

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2009年4月2日 (02.04.2009)

PCT

(10) 国際公開番号
WO 2009/041402 A1

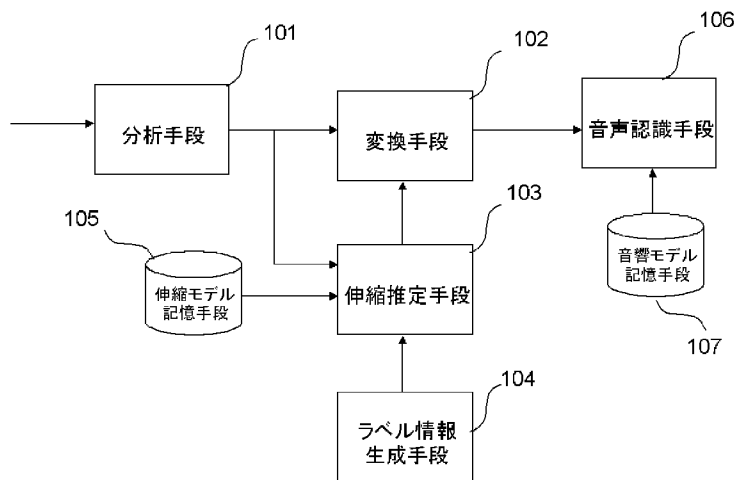
- (51) 国際特許分類:
G10L 15/20 (2006.01) *G10L 15/04* (2006.01)
G10L 13/06 (2006.01) *G10L 15/06* (2006.01)
G10L 15/02 (2006.01)
- (21) 国際出願番号: PCT/JP2008/067125
- (22) 国際出願日: 2008年9月22日 (22.09.2008)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
 特願2007-247572 2007年9月25日 (25.09.2007) JP
- (71) 出願人 (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).
- (74) 代理人: 加藤 朝道 (KATO, Asamichi); 〒2220033 神奈川県横浜市港北区新横浜3丁目20番12号加藤内外特許事務所内 Kanagawa (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE,

[続葉有]

(54) Title: FREQUENCY AXIS ELASTIC COEFFICIENT ESTIMATION DEVICE, SYSTEM METHOD AND PROGRAM

(54) 発明の名称: 周波数軸伸縮係数推定装置とシステム方法並びにプログラム

[図1]



101... ANALYZING MEANS
 102... CONVERTING MEANS
 106... SOUND RECOGNIZING MEANS
 105... ELASTIC MODEL MEMORY MEANS
 103... ELASTICITY ESTIMATING MEANS
 107... ACOUSTIC MODEL MEMORY MEANS
 104... LABEL INFORMATION GENERATING MEANS

(57) Abstract: It is an object to provide an elastic coefficient estimation system configured to suppress costs for the estimation of an elastic coefficient and estimates a high accuracy value. The elastic coefficient estimation system is provided with a label information generating means (104) that outputs sound or non-sound label information, an elastic model memory means (105) that stores a probability model to express an emergence probability of the sound or non-sound, an elasticity estimating means (103) that calculates an elastic coefficient in a frequency axis direction by using the probability model to express a sound or non-sound emergence probability, a sound or non-sound label and a cepstrum.

(57) 要約: 本発明は、伸縮係数の推定のためのコストを抑え、かつ高精度な値を推定する、伸縮係数推定システムを提供することを目的とする。音声・非音声のラベル情報を出力するラベル情報生成手段104と、音声・非音声の出現確率を表す確率モデルを記憶する伸縮モデル記憶

手段105と、音声・非音声の出現確率を表す確率モデルと音声・非音声のラベルとケプストラムを用いて

[続葉有]

WO 2009/041402 A1



SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE,
SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

(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,

添付公開書類:
— 国際調査報告書

明 細 書

周波数軸伸縮係数推定装置とシステム方法並びにプログラム 技術分野

[0001] (関連出願についての記載)

本願は、先の日本特許出願2007-247572号(2007年9月25日出願)の優先権を主張するものであり、前記先の出願の全記載内容は、本書に引用をもって繰込み記載されているものとみなされる。

本発明は、音声処理技術に関し、特に、音声の周波数軸上の伸縮で表される話者性を正規化するための伸縮係数推定装置、ケプストラム変換システム、音声認識システム、波形合成システム、方法並びにプログラムに関する。

背景技術

[0002] 音声認識の分野において、認識性能の高精度化のため、話者の声道長の違いに起因するスペクトルの共鳴周波数の揺らぎを補正する方法が提案されている。このような技術は、声道長正規化と呼ばれている。声道長正規化の方法として、例えば、特許文献1(特許3632529号公報)では、共鳴周波数のずれをケプストラム空間上の線形変換を用いてモデル化することで、効率的にずれの度合いの推定している。

[0003] 特許文献1に提案されている声道長正規化の構成は、音声を分析しケプストラムを出力する分析部と、ケプストラムから伸縮の度合いを表す伸縮係数の値を推定する伸縮推定部と、伸縮係数の値を用いてケプストラムを線形変換する変換部で構成されている。

[0004] 変換部で用いられる線形変換は、オールパスフィルタの逆変換を用いることで、周波数軸上の変換をケプストラムの線形変換として表すことを利用している。この変換に用いられるパラメータの数は1つである。

[0005] 特許文献1では、伸縮係数の推定に認識で用いるHMM(Hidden Markov Model)を用いている。認識で用いるHMMは、特許文献1でも記述されているように、単語や音素などを単位とした音韻情報の出力確率をモデル化したものである。

[0006] 例えば「はい」を音素に分解し、「h a i」とし、これらh, a, iごとに、出現確率をモデ

ル化する。しばしば用いられる出現確率の形態として正規分布があげられ、この場合、音素ごとにケプストラム等の特徴量の値の平均と分散が事前に計算されて認識等の際に用いられる。特許文献1では、伸縮係数を推定するため、次式(1)が用いられる。

$$[0007] \quad \alpha = \frac{\sum_{j=1}^J \sum_{t=1}^T \gamma_t(j) \left[\sum_{m=1}^M \frac{(c_{mt} - \mu_{mj}) \{ (m-1)c_{(m-1)t} - (m+1)c_{(m+1)t} \}}{\sigma_{mj}^2} \right]}{\sum_{j=1}^J \sum_{t=1}^T \gamma_t(j) \left[\sum_{m=1}^M \frac{\{ (m-1)c_{(m-1)t} - (m+1)c_{(m+1)t} \}^2}{\sigma_{mj}^2} \right]} \quad (1)$$

[0008] ここで、J, jは音素や状態の数と音素や状態を同定するIDを表し、tは時刻、M, mはケプストラムの次元数と次元、 c_{mt} は時刻tのm次元目のケプストラム係数、 μ_{mj} 、 σ_{mj} は、音素j、m次元目のHMMにおける平均ベクトルと標準偏差の値を表す。

[0009] この式を用いて伸縮係数の値を推定する場合、平均ベクトルや分散を同定するための情報が必要となる。

[0010] 同定するためのID情報は、例えば音声の内容を記述する単語情報などを与えることで計算可能である。すなわち、前述の例の場合、「はい」が与えられると、「h a i」のような音素列情報を特定でき、さらにh, a, iそれぞれの状態系列に展開でき、状態に属する確率分布の特定もできる。

[0011] 特許文献1:特許第3632529号公報

非特許文献1:HTKBook Ver. 3. 3、第35～40頁、第54～64頁、第127～130頁

発明の開示

発明が解決しようとする課題

[0012] 上記の特許文献1、非特許文献1の各開示を本書に引用をもって繰り込むものとする。以下に本発明による関連技術の分析を与える。

[0013] 音素列等の取得方法として、予め発声内容を知ることの出来る場合は、人手により与えることが可能である。

[0014] 一方、話し言葉など事前に発話の内容を知りえない場合には、人手により音素情報を与えるには大きなコストがかかり現実的では無い。このような場合には、一旦、音声

認識を行うことで自動的に得る方法が取られることが多い。

[0015] しかしながら、音声認識の処理は比較的大きな演算量が必要であるため、限られた演算量の中で推定のための音素列情報生成だけのために、その演算を行うことは避ける必要がある。

[0016] また、音声認識から得られる音素列情報は間違いが含まれ、そのために推定精度の劣化を避けることが出来ない。

[0017] したがって、本発明の目的は、声道長正規化処理を行う際の伸縮係数の推定を少ない演算量で可能たらしめる伸縮係数推定装置、該装置を備えたケプストラム変換システム、音声認識システム、波形合成システム、方法並びにプログラムを提供することにある。

[0018] 本発明の他の目的は、声道長の違いを表す伸縮係数を推定の精度を向上することを可能とする伸縮係数推定装置、該装置を備えたケプストラム変換システム、音声認識システム、波形合成システム、方法並びにプログラムを提供することにある。

課題を解決するための手段

[0019] 本願で開示される発明は、前記課題を解決するため、概略以下の構成とされる。

[0020] 本発明の1つの側面(アスペクト)によれば、音声・非音声のラベル情報を出力するラベル情報生成手段と、音声・非音声の出現確率を表す確率モデルを記憶する伸縮モデル記憶手段と、音声・非音声の出現確率を表す確率モデルと音声・非音声のラベルとケプストラムを用いて周波数軸方向の伸縮係数を計算する伸縮推定手段と、を備えた伸縮係数推定装置が提供される。

[0021] 本発明の他の側面によれば、前記伸縮係数推定装置に、伸縮係数を用いてケプストラムの変換を行う変換手段を備えたケプストラム変換システムが提供される。

[0022] 本発明の他の側面によれば、前記ケプストラム変換システムに、音素単位などの音韻情報の出現確率を表す確率モデルを記憶する音響モデル記憶手段と、変換されたケプストラムと、音素単位などの音韻情報の出現確率を表す確率モデルを用いて音声認識を行う音声認識手段とを備えた音声認識システムが提供される。

[0023] 本発明のさらに他の側面によれば、前記ケプストラム変換システムによって変換されたケプストラムを用いて波形を合成する波形合成手段を備えた波形合成システム

が提供される。

[0024] 本発明のさらに別の側面においては、前記伸縮モデル記憶手段は、音声・非音声の出現確率を表す確率モデルを複数備え、前記伸縮推定手段は、前記伸縮モデル記憶手段に記憶されているモデル数の回数だけ、伸縮係数を求め、前記伸縮推定手段で求めた伸縮係数の中から伸縮係数を決定し、該伸縮係数に対応するモデル情報とともに出力する伸縮係数決定手段と、

前記伸縮推定手段で求めた伸縮係数の中から伸縮係数、モデル情報を決定する伸縮係数決定手段と、

前記モデル情報を基に音響モデル記憶手段に記憶されている音響モデルを選択して音声認識手段に供給する音響モデル切替手段と、を備え、

前記変換手段は、前記伸縮係数決定手段で決定された伸縮係数を用いてケプストラムの変換を行う構成としてもよい。

[0025] 本発明によれば、音声・非音声の出現確率を表す確率モデルを記憶手段に用意しておき、

音声・非音声のラベル情報を出力する工程と、

前記音声・非音声の出現確率を表す確率モデルと、前記音声・非音声のラベルと、ケプストラムと、を用いて周波数軸方向の伸縮係数を計算する工程と、

を含む、伸縮係数推定方法が提供される。

[0026] 本発明によれば、前記伸縮係数推定方法で求めた前記伸縮係数を用いてケプストラムの変換を行う工程を含むケプストラム変換方法が提供される。

[0027] 本発明によれば、前記ケプストラム変換方法で変換された前記ケプストラムと、音韻情報の出現確率を表す確率モデルを用いて、音声認識を行う工程と

を含む音声認識方法が提供される。

[0028] 本発明によれば、前記ケプストラム変換方法で変換された前記ケプストラムを用いて波形を合成する工程を含む波形合成方法が提供される。

本発明によれば、音声・非音声の出現確率を表す確率モデルを記憶手段に用意しておき、

音声・非音声のラベル情報を出力する処理と、

前記音声・非音声の出現確率を表す確率モデルと、前記音声・非音声のラベルと、ケプストラムと、を用いて周波数軸方向の伸縮係数を計算する処理と、をコンピュータに実行させる伸縮係数推定プログラムが提供される。

[0029] 本発明によれば、前記伸縮係数推定プログラムで求めた伸縮係数を用いて、ケプストラムの変換を行う処理をコンピュータに実行させるケプストラム変換プログラムが提供される。

[0030] 本発明によれば、前記ケプストラム変換プログラムで変換された前記ケプストラムと、音韻情報の出現確率を表す確率モデルを用いて音声認識を行う処理をコンピュータに実行させる音声認識プログラムが提供される。

[0031] 本発明によれば、ケプストラム変換プログラムで変換された前記ケプストラムを用いて波形を合成する処理をコンピュータに実行させる波形合成プログラムが提供される。

発明の効果

[0032] 本発明によれば、話者の大域的な情報を表すように設計されたHMM、すなわち音声と非音声の2クラスに絞った確率モデルを用いることで、声道長正規化処理を行う際の伸縮係数の推定を少ない演算量で実行可能としている。

[0033] 本発明によれば、声道長の違いを表す伸縮係数の推定精度を向上させることができる。

図面の簡単な説明

[0034] [図1]本発明の第1の実施例の構成を示すブロック図である。

[図2]本発明の第1の実施例の動作を示す流れ図である。

[図3]本発明の第2の実施例の構成を示すブロック図である。

[図4]本発明の第2の実施例の動作を示す流れ図である。

符号の説明

[0035] 101 分析手段

102 変換手段

103 伸縮推定手段

104 ラベル情報生成手段

105 伸縮モデル記憶手段

106 音声認識手段

107 音響モデル記憶手段

201 伸縮係数決定手段

202 音響モデル切り替え手段

発明を実施するための最良の形態

[0036] 本発明は、音声・非音声のラベル情報を出力するラベル情報生成手段(104)と、音声・非音声の出現確率を表す確率モデルを記憶する伸縮モデル記憶手段(107)と、音声・非音声の出現確率を表す確率モデルと音声・非音声のラベルとケプストラムを用いて周波数軸方向の伸縮係数を計算する伸縮推定手段(103)とを備えている。

[0037] 本発明においては、推定話者の声道長の違いから生じるフォルマンと周波数の揺らぎを抑えるために用いられる周波数軸上の伸縮関数の伸縮係数の推定において、話者の大域的な情報を表すように設計されたHMM、すなわち音声と非音声の2クラスに絞った確率モデルを用いることで、少ない演算量で伸縮係数の推定が可能になる。

[0038] 本発明においては、音声・非音声の2つのクラスに絞ることで、不確定な音素情報を推定に使う必要が無くなる。本発明によれば、音声・非音声の音素情報は、事前に音声認識を行うなどの処理を必要としないため演算量を必要としない。

[0039] 話者の大域的な情報である、声道長の違いの推定には、細かく音素単位でその差異を比較するよりも、音声と非音声のように大雑把な単位で推定したほうが良いと考えられる。また、与えられる音素情報等も、細かくすればするほど、間違いが含まれるのに対し、音声と非音声では間違いようがないことから、与えられる音素情報に影響されずに、高精度に推定可能である。

[0040] このため、従来の方法を用いた場合と比べて高精度に声道長の違いを表す伸縮係数を推定することができる。以下実施例に即して詳細に説明する。

実施例

[0041] 図1は、本発明の一実施例の構成を示す図である。以下では、音声認識に用いる

特徴量変換手段について説明する。図1を参照すると、本実施例は、分析手段101と、変換手段102と、伸縮推定手段103と、ラベル情報生成手段104と、伸縮モデル記憶手段105と、音声認識手段106と、音響モデル記憶手段107を備えている。

[0042] 特に制限されないが、分析手段101は、例えばサンプリング周波数を44.1kHz、1サンプルあたり16ビットでA/D変換された入力音声信号を、10msec程度の一定区間を切り出し、プリエンファシス、FFT(高速フーリエ変換)、フィルタバンクの処理を行った後、コサイン変換を行うことで計算されるメルケプストラムを出力する。

[0043] 音響モデル記憶手段107は、例えばケンブリッジ大学より公表されている、非特許文献1(HTKBook Ver. 3. 3の35ページから40ページ)に記載されている、前後の音素コンテキストを考慮したトライフォンを音素として持つHMMなどで表される音響モデルが記憶されている。

[0044] 伸縮モデル記憶手段105は、音声／非音声の2つのクラスだけを現すHMMを用いる。音響モデル記憶手段107に記憶されているHMMと区別するために、GMM(Gaussian Mixture Model)という。

[0045] HMMが音素コンテキストを考慮したトライフォンなどの音素を用いているのに対し、GMMは、話者ごとに周波数軸上の伸縮を用いて補正を行う声道長正規化の特徴をふまえ、細かな音素のクラスをモデル化したHMMを用いるのではなく、2つのクラスだけを保持する。

[0046] GMMの学習について述べる。まず、音声の音響的特長は前述のケプストラムや非特許文献1(54ページから64ページ)に記載されているような、パワーや前後の時刻の差分を用いる。

[0047] 音響的特徴とそれに対応する書き起こしテキストとを用いて得られたラベルデータを用い、非特許文献1(127ページから130ページ)に記載されているフォワード・バックワード確率を計算することで特徴とラベルデータ対応付ける。

[0048] 本実施例では、「sil voice sil」のような2クラスのラベル情報を用いる。ここで、「sil」は、非音声を表し、「voice」は音声を表す。

[0049] 対応付けられた特徴量とラベルデータを用いて、対応するラベルごとに特徴量の分布を表すパラメータを計算する。

- [0050] 分布を表す関数として正規分布が用いられることが多い。この場合、パラメータは、平均と分散である。
- [0051] GMMの学習に使われる音声は、後のHMMを用いた音声認識を考慮して、HMMの学習に用いた音声データを用いるのが望ましい。
- [0052] ラベル情報生成手段104は、学習時と同様に、「sil voice sil」のラベルデータを生成し、出力する。
- [0053] このとき、音声の長さに応じて、voiceラベルの数を変えることで、より正確な対応付けを行うことが出来る。これは、音声データが長い場合には、より多くの音声区間があると考えられ、音声データの長さに応じてvoiceラベルの数を増やすことで、不要な遷移をおこす確率を減らす効果がある。
- [0054] 伸縮推定手段103は、上記式(1)を用いて、周波数軸上の伸縮を表すパラメータの値を計算する。
- [0055] 式(1)における、Jはクラスの数を表し、本実施例では、1である。
- [0056] 式(1)において、Tは特徴量の数を表す。この場合、音声の長さを切り出された時間の単位で割った場合の数である。
- [0057] $\gamma_t(j)$ は、時刻tの音声クラスの占有度数を表す。この値は、前述のフォワード・バックワードアルゴリズムで計算可能である。
- [0058] ケプストラム係数の次元数である c_{mt} は、時刻tにおけるm次元目のケプストラム係数の値である。
- [0059] μ_{jm} は、音声クラスのm次元目の平均ベクトル、 σ_{mj} は、音声クラスのm次元目の標準偏差を表す。
- [0060] 伸縮係数 α は、特許文献1同様、周波数軸の伸縮の度合いを表すワーピングパラメータであり、伸縮推定手段103で計算する値である。
- [0061] 変換手段102は、伸縮推定手段103によって計算された伸縮係数 α を用いてケプストラムを変換し、出力する。ケプストラムの変換は、特許文献1同様、次式(2)で行われる。

$$\hat{c}_m^{(i)} = \begin{cases} c_{-i} + \alpha \hat{c}_0^{(i-1)}, & m = 0 \\ (1 - \alpha^2) \hat{c}_{m-1}^{(i-1)} + \alpha \hat{c}_1^{(i-1)}, & m = 1 \\ \hat{c}_{m-1}^{(i-1)} + \alpha (\hat{c}_m^{(i-1)} - \hat{c}_{m-1}^{(i)}), & m \geq 2 \end{cases}$$

$$i = -\infty, \dots, -1, 0. \quad (2)$$

[0062] ここで、 $\hat{\cdot}$ のついた c が変換後のケプストラムを表す。

[0063] 音声認識手段106は、変換手段102によって変換されたケプストラムを特徴量としてHMMを用いて認識を行う。

[0064] 本実施例の変形として、変換手段102で変換されたケプストラムを用いて波形の合成を行う波形合成手段を備えた構成としてもよい。すなわち、この伸縮推定手段103は、音声合成装置にも適用可能である。

[0065] なお、分析手段101、変換手段102、伸縮推定手段103、音声認識手段106は、コンピュータ(プロセッサ、あるいはデジタル信号プロセッサ)で実行されるプログラムによりその機能を実現するようにしてもよいことは勿論である。

[0066] 図2は、本発明の第1の実施例の処理を説明するための流れ図である。図1及び図2を参照して、本発明の第1の実施例の全体の動作について詳細に説明する。まず、音声データ信号を入力し(図2のステップA101)、分析手段101でメルケプストラムの計算を行う(ステップA102)。

[0067] 伸縮推定手段103が、分析手段101から出力されるメルケプストラムと、伸縮モデル記憶手段105のGMM(A104)と、ラベル情報生成手段104からの音声・非音声のラベル情報(A105)とを用いて、伸縮パラメータを計算する(ステップA103)。

[0068] 変換手段102が、式(2)のいずれかの変換の関数を用い、伸縮係数 α を用いてケプストラムを変換し、出力する(ステップA106)。

[0069] 音声認識手段106は、変換手段102によって変換されたケプストラムを特徴量として音響モデル記憶手段107のHMMを用いて認識を行う(ステップA107)。

[0070] 図3は、本発明の第2の実施例の構成を示す図である。本実施例は、前記第1の実施例に加え、伸縮係数決定手段201と、音響モデル切り替え手段202を備えている。

[0071] 伸縮モデル記憶手段105は、前記第1の実施例で説明されたGMMと同様のモデ

ルを複数記憶する。但し、複数とは学習された音声データの違うGMMである。例えば、男性と女性の2つのGMMの場合や、話者ごとに学習されたものなどが考えられる。

[0072] 伸縮推定手段103は、前記第1の実施例と同様、伸縮係数(ワーピングパラメータ) α を計算する。

[0073] 本実施例の場合、伸縮推定手段103の伸縮推定の動作は、複数回繰り返され、伸縮モデル記憶手段105に記憶されているGMMの数の回数だけ、ワーピングパラメータが計算される。また、ワーピングパラメータを計算するために、フォワード・バックワード確率も複数回計算される。

[0074] 伸縮係数決定手段201は、伸縮推定手段103で計算されたGMMごとのフォワード・バックワード確率のうち、一番高いものを選択し、GMM情報とワーピングパラメータ α の値を出力する。

[0075] GMM情報とは、選択されたGMMの情報であり、例えば複数のGMMが男性と女性の場合、男性、女性のいずれかを出力する。

[0076] また、GMMが話者ごとの場合、話者名を出力する。音響モデル切り替え手段202は、伸縮係数決定手段201で出力されたGMM情報を基に、音響モデル記憶手段107に記憶されているHMMを選択して出力する。

[0077] 図4は、本発明の第2の実施例の処理を説明するための流れ図である。図3及び図4を参照して、本発明の第2の実施例の全体の動作について詳細に説明する。まず、音声データ信号を入力し(図4のステップA201)、分析手段101でメルケプストラムの計算を行う(ステップA202)。

[0078] 伸縮推定手段103が、分析手段101から出力されるパターン(メルケプストラム)と、ラベル情報生成手段104からの音声・非音声のラベル情報(A205)とを用いて、伸縮モデル記憶手段105に記憶されているGMMの数(A204)の回数だけ、伸縮パラメータを計算する(ステップA203)。

[0079] 伸縮係数決定手段201が、選択したGMM情報と伸縮係数(ワーピングパラメータ) α の値を出力する(ステップA206)。

[0080] 変換手段102が、式(2)のいずれかの変換の関数を用い、伸縮係数 α を用いてケ

プストラムを変換し、出力する(ステップA207)。

- [0081] 音響モデル切り替え手段202は、伸縮係数決定手段201で出力されたGMM情報を基に、音響モデル記憶手段107に記憶されているHMMを選択する(ステップA208)。
- [0082] 音声認識手段106は、変換手段102によって変換されたケプストラムを特徴量として音響モデル記憶手段107のHMMを用いて認識を行う(ステップA209)。
- [0083] 本実施例によれば、このような構成をとることで、特定話者向けのモデルの選択と、ワーピングパラメータの推定を同時に行うことが出来る。
- [0084] 本実施例の変形として、変換手段102で変換されたケプストラムを用いて波形の合成を行う波形合成手段を備えた構成としてもよい。すなわち、この伸縮推定手段103は、音声合成装置にも適用可能である。
- [0085] なお、本実施例において、分析手段101、変換手段102、伸縮推定手段103、音声認識手段106、伸縮係数決定手段201、音響モデル切り替え手段202は、コンピュータ(プロセッサ、あるいはデジタル信号プロセッサ)で実行されるプログラムによりその機能を実現するようにしてもよいことは勿論である。

産業上の利用可能性

- [0086] 本発明によれば、音声認識を行う音声認識装置や、音声認識装置をコンピュータに実現するためのプログラムといった用途に適用できる。
- [0087] なお、本発明の全開示(請求の範囲を含む)の枠内において、さらにその基本的技術思想に基づいて、実施形態ないし実施例の変更・調整が可能である。また、本発明の請求の範囲の枠内において種々の開示要素の多様な組み合わせないし選択が可能である。すなわち、本発明は、請求の範囲を含む全開示、技術的思想にしたがって当業者であればなし得るであろう各種変形、修正を含むことは勿論である。

請求の範囲

- [1] 音声・非音声のラベル情報を出力するラベル情報生成手段と、
音声・非音声の出現確率を表す確率モデルを記憶する伸縮モデル記憶手段と、
前記音声・非音声の出現確率を表す確率モデルと、前記音声・非音声のラベルと、
ケプストラムとを用いて周波数軸方向の伸縮係数を計算する伸縮推定手段と、
を備える、ことを特徴とする伸縮係数推定装置。
- [2] 請求項1記載の伸縮係数推定装置と、
前記伸縮係数推定装置で求めた伸縮係数を用いてケプストラムの変換を行う変換
手段と、
を備える、ことを特徴とするケプストラム変換システム。
- [3] 入力音声信号を所定区間切り出して所定の変換を施しケプストラムを導出する分析
手段を備え、
前記変換手段は、前記分析手段で求めたケプストラムを、前記伸縮推定手段から
の伸縮係数を用いて変換する、ことを特徴とする請求項2記載のケプストラム変換シ
ステム。
- [4] 請求項2又は3記載のケプストラム変換システムと、
音素単位などの音韻情報の出現確率を表す確率モデルを記憶する音響モデル記
憶手段と、
前記変換されたケプストラムと、前記音響モデル記憶手段に記憶されている前記音
素単位などの音韻情報の出現確率を表す確率モデルを用いて、音声認識を行う音
声認識手段と、
を備える、ことを特徴とする音声認識システム。
- [5] 請求項2又は3記載のケプストラム変換システムと、
前記変換されたケプストラムを用いて波形を合成する波形合成手段と、
を備える、ことを特徴とする波形合成システム。
- [6] 請求項5記載の波形合成システムを備えた音声合成システム。
- [7] 前記伸縮モデル記憶手段は、音声・非音声の出現確率を表す確率モデルを複数
備え、

前記伸縮推定手段は、前記伸縮モデル記憶手段に記憶されているモデル数の回数だけ、伸縮係数を求め、

前記伸縮推定手段で求めた伸縮係数の中から伸縮係数を決定し、該伸縮係数に対応するモデル情報とともに出力する伸縮係数決定手段と、

前記モデル情報を基に、前記音響モデル記憶手段に記憶されている音響モデルを選択して音声認識手段に供給する音響モデル切替手段と、

をさらに備え、

前記変換手段は、前記伸縮係数決定手段で決定された伸縮係数を用いてケプストラムの変換を行う、ことを特徴とする請求項4記載の音声認識システム。

- [8] 音声・非音声の出現確率を表す確率モデルを記憶手段に用意しておき、
音声・非音声のラベル情報を出力する工程と、
前記音声・非音声の出現確率を表す確率モデルと、前記音声・非音声のラベルと、ケプストラムと、を用いて周波数軸方向の伸縮係数を計算する工程と、
を含む、ことを特徴とする伸縮係数推定方法。
- [9] 請求項8記載の伸縮係数推定方法で求めた前記伸縮係数を用いてケプストラムの変換を行う工程を含む、ことを特徴とするケプストラム変換方法。
- [10] 入力音声信号を所定区間切り出して所定の変換を施しケプストラムを導出する分析工程を含み、
前記ケプストラムを変換する工程は、前記分析工程で求めたケプストラムを、前記伸縮係数を用いて変換する、ことを特徴とする請求項9記載のケプストラム変換方法。
- [11] 請求項9又は10記載のケプストラム変換方法で変換された前記ケプストラムと、音韻情報の出現確率を表す確率モデルを用いて音声認識を行う工程と
を含む、ことを特徴とする音声認識方法。
- [12] 請求項9又は10記載のケプストラム変換方法で変換された前記ケプストラムを用いて波形を合成する工程を含む、ことを特徴とする波形合成方法。
- [13] 前記音声・非音声の出現確率を表す確率モデルを記憶手段に複数用意し、
前記モデル数の回数だけ、伸縮係数を求める工程と、

複数の伸縮係数の中から伸縮係数を決定し、該伸縮係数に対応するモデル情報を選択する工程と、

前記決定された伸縮係数を用いてケプストラムの変換を行う工程と、

選択されたモデル情報に基づき、音響モデルを選択して音声認識に供給する工程と、

を含む、ことを特徴とする請求項11記載の音声認識方法。

[14] 音声・非音声の出現確率を表す確率モデルを記憶手段に用意しておき、

音声・非音声のラベル情報を出力する処理と、

前記音声・非音声の出現確率を表す確率モデルと、前記音声・非音声のラベルと、ケプストラムと、を用いて周波数軸方向の伸縮係数を計算する処理と、

をコンピュータに実行させる伸縮係数推定プログラム。

[15] 請求項14記載の伸縮係数推定プログラムで求めた伸縮係数を用いて、ケプストラムの変換を行う処理をコンピュータに実行させるケプストラム変換プログラム。

[16] 入力音声信号を所定区間切り出して所定の変換を施しケプストラムを導出する分析処理で求めたケプストラムを、前記伸縮係数を用いて変換する、ことを特徴とする請求項15記載のケプストラム変換プログラム。

[17] 請求項15又は16記載のケプストラム変換プログラムで変換された前記ケプストラムと、音韻情報の出現確率を表す確率モデルを用いて音声認識を行う処理をコンピュータに実行させる音声認識プログラム。

[18] 請求項15又は16記載のケプストラム変換プログラムで変換された前記ケプストラムを用いて波形を合成する処理をコンピュータに実行させる波形プログラム。

[19] 請求項17記載の音声認識プログラムにおいて、

前記音声・非音声の出現確率を表す確率モデルを記憶手段に複数用意しておき、前記モデル数の回数だけ、伸縮係数を求める処理と、

複数の伸縮係数の中から伸縮係数を決定し、該伸縮係数に対応するモデル情報を選択する処理と、

前記決定された伸縮係数を用いてケプストラムの変換を行う処理と、

選択されたモデル情報に基づき、音響モデルを選択して音声認識に供給する処理

と、

を前記コンピュータに実行させる音声認識プログラム。

- [20] 入力音声信号を所定区間切り出して変換を施しケプストラムを導出する分析部と、
音声・非音声のラベル情報を出力するラベル情報生成部と、
音声・非音声の出現確率を表す確率モデルを記憶する伸縮モデル記憶部と、
前記伸縮モデル記憶部に記憶されている前記音声・非音声の出現確率を表す確率モデルと、前記ラベル情報生成部から出力される前記音声・非音声のラベルと、前記分析部から出力されるケプストラムとを用いて周波数軸方向の伸縮係数を計算する伸縮推定部と、

前記分析部から出力されるケプストラムを、前記伸縮推定部で求めた伸縮係数を用いて変換を行う変換部と、

音素単位などの音韻情報の出現確率を表す確率モデルを記憶する音響モデル記憶部と、前記変換部で変換されたケプストラムを特徴量として前記音響モデル記憶部に記憶されている音素単位などの音韻情報の出現確率を表す確率モデルを用いて、音声認識を行う音声認識部と、

を備える、ことを特徴とする音声認識装置。

- [21] 前記伸縮モデル記憶部は、音声・非音声の出現確率を表す確率モデルを複数備え、

前記伸縮推定部は、前記伸縮モデル記憶部に記憶されているモデル数の回数だけ、伸縮係数を求め、

前記音声認識装置はさらに、

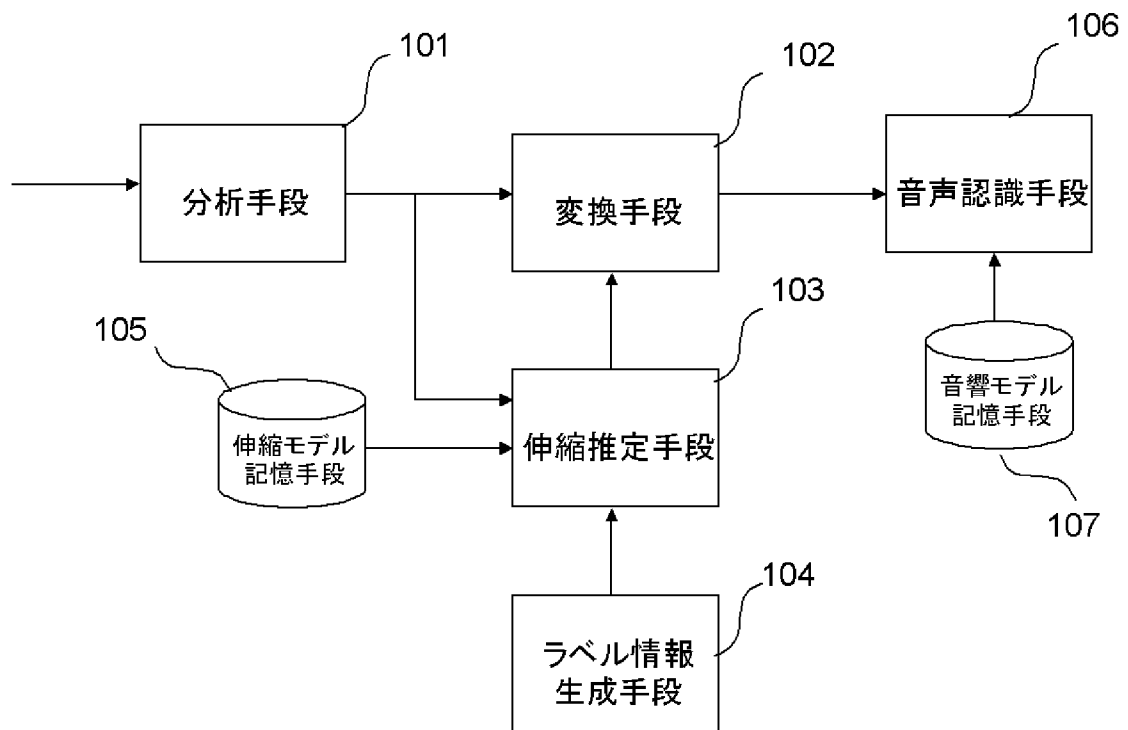
前記伸縮推定部で求めた伸縮係数の中から伸縮係数を決定し、該伸縮係数に対応するモデル情報とともに出力する伸縮係数決定部と、

前記モデル情報を基に、前記音響モデル記憶部に記憶されている音響モデルを選択して音声認識部に供給する音響モデル切替部と、

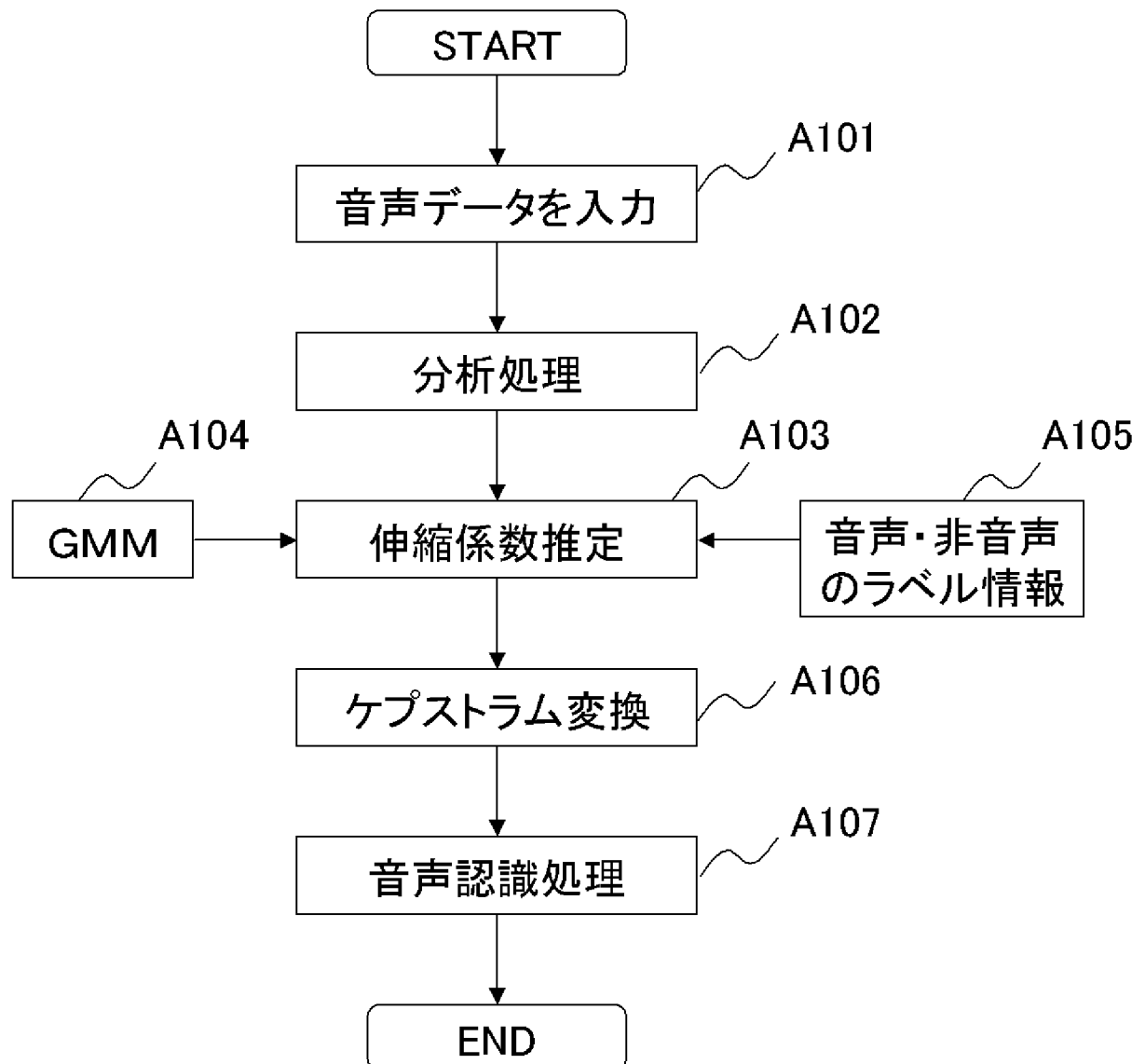
を備え、

前記変換部は、前記伸縮係数決定部で決定された伸縮係数を用いて、前記ケプストラムの変換を行う、ことを特徴とする請求項20記載の音声認識装置。

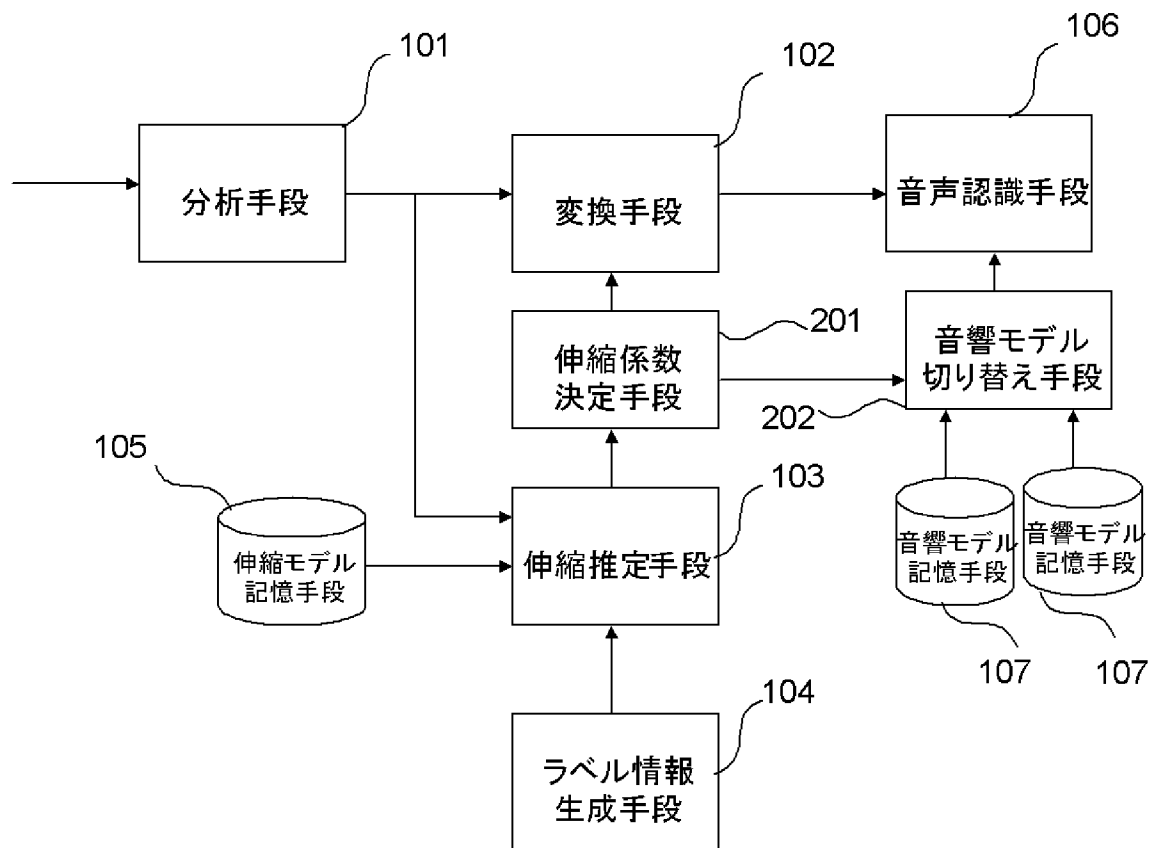
[図1]



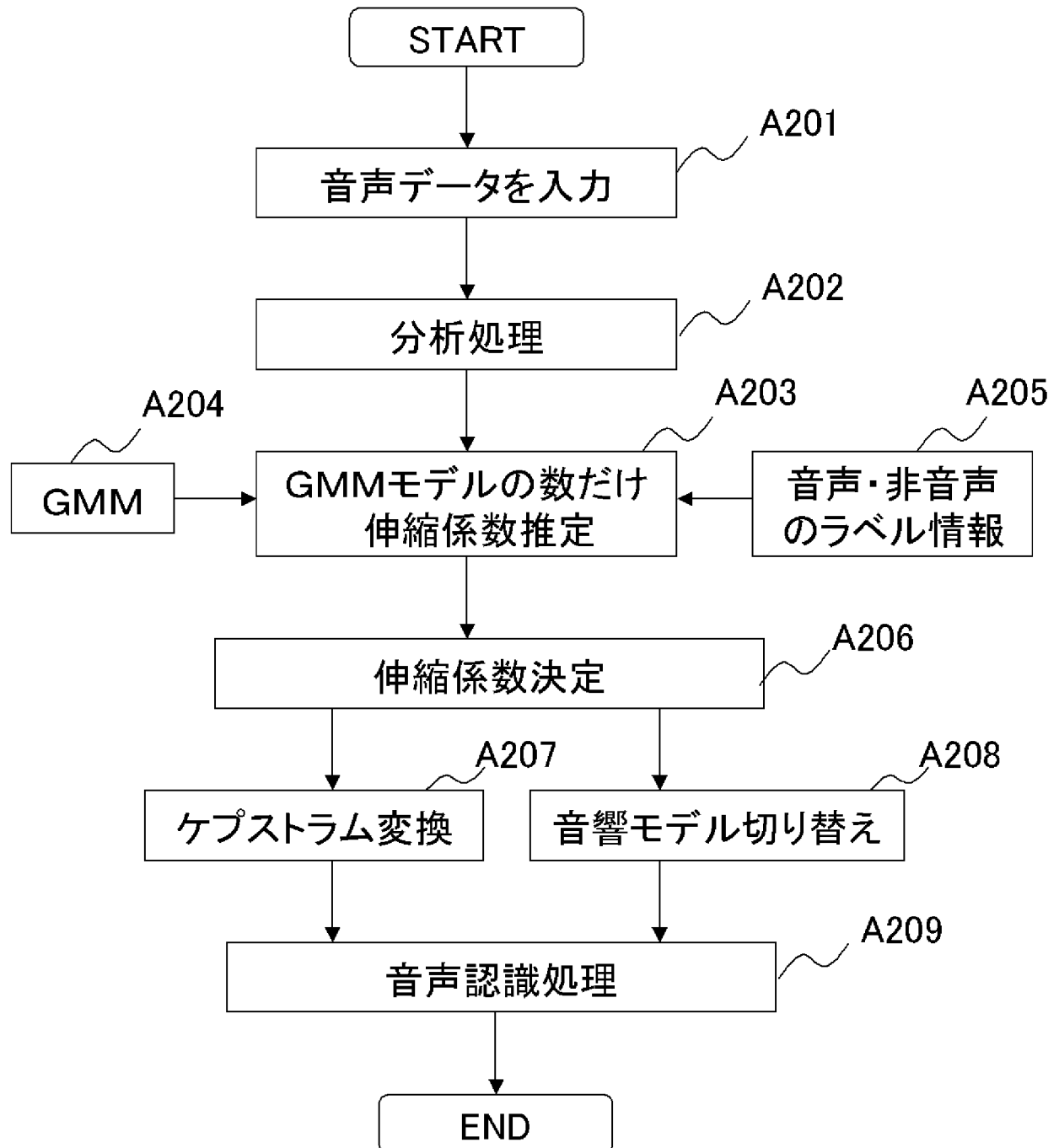
[図2]



[図3]



[図4]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/067125

A. CLASSIFICATION OF SUBJECT MATTER

G10L15/20(2006.01)i, G10L13/06(2006.01)i, G10L15/02(2006.01)i, G10L15/04(2006.01)i, G10L15/06(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)

G10L15/20, G10L13/06, G10L15/02, G10L15/04, G10L15/06

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	1996-2008
Kokai Jitsuyo Shinan Koho	1971-2008	Toroku Jitsuyo Shinan Koho	1994-2008

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 3632529 B2 (NEC Corp.), 07 January, 2005 (07.01.05), Full text; all drawings & US 6934681 B1 & EP 1096475 A2 & DE 60036522 D	1-21
Y	Keisuke NAKAMURA, "Identification of environmental noise and unnecessary utterance on a real information guidance system with spoken dialogue interface", IEICE Technical Report, 22 January, 2004 (22.01.04), Vol.103, No.632, pages 13 to 18	1-21
A	JP 6-214596 A (Ricoh Co., Ltd.), 05 August, 1994 (05.08.94), Full text; all drawings (Family: none)	1-21

☒ 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
15 October, 2008 (15.10.08)

Date of mailing of the international search report
28 October, 2008 (28.10.08)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/067125

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 11-327592 A (ATR Interpreting Telecommunications Research Laboratories), 26 November, 1999 (26.11.99), Full text; all drawings & US 6236963 B1	1-21

A. 発明の属する分野の分類 (国際特許分類 (I P C))

Int.Cl. G10L15/20(2006.01)i, G10L13/06(2006.01)i, G10L15/02(2006.01)i, G10L15/04(2006.01)i, G10L15/06(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 (I P C))

Int.Cl. G10L15/20, G10L13/06, G10L15/02, G10L15/04, G10L15/06

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 0 8 年
日本国実用新案登録公報	1 9 9 6 - 2 0 0 8 年
日本国登録実用新案公報	1 9 9 4 - 2 0 0 8 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 請求の範囲の番号
Y	JP 3632529 B2 (日本電気株式会社) 2005.01.07, 全文、全図 & US 6934681 B1 & EP 1096475 A2 & DE 60036522 D	1 - 2 1
Y	中村 敬介 Keisuke NAKAMURA, 実環境音声情報案内システムに おける環境雑音及び不要発話の識別 Identification of environmental noise and unnecessary utterance on a real information guidance system with spoken dialogue interface, 電 子情報通信学会技術研究報告 V o l . 1 0 3 N o . 6 3 2 IEICE Technical Report, 2004.01.22, 第103巻, 第632号, p.13-18	1 - 2 1

☒ 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

1 5 . 1 0 . 2 0 0 8

国際調査報告の発送日

2 8 . 1 0 . 2 0 0 8

国際調査機関の名称及びあて先

日本国特許庁 (I S A / J P)

郵便番号 1 0 0 - 8 9 1 5

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)

井上 健一

5 Z

9 3 7 3

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 4 1

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 6-214596 A (株式会社リコー) 1994.08.05, 全文、全図 (ファミリーなし)	1 - 2 1
A	JP 11-327592 A (株式会社エイ・ティ・アール音声翻訳通信研究所) 1999.11.26, 全文、全図 & US 6236963 B1	1 - 2 1