(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
2005 年10 月20 日 (20.10.2005)

PCT

(10) 国際公開番号
WO 2005/098820 A1

(51) 国際特許分類[7]: G10L 15/06, 15/28

(21) 国際出願番号: PCT/JP2005/005052

(22) 国際出願日: 2005 年3 月15 日 (15.03.2005)

(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2004-102904 2004 年3 月31 日 (31.03.2004) JP

(71) 出願人 (米国を除く全ての指定国について): パイオニア株式会社 (PIONEER CORPORATION) [JP/JP]; 〒1538654 東京都目黒区目黒１丁目４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): 小林 載 (KOBAYASHI, Hajime) [JP/JP]; 〒3502288 埼玉県鶴ヶ島市富士見６丁目１番１号 パイオニア株式会社 総合研究所内 Saitama (JP). 外山 聡一 (TOYAMA, Soichi) [JP/JP]; 〒3502288 埼玉県鶴ヶ島市富士見６丁目１番１号 パイオニア株式会社 総合研究所内 Saitama (JP). 鈴木 康悟 (SUZUKI, Yasunori) [JP/JP]; 〒3502288 埼玉県鶴ヶ島市富士見６丁目１番１号 パイオニア株式会社 総合研究所内 Saitama (JP).

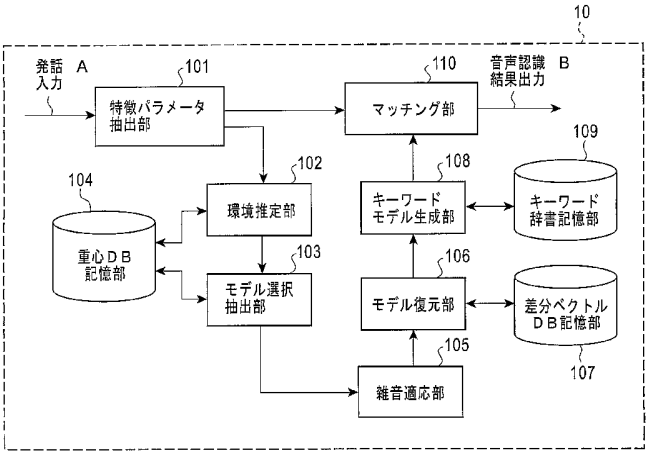
(74) 代理人: 藤村 元彦 (FUJIMURA, Motohiko); 〒1040045 東京都中央区築地４丁目１番１７号 銀座大野ビル 藤村国際特許事務所 Tokyo (JP).

(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, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU,

[続葉有]

(54) Title: SPEECH RECOGNITION DEVICE AND SPEECH RECOGNITION METHOD

(54) 発明の名称: 音声認識装置及び音声認識方法



A SPEECH INPUT
101 CHARACTERISTIC PARAMETER EXTRACTION UNIT
110 MATCHING UNIT
B SPEECH RECOGNITION RESULT OUTPUT
102 ENVIRONMENT ESTIMATION UNIT
108 KEYWORD MODEL GENERATION UNIT
109 KEYWORD DICTIONARY STORAGE UNIT
104 CENTER-OF-GRAVITY DB STORAGE UNIT
103 MODEL SELECTION EXTRACTION UNIT
106 MODEL RESTORATION UNIT
107 DIFFERENCE VECTOR DB STORAGE UNIT
105 NOISE PROCESSING UNIT

(57) Abstract: There are provided a speech recognition device and a speech recognition method capable of improving the noise processing function in speech recognition and reducing the memory amount used. In advance, an acoustic model is subjected to clustering processing, and a center of gravity of each cluster and a difference vector between the center of gravity and each model are calculated. Estimated noise models

[続葉有]

WO 2005/098820 A1

ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, 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, MC, 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

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。

are model-combined with the calculated center of gravity and the combined center of gravity and the difference vector are stored in a memory. In actual recognition, an optimal center of gravity for the environment estimated by an utterance environment estimation is extracted from the memory. By using the difference vector stored in the memory on the center of gravity extracted, a model restoration is performed and noise processing is executed by using the model.

(57) 要約: 音声認識処理における雑音適応処理の機能を向上させ、かつ使用メモリの量を低減させた音声認識装置及び音声認識方法を提供する。予め、音響モデルをクラスタリング処理して各クラスタの重心とその重心と各モデルとの差分ベクトルを算出し、想定される各種雑音モデルと算出した重心とのモデル合成を行って合成された各々の重心と差分ベクトルをメモリに格納する。実際の認識処理において、発話環境推定によって推定された環境に最適な重心を同メモリより抽出して、該抽出された重心に同メモリに記憶された差分ベクトルを用いてモデル復元を行い同モデルにより雑音適応処理を実行する。

明細書

音声認識装置及び音声認識方法

技術分野

本発明は、例えば、発話音声を認識する音声認識装置及び音声認識方法等に関する。

背景技術

音声認識処理の実用性を向上させるためには、発話音声自体に対する認識技術の向上みならず、発話音声に重畳された環境雑音に対する処理技術の向上が重要となる。従来、かかる環境雑音に対する耐性を向上させて音声認識処理の実用性を高める、いわゆる雑音適応処理技術が開発されており、例えば、特許文献１や非特許文献１に示されるような技術が開示されている。

前者の場合は、音響モデルの分布（以下、単に“分布”と言う）を予め決められた数のグループ（以下“クラスタ”と言う）で分類する。具体的には、類似した複数の分布が同一のグループに割り当てられるようにクラスタリング処理を行い、所定の数だけのクラスタが生成されるようにする。そして、各クラスタの重心に対してモデル合成処理を行う。

このため、一般に知られるモデル合成処理に比較して、高速度のモデル合成が可能となる。しかしながら、１つの分布に対するモデル合成処理に必要とされる演算量が多く、例えば、カーナビゲーション装置などのいわゆる組込み機器へ応用する際に、その性能維持を前提とすると、処理時間や機器実装スペー

ス上の制限から機器への実装が困難となるおそれがある。

一方、後者の場合は、予め種々の雑音環境カテゴリについて複数の雑音モデルをメモリ内に準備しておき、音声入力信号の雑音環境カテゴリに応じた雑音モデルを選択して、雑音環境に応じた雑音適応処理を行うことができる。しかし、予め準備する雑音環境カテゴリや雑音モデルの数が増加するとそれを記憶するメモリ量も飛躍的に増大するので、携帯機器や車両搭載機器に組み込まれる組込み型の音声認識装置での利用が困難であった。

【特許文献１】特開平１０-１６１６９２

【非特許文献１】日本音響学会講演論文集（１-８-４；２０００年３月）「雑音環境へのヤコビ適応法の拡張」（赤江俊彦　他３名）

発明の開示

本発明が解決しようとする課題には、音声認識処理における雑音適応処理の機能を向上させ、かつ使用メモリの量を低減させた音声認識装置、及び音声認識方法を提供することが一例として挙げられる。

請求項１に記載の発明は、入力音声信号について雑音モデルに基づく雑音適応処理を施して前記入力音声信号に対する音声認識を行う音声認識装置であって、複数の雑音環境カテゴリの各々に含まれる複数の雑音モデルを代表する第１パラメータを予め算出してこれを前記雑音環境カテゴリ毎に記憶する第１記憶手段と、前記複数の雑音モデルの各々と前記第１パラメータとの相対的な位置情報を表す第２パラメータを予め算出してこれを記憶する第２記憶手段と、前記入力音声信号に重畳された環境雑音の特性に基づいて前記環境雑音の属する雑音環境カテゴリを推定する推定手段と、前記推定手段によって推定された

雑音環境カテゴリに対応する前記第１パラメータを前記第１記憶手段から選択抽出する選択手段と前記選択手段によって抽出された第１パラメータと前記第２記憶手段から読み出した前記第２パラメータとを用いて前記環境雑音に適合する雑音モデルを復元し、これを用いて前記入力音声信号に対する雑音適応処理を行う雑音適応手段とを含むことを特徴とする。

また、請求項７に記載の発明は、入力音声信号について雑音モデルに基づく雑音適応処理を施して前記入力音声信号に対する音声認識を行う音声認識方法であって、複数の雑音環境カテゴリの各々に含まれる複数の雑音モデルを代表する第１パラメータを予め算出してこれを前記雑音環境カテゴリ毎に第１のメモリに記憶するステップと、前記複数の雑音モデルの各々と前記第１パラメータとの相対的な位置情報を表す第２パラメータを予め算出してこれを第２のメモリに記憶するステップと、前記入力音声信号に重畳された環境雑音の特性に基づいて前記環境雑音の属する雑音環境カテゴリを推定するステップと、該推定された雑音環境カテゴリに対応する前記第１パラメータを前記第１のメモリから選択抽出するステップと該選択抽出された第１パラメータと前記第２のメモリから読み出した前記第２パラメータとを用いて前記環境雑音に適合する雑音モデルを復元し、これを用いて前記入力音声信号に対する雑音適応処理を行うステップとを含むことを特徴とする。

図面の簡単な説明

図１は、本発明による音声認識装置の実施例を示すブロック図である。

図２は、図１の音声認識装置における準備段階処理の過程を示すフローチャートである。

図３は、図２の処理過程で形成されるクラスタの構成を示す模式図である。

図４は、図１の音声認識装置における重心データベース記憶部１０４の内容を示す構成図である。

図５は、図１の音声認識装置における音声認識処理の過程を示すフローチャートである。

図６は、図５の環境推定処理における雑音カテゴリの選択の様子を示す説明図である。

図７は、本発明による音声認識装置の第２の実施例を示すブロック図である。

図８は、非定常パラメータ除外処理の一例を示すフローチャートである。

図９は、非定常パラメータ除外処理を適用例を示す雑音パラメータの図である。

発明を実施するための形態

図１に本発明の実施例である音声認識装置を示す。同図に示される音声認識装置１０は、例えば、同装置のみが単体で用いられる構成であっても良いし、或いは、携帯電話やカーナビゲーション機器等の他の装置に内蔵される構成としても良い。

図１において、特徴パラメータ抽出部１０１は、入力された発話音声信号を発話区間、及びその前後の非発話区間を含めて、これらの区間に存する音響信号をその音響的特徴を表す特徴パラメータに変換して抽出する部分である。

環境推定部１０２は、非発話区間の特徴パラメータに基づいて、入力された発話音声信号に重畳された雑音の環境カテゴリを推定する部分である。なお、

雑音の環境カテゴリとは、例えば、自動車に関係する雑音の場合であれば、エンジンによる雑音が１つの雑音カテゴリに相当し、また、カーエアコンによる雑音も１つの雑音カテゴリに相当するモデル選択抽出部１０３は、環境推定部１０２で推定されたカテゴリに含まれる雑音モデルに関する各種データを、重心データベース記憶部１０４（以下、単に“記憶部１０４”と言う）から選択抽出する部分である。また、雑音適応処理部１０５は、上記の選択抽出された各種データを用いて、例えば、ヤコビ適応手法などの手法による雑音適応処理を実行する部分である。

モデル復元部１０６は、上記の雑音適応処理の結果に基づいて、差分ベクトルデータベース記憶部１０７（以下、単に“記憶部１０７”と言う）に予め記憶されている差分ベクトルを用いてモデル復元処理を行う部分である。なお、記憶部１０４、記憶部１０７に予め記憶されている各種データや差分ベクトルの詳細については後述する。

キーワードモデル生成部１０８は、モデル復元部１０６から出力される音響モデルを基に、キーワード辞書記憶部１０９（以下、単に“記憶部１０９”と言う）にストアされている語彙の中から認識候補となるものを抽出して、音響パターンとしてのキーワードモデルを生成する部分である。

マッチング部１１０は、特徴パラメータ抽出部１０１から供給される発話区間の特徴パラメータを、キーワードモデル生成部１０８で生成されたキーワードモデルの各々に当て嵌め、各々についてのマッチング尤度を求めて、入力された発話音声信号に対する音声認識処理を行う部分である。

なお、図１のブロック図における信号の流を示す矢印の向きは、各構成要素

間の主要な信号の流れを示すものであり、例えば、かかる信号に付随する応答信号や監視信号等の各種信号に関しては、矢印の向きと逆方向に伝達される場合をも含むものとする。また、図中に示される各構成要素の区分や信号経路は、動作説明の便宜上為されているものであり実際の装置においてかかる表記どおりの構成が実現される必要はない。

次に、記憶部１０４、及び記憶部１０７に予め記憶される、重心データ等の各種データ、及び差分ベクトルデータの生成処理について説明を行う。

本実施例では、実施の準備段階として上記の各々のデータを予め生成し、上記の各記憶部のメモリ内に準備しておく必要がある。かかる準備段階におけるデータ生成処理の様子を、図２のフローチャートに基づいて説明する。

先ず、図２のステップＳ２０１において、入力された音響モデルに対し類似した音響モデルの分布（以下“分布”と言う）をグループ化するクラスタリング処理が行われる。クラスタリング処理によって形成されるグループ（以下“クラスタ”と言う）の数は予め設定されており、生成されるクラスタの数が、かかる定数に達するまでクラスタリング処理が継続される。なお、ステップＳ２０１では、クラスタリング処理の進行と同時に、各々の分布が何れのクラスタに属するかなどの情報を示すクラスタ情報が生成される。

次のステップＳ２０３において、各々のクラスタに属する分布を全て使用して、それぞれのクラスタにおける重心計算が行われる。重心計算の処理を具体的に示せば、１つのクラスタｉに属する音響モデルをｍ(ｎ)（ｎ＝１，２，…，ｐ：ｐはクラスタｉに属する
音響モデルの個数)、計算により算出される重心をｇ(ｉ)とすると、重心ｇ(ｉ)

は、

【式１】

$$g(i) = \frac{1}{p}\sum_{n=1}^{p} m(n)$$

として求められる。

ステップＳ２０３の処理が終了すると、次のステップＳ２０５において、各クラスタに属する音響モデルの分布と、クラスタにおける重心との差分が計算される。具体的には、クラスタｉに属する音響モデルをｍ(ｎ)、クラスタの重心をｇ(ｉ)、及び差分を表す差

分ベクトルをｄ(ｎ) とすると、差分ベクトルｄ(ｎ) は、

$$d(n) = m(n) - g(i)$$

として求めることができる。

ステップＳ２０５において算出された差分ベクトルの値は、音声認識装置１０の記憶部１０７に格納される。なお、クラスタｉに含まれる各々の分布ｍ(ｎ)と、クラスタの重

心ｇ(ｉ)、及び差分ベクトルｄ(ｎ) との相互関係を、図３のクラスタ構造の模式図に示す。

次に、ステップＳ２０７において所定の雑音モデルを用意し、これとステップＳ２０３で求めた各クラスタの重心とのモデル合成処理が行われる。モデル合成後の各クラスタの重心は、音声認識装置１０の記憶部１０４に格納される。

本実施例では、環境雑音カテゴリ毎に雑音モデルを複数準備して、各雑音モデルに対してモデル合成処理が行われる。それ故、モデル合成処理前の１つのクラスタ重心から、モデル合成処理に供された雑音モデルの個数分の合成後クラスタが生成される。

なお、ステップＳ２０７のモデル合成処理によって求められた各種データが、記憶部１０４の内部に格納される様子を図４に示す。同図に示される如く、記憶部１０４では各々の環境雑音カテゴリ毎に

(1) １つの雑音モデル

(3) １つの音響モデルに対して、図２に示される処理を施すことによって得られるクラスタの重心

(2) 雑音適応部１０５における雑音適応処理に必要とされる、各クラスタの重心に対応する情報

の３種類のデータが格納されることになる。

以上に説明した準備段階の処理によって、本実施例に基づく音声認識装置１０の、記憶部１０４、及び記憶部１０７には、それぞれ所定のデータが予め格納される。

次に、音声認識装置１０による、発話音声の音声認識処理時における動作について、図１のブロック図、及び図５のフローチャートに基づいて説明を行う。

先ず、音声認識装置１０に発話音声信号が入力されると、図５に示されるステップＳ３０１の特徴パラメータ変換処理が実行される。すなわち、音声認識装置１０の特徴パラメータ抽出部１０１は、入力された発話音声信号を発話区間、及びその前後の非発話区間を含めて、ＬＰＣケプストラムやＭＦＣＣ（メ

ル周波数ケプストラム係数）などの発話パラメータに変換する。なお、用いられる発話パラメータの種類は、かかる事例に限定されるものではなく、発話信号の音響的特徴を表現したパラメータであり、かつこれらの音響パラメータのフォーマットと同じパラメータを用いて表現された音響モデルであれば同様に用いることが可能である。

ステップＳ３０１の特徴パラメータ変換処理が為されると、特徴パラメータ抽出部１０１から、発話区間の特徴パラメータがマッチング部１１０に、非発話区間の特徴パラメータが環境推定部１０２にそれぞれ供給される。

次のステップＳ３０３の環境推定処理は、記憶部１０４に格納されている複数の環境雑音カテゴリの中から所定のカテゴリを選択するため、入力された発話音声に重畳される環境雑音を推定する処理である。すなわち、環境推定部１０２は、非発話区間の特徴パラメータに基づいて入力された発話信号の環境雑音を推定し、これに該当する雑音カテゴリを求める。前述した如く、記憶部１０４のデータベースには、異なる環境雑音のカテゴリの代表にあたる雑音モデルが記憶されている。

環境推定部１０２は、先ず、これらの雑音モデルと、上記の非発話区間の特徴パラメータに基づいて、各雑音カテゴリ毎の雑音尤度を算出する。

これらの雑音モデルには、ターゲットとされる多数の環境雑音のデータベースを利用して計算された特徴パラメータの平均値や共分散が含まれている。したがって、各環境カテゴリ毎の雑音尤度は、雑音モデルの平均・分散から得られる正規分布に、上記非発話区間の特徴パラメータである発話パラメータを当て嵌めることによって求めることができる。

かかる事例として、雑音カテゴリ１乃至３の３つの雑音モデルの正規分布に、上記非発話区間の特徴パラメータである発話パラメータを当て嵌めて、雑音尤度を求める場合の事例を図６に示す。図６の事例では、雑音カテゴリ１ー３の雑音モデルに、入力発話信号の環境雑音を示す発話パラメータを当て嵌めたとき、雑音カテゴリ２の雑音尤度が他の２つ比較して高いことを表している。それ故、同図においては、環境雑音カテゴリの推定結果として雑音カテゴリ２が選択されることになる。

ステップＳ３０３の環境推定処理によって、入力発話音声に重畳された環境雑音のカテゴリが推定されると、次のステップＳ３０５において雑音モデルの選択抽出処理が行われる。すなわち、モデル選択抽出部１０３は、環境推定部１０２によって推定された雑音カテゴリに関する各種のデータを、記憶部１０４のデータベース中から選択してこれを抽出する。

図４に示される如く、記憶部１０４のデータベースには、音響モデルの分布をクラスタリング処理し、それぞれのクラスタ重心を雑音モデルで合成した重心データ、当該雑音モデル、及び各重心に対応する雑音適応処理用のデータが、各々の環境雑音のカテゴリ毎に格納されている。ステップＳ３０５において、選択された雑音カテゴリに属するこれらのデータが、記憶部１０４から雑音適応部１０５にロードされることになる。

そして、次のステップＳ３０７において、雑音適応部１０５による雑音適応処理が実行される。かかる雑音適応処理には、種々の手法を用いることが可能である。例えば、ヤコビ適応手法による雑音適応処理を行う際には、各クラスタの重心に対応するヤコビ行列についても予め重心データベースとして記憶部

１０４に格納して置く。そして、ステップＳ３０７における雑音適応処理を行う段階において、記憶部１０４から雑音適応部１０５に該当する雑音カテゴリのヤコビ行列データを読み出し、かかるデータを使用してヤコビ適応手法による雑音適応処理を行う。

ステップＳ３０７の雑音適応処理が終了すると、次のステップＳ３０９において、予め記憶部１０７に準備しておいた差分ベクトルのデータを用い、モデル復元部１０６によるモデル復元処理が行われる。かかる処理によって雑音適応後の音響モデルが準備される。

その後、ステップＳ３１１においてキーワードモデル生成処理が実行される。同処理において、キーワードモデル生成部１０８は、音声認識の認識候補となる語彙を記憶部１０９から抽出して、これを音響パターンとして定型化するキーワードモデル生成処理を実行する。

次の、ステップＳ３１３では、マッチング部１１０によるマッチング処理が実行される。すなわち、キーワードモデル生成部１０８において生成されたキーワードモデルの各々に対して、特徴パラメータ抽出部１０１から供給される発話区間の特徴パラメータが照合され、各々のキーワードに対するマッチングの度合いを示すキーワード尤度が算出される。そして、かかる処理によって求められたキーワード尤度のうちで、最も高い値を示すキーワードが、入力された発話音声に対する認識結果として、音声認識装置１０から出力される。

以上に説明したように、本実施例によれば、複数の初期合成モデルのセットを複数の初期合成クラスタに変換して記憶することによって使用メモリの使用量を削減することができ、雑音適応処理能力の高い音声認識装置を組み込み機

器への搭載することが容易となる。また、準備段階におけるクラスタリングをモデル合成を行う前に実施するので差分ベクトルを共用化することが可能となり、必要とされる構成の簡易化と性能向上を同時に達成することができる。

また、本実施例に話者適応機能を追加して、かつ話者適応を差分ベクトルで行う構成とした場合、重心データベースの内容をグレードアップする際に、発話者の特徴を反映させた状態でグレードアップされた環境での発話認識が可能となる。

図７は、本発明による音声認識装置の第２の実施例を示すブロック図である。

図において、図１に示された音声認識装置１０の各構成部分と同一の部分には同様の参照番号を付し、説明は繰り返さない。

図示された音声認識装置２０は図１に示された音声認識装置１０の各構成部分に加えて特徴パラメータ抽出部１０１と環境推定部１０２との間に設けられた非定常パラメータ除外処理部１１１を有することが特徴である。

非定常パラメータ除外処理部１１１は、特徴パラメータ抽出部１０１から供給される特徴パラメータの集合（雑音集合と称する）の中から非定常パラメータに該当するものを除外する動作を行う。

図８は、非定常パラメータ除外処理部１１１によって行われる非定常パラメータ除外処理の一例を示すフロー図である。

図示の如く、非定常パラメータ除外処理部１１１はステップＳ４０１において入力された１つの雑音集合を複数のグループに分類するクラスタリング処理を行う。次に、ステップＳ４０２において各クラスタ間の重心間類似度を求める。続いて、ステップＳ４０３においてクラスタ間の類似度の最も低いものが

所定の閾値以下であるか否かが判別される。

ステップＳ４０３において類似度が所定の閾値以下であれば、処理はステップＳ４０４に進み要素数（集合に属する特徴パラメータの数を指す）の少ない方のクラスタに属する特徴パラメータを除外する。続いて処理はステップＳ４０５に進み適応雑音モデルの生成（図６のステップＳ３０３の環境推定処理に対応）を行う。またステップＳ４０３において類似度が所定の閾値より大であれば、ステップＳ４０４の処理は実行せずにステップＳ４０５に進む。

ステップＳ４０５の実行に続いてステップＳ４０６において環境カテゴリ選択処理（図６のステップＳ３０５のモデル選択抽出処理に対応）が行われる。

次に、非定常パラメータ除外処理の適用例について説明する。

図９は、環境Ａに該当する雑音パラメータに突発性雑音の影響を受けた雑音パラメータが混在する雑音集合を示している。

非定常パラメータ除外処理を実行しない場合には雑音集合全体に対する重心を求め、この重心と環境Ａ、環境Ｂをそれぞれあらわす雑音モデルとの類似度を求めることにより環境判別が行われる。このため、図示の如く突発性雑音の影響を受け、雑音集合の重心は環境Ａの雑音モデルよりも環境Ｂの雑音モデルの方が類似性が高くなり、環境Ｂとして誤判定されてしまう。

一方、第２の実施例の如く、非定常パラメータ除外処理を実行する場合には突発性雑音と判断される雑音パラメータが除外され、かかる雑音パラメータを除外した雑音パラメータの重心が求められる。これにより。求められた重心と環境Ａの雑音モデルとの類似度が高くなり、環境カテゴリをより適切に選択することが可能となる。

このように非定常パラメータ除外処理を行うことにより音声認識の性能を向上させることができる。

なお、上記の第１及び第２の実施例において、発話音声入力に重畳された環境雑音が予め準備された雑音モデルに該当しない場合は、この環境雑音を、例えば、ＨＤＤ等の別記憶媒体に所定期間蓄積した後、かかる蓄積データを基にしてこれに対する重心値等のデータを算出するようにしても良い。これによって、重心データベースにおける新たな雑音カテゴリを環境に合わせて増加させることが可能となる。

さらに、通信ネットワークを介して大規模なデータベースを擁するサーバーとデータの授受を行う通信手段を設けることにより、重心データベースや差分ベクトルデータベースのメモリの一部若しくは全部を、これらのサーバーのデータベースに置き換えるようにしても良い。

請求の範囲

【請求項１】

入力音声信号について雑音モデルに基づく雑音適応処理を施して前記入力音声信号に対する音声認識を行う音声認識装置であって、

複数の雑音環境カテゴリの各々に含まれる複数の雑音モデルを代表する第１パラメータを予め算出してこれを前記雑音環境カテゴリ毎に記憶する第１記憶手段と、

前記複数の雑音モデルの各々と前記第１パラメータとの相対的な位置情報を表す第２パラメータを予め算出してこれを記憶する第２記憶手段と、

前記入力音声信号に重畳された環境雑音の特性に基づいて前記環境雑音の属する雑音環境カテゴリを推定する推定手段と、

前記推定手段によって推定された雑音環境カテゴリに対応する前記第１パラメータを前記第１記憶手段から選択抽出する選択手段と

前記選択手段によって抽出された第１パラメータと前記第２記憶手段から読み出した前記第２パラメータとを用いて前記環境雑音に適合する雑音モデルを復元し、これを用いて前記入力音声信号に対する雑音適応処理を行う雑音適応手段と、を含むことを特徴とする音声認識装置。

【請求項２】

前記第１パラメータは、音響モデルについてクラスタリング処理を施して得られる重心値と１つの雑音環境カテゴリに含まれる複数の雑音モデルの各々に対して施されたモデル合成処理によって得られる前記雑音モデル毎の重心値、及びこれに付随するデータであることを特徴とする請求項１に記載の音声認識

装置。

【請求項３】

前記入力音声信号に重畳された環境雑音の特徴パラメータの集合の中から非定常パラメータを除外する非定常パラメータ除外処理部を更に有することを特徴とする請求項１に記載の音声認識装置。

【請求項４】

前記クラスタリング処理は、当該処理によって形成される音響モデルのグループ数が所定のグループ数に到達するまで継続されることを特徴とする請求項２に記載の音声認識装置。

【請求項５】

前記第２パラメータは、前記重心値と前記複数の雑音モデルの各々との間における差分ベクトルであることを特徴とする請求項２乃至４のいずれか１項に記載の音声認識装置。

【請求項６】

前記推定手段は、前記環境雑音が予め準備された雑音環境カテゴリに該当しないことを検知した場合に、前記環境雑音から前記第１パラメータを抽出してこれを前記第１記憶手段に追加して記憶する記憶追加手段、をさらに含むことを特徴とする請求項１ないし５のいずれか１項に記載の音声認識装置。

【請求項７】

通信ネットワークを介してデータベースを含むサーバーと前記第１及び第２記憶手段に含まれるメモリとの間におけるデータを中継する通信手段をさらに含み、前記データベースを前記メモリの一部若しくは全部として利用すること

を特徴とする請求項１ないし６のいずれか１項に記載の音声認識装置。

【請求項８】

入力音声信号について雑音モデルに基づく雑音適応処理を施して前記入力音声信号に対する音声認識を行う音声認識方法であって、

複数の雑音環境カテゴリの各々に含まれる複数の雑音モデルを代表する第１パラメータを予め算出してこれを前記雑音環境カテゴリ毎に第１のメモリに記憶するステップと、

前記複数の雑音モデルの各々と前記第１パラメータとの相対的な位置情報を表す第２パラメータを予め算出してこれを第２のメモリに記憶するステップと、

前記入力音声信号に重畳された環境雑音の特性に基づいて前記環境雑音の属する雑音環境カテゴリを推定するステップと、

該推定された雑音環境カテゴリに対応する前記第１パラメータを前記第１のメモリから選択抽出するステップと

該選択抽出された第１パラメータと前記第２のメモリから読み出した前記第２パラメータとを用いて前記環境雑音に適合する雑音モデルを復元し、これを用いて前記入力音声信号に対する雑音適応処理を行うステップと、を含むことを特徴とする音声認識方法。

【請求項９】

前記入力音声信号に重畳された環境雑音の特徴パラメータの集合の中から非定常パラメータを除外するステップを更に有することを特徴とする請求項８に記載の音声認識方法。

FIG. 1

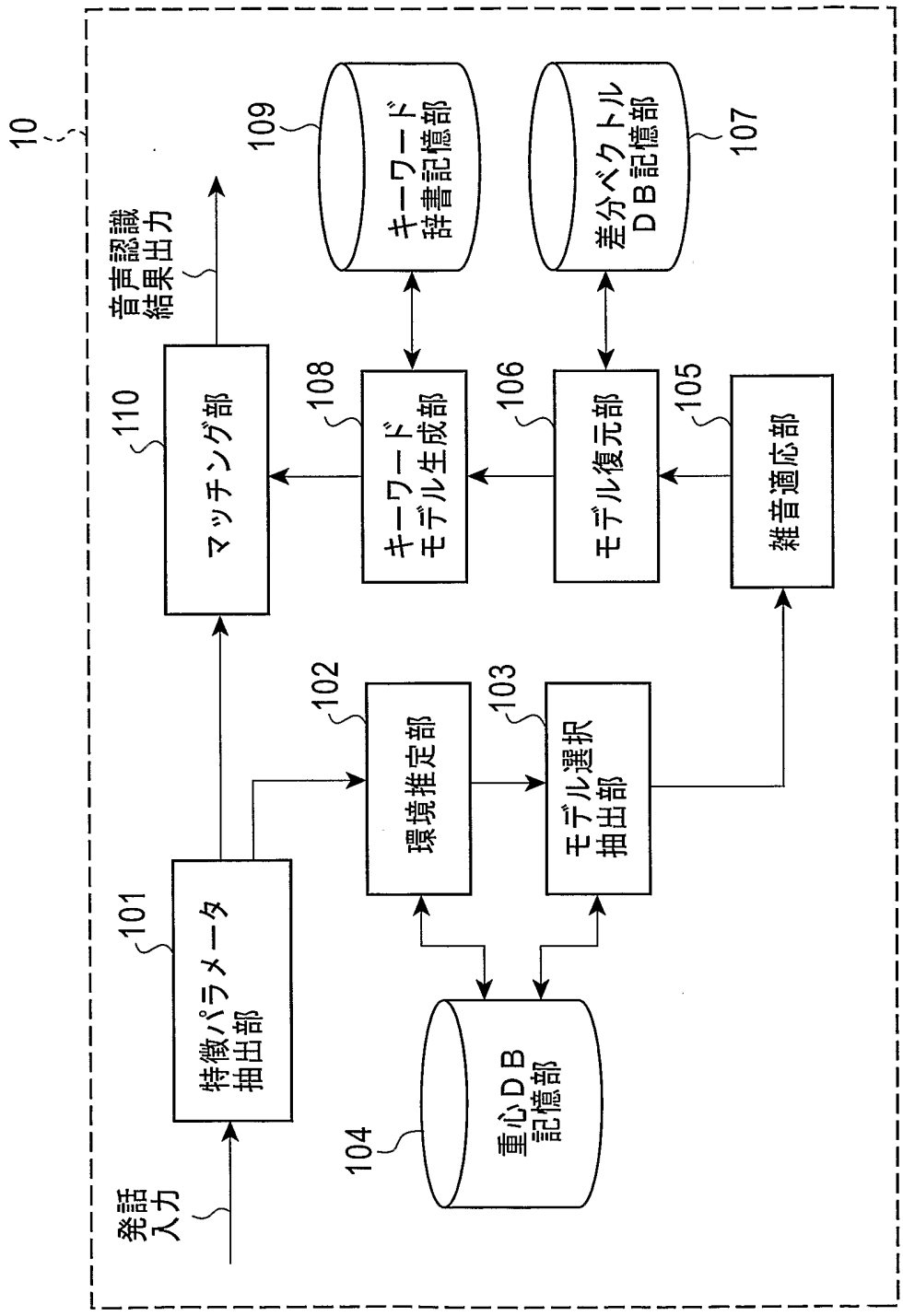

10
発話入力
特徴パラメータ抽出部
101
環境推定部
102
モデル選択抽出部
103
重心DB記憶部
104
雑音適応部
105
モデル復元部
106
差分ベクトルDB記憶部
107
キーワード辞書記憶部
109
キーワードモデル生成部
108
マッチング部
110
音声認識結果出力

FIG. 2

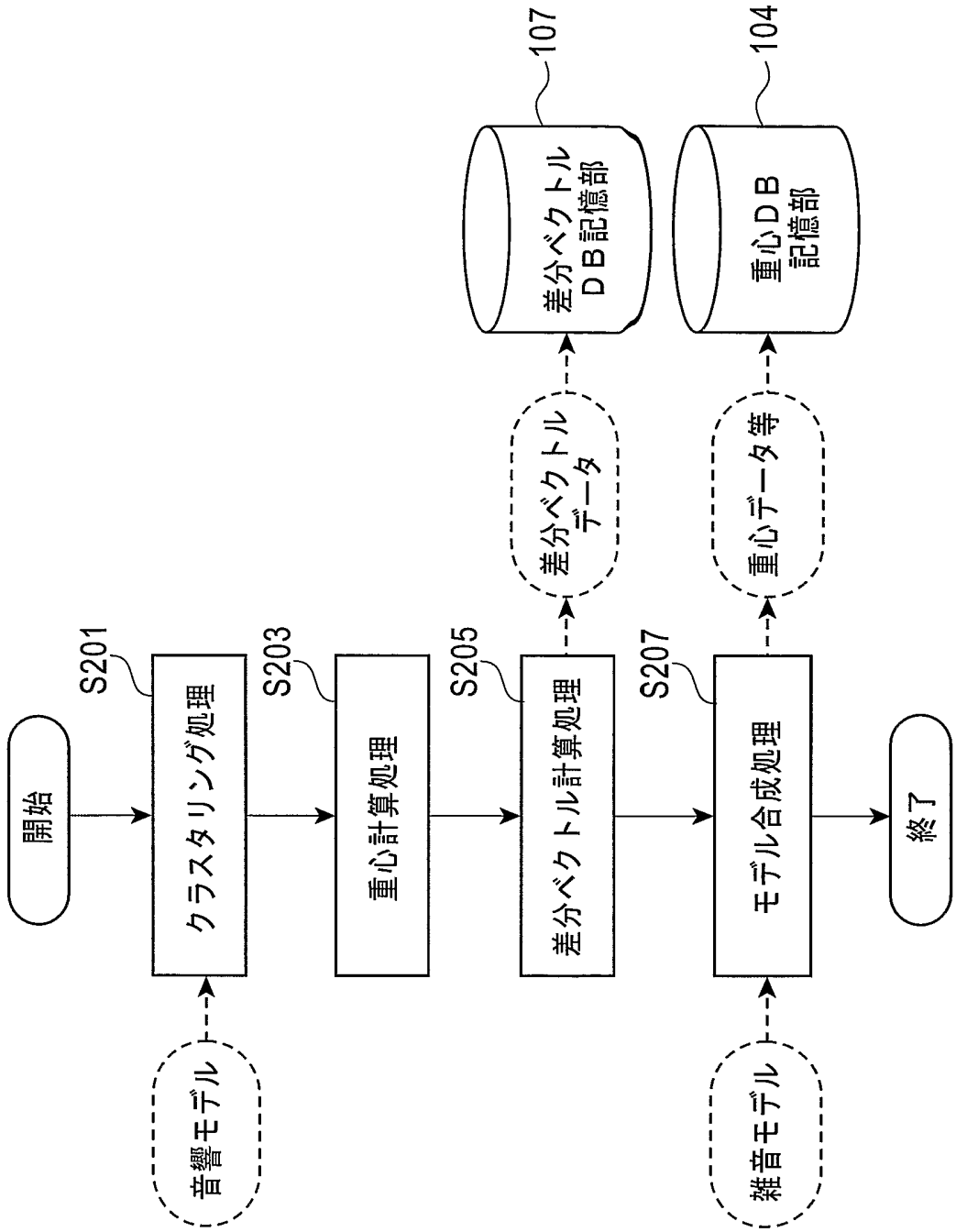

開始
S201
クラスタリング処理
S203
重心計算処理
S205
差分ベクトル計算処理
S207
モデル合成処理
終了
音響モデル
差分ベクトルデータ
重心データ等
雑音モデル
107
差分ベクトルDB記憶部
104
重心DB記憶部

FIG. 3

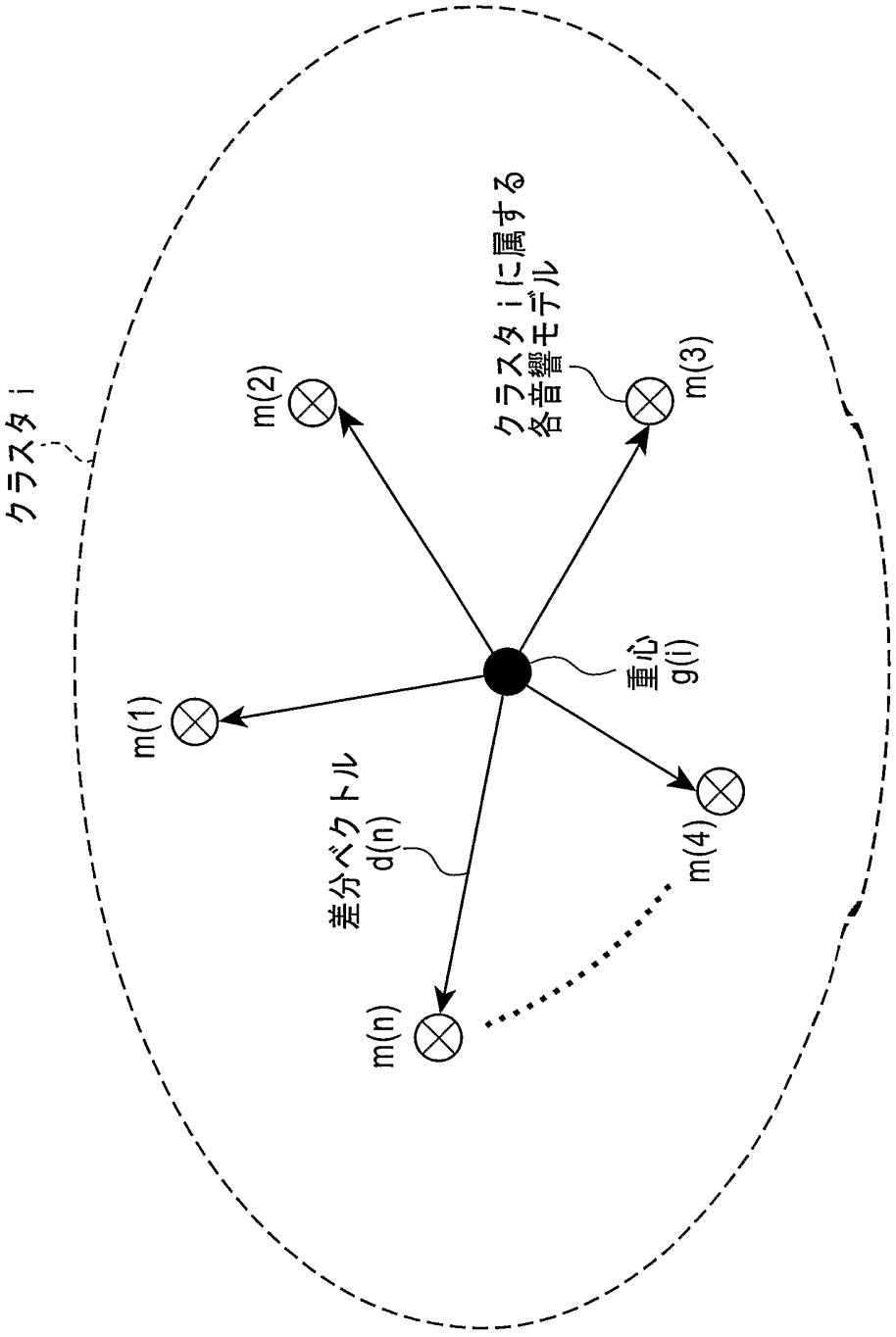


FIG. 4

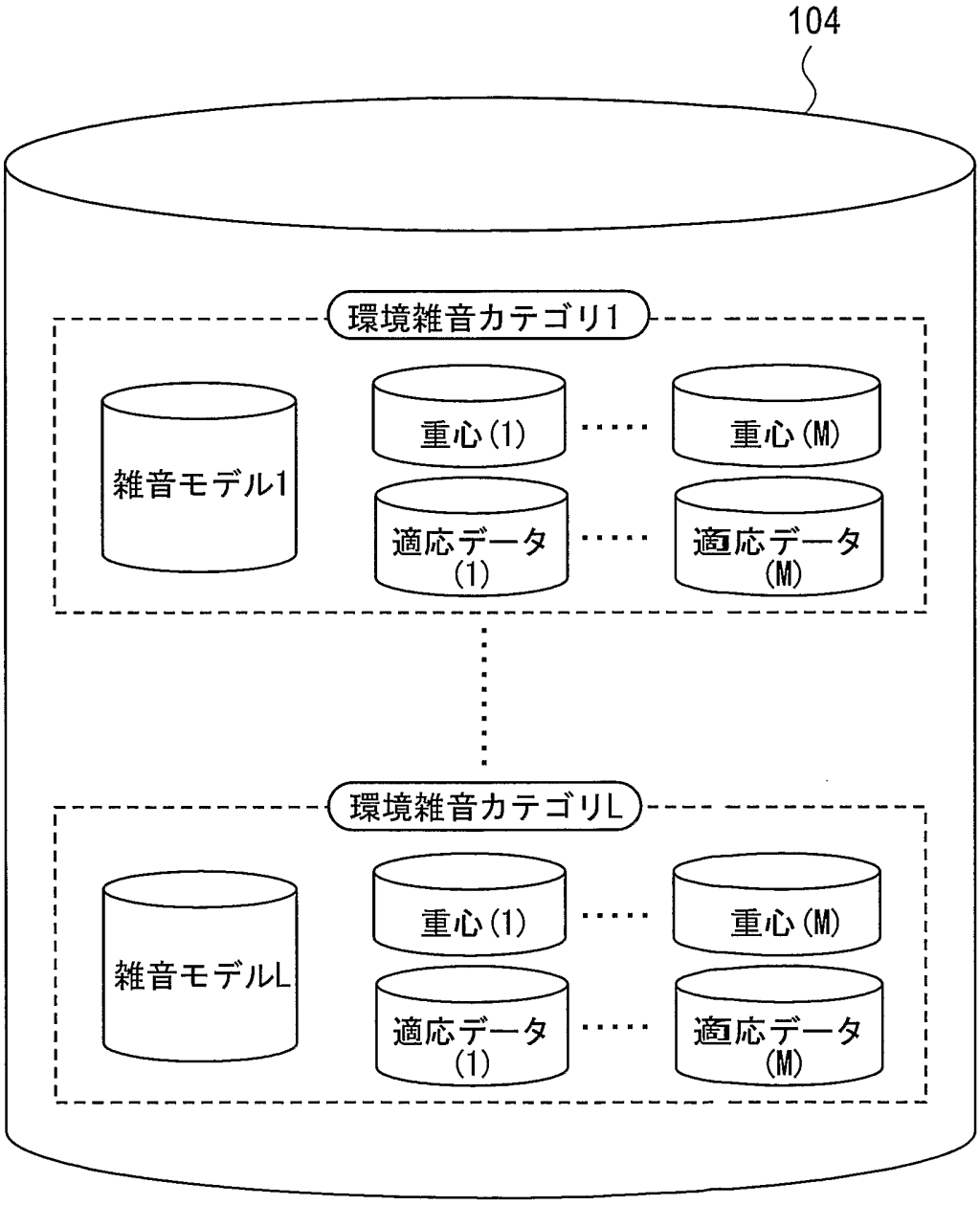

104
環境雑音カテゴリ1
雑音モデル1
重心(1)
重心(M)
適応データ(1)
適応データ(M)
環境雑音カテゴリL
雑音モデルL
重心(1)
重心(M)
適応データ(1)
適応データ(M)

FIG. 5

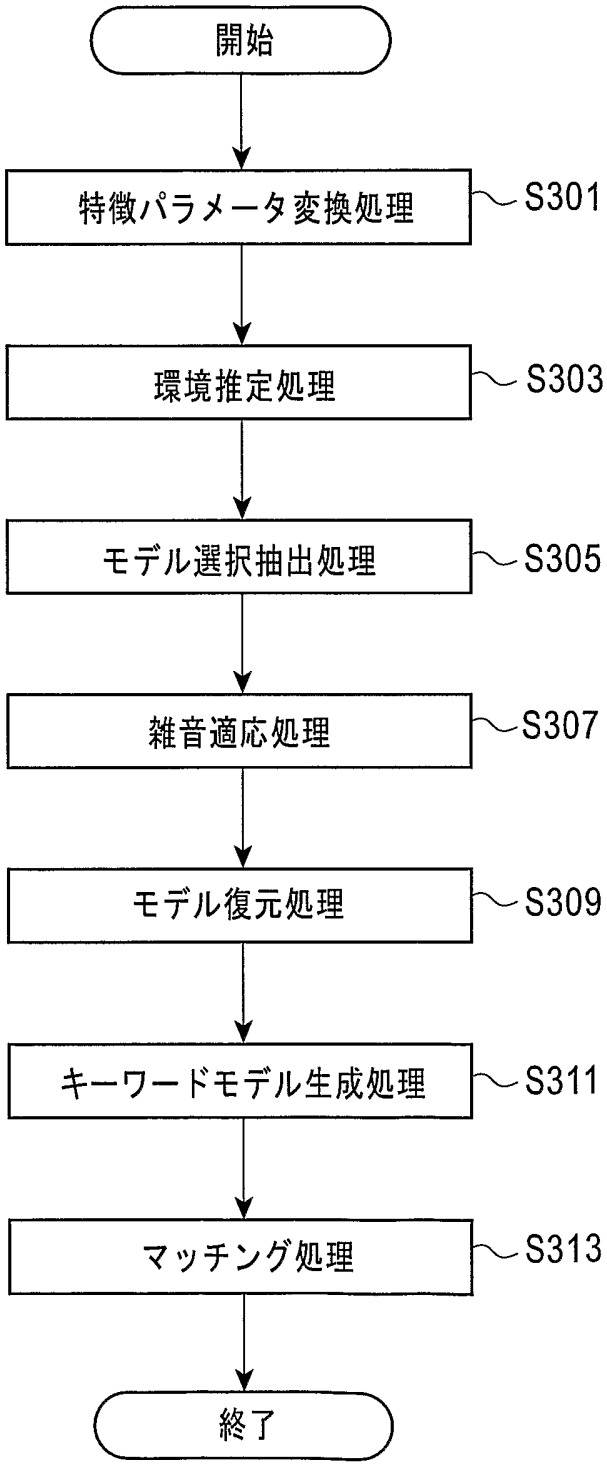

開始
特徴パラメータ変換処理
S301
環境推定処理
S303
モデル選択抽出処理
S305
雑音適応処理
S307
モデル復元処理
S309
キーワードモデル生成処理
S311
マッチング処理
S313
終了

FIG. 6

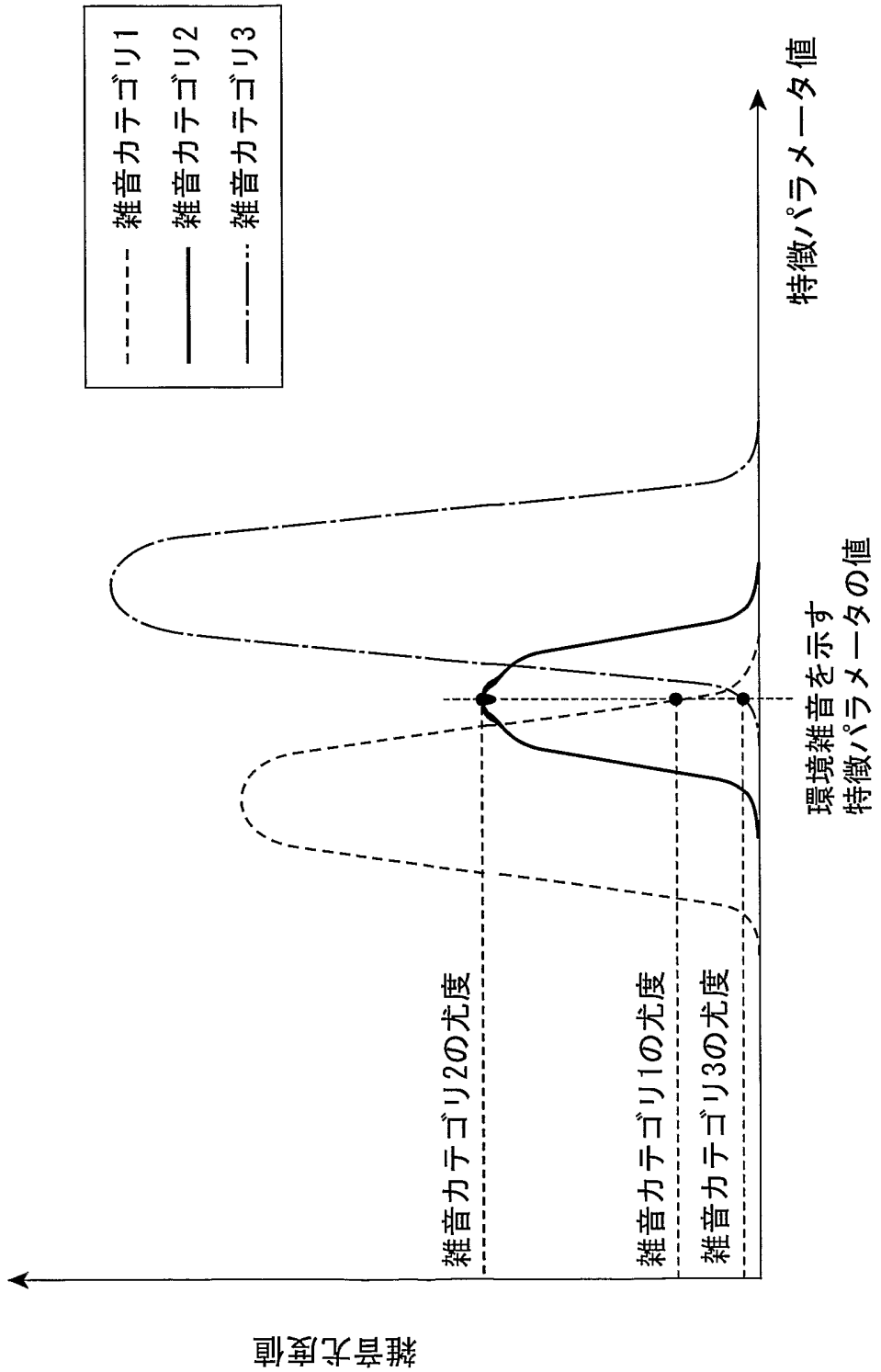

雑音カテゴリ1
雑音カテゴリ2
雑音カテゴリ3
特徴パラメータ値
雑音尤度値
雑音カテゴリ2の尤度
雑音カテゴリ1の尤度
雑音カテゴリ3の尤度
環境雑音を示す
特徴パラメータの値

FIG. 7

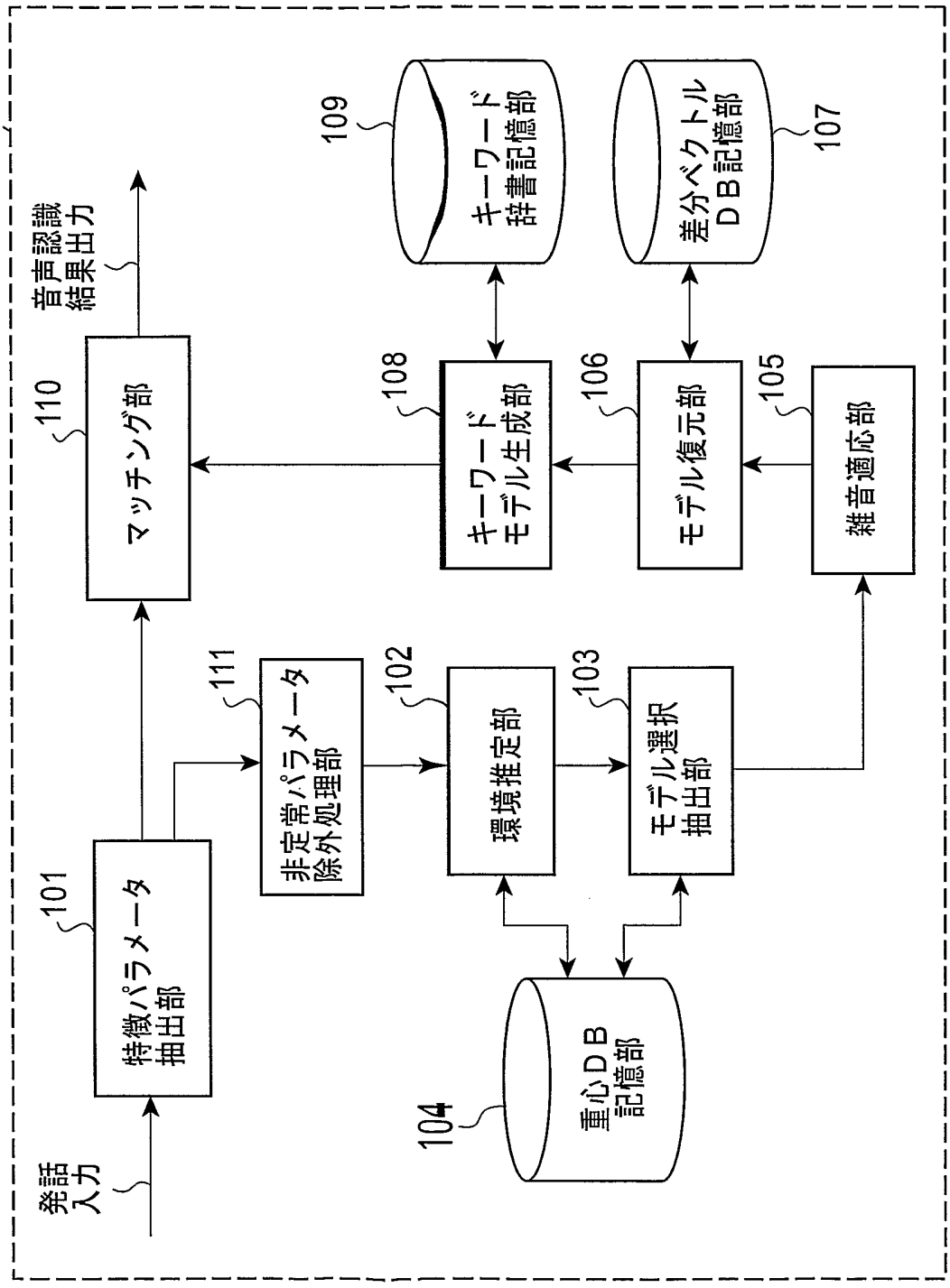

20
発話入力
特徴パラメータ抽出部
101
非定常パラメータ除外処理部
111
環境推定部
102
モデル選択抽出部
103
重心DB記憶部
104
雑音適応部
105
モデル復元部
106
キーワードモデル生成部
108
マッチング部
110
音声認識結果出力
キーワード辞書記憶部
109
差分ベクトルDB記憶部
107

FIG. 8

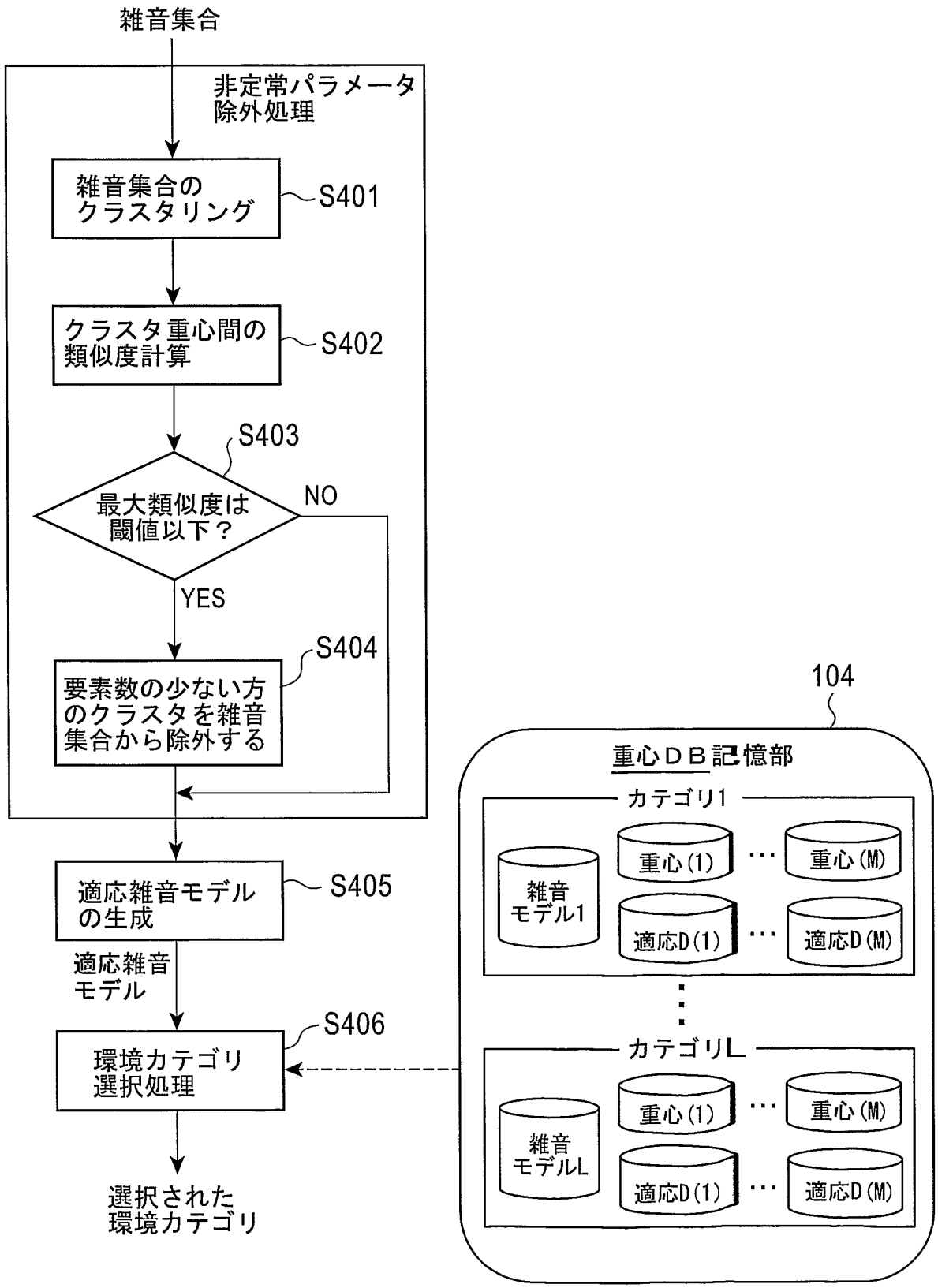
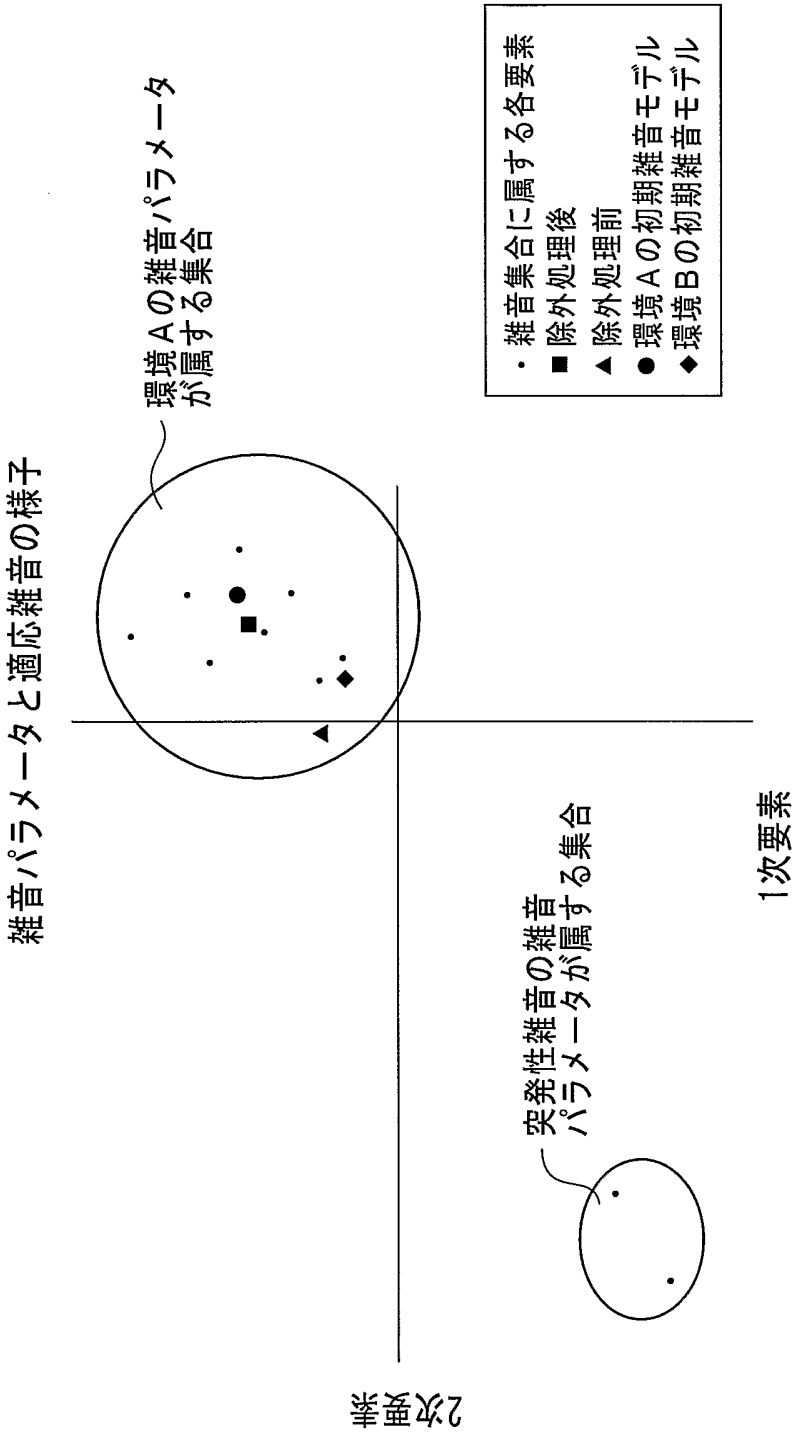

雑音集合
非定常パラメータ除外処理
雑音集合のクラスタリング
S401
クラスタ重心間の類似度計算
S402
S403
最大類似度は閾値以下？
NO
YES
S404
要素数の少ない方のクラスタを雑音集合から除外する
適応雑音モデルの生成
S405
適応雑音モデル
環境カテゴリ選択処理
S406
選択された環境カテゴリ
104
重心ＤＢ記憶部
カテゴリ1
雑音モデル1
重心(1)
重心(M)
適応D(1)
適応D(M)
カテゴリL
雑音モデルL
重心(1)
重心(M)
適応D(1)
適応D(M)

FIG. 9


雑音パラメータと適応雑音の様子
環境Ａの雑音パラメータ
が属する集合
・雑音集合に属する各要素
■ 除外処理後
▲ 除外処理前
● 環境Ａの初期雑音モデル
◆ 環境Ｂの初期雑音モデル
突発性雑音の雑音
パラメータが属する集合
1次要素
2次要素

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2005/005052

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl[7] G10L15/06, 15/28

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)
Int.Cl[7] G10L15/00-15/28

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 1996-2005
Kokai Jitsuyo Shinan Koho 1971-2005 Toroku Jitsuyo Shinan Koho 1994-2005

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 1400952 A1 (Pioneer Corp.), 24 March, 2004 (24.03.04), Full text; Figs. 1 to 9 & JP 2004-109464 A	1-9
A	JP 2003-330484 A (Pioneer Corp.), 19 November, 2003 (19.11.03), Full text; Figs. 1 to 4 & US 2003-220791 A1 & EP 1357541 A2 & CN 1453767 A	1-9
A	JP 2002-91485 A (Pioneer Corp.), 27 March, 2002 (27.03.02), Full text; Figs. 1 to 4 & US 2002-52741 A1 & EP 1189205 A2 & CN 1345030 A	1-9

☒ 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
14 June, 2005 (14.06.05)	05 July, 2005 (05.07.05)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2005/005052

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 2002-14692 A (Matsushita Electric Industrial Co., Ltd.), 18 January, 2002 (18.01.02), Full text; Figs. 1 to 9 & US 2002-55840 A1 & EP 1168301 A2 & CN 1331467 A	1-9
A	JP 10-161692 A (Canon Inc.), 19 June, 1998 (19.06.98), Full text; Figs. 1 to 7 & US 5956679 A1 & EP 847041 A3	1-9
A	JP 10-149191 A (Nippon Telegraph And Telephone Corp.), 02 June, 1998 (02.06.98), Full text; Figs. 1 to 8 & EP 831461 A2	1-9
A	JP 2000-330587 A (NEC Corp.), 30 November, 2000 (30.11.00), Full text; Figs. 1 to 9 (Family: none)	3,9
A	JP 9-258765 A (Kokusai Denshin Denwa Co., Ltd. (KDD)), 03 October, 1997 (03.10.97), Full text; Figs. 1 to 14 (Family: none)	3,9
A	AKAE et al., "Zatsuon Kankyo eno Jacobi Tekioho no Kakucho", The Acoustical Society of Japan (ASJ) Koen Ronbunshu, 2000 Nen Shunki I, 15 March, 2000 (15.03.00), 1-8-4, pages 7 to 8	1-9
A	IDA et al., "Zatsuon DB to Model Tekioka o Mochiita HMM Goseiho ni okeru Zatsuon Hendo Taisei no Hyoka", The Acoustical Society of Japan (ASJ) Koen Ronbunshu, 2004 Nen Shuki I, 02 October, 2001 (02.10.01), 1-1-17, pages 33 to 34	1-9
A	NOGUCHI et al., "1 Channel Nyuryoku Shingochu no Toppatsusei Zatsuon no Hanbetsu to Jokyo", The Acoustical Society of Japan (ASJ) Koen Ronbunshu, 2004 Nen Shunki I, 17 March, 2004 (17.03.04), 3-P-30, pages 655 to 656	3,9

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２００５／００５０５２

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.[7]　G10L15/06, 15/28

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.[7]　G10L15/00-15/28

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	１９７１－２００５年
日本国実用新案登録公報	１９９６－２００５年
日本国登録実用新案公報	１９９４－２００５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）

IEEE
JICST ファイル（JOIS）

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	EP 1400952 A1 (Pioneer Corporation) 2004.03.24, 全文、第1-9図 & JP 2004-109464 A	1-9
A	JP 2003-330484 A (パイオニア株式会社) 2003.11.19, 全文、第1-4図 & US 2003-220791 A1 & EP 1357541 A2 & CN 1453767 A	1-9
A	JP 2002-91485 A (パイオニア株式会社) 2002.03.27, 全文、第1-4図 & US 2002-52741 A1 & EP 1189205 A2 & CN 1345030 A	1-9

☑ 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 １４．０６．２００５	国際調査報告の発送日 05.7.2005
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　9559 石丸　昌平 電話番号　０３－３５８１－１１０１　内線　３５４１

様式ＰＣＴ／ＩＳＡ／２１０（第２ページ）（２００４年１月）

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 2002-14692 A(松下電器産業株式会社)2002.01.18, 全文、第1-9図 & US 2002-55840 A1 & EP 1168301 A2 & CN 1331467 A	1-9
A	JP 10-161692 A(キャノン株式会社)1998.06.19, 全文、第1-7図 & US 5956679 A1 & EP 847041 A3	1-9
A	JP 10-149191 A(日本電信電話株式会社)1998.06.02, 全文、第1-8図 & EP 831461 A2	1-9
A	JP 2000-330587 A(日本電気株式会社)2000.11.30, 全文、第1-9図 (ファミリーなし)	3, 9
A	JP 9-258765 A(国際電信電話株式会社)1997.10.03, 全文、第1-14図 (ファミリーなし)	3, 9
A	赤江他, 雑音環境へのヤコビ適応法の拡張, 日本音響学会講演論文集 2000年春季Ⅰ, 2000.03.15, 1-8-4, pp.7-8	1-9
A	伊田他, 雑音DBとモデル適応化を用いたHMM合成法における雑音変動耐性の評価, 日本音響学会講演論文集 2001年秋季Ⅰ, 2001.10.02, 1-1-17, pp.33-34	1-9
A	野口他, 1チャネル入力信号中の突発性雑音の判別と除去, 日本音響学会講演論文集 2004年春季Ⅰ, 2004.03.17, 3-P-30, pp.655-656	3, 9