedure fo r Finding the N Most Likely Sentence Hyp otheses," in Proceedings of ICASSP1990, 1990, vol. 1, pp. 81 94.)

【非特許文献3】フランク・K・スーン及びエン フォン ファン、「ツリー・トレリス に基づく、連続音声認識におけるNベストの文仮説の発見のための高速探索」、ICASS P1991予稿集、1991年、第1巻、pp. 705 708 (Frank K. Soong and Eng Fong Huang, "A Tree Trellis Based Fast Search for Finding the N Bes t Sentence Hypotheses in Continuous Spee ch Recognition," in Proceedings of ICASS P1991, 1991, vol. 1, pp. 705 708.)

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかし、上記した非特許文献1~3には、単語事後確率をどのように用いれば、音声認識の受入/拒否を容易に、かつ信頼性高く判定できるようになるかについての具体的な記載はない。音声認識の精度が高くないような応用分野では、受入/拒否の判定を行なう際の制約がきつく、その判定は容易ではない。そのため、容易に、かつ信頼性高く音声認識結果の受入/拒否を判定できる技術が、特に音声認識技術の新しい適用分野において必要とされている。

【0006】

また、従来から音声認識技術が使用されている技術分野でも、音声認識結果の受入/拒否を自動的に判定可能になれば、人間の介入なしに様々な処理を自動的に実行できる。例えば、音声認識の信頼度に基づいて音声認識環境を自動的に調整したり、何らかの形でコ

10

20

30

40

50

ンピュータが自動的なトレーニングを行ったり、語学の発声の練習に用いたりすることもできる。

【0007】

したがって、本発明の目的は、音声認識結果の受入/拒否を容易に、かつ信頼性高く行なうことができる音声認識結果の検証装置を提供することである。

【0008】

本発明の他の目的は、音声認識結果である単語列の単語ごとに、受入/拒否を容易に、かつ信頼性高く行なうことができる音声認識結果の検証装置を提供することである。

【課題を解決するための手段】

【0009】

本発明の第1の局面にかかる音声認識結果の信頼度検証装置は、音声認識デコーダから出力される、各々単語事後確率が付与された単語からなる、複数の仮説単語列をあらゆる音声認識結果を受け、単語事後確率に基づいて当該音声認識結果の信頼度を検証するための、音声認識結果の信頼度検証装置であって、音声認識結果に含まれる各単語について、音声認識結果に含まれる単語の単語事後確率に基づいて一般化単語事後確率を算出するための一般化単語事後確率算出手段と、音声認識結果に含まれる各単語の単語事後確率を、一般化単語事後確率算出手段により算出された一般化単語事後確率で更新するための更新手段と、更新手段により単語事後確率が更新された音声認識結果に基づき、複数の仮説単語列の中で、当該仮説単語列に含まれる単語の単語事後確率の和が最大となるものを探索するための探索手段と、探索手段により探索された仮説単語列の単語事後確率の和が所定の条件を充足するか否かを判定することにより、音声認識結果の信頼度を検証するための判定手段とを含む。

【0010】

好ましくは、音声認識結果の信頼度検証装置は、一般化単語事後確率算出手段による一般化単語確率の算出に先立って、音声認識結果のうち、所定の基準により定められるしきい値よりも尤度が高いものからなる単語列のみを選択して一般化単語事後確率算出手段に与えるための手段をさらに含む。

【0011】

より好ましくは、音声認識結果に含まれる各単語には、さらに音声認識デコーダへの入力発話中における時間期間を定める情報が付されており、一般化単語事後確率算出手段は、音声認識結果中に含まれる各単語について、当該単語の時間期間と重なる時間期間であって、かつ当該単語と一致する単語を音声認識結果中で検索するための単語検索手段と、単語検索手段により検索された単語の単語事後確率の総和と、音声認識結果に含まれる全ての単語の単語事後確率の総和とに基づいて、各単語の一般化単語事後確率を算出するための手段とを含む。

【0012】

さらに好ましくは、一般化単語事後確率を算出するための手段は、単語検索手段により検索された単語の単語事後確率の総和と、音声認識結果に含まれる全ての単語の単語事後確率の総和との比率によって、各単語の一般化単語事後確率を算出するための手段を含む。

【0013】

好ましくは、仮説単語列中の単語 w の一般化単語事後確率 $p([w; s, t] | x_1^T)$ (ただし s 及び t はそれぞれ単語 w の時間期間の開始時刻及び終了時刻) は次の式

【0014】

10

20

30

40

【数1】

$$\begin{aligned}
 & p([w; s, t] | x_1^T) \\
 &= \sum_{\substack{M, [w; s, t]_1^M \\ \exists n, 1 \leq n \leq M \\ w = w_n \\ [s, t] \cap [s_n, t_n] \neq \emptyset}} \frac{\prod_{m=1}^M p^\alpha(x_{s_m}^{t_m} | w_m) p^\beta(w_m | w_1^M)}{p(x_1^T)}
 \end{aligned}$$

10

で与えられ、ただし $x_1^T = x_1, \dots, x_T$ は観測された音声シーケンスであり、 M は音声認識結果の仮説に含まれる単語数であり、 s_n 及び t_n はそれぞれ、単語 w と一致する n 番目の単語 w_n の開始時刻及び終了時刻であり、 $p(x_{s_m}^{t_m} | w_m)$ は音響尤度であり、 $p(w_m | w_1^M)$ は言語尤度であり、 $p(x_1^T)$ は音響観測尤度であり、及び はそれぞれ所定の定数である。

【0015】

本発明の第2の局面にかかるコンピュータプログラムは、コンピュータにより実行されると、上記したいずれかの音声認識結果の信頼度検証装置の各手段を実現するよう、当該コンピュータを動作させる。

20

【0016】

本発明の第3の局面にかかるコンピュータは、上記したコンピュータプログラムによりプログラムされたものである。

【発明を実施するための最良の形態】

【0017】

〔導入〕

本実施の形態では、連続音声認識により認識された各単語を受入れるか、拒絶するかを判定するという問題を、注目単語の位置の特定という考え方を導入することで解決する。注目単語以外の単語（非注目単語）については、互いに区別せずいずれも単にそれぞれの場所を占めるだけのものとして取り扱って、注目単語の事後確率を算出する。

30

【0018】

このように注目単語／非注目単語という二分法を採用することにより、動的計画法に基づく文字列のアライメント等の複雑な処理を行なう必要が回避できる。

【0019】

まず、以下の概念を導入し、それらについて説明する。すなわち、それらは、(1) 音声認識結果の単語ラティス又は N ベストリスト中における、注目単語の位置決定を行なうための、仮説（候補）となる文字列の探索空間の削減、(2) ある候補単語の複数の出現個所における事後確率をグループ化する際の時間的制約の緩和、及び(3) 音響モデル及び言語モデルによる寄与に対する適切な重み付け、である。

【0020】

文字列と単語の事後確率

HMM (Hidden Markov Model : 隠れマルコフモデル) を用いる音声認識装置では、所与の音響観測データ $x_1^T = x_1, \dots, x_T$ に対する、最適な単語シーケンス $w_1^{M^*} = w_1^*, \dots, w_M^*$ を、以下に示すように、可能な全ての単語シーケンスからなる空間を探索して、最大事後確率 (MAP) を与えるものとして求める。

40

【0021】

【数2】

$$w_1^{M*} = \arg \max_{\{M, w_1^M\}} p(w_1^M | x_1^T) \quad (1)$$

$$= \arg \max_{\{M, w_1^M\}} \frac{p(x_1^T | w_1^M) p(w_1^M)}{p(x_1^T)} \quad (2)$$

$$= \arg \max_{\{M, w_1^M\}} p(x_1^T | w_1^M) p(w_1^M) \quad (3)$$

10

ただし、 $p(x_1^T | w_1^M)$ は音響モデルの確率、 $p(w_1^M)$ は言語モデルによる確率、 $p(x_1^T)$ は音響の観測確率である。

【0022】

トレーニング環境とテスト環境、話者、ノイズ等の相違により「最適な」単語シーケンスであっても誤りを含むことがある。そこで、数学的に扱いやすく、かつ統計的に好ましい何らかの信頼度尺度を採用すべきである。

【0023】

単語列の事後確率 $p(w_1^M | x_1^T)$ は、観測された音響 x_1^T に対し、認識された単語列 w_1^M の尤度を測るものであるが、これは対応する時間的セグメンテーション

20

【0024】

【数3】

$$[w; s, t]_1^M = [w_1; s_1, t_1] \cdots [w_M; s_M, t_M] \quad (4)$$

を仮定することで算出される。ただし、 s 及び t は単語 w の始点及び終点の時刻を示し、 $s_1 = 1$ 、 $t_M = T$ 、 $1 \leq m \leq M-1$ の m に対し $t_m + 1 = s_{m+1}$ である。

【0025】

これを用いて、式(2)を次のように書き換えることができる。

【0026】

30

【数4】

$$p(w_1^M | x_1^T) = \frac{p(x_1^T | [w; s, t]_1^M) \cdot p(w_1^M)}{p(x_1^T)} \quad (5)$$

$$= \frac{\prod_{m=1}^M p(x_{s_m}^{t_m} | w_m) \cdot p(w_m | w_1^{m-1})}{p(x_1^T)} \quad (6)$$

認識された単語列の全体の信頼性を測るためには、この単語列事後確率 $p(w_1^M | x_1^T)$ を採用するのが自然である。

40

【0027】

単語の信頼性を測るために適切な信頼度尺度は、単語事後確率 $p([w_m; s_m, t_m] | x_1^T)$ である。これは特定の単語を含む単語列の事後確率を全て合計することにより算出される。

【0028】

【数5】

$$\begin{aligned}
 & p([w; s, t] | x_1^T) \\
 &= \sum_{\substack{M, [w; s, t]_1^M \\ \exists n, 1 \leq n \leq M \\ [w_n; s_n, t_n] = [w; s, t]}} \frac{\prod_{m=1}^M p(x_{s_m}^{t_m} | w_m) p(w_m | w_1^{m-1})}{p(x_1^T)} \quad (7)
 \end{aligned}$$

この単語事後確率を実際に有効な信頼度尺度として用いるためには、さらにいくつかの問題を解決する必要がある。 10

【0029】

[単語事後確率の修正]

考慮すべき仮説数

大語彙の連続音声認識装置(LVCSR)においては、可能な単語列の探索空間は膨大である。しかし、各単語列の事後確率の値には大きな相違があり、比較的低い尤度の単語列については刈込みしても差し支えない。このようにして得た、単語列の仮説の部分集合のみを用いて単語ラティス/グラフ又はN ベスト単語列リストを得ることができる。以下の実施の形態では、そのように部分集合を用いて得た単語ラティス/グラフを使用するものとする。 20

【0030】

仮説内の単語の時間的なレジストレーション

単語の時間的位置決め(レジストレーション)を[w; s, t]で表わす。別々の仮説中にある同一の単語が出現する場合でも、その位置は仮説によって多少異なることがあり得る。自動音声認識(ASR)の最終的目標は発話中の単語からなる内容を認識することであるから、厳密な時間的制約を多少緩和することにする。ここでは、ある単語がある単語列中において出現する期間が、基準となる単語の期間[s, t]と重なっており(オーバーラップしている)、かつその単語が基準となる単語と一致しているような単語を検索し、それら単語をその基準となる単語の事後確率の計算に含める。その結果式(7)は以下のように書き換えられる。 30

【0031】

【数6】

$$\begin{aligned}
 & p([w; s, t] | x_1^T) \\
 &= \sum_{\substack{M, [w; s, t]_1^M \\ \exists n, 1 \leq n \leq M \\ w = w_n \\ [s, t] \cap [s_n, t_n] \neq \emptyset}} \frac{\prod_{m=1}^M p(x_{s_m}^{t_m} | w_m) p(w_m | w_1^{m-1})}{p(x_1^T)} \quad (8)
 \end{aligned}$$

40

音響尤度と言語尤度との比重

本実施の形態では、音響尤度と言語尤度とは、それぞれ及びで示される重みによって指数的な重み付けがなされる。式(8)にこれを適用すると次式となる。

【0032】

【数7】

$$\begin{aligned}
 & p([w; s, t] | x_1^T) \\
 &= \sum_{\substack{M, [w; s, t]_1^M \\ \exists n, 1 \leq n \leq M \\ w = w_n \\ [s, t] \cap [s_n, t_n] \neq \emptyset}} \frac{\prod_{m=1}^M p^\alpha(x_{s_m}^{t_m} | w_m) p^\beta(w_m | w_1^M)}{p(x_1^T)} \quad (9)
 \end{aligned}$$

10

〔注目単語の抽出〕

ここで、本実施の形態に係る単語抽出方式により抽出された注目単語の受入／拒否について検討する。図1に本実施の形態で使用する、音声認識の結果得られる単語ラティス／グラフの例を、図2に同様に音声認識の結果得られる単語列のN ベストリスト（仮説の単語列のうち、尤度の高いN個からなるリスト）の模式的な例を、それぞれ示す。

【0033】

図1を参照して、本実施例で使用する単語ラティス／グラフ20は、従来のものと異なり、注目単語（「w」で示す。）以外の単語については個々の単語ラベルを付さず、いずれも単に「*」というラベルを付してあるだけである。

20

【0034】

この単語wの出現個所の各々に対し、前方 後方アルゴリズムを用いて単語事後確率を効率的に計算できる。その後、この特定の単語w（たとえば単語30、32、34）を通るパスの全てについての尤度を合計し、その合計をこの単語ラティス／グラフ内の全てのパスの尤度の合計で除算し正規化することによって、一般化された単語事後確率（以下「一般化単語事後確率」と呼ぶ。）が算出できる。さらにこの際、単語の時間的レジストレーション（単語開始及び終了時刻の一致）の条件を緩和する。すなわち、各パスの単語wの期間が正確に一致する必要はなく、時間的にオーバーラップしているものの事後確率の合計を計算する。

【0035】

30

同様に、図2に示すようなN ベストリストでも単語wの一般化単語事後確率を算出できる。ここでも、注目単語70（単語w）以外の単語には「*」というマークが付され、その単語が何であるかは問わない。図2に示すように仮説50、…、62が存在している場合を考える。単語70と時間的にオーバーラップしている単語wとして仮説54、56、58、60の単語72、74、76、78等が考えられる。仮説62の単語80は、単語70の期間とオーバーラップしていないのでこの場合の単語事後確率の算出には用いられない。

【0036】

上記したように単語70及びこれと時間的にオーバーラップしている、単語70と同じ単語72、74、76、78が出現する仮説50、54、56、58、60の尤度の合計を算出し、それをN ベストリスト中の全ての仮説の尤度の合計で除算し正規化することによって、この単語の一般化単語事後確率を算出できる。ここでも、時間的レジストレーションの制約を緩和している。

40

【0037】

なお、上記したようにして注目単語を抽出して一般化単語事後確率を計算する場合、単語のアライメントは不要である。また動的プログラム法により仮説のアライメントを求める必要もない。

【0038】

〔本実施の形態に係る装置構成〕

図3に、本実施の形態に係る仮説検証装置94を含む音声機械翻訳装置80のブロック

50

図を示す。図3を参照して、この音声機械翻訳装置80は、入力される音声 $s(t)$ 100の音声認識を行ない、認識結果を図1に示す単語ラティス/グラフ20のような単語グラフ104として出力するためのASRデコーダ90と、ASRデコーダ90から出力される単語グラフ104の中の各単語に対して上記したように一般化単語事後確率を用いた仮説の検証を行ない、その結果最も尤度の高い単語列106を、一般化単語事後確率を付して出力するための仮説検証装置94と、仮説検証装置94の出力する単語列106を入力として機械翻訳を行ない、翻訳結果110を出力するための機械翻訳装置92とを含む。

【0039】

音声機械翻訳装置80はさらに、仮説検証装置94の出力する単語列106を受け、当該単語列の各単語に付されている単語事後確率に基づいて、この単語列106を認識結果として受入れるか拒否するかを決定してその結果をユーザインタフェース(以下「ユーザI/F」と呼ぶ。)82を用いてユーザにフィードバックするとともに、判定結果に基づいて機械翻訳装置92の制御を行なうための受入/拒否判定装置96を含む。

【0040】

図4に仮説検証装置94の詳細を示す。図4を参照して、仮説検証装置94は、ASRデコーダ90から与えられる単語グラフ104を記憶するための単語グラフ記憶部120と、単語グラフ記憶部120に記憶された単語グラフ中において、尤度の低い単語列を除く単語列に含まれる各単語について、一般化単語事後確率を算出するために、単語グラフ中で当該単語の期間とオーバーラップする、同じ単語を検索するための対象単語検索部122と、対象単語検索部122により検索された単語群に対し、前述した算出方法によりその一般化単語事後確率を算出するための事後確率算出部124と、事後確率算出部124により単語ごとに算出された一般化単語事後確率を、単語グラフ記憶部120に記憶された単語グラフの各単語に再付与して単語グラフを更新するための単語グラフ更新部126と、このようにして単語ごとに一般化単語事後確率が再付与された単語グラフの中で、最も高い一般化単語事後確率を示すパス(最尤パス)を探索してそのパスに含まれる単語列を一般化単語事後確率とともに単語列106として出力するための最尤パス探索部128とを含む。

【0041】

[動作]

この装置は以下のように動作する。図3を参照して、入力音声100が与えられると、ASRデコーダ90は音声認識を行ない、その結果を単語グラフ104として出力する。この単語グラフ104の各単語には、それぞれ認識時に単語事後確率が得られ付与されている。

【0042】

図4を参照して、仮説検証装置94の単語グラフ記憶部120はこの単語グラフのうち、尤度の低いものを除くサブセットを記憶する。対象単語検索部122は、単語グラフ記憶部120に記憶されている単語グラフの単語列のうち、単語ごとに、一般化単語事後確率の計算対象となる単語群(当該単語と同一の単語で、当該単語の出現期間と重なる期間に他のパス上に出現している単語)を検索し、事後確率算出部124に与える。

【0043】

事後確率算出部124は、対象単語検索部122により検索された単語群を対象にして前述したとおり、単語事後確率を合計し、その合計を全パスの単語事後確率で除算し正規化することで、対象単語の一般化単語事後確率を算出する。

【0044】

単語グラフ更新部126は、単語ごとに、単語グラフ中のその単語に対し、事後確率算出部124で算出された一般化単語事後確率を再付与する。

【0045】

全ての単語に対し一般化単語事後確率の再付与が行われると、最尤パス探索部128が最も高い一般化単語事後確率を示す単語列を単語グラフ中で探索し、その結果見出された

10

20

30

40

50

パスに含まれる単語列を、一般化単語事後確率とともに単語列 106 として図 3 に示す機械翻訳装置 92 および受入 / 拒否判定装置 96 に与える。

【0046】

受入 / 拒否判定装置 96 は、単語列 106 から与えられる単語列の単語事後確率に基づき、当該単語列を認識結果として受入れるか、拒否するかを判定する。この場合、この単語列の尤度を所定のしきい値と比較し、しきい値以上であれば受入れ、しきい値未満であれば拒否する。受入 / 拒否判定装置 96 は、認識結果を受入れる場合には機械翻訳装置 92 を制御して単語列 106 に対する機械翻訳を実行させる。拒否する場合には機械翻訳装置 92 による機械翻訳を停止させるとともに、ユーザ I / F 82 を用いて、ユーザに対して認識結果が拒否されたことを伝える。

10

【0047】

機械翻訳装置 92 は、受入 / 拒否判定装置 96 から翻訳の開始を指示されたことに応答して単語列 106 に対する機械翻訳を行ない、翻訳結果 110 を出力する。この際、単語列 106 の各単語に事後確率が付与されているため、翻訳においてこの事後確率を考慮した翻訳を行なうことができる。

【0048】

[実験]

実験システムの構成

上記した実施の形態に従った装置をセットアップし、実験を行なった。この実験では、単語グラフではなく N ベストリストを用いる方式を採用した。実験では、出願人において作成した日本語基本旅行表現コーパスを使用した。テストセットとしてセット 01 およびセット 02 の二つを用いた。これらテストセットは、それぞれ 510 発話および 508 発話からなる。これら二つのテストセットは、セットごとに、10 人の話者による種類の発話を録音したものである。なお、ASR デコーダ 90 としては、出願人において開発したものをを用いた。

20

【0049】

本実験では、発話ごとに ASR デコーダ 90 により 100 ベストの認識結果の仮説を出力し、探索のビーム幅を狭くした。

【0050】

性能評価

実験では、信頼度尺度の誤り率 (Confidence Error Rate: CER) を、誤り拒否数 (FR) および誤り受入数 (FA) を採用して以下のように定義し実験システムの性能を評価した。

30

【0051】

【数 8】

$$CER = \frac{\text{誤り受入数} + \text{誤り拒否数}}{\text{認識された単語の総数}} \times 100\%$$

結果

仮に認識結果を全て受入れることにし、拒否しない場合には、判別可能な誤りは挿入と置換だけとなる。実験では、この誤りレベルをベースラインとした。実験で使用したテストセットに対する認識結果の、ベースラインの CER を次のテーブル 1 に「ベースライン」として示す。

40

【0052】

【表 1】
テーブル1

	SET01	SET02
ベースライン	25.26%	22.32%
再出現率 ($\alpha=0, \beta=0$)	23.05%	20.69%
最適 α, β における単語事後確率	18.48%	16.82%

単語事後確率の簡単な算出方法は、ある特定の単語について、その単語が出現する仮説数を数えることである。この数と、仮説の総数との比によって、その単語の一般化単語事後確率の大まかな値を算出できる。これは、式(9)において $\alpha = 0$ とした場合に相当する。この大まかな算出方法を使用した場合の結果をテーブル1において「再出現率」として示す。テーブル1から分かるように、この方法を用いるとベースラインに対して2ポイント程度のCERの改善が得られた。

【0053】

一般化単語事後確率に対する音響尤度と言語尤度とによる寄与の割合は明確には分からないが、一般化単語事後確率をできるだけ精度高く算出するためには、 α の値を適切に決めることが有用である。式(9)によって算出した一般化単語事後確率に対する、 α および β の種々の値による影響を調べるために、 α 及び β の広い範囲にわたって単一しきい値の分類器の性能をテストした。結果を図5(セット01)及び図6(セット02)の等高線図により示す。図5及び図6において、色の濃い領域に属する α, β の組合せを用いた場合に、高い性能が得られた。

【0054】

この結果から、セット01とセット02とのいずれの場合にも、高い性能を示す α と β の組合せはグラフ上のほぼ直線上に存在していること、さらに α 及び β の値が比較的小さい領域に性能の高い部分があることが分かった。そこで、この最適と思われる領域 ($\alpha \in [0.01, 0.2]$ 、 $\beta \in [0.1, 1.5]$) 上でより詳細な調査を行なった。その結果を図7(セット01)及び図8(セット02)に示す。

【0055】

図7及び図8から、セット01に対しては $\alpha = 0.06$ 、 $\beta = 0.3$ において最も高い性能が得られ、セット02に対しては $\alpha = 0.03$ 、 $\beta = 0.3$ において最も高い性能が得られることが分かった。これら最適点におけるCERをテーブル1の最下行に示す。

【0056】

図9に、最適な α と β を用いたときの、再出現率と単語事後確率 (wpp) のROC (Receiver Operating Characteristics) 曲線を示す。図9(A)はセット01、図9(B)はセット02に対するものである。

【0057】

図9からは、テストセットの一方を用いて得られた最適な α 及び β を他方のテストセットに適用した場合、性能低下はごくわずかであることが分かる。テーブル2は、このクロス検証によるCER性能を示す。

【0058】

【表 2】
テーブル2

	$\alpha=0.06, \beta=0.3$ (optimal for set01)	$\alpha=0.03, \beta=0.3$ (optimal for set02)
SET01	18.48%	18.78%
SET02	17.53%	16.82%

表2からは、これらパラメータが非常に安定したものであることが分かる。 α 、 β のいずれを多少変更しても性能の急激な低下は生じない。

10

20

30

40

50

【 0 0 5 9 】

図 1 0 は、これら二つのテストセットに対する最適点の近傍での、システム性能の挙動をより詳細に示したものである。図 1 0 (C) 及び (D) から分かるように、 の変動に対して C E R は比較的安定している。また図 1 0 (A) 及び (B) から分かるように、 の変動に対しては の変動ほど C E R は安定しない。しかしそれでも急激な C E R の変動はなく、比較的安定しているといえる。

【 0 0 6 0 】

2 クラスガウス分布分類器

比較のため、単語スポッティングによる仮説検証方式を 2 クラスへの分類処理に変換した。この分類器を所与のデータセットを用いてトレーニングした。データを最初に音声認識し、100 ベストリストを生成した。この100 ベストリストに基づき、認識結果の単語列中の各単語の事後確率を計算し、正確度を示すラベルでタグ付けした。

【 0 0 6 1 】

正確な単語と不正確な単語との単語事後確率を二つのクラスに分類し、二つのガウス分布モデルをトレーニングした。トレーニングセット及びテストセットとして、セット 0 1 及びセット 0 2 を交代で用いた。

【 0 0 6 2 】

得られた性能をテーブル 3 に示す。

【 0 0 6 3 】

【表 3】

テーブル 3

トレーニングセット	SET02	SET01
テストセット	SET01	SET02
CER(ガウスモデル分類器)	19.09%	17.80%
CER(単一しきい値)	18.78%	17.35%

テーブル 3 から分かるように、いずれの場合も、性能はテーブル 1 に示した単一しきい値による仮説分類の場合とほぼ等しい。もっとも、いずれの場合もテーブル 1 の場合と比較してやや性能が落ちている。

【 0 0 6 4 】

以上のとおり、本発明の実施の形態に係る仮説検証装置 9 4 を用いると、A S R デコーダの出力する各単語について、一般化単語事後確率を算出できる。その際、(1) 探索のための仮説数を少なくして探索空間を小さくすることにより処理を高速に行なうことができ、(2) ある単語の単語事後確率を算出する際、同一単語を選択するための時間的制約を緩和したことにより、単語事後確率の安定した値を算出することができ、(3) 単語事後確率の算出にあたって、音響尤度と言語尤度との寄与の度合いを 及び で反映させるようにし、かつそれらの最適な範囲の値を特定したことにより、性能の安定した仮説検証装置を得ることができる。

【 0 0 6 5 】

以上ブロック図形式で説明した各機能部は、いずれもコンピュータハードウェア及び当該コンピュータ上で実行されるソフトウェア(コンピュータプログラム)により実現することができる。このコンピュータハードウェアとしては、音声を扱う設備を持ったものであれば、汎用のハードウェアを有するものを用いることができる。そうしたソフトウェアもまた一つのデータであり、記憶媒体に記憶させて流通させることができる。

【 0 0 6 6 】

当該ソフトウェアには、上記した仮説検証装置 9 4 の機能を実現するために必要な全ての命令を含んでいる必要はなく、例えばオペレーティングシステムに備えられている命令を呼び出すことにより、所望の機能を実現するものでもよい。すなわち、コンピュータのハードウェア及びソフトウェア資源を利用して上記した仮説検証装置 9 4 の各機能を実

10

20

30

40

50

現するものであればよい。

【0067】

また、図3に示す音声機械翻訳装置80も、マイクロフォン及び音声処理専用のボード等を除き、一般的な構成のコンピュータ及びソフトウェアにより実現可能である。

【0068】

そして、そうしたソフトウェアによりプログラムされたコンピュータは、本発明に係る音声認識結果の信頼度検証装置となる。

【0069】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

【図面の簡単な説明】

【0070】

【図1】本発明の第1の実施の形態に係る仮説検証装置94の動作原理を説明するための、単語ラティス/グラフの模式図である。

【図2】本発明の第1の実施の形態に係る仮説検証装置94の原理を説明するための、Nベクトリストの模式図である。

【図3】本発明の第1の実施の形態に係る仮説検証装置94を用いた音声機械翻訳装置80のブロック図である。

【図4】図3に示す仮説検証装置94の詳細なブロック図である。

【図5】セット01を用い、及びの広い範囲にわたって単一しきい値の分類器の性能をテストした結果を示す等高線図である。

【図6】セット02を用い、及びの広い範囲にわたって単一しきい値の分類器の性能をテストした結果を示す等高線図である。

【図7】セット01を用い、及びの微小な範囲における単一しきい値の分類器の性能をテストした結果を示す等高線図である。

【図8】セット02を用い、及びの微小な範囲における単一しきい値の分類器の性能をテストした結果を示す等高線図である。

【図9】最適なとを用いたときの、再出現率と一般化単語事後確率のROC曲線を比較して示すグラフである。

【図10】図10は、二つのテストセット(01、02)に対する最適点の近傍での、及びの変動に対するシステム性能の挙動を示したグラフである。

【符号の説明】

【0071】

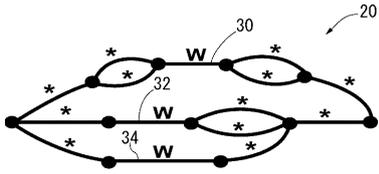
80 音声機械翻訳装置、90 ASRデコーダ、92 機械翻訳装置、94 信頼度尺度方式仮説検証装置、96 受入/拒否判定装置、120 単語グラフ記憶部、122 対象単語検索部、124 事後確率算出部、126 単語グラフ更新部、128 最尤パス探索部

10

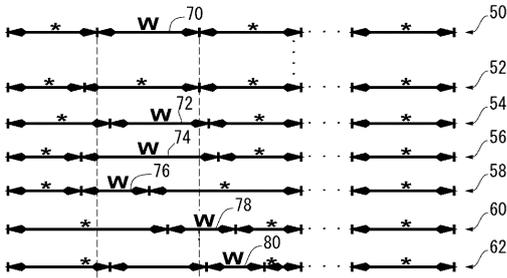
20

30

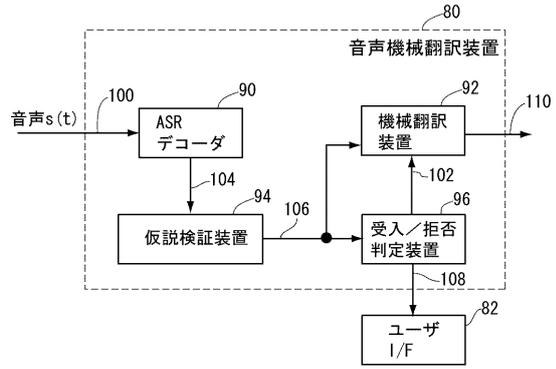
【図1】



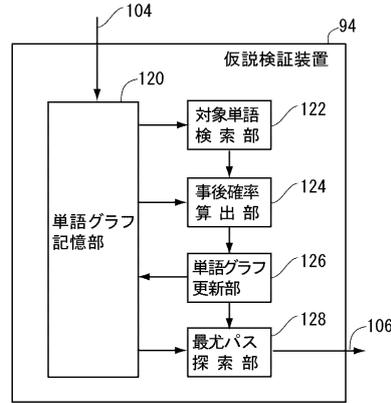
【図2】



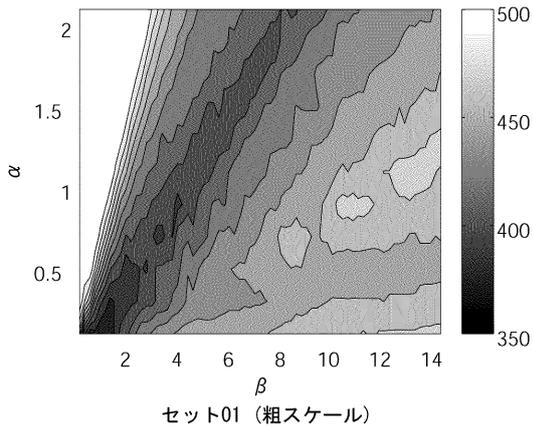
【図3】



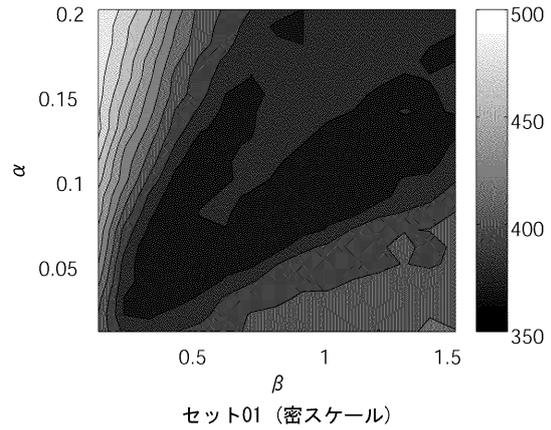
【図4】



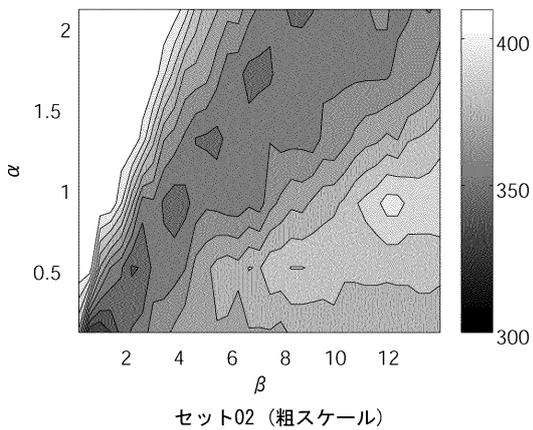
【図5】



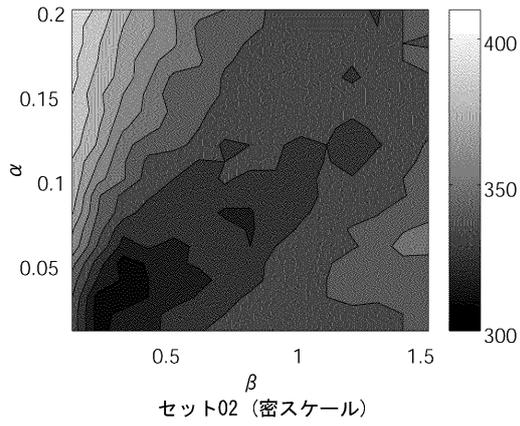
【図7】



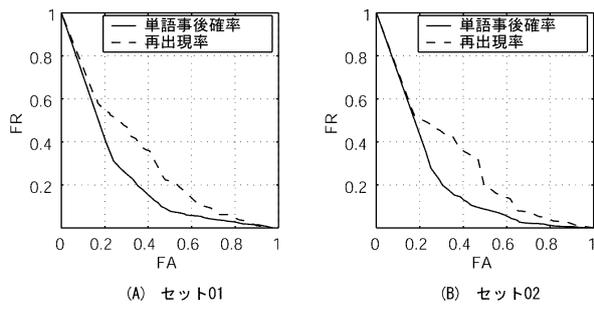
【図6】



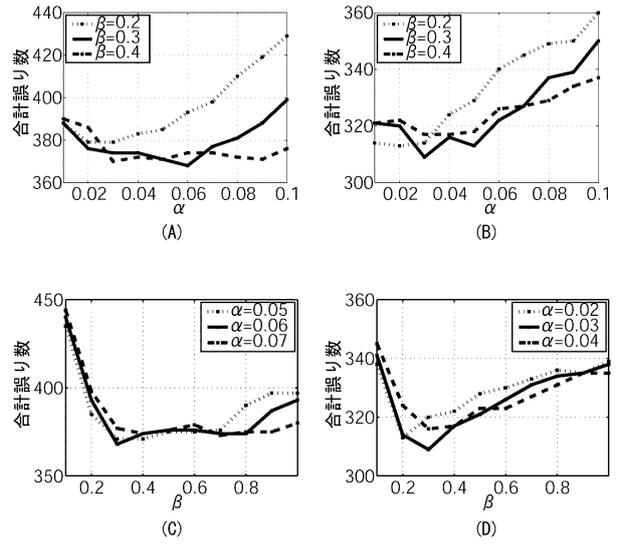
【図8】



【図9】



【図10】



フロントページの続き

審査官 毛利 太郎

(56)参考文献 国際公開第03/088209(WO, A1)

李 晃伸、鹿野 清宏、河原 達也、音声認識エンジンJuliusにおける単語事後確率を用いた信頼度算出、日本音響学会2003年秋季研究発表会講演論文集、日本、社団法人日本音響学会、2003年9月17日、Vol.1, pp.117-118

山本 博史、マルコフ コンスタンチン、奥田 浩三、単語適合率最大基準に基づく複数システムの統合、電子情報通信学会技術研究報告、日本、社団法人電子情報通信学会、2002年6月21日、Vol.102 No.160, pp.43-47

緒方 淳、有木 康雄、音声認識精度向上のための信頼度尺度の比較、情報処理学会研究報告、日本、社団法人情報処理学会、2000年12月22日、Vol.2000 No.119, pp.113-118

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/28