(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4274962号
(P4274962)

(45)発行日　平成21年6月10日(2009.6.10)　(24)登録日　平成21年3月13日(2009.3.13)

(51)Int.Cl.		F I		
G10L 15/06	(2006.01)	G10L	15/06	310T
G10L 15/14	(2006.01)	G10L	15/06	400V
G10L 15/28	(2006.01)	G10L	15/14	200C
		G10L	15/28	360C
		G10L	15/28	370Z

請求項の数 13　(全 26 頁)

(21)出願番号　特願2004-28542(P2004-28542)
(22)出願日　平成16年2月4日(2004.2.4)
(65)公開番号　特開2005-221678(P2005-221678A)
(43)公開日　平成17年8月18日(2005.8.18)
審査請求日　平成17年6月13日(2005.6.13)

(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

前置審査

(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
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　松田　繁樹
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
(72)発明者　賈廣　貴敏
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
(72)発明者　コンスタンティン・マルコフ
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

最終頁に続く

(54)【発明の名称】音声認識システム

(57)【特許請求の範囲】
【請求項１】
それぞれ所定の音響特徴量をパラメータとする複数の音響モデル群を記憶するための記憶手段を含む音声認識システムであって、前記複数の音響モデル群の各々は、それぞれ異なる発話環境での発話音声のデコードに最適化された、同種の複数の音響モデルを含み、
前記音声認識システムはさらに、
入力される音声から前記所定の音響特徴量を算出するための特徴量算出手段と、
前記入力される音声のうち、発話がない部分の前記音響特徴量に基づいて、前記複数の音響モデル群の各々に対して、それぞれ前記複数の音響モデル群の混合重み適応化により、前記入力される音声の発話環境に適応化された複数の適応化音響モデルを作成するためのモデル適応化手段と、
前記複数の音響モデル群の各々に対して設けられ、前記入力される音声の発話部分に応答し、前記複数の適応化音響モデルをそれぞれ用いて、前記入力される音声の前記発話部分の前記所定の音響特徴量をデコードし、複数の仮説を出力するための複数のデコード手段と、
前記複数のデコード手段が出力する前記複数の仮説を統合することにより音声認識結果を出力するための統合手段とを含み、
前記統合手段は、
前記複数の音響モデル群の各々に対し、前記複数のデコード手段により得られる前記複数の仮説から、各単語の音響言語尤度の和が最大となる仮説を選択するための仮説選択手

段と、
前記仮説選択手段によって前記複数の音響モデル群の各々に対して選択された仮説から、単語ラティスを作成するためのラティス作成手段と、
前記ラティス内の単語列の中で、単語の音響尤度とＮ－グラム単語列（Ｎは１以上の整数）の言語尤度とから算出される値が所定条件を満足するものを認識結果として選択するための単語列選択手段とを含む、音声認識システム。
【請求項２】
前記複数の音響モデル群の各々は、それぞれ異なる雑音が重畳された発話音声のデコードに最適化された複数の音響モデルを含む、請求項１に記載の音声認識システム。
【請求項３】
前記モデル適応化手段は、
前記入力される音声の前記音響特徴量に基づいて、前記複数の音響モデル群の各々について、当該音響モデル群に含まれる複数の音響モデルのうち、前記入力される音声の前記音響特徴量に関連する予め定める条件を充足する、所定個数の音響モデルを選択するための手段と、
前記選択するための手段により前記複数の音響モデル群の各々について選択された前記所定個数の音響モデルから、雑音ガウス混合分布の混合重み適応化手法により前記適応化音響モデルを作成するための手段とを含む、請求項２に記載の音声認識システム。
【請求項４】
前記複数の音響モデル群は、
互いに異なる複数種類の雑音が第１のＳＮＲ（信号対雑音比）で重畳された発話音声のデコードに最適化された複数の音響モデルを含む第１の音響モデル群と、
前記複数種類の雑音が、前記第１のＳＮＲと異なる第２のＳＮＲで重畳された発話音声のデコードに最適化された複数の音響モデルを含む第２の音響モデル群とを含む、請求項１に記載の音声認識システム。
【請求項５】
前記複数の音響モデル群は、
それぞれ異なる発話環境での発話音声のデコードに最適化された、第１の音響モデル構造に基づく第１の音響モデル群と、
それぞれ異なる発話環境での発話音声のデコードに最適化された、前記第２の音響モデル構造とは異なる第２の音響モデル構造に基づく第２の音響モデル群とを含む、請求項１に記載の音声認識システム。
【請求項６】
前記第１の音響モデル構造は、通常発話に対して想定される音響モデル構造である、請求項５に記載の音声認識システム。
【請求項７】
前記第２の音響モデル構造は、言直し発話に対して想定される音響モデル構造である、請求項５又は請求項６に記載の音声認識システム
【請求項８】
前記複数の音響モデル群は、
それぞれ異なる発話環境での発話音声のデコードに最適化された、第１の種類の音響特徴量をパラメータとする第１の音響モデル群と、
それぞれ異なる発話環境での発話音声のデコードに最適化された、前記第１の種類の音響特徴量と異なる第２の音響特徴量をパラメータとする第２の音響モデル群とを含む、請求項１に記載の音声認識システム。
【請求項９】
前記第１の種類の音響特徴量はＭＦＣＣ（メル周波数ケプストラム係数）である、請求項８に記載の音声認識システム。
【請求項１０】
前記第２の種類の音響特徴量はＤＭＦＣＣ（差分メル周波数ケプストラム係数）である、

請求項８又は請求項９に記載の音声認識システム。
【請求項１１】
前記モデル適応化手段は、
前記入力される音声の前記音響特徴量に基づいて、前記第１の音響モデル群に含まれる音響モデルのうち、前記入力される音声の前記音響特徴量に関連する予め定める条件を充足する、所定個数の音響モデルを選択するための第１の音響モデル選択手段と、
前記第１の音響モデル選択手段により選択された音響モデルから、雑音ＧＭＭの混合重み適応化手法により第１の適応化音響モデルを作成するための手段と、
前記入力される音声の前記音響特徴量に基づいて、前記第２の音響モデル群に含まれる音響モデルのうち、前記入力される音声の前記音響特徴量に関連する予め定める条件を充足する、所定個数の音響モデルを選択するための第２の音響モデル選択手段と、
前記第２の音響モデル選択手段により選択された音響モデルから、雑音ＧＭＭの混合重み適応化手法により第２の適応化音響モデルを作成するための手段とを含む、請求項４～請求項１０のいずれかに記載の音声認識システム。
【請求項１２】
前記単語列選択手段は、前記ラティス内の単語列の中で、前記算出される値が最大となるものを認識結果として選択するための手段を含む、請求項１～請求項１１のいずれかに記載の音声認識システム。
【請求項１３】
前記単語列選択手段は、
前記ラティス内の単語の音響尤度と、Ｎ－グラム単語列の言語尤度とを、それぞれ所定の正規化方式により正規化するための正規化手段と、
前記ラティス内の単語列ごとに、前記正規化手段により正規化された当該単語列内の単語の音響尤度と前記Ｎ－グラム単語列の言語尤度とにそれぞれ所定の重みを加算して得られる値が前記所定条件を満足するものを認識結果として選択するための手段とを含む、請求項１～請求項１１のいずれかに記載の音声認識システム。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は大語彙の連続音声認識装置及び方法に関し、特に、雑音に強く、発話スタイルの変動に対しても頑健に音声を認識することが可能な連続音声認識システムに関する。
【背景技術】
【０００２】
近年、雑音又は発話スタイルに対して頑健な音声認識の研究が盛んに行なわれている。実環境において音声認識を使用するためには、通行する自動車などの乗り物から発せられるエンジン雑音や風切り音、駅、オフィス内などの人の声、コンピュータからのファンの音など、多種多様な雑音環境において高精度な音声認識が実現されなければならない。
【０００３】
さらに雑音だけでなく、使用者の年齢や性別、また感情や体調によってその発話スタイルは刻一刻と変化する。音声認識装置は、そのような発話スタイルの変動に対しても雑音と同様に頑健でなければならない。
【０００４】
雑音又は発話スタイルなど個別の変動に対する頑健化手法が従来から数多く提案されてきた。これについては後掲の非特許文献１を参照されたい。本明細書では以下、音声の音響的言語的特徴に影響する要因のことを総じて「発話環境」と呼ぶこととする。
【０００５】
雑音に対して頑健な音響特徴量の分析手法として、「ＳＳ（Ｓｐｅｃｔｒａｍ　Ｓｕｂｔｒａｃｔｉｏｎ）法（後掲の非特許文献２を参照されたい）」を音声認識の前処理として用いる手法が提案されている。これ以外にも、ＲＡＳＴＡ（ＲｅｌＡｔｉｖｅ　ＳｐｅｃＴｒＡ）、ＤＭＦＣＣ（Ｄｉｆｆｅｒｅｎｔｉａｌ　Ｍｅｌ　Ｆｒｅｑｕｅｎｃｙ　Ｃ

ｅｐｓｔｒｕｍ　Ｃｏｅｆｆｉｃｉｅｎｔ）など、いくつかの音響分析手法が提案されている。

【０００６】

ＳＳ法では、雑音重畳音声のスペクトルに対して雑音スペクトルを減算することにより、ＳＮＲ（信号対雑音比）を改善している。ＲＡＳＴＡ法では、個々の周波数バンドの値の変化に対して、音声情報が多く含まれている１から１２Ｈｚの変調スペクトラム成分を抽出することにより雑音の影響を軽減している。またＤＭＦＣＣはＦＦＴ（高速フーリエ変換）によって得られるフーリエ係数に対して、隣り合う係数間で差分をとり、音声などのピッチを持つスペクトルを強調することによって耐雑音性を改善している。

【０００７】

雑音に頑健な音響モデルの研究としては、ＰＭＣ（Ｐａｒａｌｌｅｌ　Ｍｏｄｅｌ　Ｃｏｍｂｉｎａｔｉｏｎ）法（後掲の非特許文献５を参照されたい。）、ヤコビ適応法（後掲の非特許文献６を参照されたい。）、ＭＬＬＲ（Ｍａｘｉｍｕｍ　Ｌｉｋｅｌｉｈｏｏｄ　Ｌｉｎｅａｒ　Ｒｅｇｒｅｓｓｉｏｎ）（後掲の非特許文献７を参照されたい。）による雑音適応などが提案されている。

【０００８】

これらのうち、ＰＭＣ法は、ＨＭＭ（隠れマルコフモデル）の出力確率分布を線形スペクトル領域に変換し雑音スペクトルを重畳することにより、環境雑音への適応を行なう手法である。このＰＭＣ法につき簡単に説明する。

【０００９】

ＰＭＣ法の概念を図２８を参照して説明する。図２８を参照して、ＰＭＣ法の対象となるもとの音響モデルが、音響の特徴量からなる音響空間６００において領域６１０の付近に存在する音響をモデル化したものであるものとする。このとき、音声認識対象の雑音を含んだ音声データ領域６１２は、雑音のためにもとの領域６１０からずれたものとなる。そこで、領域６１２と領域６１０との差分を考え、この差分に相当する量を音響モデル６１０に加えることにより音響モデルの音響空間６００内における位置を領域６１２まで移動するよう音響モデルを変換する。

【００１０】

このようにして変換した後の音響モデルを用いれば、領域６１２の付近に存在する雑音を含んだ音声については、もとの音響モデルを用いたものより高い精度で認識できる。

【００１１】

ヤコビ適応法は、雑音の変化に伴う出力確率分布の非線形変換を線形近似することにより、雑音環境へ高速に適応する手法である。

【００１２】

ＭＬＬＲを用いた雑音適応は、無雑音音声と雑音重畳音声との間の分布移動を回帰行列を用いて表現し、音響モデル全体を雑音モデルに適応化する手法である。

【００１３】

さらに、雑音の分布の時間変動を逐次的に推定することにより、非定常雑音に対する認識精度を改善する手法（後掲の非特許文献９を参照されたい。）が提案されている。

【００１４】

発話スタイルに対する頑健性の改善手法としては、発話スタイル依存の音響モデルを用いる手法の他、ロンバード効果によるスペクトルの変形を考慮した手法（非特許文献８を参照されたい。）及び個々の母音ＨＭＭの最後に無音状態を追加することにより音声強調発声や言直し発話に頑健な音響モデルを構築する手法（非特許文献１０を参照されたい。）などが提案されている。そのほかにも、講演音声などの音素継続時間の短い発声を含む音声に対して、分析フレーム周期又はウィンドウ幅を自動選択することにより認識精度を改善する手法（非特許文献１１、１２参照）が提案されている。

【００１５】

これらの頑健化手法は主として、雑音や発話スタイルなどの個別の変動に対する頑健化である。音声認識を実環境で用いるためには、複数の発話環境が刻一刻と変化する状況で

あっても頑健に音声を認識することができなければならない。このような種々の外乱に対して頑健な音声認識を実現するための方法は大きく２つに分類することができると考えられる。発話環境の変動に頑健な音響モデル及び言語モデルを用いて単数のデコーダで認識を行なうシングルタイプの方法と、お互いに異なる環境に適応化された複数の音響モデル及び言語モデルを使用して得られた複数の仮説を統合するパラレルタイプの手法とである。

【００１６】

シングルタイプの音声認識システムを構築するためには、広い発話環境の音声を頑健に認識する音響モデル及び言語モデルが必要である。そのために、男性及び女性双方の学習データから性別独立な音響モデルを推定するなど、複数の発話環境のデータを用いてＨＭＭのモデルパラメータ推定を行なうことにより頑健性を改善する手法がある。しかし、男性女性などのお互いの音響的特徴が大きく異なる場合ではなく、種々のＳＮＲのデータを用いて学習する場合、個々の音素モデルの分布が過度に広がることにより音素分類精度の低下が懸念される。従って、このようなモデル化法には頑健化の限界があると考えられる。

【００１７】

セグメントモデル（非特許文献１３参照）では、時間的に離れた音響特徴ベクトル間の相関を計算することで音声の非定常な振舞いのモデル化を試みている。時間的に離れた特徴ベクトル間の相関として発話環境の変動をモデル化することができるならば、セグメントモデルにおいて広い発話環境の音声を頑健に認識できる可能性がある。しかし、効率的な相関の計算方法やモデルパラメータの増大などの問題により十分な精度は得られていない。

【００１８】

一方、パラレルタイプによる音声認識は、個々の音響モデルや言語モデルの利用可能な発話環境が限られていたとしても、それらを複数個使用しパラレルにデコーディングすることにより、個々の音素間の分類精度を低下させることなく広い発話環境の音声を頑健に認識できる可能性がある。

【００１９】

このような音声認識の例としては、ＳＮＲに依存した音響モデルを用いて得られた複数の仮説を最大尤度基準で選択する手法、複数のお互いに異なる音響特徴量を用いて音声認識を行ない、得られた複数の仮説を単語単位で統合する仮説統合法（非特許文献１５参照）が提案されている。

【００２０】

【非特許文献１】中村、『実音響環境に頑健な音声認識を目指して』、信学技報、ＥＡ２００２－１２、ｐｐ.３１－３６、２００２。

【非特許文献２】Ｓ．Ｆ．ボル、『スペクトル減算を用いた音声中の音響雑音の抑制』、ＩＥＥＥ音響音声信号処理論文集、第ＡＳＳＰ－２７巻、第１１３－１２０頁、１９７９年。（Ｓ．Ｆ．Ｂｏｌｌ，『Ｓｕｐｐｒｅｓｓｉｏｎ ｏｆ Ａｃｏｕｓｔｉｃ Ｎｏｉｓｅ ｉｎ Ｓｐｅｅｃｈ Ｕｓｉｎｇ Ｓｐｅｃｔｒａｌ Ｓｕｂｔｒａｃｔｉｏｎ，』ＩＥＥＥ Ｔｒａｎｓ．Ａｃｏｕｓｔ．，Ｓｐｅｅｃｈ，Ｓｉｇｎａｌ Ｐｒｏｃｅｓｓｉｎｇ，ｖｏｌ．ＡＳＳＰ－２７，ｐｐ．１１３－１２０，１９７９．）

【非特許文献３】Ｈ．ヘルマンスキ及びＮ．モーガン、「音声のＲＡＳＴＡ処理」、ＩＥＥＥ音声及び音響処理トランザクション、第２巻、第４号、第５８７－５８９頁（Ｈ．Ｈｅｒｍａｎｓｋｙ ａｎｄ Ｎ．Ｍｏｒｇａｎ，『ＲＡＳＴＡ Ｐｒｏｃｅｓｓｉｎｇ ｏｆ Ｓｐｅｅｃｈ，』ＩＥＥＥ Ｔｒａｎｓ．Ｓｐｅｅｃｈ ａｎｄ Ａｕｄｉｏ Ｐｒｏｃｃｅｓｓｉｎｇ，ｖｏｌ．２，ｎｏ．４，ｐｐ．５８７－５８９，１９９４．）

【非特許文献４】Ｊ．チェン、Ｋ．Ｋ．パリワル、Ｓ．ナカムラ、『頑健な音声認識のための差分パワースペクトル由来のケプストラム』、音声コミュニケーション、第４１巻第２－３号、第４６９－４８４頁、２００３年。（Ｊ．Ｃｈｅｎ，Ｋ．Ｋ．Ｐａｌｉｗａｌ，Ｓ.Ｎａｋａｍｕｒａ，『Ｃｅｐｓｔｒｕｍ Ｄｅｒｉｖｅｄ ｆｒｏｍ Ｄｉｆｆｅ

rentiated Power Spectrum for Robust Speech Recognition,』Speech Communication, vol. 41, no. 2－3, pp. 469－484, 2003.）
【非特許文献５】M. ゲールズ及びS. ヤング、『パラレルモデルの組合せを用いた頑健な連続音声認識』、ＩＥＥＥ音声及び音響処理論文集、第４巻、第５号、第３５２－３５９頁、１９９６年。（M. Gales and S. Young,『Robust Continuous Speech Recognition Using Parallel Model Combination,』 IEEE Trans. on Speech and Audio Processing, vol. 4, No. 5, pp. 352－359, 1996.）
【非特許文献６】Y. ヤマグチ、S. タカハシ及びS. サガヤマ、『ヤコビアン適応アルゴリズムを用いた環境雑音への音響モデルの高速適応』、ユーロスピーチ予稿集、９７、第２０５１－２０５４頁、１９９７年。（Y. Yamaguchi, S.Takahashi and S. Sagayama,『Fast Adaptation of Acoustic Models to Environmental Noise Using Jacobian Adaptation Algorithm,』 Proc. Eurospeech, 97, pp. 2051－2054, 1997.）
【非特許文献７】C. J. レゲッタ及びP. C. ウッドランド、『連続密度隠れマルコフモデルの話者適応のための最大尤度線形回帰』、コンピュータ音声及び言語、第９巻、第１７１－１８５頁、１９９５年。（C. J. Leggetter and P. C. Woodland,『Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models,』 Computer Speech and Language, vol. 9, pp. 171－185, 1995.）
【非特許文献８】J. C. ジャンカ、『ロンバード効果とその聴者及び自動音声認識装置に対する役割』、アメリカ音響学会誌、第９３巻、第５１０－５２４頁、１９９３年。（J. C. Junqua,『The Lombard Reflex and its Role on Human Listeners and Automatic Speech Recognizer,』 J. Acoustic Soc. Amer., vol. 93, pp. 510－524, 1993.）
【非特許文献９】K. ヤオ、B. E. シー、S. ナカムラ及びZ. カオ、『非定常雑音における頑健な音声認識のための連続ＥＭアルゴリズムによる残存雑音の補償』、ＩＣＳＬＰ２０００予稿集、第１巻、第７７０－７７３頁、２０００年。（K. Yao, B. E. Shi, S. Nakamura and Z. Cao,『Residual Noise Compensation by a Sequential EM Algorithm for Robust Speech Recognition in Nonstationary Noise,』 Proc. ICSLP2000, vol. 1, pp. 770－773, 2000.）
【非特許文献１０】奥田、松井、中村、『誤認識時の言い直し発話における発話スタイルの変動に頑健な音響モデル構築法』信学論, vol. J86－DII, no. 1, pp. 42－51, 2003.
【非特許文献１１】奥田、河原、中村、『ゆう度基準による分析周期・窓長の自動選択手法を用いた発話速度の補正と音響モデルの構築』信学論、vol. J86－DII, no. 2, pp. 204－211, 2003.
【非特許文献１２】南條、河原、『発話速度に依存したデコーディングと音響モデルの適応』信学技報、SP2001－103、2001.
【非特許文献１３】M. オステンドルフ、V. ディガラキス及びO. キンバル、『HMMからセグメントモデルへ：音声認識のためのストカスティックモデリングの統一見解』、ＩＥＥＥ音声及び音響処理論文集、第４巻、第５号、第３６０－３７８頁、１９９６年。

（M．Ｏｓｔｅｎｄｏｒｆ，Ｖ．Ｄｉｇａｌａｋｉｓ ａｎｄ Ｏ．Ｋｉｍｂａｌｌ，『Ｆｒｏｍ ＨＭＭｓ ｔｏ Ｓｅｇｍｅｎｔ Ｍｏｄｅｌｓ：Ａ Ｕｎｉｆｉｅｄ Ｖｉｅｗ ｏｆ Ｓｔｏｃｈａｓｔｉｃ Ｍｏｄｅｌｉｎｇ ｆｏｒ Ｓｐｅｅｃｈ Ｒｅｃｏｇｎｉｔｉｏｎ，』ＩＥＥＥ Ｔｒａｎｓ．Ｓｐｅｅｃｈ ａｎｄ Ａｕｄｉｏ Ｐｒｏｃ．，ｖｏｌ．４，ｎｏ．５，ｐｐ．３６０－３７８，１９９６．）
【非特許文献１４】伊田、中村、『雑音ＧＭＭの適応化とＳＮ比別マルチパスモデルを用いたＨＭＭ合成による高速な雑音環境適応化』信学論、ｖｏｌ．Ｊ８６－Ｄ－ＩＩ，ｎｏ．２，ｐｐ．１９５－２０３，２００３．
【非特許文献１５】Ｋ．マルコフ、Ｔ．マツイ、Ｒ．グルーン、Ｊ．ツァン、Ｓ．ナカムラ、『ＤＡＲＰＡ ＳＰＩＮＥ２用の雑音及びチャネル歪に頑健なＡＳＲシステム』、ＩＥＩＣＥ情報及びシステム論文集、第Ｅ８６－Ｄ巻、第３号、２００３年。（Ｋ．Ｍａｒｋｏｖ，Ｔ．Ｍａｔｓｕｉ，Ｒ．Ｇｒｕｈｎ，Ｊ．Ｚｈａｎｇ，Ｓ．Ｎａｋａｍｕｒａ，『Ｎｏｉｓｅ ａｎｄ Ｃｈａｎｎｅｌ Ｄｉｓｔｏｒｔｉｏｎ Ｒｏｂｕｓｔ ＡＳＲ Ｓｙｓｔｅｍ ｆｏｒ ＤＡＲＰＡ ＳＰＩＮＥ２ Ｔａｓｋ，』ＩＥＩＣＥ Ｔｒａｎｓ．Ｉｎｆ．＆Ｓｙｓｔ．，ｖｏｌ．Ｅ８６－Ｄ，ｎｏ．３，２００３．）
【非特許文献１６】Ｍ．オステンドルフ及びＨ．シンガー、『最大尤度連続状態分割を用いたＨＭＭトポロジー設計』、コンピュータ音声及び言語、第１１巻、第１号、第１７－４１頁１９９７年。（Ｍ．Ｏｓｔｅｎｄｏｒｆ ａｎｄ Ｈ．Ｓｉｎｇｅｒ，『ＨＭＭ Ｔｏｐｏｌｏｇｙ Ｄｅｓｉｇｎ Ｕｓｉｎｇ Ｍａｘｉｍｕｍ Ｌｉｋｅｌｉｈｏｏｄ Ｓｕｃｃｅｓｓｉｖｅ Ｓｔａｔｅ Ｓｐｌｉｔｔｉｎｇ，』Ｃｏｍｐｕｔｅｒ Ｓｐｅｅｃｈ ａｎｄ Ｌａｎｇｕａｇｅ，ｖｏｌ．１１，ｎｏ．１，ｐｐ．１７－４１，１９９７．）

【発明の開示】
【発明が解決しようとする課題】
【００２１】
しかしながら上述したいずれの方法においても、例えばモデルの変換に時間を要すること、雑音又は発話スタイルなど、個別の要素の変動に的確に対応することが難しいこと、などから、実環境における雑音を含んだ音声や、発話スタイルが変動する音声に対して音声認識を精度よく行なうことは未だ可能でない。
【００２２】
それゆえにこの発明の目的は、雑音などの個別の変動に実時間で追従して精度高く認識することができる音声認識システムを提供することである。
【００２３】
この発明の他の目的は、雑音などの個別の変動だけでなく、発話スタイルの変動に対しても頑健に音声を認識することができる音声認識システムを提供することである。
【課題を解決するための手段】
【００２４】
本発明に係る音声認識システムは、それぞれ異なる発話環境での発話音声のデコードに最適化された、それぞれ所定の音響特徴量をパラメータとする複数の音響モデル群を記憶するための記憶手段と、入力される音声から所定の音響特徴量を算出するための特徴量算出手段と、入力される音声の音響特徴量に基づいて、それぞれ複数の音響モデル群の混合重み適応化により、入力される音声の発話環境に適応化された複数の適応化音響モデルを作成するためのモデル適応化手段と、複数の適応化音響モデルを用いて、入力される音声の所定の音響特徴量をデコードし複数の仮説を出力するためのデコード手段と、デコード手段が出力する複数の仮説を統合することにより音声認識結果を出力するための統合手段とを含む。
【００２５】
デコード手段により出力される複数の仮説が互いに相補的である場合、統合手段により仮説を統合することにより、より精度の高い音声認識結果を得られる可能性が高い。
【００２６】

好ましくは、複数の音響モデル群の各々は、それぞれ異なる雑音が重畳された発話音声のデコードに最適化された複数の音響モデルを含む。

【００２７】

複数の音響モデル群の各々が含む音響モデルは、互いに異なる雑音が重畳された発話音声のデコードに適している。種々の雑音について適応化音響モデルが得られるので、雑音が異なる発話環境での音声認識の精度が向上することが期待できる。

【００２８】

より好ましくは、モデル適応化手段は、入力される音声の音響特徴量に基づいて、複数の音響モデル群の各々について、当該音響モデル群に含まれる複数の音響モデルのうち、入力される音声の音響特徴量に関連する予め定める条件を充足する、所定個数の音響モデルを選択するための手段と、選択するための手段により複数の音響モデル群の各々について選択された所定個数の音響モデルから、雑音ガウス混合分布の混合重み適応化手法により適応化音響モデルを作成するための手段とを含む。

【００２９】

モデル適応化にあたって、入力音声の発話環境と所定の関係にある発話環境に対応する音響モデルを複数個選び、それらに対し混合重み適応化を行なって適応化環境モデルを作成する。適応化モデル作成時の計算量が少なくて済み、またモデルの精度は十分に高くなる。

【００３０】

複数の音響モデル群は、互いに異なる複数種類の雑音が第１のＳＮＲで重畳された発話音声のデコードに最適化された複数の音響モデルを含む第１の音響モデル群と、複数種類の雑音が、第１のＳＮＲと異なる第２のＳＮＲで重畳された発話音声のデコードに最適化された複数の音響モデルを含む第２の音響モデル群とを含んでもよい。

【００３１】

種々のＳＮＲに最適化された音響モデルに基づいて、それぞれ適応化モデルが作成される。より広い発話環境に対し適応化モデルが作成されるので、入力音声の発話環境に近いものが得られる可能性が高くなる。その結果、音声認識精度の向上が期待できる。

【００３２】

好ましくは、複数の音響モデル群は、それぞれ異なる発話環境での発話音声のデコードに最適化された、第１の音響モデル構造に基づく第１の音響モデル群と、それぞれ異なる発話環境での発話音声のデコードに最適化された、第２の音響モデル構造とは異なる第２の音響モデル構造に基づく第２の音響モデル群とを含む。

【００３３】

第１及び第２の音響モデル構造に基づく音響モデル群を用いることにより、雑音以外の要因の変化に対しても頑健に音声認識を行なうことができる。

【００３４】

好ましくは、第１の音響モデル構造は、通常発声に対して想定される音響モデル構造であり、さらに好ましくは第２の音響モデル構造は、言直し発話に対して想定される音響モデル構造である。

通常発声と言直し発話という二種類の発話に対応する音響モデル構造を用いることにより、話者の発話スタイルの変化に対しても頑健に音声認識を行なうことができる。

【００３５】

複数の音響モデル群は、それぞれ異なる発話環境での発話音声のデコードに最適化された、第１の種類の音響特徴量をパラメータとする第１の音響モデル群と、それぞれ異なる発話環境での発話音声のデコードに最適化された、第１の種類の音響特徴量と異なる第２の音響特徴量をパラメータとする第２の音響モデル群とを含んでもよい。

【００３６】

好ましくは、第１の種類の音響特徴量はＭＦＣＣであり、さらに好ましくは、第２の種類の音響特徴量はＤＭＦＣＣである。

【００３７】

第１及び第２の音響特徴量を用いる音響モデル群を用いることにより、種々発話環境の変化に対しても頑健に音声認識を行なうことができる。

【００３８】

デコード手段は、第１及び第２の適応化音響モデルの各々に対し、入力される音声の所定の音響特徴量をデコードし複数の仮説を出力するための手段を含んでもよい。

【００３９】

好ましくは、デコード手段はさらに、第１及び第２の適応化音響モデルの各々について、出力するための手段により出力された複数の仮説のうち、各単語の音響言語尤度の和が最大となる仮説を選択し、統合手段に与えるための手段を含む。

【００４０】

統合手段に与える前に、適応化音響モデルごとに音響言語尤度の和が最大となる仮説を選択することにより、統合処理の際の探索空間が小さくなり統合処理が容易かつ高速になる。

【００４１】

さらに好ましくは、適応化音響モデルを作成するための手段は、入力される音声の音響特徴量に基づいて、第１の音響モデル群に含まれる音響モデルのうち、入力される音声の音響特徴量に関連する予め定める条件を充足する、所定個数の音響モデルを選択するための第１の音響モデル選択手段と、第１の音響モデル選択手段により選択された音響モデルから、雑音ＧＭＭ（Ｇａｕｓｓｉａｎ　Ｍｉｘｔｕｒｅ　Ｍｏｄｅｌ）の混合重み適応化手法により第１の適応化音響モデルを作成するための手段と、入力される音声の音響特徴量に基づいて、第２の音響モデル群に含まれる音響モデルのうち、入力される音声の音響特徴量に関連する予め定める条件を充足する、所定個数の音響モデルを選択するための第２の音響モデル選択手段と、第２の音響モデル選択手段により選択された音響モデルから、雑音ＧＭＭの混合重み適応化手法により第２の適応化音響モデルを作成するための手段とを含む。

【００４２】

複数の音響モデル群からそれぞれ適応化音響モデルが作成され、それらを用いてデコードが行われ、かつそれらのデコード結果が統合されるので、それぞれの音響モデルの特徴群の特長を組み合わせた頑健な音声認識を行なうことができる。

【００４３】

より好ましくは、仮説統合手段は、デコード手段が出力する複数の仮説から、単語ラティスを作成するためのラティス作成手段と、ラティス内の単語列の中で、単語の音響尤度とＮ－グラム単語列（Ｎは１以上の整数）の言語尤度とから算出される尤度が所定条件を満足するもの、例えば最大となるもの、を認識結果として選択するための選択手段とを含む。

【００４４】

さらに好ましくは、選択手段は、ラティス内の単語の音響尤度と、Ｎ－グラム単語列の言語尤度とを、それぞれ所定の正規化方式により正規化するための正規化手段と、単語列ごとに、正規化手段により正規化された当該単語列内の単語の音響尤度とＮ－グラム単語列の言語尤度とにそれぞれ所定の重みを加算して得られる尤度が所定条件を満足するものを認識結果として選択するための手段とを含む。

【００４５】

音響尤度と言語尤度とはそれぞれ別の概念に基づく尤度であるから、両者の尤度を単純に加算するのは意味がない。両者を予め正規化し、正規化後の尤度から算出される尤度によって単語列を選択することにより、尤度が正しく算出されることになり、頑健な音声認識を行なうことができる。

【００４６】

この発明の他の局面に係る音声認識システムは、それぞれ異なる発話環境での発話音声のデコードに最適化された、ＤＭＦＣＣをパラメータとする音響モデル群を記憶するための記憶手段と、入力される音声からＤＭＦＣＣを算出するためのＤＭＦＣＣ算出手段と、

入力される音声から算出されたＤＭＦＣＣに基づいて、音響モデル群の混合重み適応化により、入力される音声の発話環境に適応化された適応化音響モデルを作成するためのモデル適応化手段と、適応化音響モデルを用いて、入力される音声から算出されるＤＭＦＣＣをデコードするためのデコード手段とを含む。

【００４７】

混合重み適応化を用いるため、実際の適応化時には音響モデル群を混合するための重みを算出するだけでよく、適応を高速に行なえる。その結果、ＤＭＦＣＣを音響特徴量として、発話環境の変化に対して頑健な音声認識を行なうことができる。

【発明を実施するための最良の形態】

【００４８】

雑音環境が頻繁に変動する状況では、音響モデルを高速に雑音環境に適応させることが可能でなければならない。以下に述べる本発明の一実施の形態では、高速な雑音環境適応として、非特許文献１４において提案されている雑音ＧＭＭの混合音適応化によるＨＭＭ合成法を用いる。

【００４９】

図１～図３を参照して、この手法の概略について説明する。図１を参照して、あらかじめ準備した種々の雑音からなる雑音ＤＢ１００から、個々の雑音を混合成分とする雑音ＧＭＭ１０２と、個々の雑音に対して別々に適応化された雑音重畳音声用ＨＭＭ１０４，１０６，…とを推定する。次に図２に示すように、短時間の未知雑音１１０を用いて雑音ＧＭＭ１０２の混合ウェイトW_{N1}，W_{N2}，…のみを推定する。そして、図３に示すように、この混合ウェイトW_{N1}、W_{N2}，…を用いて、雑音重畳音声用ＨＭＭ１０４，１０６，…を状態レベルで複数混合化する。例えばＨＭＭ１０４の状態１２０と、ＨＭＭ１０６の状態１２２とに対して、それぞれのガウス混合分布に対し図２に示すステップにより計算された混合ウェイトを乗算して足し合わせて状態出力確率分布１２４を算出し、雑音適応されたＨＭＭの状態１２６の状態出力確率分布とする。

【００５０】

図１～図３においてN_iは第ｉ番目の雑音、λ_iは第ｉ番目の雑音に対する雑音重畳音声用ＨＭＭを表す。P_{Ni}とw_{Ni}は雑音ＧＭＭにおける第ｉ番目の雑音の分布とその分布に対する混合ウェイトとをそれぞれ示す。さらに$w_{\lambda ij}$と$p_{\lambda ij}$は第ｉ番目の雑音用の雑音重畳音声用ＨＭＭにおける第ｊ番目の混合分布Ｎの分岐確率と混合成分とを表す。

【００５１】

この手法の利点として、適応の計算時間がＧＭＭの混合ウェイトの推定時間のみであり大変高速である点と、雑音適応されたＨＭＭが複数の雑音環境の分布を含んでおり、単一の雑音から推定された音響モデルよりも雑音の短時間の変動に対する頑健性が高い点とを挙げることができる。

【００５２】

上記した混合重み適応化によるＨＭＭ合成法を用いる場合、音響特徴量としてはＭＦＣＣを用いることが考えられる。しかし、ＭＦＣＣのみでは認識精度を高めることが難しいことが実験的に判明した。そこで本実施の形態では、ＭＦＣＣとは異なる音響特徴量を用いた音声認識を行ない、その結果とＭＦＣＣによる音声認識の結果とを統合することを考える。本実施の形態では、雑音の変動に対して頑健な特徴量として非特許文献４において提案されたＤＭＦＣＣ特徴量を用いることとする。以下、ＤＭＦＣＣ特徴について述べる。なお、以下の処理では、音声データは所定サンプリング周波数及び所定窓長でサンプリングしたフレームとして準備されているものとする。

【００５３】

ＤＭＦＣＣ特徴量は、式（１）に示すＤＰＳ（ｄｉｆｆｅｒｅｎｔｉａｌ ｐｏｗｅｒ ｓｐｅｃｔｒｕｍ）を基礎とする特徴量である。式（１）中のＹ（ｉ，ｋ）は、第ｉ番目のフレームにおける第ｋ番目のパワースペクトラム係数を表す。同様にＤ（ｉ，ｋ）は第ｉ番目のフレームにおける第ｋ番目のＤＰＳ係数を表す。ＤＭＦＣＣ特徴量は、このＤＰＳ係数に対してＤＣＴ（ｄｉｓｃｒｅｔｅ ｃｏｓｉｎｅ ｔｒａｎｓｆｏｒｍ）を行

なうことにより抽出される。

$$D(i,k)=|Y(i,k)-Y(i,k+1)| \qquad (1)$$

有声母音などのピッチを含む音声から抽出されたパワースペクトラムは、基本周波数の高調波の影響によって櫛型の形状を持つ。このようなパワースペクトラムからＤＰＳ係数を計算した場合、隣り合うパワースペクトラム係数間の差が大きいため、ＤＰＳ係数の値も同様に大きなパワーとして計算される。一方、雑音などの特徴を持たない波形のパワースペクトラムから計算されるＤＰＳ係数は、隣り合うパワースペクトラム係数間の差が小さいため、ＤＰＳ係数の値も小さくなると考えられる。雑音重畳音声のパワースペクトラムを無雑音音声のパワーと雑音のパワーの和であると仮定した場合、ＤＰＳ係数を計算することによって、音声と比較してなだらかに変化する雑音のパワー成分を減衰させることができると考えられる。

【００５４】

本実施の形態では、上述のようにＭＦＣＣ特徴量とＤＭＦＣＣ特徴量とを用いて、パラレルにデコーディングを行ない、得られた仮説の統合による音声認識精度の改善を試みている。

【００５５】

図４に、本実施の形態に係る音声認識システム１３０の概略ブロック図を示す。図４を参照して、このシステム１３０は、初期ＨＭＭ１５０と、雑音データベース（ＤＢ）１５２と、雑音が重畳された学習データ１５３とから、パラレルに音声をデコードするためのＭＦＣＣ・ＨＭＭ群１５６及びＤＭＦＣＣ・ＨＭＭ群１５８を作成するためのＨＭＭ作成部１５４と、ＨＭＭ作成部１５４により作成されたＭＦＣＣ・ＨＭＭ群１５６及びＤＭＦＣＣ・ＨＭＭ群１５８を用いて、入力音声１４４に対する音声認識を行ない、音声認識結果１４６を出力するための認識処理部１４２とを含む。

【００５６】

図５はＨＭＭ作成部１５４のブロック図である。図５を参照して、ＨＭＭ作成部１５４は、初期ＨＭＭ１５０と雑音ＤＢ１５２とから、前述したＰＭＣ法を用いて雑音重畳音声用ＭＦＣＣ・ＨＭＭ群１５６を作成するためのＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７０と、雑音重畳済みの学習データ１５３を用いて初期ＨＭＭ１５０に対する学習を行なうことにより、雑音重畳音声用ＤＭＦＣＣ・ＨＭＭ群１５８を作成するためのＤＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７２とを含む。

【００５７】

本実施の形態では、雑音ＤＢ１５２としては１２種類の異なる雑音を用いる。学習データ１５３についても、無雑音学習データに上記したものと同種の雑音を重畳したものを用いる。なお、雑音の重畳に際しては、１０ｄＢ，２０ｄＢ及び３０ｄＢの三種のＳＮＲを用いている。初期ＨＭＭ１５０としては、無雑音音響モデルとして学習済みのものを準備する。

【００５８】

ＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７０は、従来技術の項で説明した通りのＰＭＣ法を用いて各雑音に対応する雑音重畳音声用ＨＭＭを推定する機能を持つ。同様にＤＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７２は、学習データ１５３を用いて最尤推定を行なうことにより雑音重畳音声用ＤＭＦＣＣ・ＨＭＭ群１５８の学習を行なう。ＤＭＦＣＣ特徴量に対しては、ＭＦＣＣ特徴量と異なりＰＭＣ法が適用できないためである。

【００５９】

図６に、ＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７０による雑音重畳音声用ＭＦＣＣ・ＨＭＭ群１５６の概念について示す。図６を参照して、ＭＦＣＣ用の初期ＨＭＭ１８０は、無雑音通常発声用ＭＦＣＣ・ＨＭＭ１９０と、無雑音言直し発話用ＭＦＣＣ・ＨＭＭ１９２とを含む。本実施の形態では、発話スタイルの変動への対応としてシステムへの言直し時に頻繁に観測される音節強調発話に対する頑健性の改善を試みている。言直し発話用

のＨＭＭはこのためのものである。

【００６０】

音声認識ソフトウェアが認識誤りを起こした場合、そのソフトウェアの使用者はもう一度同じ発声を繰返さなければならない。このような言直し発話では、母音の後に短時間のポーズが挿入されるなど、通常発声とは異なる音響的特徴を持つことが報告されている。この言直し発話を頑健に認識するため、図１７に示すような構造を持つ音響モデル４４０が提案されている。図１７を参照して、この母音モデルは、母音の後に短時間ポーズを挿入するため、例えばｔ－ａ＋ｓｉｌの状態パス（図１７において、「ｔ－ａ＋ｋ」などの表記は、先行音素が／ｔ／、後続音素が／ｋ／、当該音素が／ａ／の環境依存音素を表す。「ｓｉｌ」は無音状態を表わす。）及び、その母音モデルの後にポーズ状態を追加した状態パスの合計３つの成分を有するマルチパス音響モデルの構造を持つ。さらに、このモデルでは、子音モデルの前に短時間ポーズの挿入を許すため、通常の子音モデルに加えてｓｉｌ－ｋ＋ｉの状態パスへの遷移が追加されている。このような音響モデルを用いることにより、通常発声の音声以外にも言直しや音節強調発声などの音声を頑健に認識することが可能となる。

【００６１】

再び図６を参照して、雑音ＤＢ１５２は、本実施の形態では１２種類の雑音データ２００，２０２，…，２０６を含む。ＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７０はこれら１２種類の雑音の各々について、３種類のＳＮＲ（１０ｄＢ、２０ｄＢ、及び３０ｄＢ）ごとにＰＭＣを用いて初期ＨＭＭ１８０を適応化することにより、雑音重畳音声用ＭＦＣＣ・ＨＭＭ群１５６を生成する。

【００６２】

生成される雑音重畳音声用ＭＦＣＣ・ＨＭＭ群１５６は、男声通常発声用ＭＦＣＣ・ＨＭＭ群２１０と、男声言直し発話用ＭＦＣＣ・ＨＭＭ群２１２と、女声通常発声用ＭＦＣＣ・ＨＭＭ群２１４と、女声言直し発話用ＭＦＣＣ・ＨＭＭ群２１６と、通常発声用無雑音ＭＦＣＣ・ＨＭＭ２１５と、言直し発話用無雑音ＭＦＣＣ・ＨＭＭ２１７とを含む。すなわち本実施の形態では、雑音重畳音声用ＭＦＣＣ・ＨＭＭ群１５６は、男声女声、１２種類の雑音、３種類のＳＮＲ、及び通常発声、言直し発話用の、２×１２×３×２＝１４４種類と通常発声用及び言直し発話用の無雑音音声用モデルの計１４６種類のＨＭＭを含む。

【００６３】

図７に、ＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７０により作成される音響モデルが、音響空間２７０中に占める領域を模式的に示す。図７に示すのは、１２個の音響モデルに対応する領域２８０～３０２のみである。しかし、上述したように作成される音響モデルは１４６種類であるので、音響空間２７０にはこれら領域２８０～３０２と同様のものが合計で１４６個作成されることになる。

【００６４】

図８に、ＤＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７２による雑音重畳音声用ＤＭＦＣＣ・ＨＭＭ群１５８の作成を概念的に示す。図８を参照して、初期ＤＭＦＣＣ・ＨＭＭ１８２は、無雑音通常発声用ＤＭＦＣＣ・ＨＭＭ２３０及び無雑音言直し発話用ＤＭＦＣＣ・ＨＭＭ２３２を含む。

【００６５】

また雑音重畳学習データ１５３は、前述した１２種類の雑音を、前述した３種類のＳＮＲで学習データに重畳したものであり、３×１２＝４２種類の雑音重畳学習データ２４０～２４６を含む。ＤＭＦＣＣ雑音重畳音声用ＨＭＭ推定部１７２は、無雑音通常発声用ＤＭＦＣＣ・ＨＭＭ２３０及び無雑音言直し発話用ＤＭＦＣＣ・ＨＭＭ２３２に対し、上記した雑音重畳学習データ１５３を用いて学習を行なうことにより、男声通常発声用ＤＭＦＣＣ・ＨＭＭ群２５０、男声言直し発話用ＤＭＦＣＣ・ＨＭＭ群２５２、女声通常発声用ＤＭＦＣＣ・ＨＭＭ群２５４、女声言直し発話用ＤＭＦＣＣ・ＨＭＭ群２５６と、通常発声用無雑音ＤＭＦＣＣ・ＨＭＭ２５５と、言直し発話用無雑音ＤＭＦＣＣ・ＨＭＭ２５７

とを生成する。

【００６６】

例えば男声通常発声用ＤＭＦＣＣ・ＨＭＭ群２５０は、各種類及び各ＳＮＲの雑音重畳学習データに対して学習した結果得られた、複数個の男声雑音重畳通常発声用ＤＭＦＣＣ・ＨＭＭ２６０，２６２，…，２６６を含む。他のＤＭＦＣＣ・ＨＭＭ群２５２、２５４、２５６も、男声か女声か、通常発声用モデルか言直し発話用モデルかを除き同様の構成である。

【００６７】

本実施の形態では、雑音重畳音声用ＤＭＦＣＣ・ＨＭＭ群１５８は雑音重畳音声用ＭＦＣＣ・ＨＭＭ群１５６と同様の構成となっている。しかし、当業者であれば容易に理解できるように、ＭＦＣＣを用いる音声認識と、ＤＭＦＣＣを用いる音声認識とで同様の構成をとる必要は全くない。それぞれ別々のデータに基づきＨＭＭを作成してもよい。最終的に作成されるＨＭＭの数が等しくなる必要もない。

【００６８】

図９は、図４に示す認識処理部１４２の詳細な構造を示すブロック図である。図９を参照して、認識処理部１４２は、入力音声１４４に対しＭＦＣＣ・ＨＭＭ群を用いて音声認識を行なうＭＦＣＣ処理部３１０と、入力音声１４４に対しＤＭＦＣＣ・ＨＭＭ群を用いた音声認識を行ない認識結果を出力するためのＤＭＦＣＣ処理部３１２と、ＭＦＣＣ処理部３１０及びＤＭＦＣＣ処理部３１２の出力を統合し、統合された認識結果を出力するための認識結果統合部３１４とを含む。

【００６９】

図１０はＭＦＣＣ処理部３１０のより詳細なブロック図である。図１０を参照してＭＦＣＣ処理部３１０は、入力音声１４４からＭＦＣＣパラメータを音響特徴量として算出するためのＭＦＣＣ算出部３２０と、ＭＦＣＣ算出部３２０から出力されるＭＦＣＣパラメータに対し、ＭＦＣＣ・ＨＭＭ群を用いて認識処理を行ない、ＨＭＭごとに認識結果を出力するためのＭＦＣＣ通常発声認識処理部３２２と、ＭＦＣＣ算出部３２０から与えられるＭＦＣＣパラメータに対し、言直し発話用ＨＭＭを用いて認識処理を行ない、ＨＭＭごとに認識結果を出力するためのＭＦＣＣ言直し発話認識処理部３２４と、ＭＦＣＣ通常発声認識処理部３２２及びＭＦＣＣ言直し発話認識処理部３２４の出力のうち、尤度が最も高いものを選択して出力するための最尤選択部３２６とを含む。

【００７０】

図１１は、ＤＭＦＣＣ処理部３１２のより詳細なブロック図である。図１１を参照してＤＭＦＣＣ処理部３１２は、入力音声１４４から音響特徴量としてＤＭＦＣＣパラメータを算出するためのＤＭＦＣＣ算出部３３０と、ＤＭＦＣＣ算出部３３０から与えられるＤＭＦＣＣパラメータに対しＤＭＦＣＣ通常発声用ＨＭＭ群を用いて認識処理を行ない、認識結果をＨＭＭごとに出力するためのＤＭＦＣＣ通常発声認識処理部３３２と、ＤＭＦＣＣ算出部３３０からＤＭＦＣＣパラメータを受取り、言直し発話用ＤＭＦＣＣ・ＨＭＭ群を用いて認識処理を行ない、ＨＭＭごとに認識結果を出力するためのＤＭＦＣＣ言直し発話認識処理部３３４と、ＤＭＦＣＣ通常発声認識処理部３３２及びＤＭＦＣＣ言直し発話認識処理部３３４から出力される認識結果のうち、尤度が最も高いものを選択して出力するための最尤選択部３３６とを含む。

【００７１】

図１０及び図１１を参照してわかるように、ＭＦＣＣ処理部３１０及びＤＭＦＣＣ処理部３１２の構造は互いにパラレルである。使用する音響特徴量がＭＦＣＣかＤＭＦＣＣかによる差異があるにすぎない。従って以下では、ＭＦＣＣ処理部３１０の構造の詳細についてのみ説明する。

【００７２】

図１２は図１０に示すＭＦＣＣ通常発声認識処理部３２２のより詳細なブロック図である。図１２を参照して、ＭＦＣＣ通常発声認識処理部３２２は、ＭＦＣＣ算出部３２０から与えられるＭＦＣＣパラメータに基づき、男声通常発声用ＭＦＣＣ・ＨＭＭ群２１０及

び女声通常発声用ＭＦＣＣ・ＨＭＭ群２１４に対する雑音ＧＭＭの混合重み適応化によるＨＭＭ合成を重畳された雑音のＳＮＲごとに行ない、男声通常発声用適応化ＭＦＣＣ・ＨＭＭ群３５４及び女声通常発声用適応化ＭＦＣＣ・ＨＭＭ群３５２を生成するための雑音適応化処理部３５０と、男声通常発声用適応化ＭＦＣＣ・ＨＭＭ群３５４を用いて、入力されるＭＦＣＣパラメータに対するデコードを行なうことにより、適応化されたＨＭＭごとにデコード結果を出力するためのＭＦＣＣ男声通常発声デコーダ部３５８と、入力されるＭＦＣＣパラメータに対し女声通常発声用適応化ＭＦＣＣ・ＨＭＭ群を用いてデコードし、ＨＭＭごとにデコード結果を出力するためのＭＦＣＣ女声通常発声デコーダ部３５６とを含む。

【００７３】

男声通常発声用適応化ＭＦＣＣ・ＨＭＭ群３５４及び女声通常発声用適応化ＭＦＣＣ・ＨＭＭ群３５２はそれぞれ、３種類のＳＮＲごとに一つ、合計三個のＨＭＭを含む。デコードには無雑音ＨＭＭも使用するので、デコーダ部３５６及び３５８はそれぞれデコード結果を４つずつ出力する。その結果、ＭＦＣＣ通常発声認識処理部３２２全体としては８つのデコード結果を出力する。

【００７４】

ここで、図１２に示す雑音適応化処理部３５０の処理について図１５及び図１６を参照して説明する。図１５を参照して、雑音適応化処理部３５０は、入力されるＭＦＣＣパラメータに基づき、音響空間２７０中における入力音声に対応する領域４２０を推定する。そしてこの領域４２０と、予め求められている各種の雑音が占める領域２８０～３０２との距離を算出する。そして、距離が最も近いものを所定個数（本実施の形態では４つ）だけ選択する。図１５の例で示せば領域２９０、２９２、２９６及び２９８により示される雑音が、入力される音声の音響空間中の領域４２０に最も近い。従って、この４つの雑音に対応する音響モデルが採用される。

【００７５】

続いて図１６を参照して、これら４つの領域２９０、２９２、２９６及び２９８に対応するＨＭＭのガウス混合分布の重みを計算し、加算することにより、入力される音声の音響空間２７０中における領域４２０をカバーするような音響モデルをＨＭＭの形で算出する。この音響モデルを用いて入力音声に対するデコードを行なう。このように各雑音に対する音響モデル自体は変化させず今後のための重みのみを計算して音声認識用のＨＭＭの適応化を行なえばよい。そのため適用の計算時間が短く、大変高速に適応化を行なうことができる。さらに、適応化されたＨＭＭが複数の雑音環境の分布を含んでいる。従って単数の雑音から推定された音響モデルを用いた場合よりも、雑音の短時間の変動に対する頑健性がより高くなるという利点がある。

【００７６】

図１３はＭＦＣＣ言直し発話認識処理部３２４の構成を示す。ＭＦＣＣ言直し発話認識処理部３２４は、入力されるＭＦＣＣパラメータを用いて、男声言直し発話用ＭＦＣＣ・ＨＭＭ群２１２及び女声言直し発話用ＭＦＣＣ・ＨＭＭ群２１６に対し雑音ＧＭＭの混合重み適応化によるＨＭＭ合成法を重畳された雑音のＳＮＲごとに行ない、男声、女声及びＳＮＲごとに適応化されたＨＭＭを出力することにより、男声言直し発話用適応化ＭＦＣＣ・ＨＭＭ群３７４及び女声言直し発話用適応化ＭＦＣＣ・ＨＭＭ群３７２を出力するための雑音適応化処理部３７０と、与えられるＭＦＣＣパラメータを、女声言直し発話用適応化ＭＦＣＣ・ＨＭＭ群３７２を用いてデコードし、ＨＭＭごとに出力するためのＭＦＣＣ女声言直し発話デコーダ部３７６と、入力されるＭＦＣＣパラメータを男声言直し発話用適応化ＭＦＣＣ・ＨＭＭ群３７４を用いてデコードし、ＨＭＭごとにデコード結果を出力するためのＭＦＣＣ男声通常発声デコーダ部３７８とを含む。

【００７７】

女声言直し発話用適応化ＭＦＣＣ・ＨＭＭ群２１６は、ＳＮＲごとに合成される３つのＨＭＭを含む。男声言直し発話用適応化ＭＦＣＣ・ＨＭＭ群も同様に、ＳＮＲごとの３つのＨＭＭを含む。また、デコードには無雑音ＨＭＭも使用される。従って、デコーダ部３

７６及び３７８はそれぞれ４つずつのデコード結果を出力する。その結果ＭＦＣＣ言直し発話認識処理部３２４の出力は８つとなる。

【００７８】

図１２及び図１３を参照して明らかなように、ＭＦＣＣ通常発声認識処理部３２２とＭＦＣＣ言直し発話認識処理部３２４との構成はパラレルである。従って以下ではＭＦＣＣ通常発声認識処理部３２２の詳細な構造のみを説明する。また図１２及びこれ以前の説明から明らかなように、ＭＦＣＣ女声通常発声デコーダ部３５６及びＭＦＣＣ男声通常発声デコーダ部３５８の構成も互いにパラレルである。従って以下では女声についてのみＭＦＣＣ通常発声認識処理部３２２の詳細な構成を説明する。

【００７９】

図１４は、ＭＦＣＣ女声通常発声デコーダ部３５６及び女声通常発声用適応化ＭＦＣＣ・ＨＭＭ群３５２の詳細な構成を示す。図１４を参照して、女声ＭＦＣＣ・ＨＭＭ群３５２は、無雑音ＨＭＭ４０２、及びそれぞれ１０ｄＢ、２０ｄＢ，及び３０ｄＢのＳＮＲで雑音が重畳された雑音重畳ＨＭＭから合成された１０ｄＢ雑音ＨＭＭ４０４、２０ｄＢ雑音ＨＭＭ４０６、及び３０ｄＢ雑音ＨＭＭ４０８とを含む。

【００８０】

ＭＦＣＣ女声通常発声デコーダ部３５６は、入力されるＭＦＣＣパラメータを、無雑音ＨＭＭ４０２、１０ｄＢ雑音ＨＭＭ４０４、２０ｄＢ雑音ＨＭＭ４０６、及び３０ｄＢ雑音ＨＭＭ４０８をそれぞれ用いてデコードし、デコード結果を出力するためのデコーダ３９０、３９２、３９４、及び３９６を含む。

【００８１】

図１８に、図９に示す認識結果統合部３１４のより詳細な構成を示す。図９に示すＭＦＣＣ処理部３１０及びＤＭＦＣＣ処理部３１２からは複数の仮説が認識結果統合部３１４に与えられる。認識結果統合部３１４は、これら複数の仮説を単語単位で統合する。その原理について図１９～図２１を参照して説明する。

【００８２】

複数の音声認識デコーダから得られた仮説が互いに相補的である場合、それぞれの仮説の正しい部分を抽出して組み合わせることにより、より正しい単語列が得られる可能性がある。ここで「相補的」とは、あるデコーダの認識結果の前半は正しいが後半は間違いであったとしても、別のデコーダの認識結果の後半部分が正しいならば、それぞれの正しい部分をつなぎあわせることによりその認識誤りを補償することができるという意味である。

【００８３】

図１９を参照して、２つの仮説４７０及び４７２が得られたものとする。仮説４７０の前半部分は誤っているが後半部分は正しい認識結果である。一方、仮説４７２については、前半の認識結果は正しいが後半は誤りである。従って仮説４７２の前半部分と仮説４７０の後半部分とをつなぎ合わせることにより、正しい結果が得られるはずである。

【００８４】

図２０を参照して、上記した結果を得るために、まず図２０に示されるような単語ラティスを、与えられた２つの仮説から再構成する。この再構成では、個々の単語の開始及び終了時間情報を用いる。

【００８５】

続いて図２１に示されるように、この単語ラティス４８０に含まれる単語列経路のうち、音響尤度と言語尤度とから算出される尤度が最も大きくなるような単語列４８２を再探索する。通常、仮説のうちでも正しい部分の尤度は高く、誤っている可能性が高い部分の尤度は低くなっている。従って、このような再探索を行なうことにより２つの仮説を統合して正しい結果を得ることができる可能性が高くなる。

【００８６】

なお本実施の形態では、ＭＦＣＣとＤＭＦＣＣ特徴量から得られた仮説に対する仮説統合を認識結果統合部３１４で行なっている。この場合、ＭＦＣＣの音響モデルから計算さ

れる音響尤度と、ＤＭＦＣＣの音響モデルから計算される尤度とを直接比較することはできない。そのため、音響モデルの尤度を比較するためには尤度の正規化が必要である。本実施の形態では、そのために、認識文全体の音響尤度で個々の単語の音響尤度を割ることにより、各単語の尤度を正規化する。さらに、仮説統合の際には、言語モデルを用いた尤度計算も行なう。この場合、音響モデルの尤度計算と言語モデルによる尤度計算との間での重み付けを考慮しなければならない。本実施の形態では、仮説統合時における言語モデルウェイトを０．０６とした。

【００８７】

図１８を参照して、認識結果統合部３１４は、上記したような機能を実現するために以下の各処理部を含む。すなわち認識結果統合部３１４は、ＭＦＣＣ及びＤＭＦＣＣのそれぞれの仮説の単語の音響尤度を正規化するための尤度正規化部４５０と、２つの仮説から個々の単語の開始及び終了時間情報を用いて単語ラティス４８０（図２０参照）を作成するための単語ラティス作成部４５２と、統合の際に参照される言語モデルを記憶するための言語モデル記憶部４５６と、統合の際の言語モデルの尤度の、音響モデルの尤度に対するウェイトを記憶するためのウェイト記憶部４５４と、単語ごとの音響尤度及び言語モデルに基づく単語列の尤度に基づいて単語ラティス４８０中の、音響尤度と言語尤度との和が最大となるような単語列を再探索することにより認識結果を統合するための最尤経路探索部４５８とを含む。

【００８８】

上記した音声認識システム１３０は以下のように動作する。図２２に、このシステムの動作の概略の流れについて示す。大きく分けて、このシステムは２つの動作局面を持つ。第一の局面は、雑音重畳音声用のＨＭＭを準備するステップ５００である。第二の局面は、このようにして準備された雑音重畳音声用のＨＭＭと無雑音用のＨＭＭとを用いて、入力される音声の認識を行なうステップ（５０２～５０８）である。

【００８９】

ステップ５００では、図４に示すような初期ＨＭＭ１５０と、雑音ＤＢ１５２とを用いて、ＭＦＣＣ・ＨＭＭ群１５６が作成され、また雑音重畳学習データ１５３を用いてＤＭＦＣＣ・ＨＭＭ群１５８が作成される。

【００９０】

このようにして、雑音重畳音声用のＨＭＭ群が作成された後は、いつでもこのＭＦＣＣ・ＨＭＭ群１５６及びＤＭＦＣＣ・ＨＭＭ群１５８を用いた音声認識を行なうことができる。図４に示す入力音声１４４が与えられると、その入力音声からＭＦＣＣパラメータ及びＤＭＦＣＣパラメータが算出される（ステップ５０２）。それらを用いて、予め準備されたＭＦＣＣ・ＨＭＭ群１５６及びＤＭＦＣＣＨＭ群１５８のうち入力音声１４４の発話環境に最も類似した発話環境に対応する所定個数（本実施の形態では４個）のＨＭＭがＭＦＣＣ及びＤＭＦＣＣのそれぞれについて選択される。これらＨＭＭからＭＦＣＣ及びＤＭＦＣＣの各々について、雑音ＧＭＭの混合重み適応化によるＨＭＭが合成される。合成されるＨＭＭは、男声・女声、通常発声・言直し発話、及び４種類のＳＮＲ（１０ｄＢ、２０ｄＢ、３０ｄＢ、無雑音）の組み合わせの各々に対してであるから、全部で２×２×４＝１６通りである。

【００９１】

続いてステップ５０４で発話入力があったか否かが判定される。発話入力があればステップ５０６に進むが、発話入力がなければ、再び重み推定５０２を行なう。本実施の形態では、発話入力があった場合には、その直前の１秒間の期間における雑音を用いて重み推定を行なっている。

【００９２】

ステップ５０６では、合成されたＨＭＭを用いた認識と、それら認識結果の統合とが行なわれる。その認識結果がステップ５０８で出力される。この後再度重み推定５０２の処理から繰り返される。

【００９３】

図２３を参照して、発話５２２に対しては、発話５２２の直前の雑音５２４を用いて合成されたＨＭＭによる音声認識が行なわれる。同様に次の発話５２６に対しては、発話５２６の直前の雑音５２８により推定されたＨＭＭを用いて音声認識が行なわれる。

【００９４】

なお、上記した男声女声、ＭＦＣＣ及びＤＭＦＣＣ、通常発声及び言直し発話などの組合せは任意に選ぶことができる。ＭＦＣＣ又はＤＭＦＣＣのいずれか一方のみを用いるシステムも可能である。

[実験１]

上記した実施の形態に係る雑音適応化手法の評価を行なうため、日本語大語彙連続音声認識実験を行なった。実験においては、予め出願人において作成した言語モデルを準備した。言語モデルの作成に使用された自然発話音声・言語データベースに含まれていた単語は６７０万語程度である。実験に使用した音声波形は、サンプリング周波数１６ｋＨｚ、分析窓長２０ｍｓ、分析周期１０ｍｓで分析を行ない、ＭＦＣＣ及びＤＭＦＣＣ特徴量を抽出した。ＭＦＣＣの音響特徴パラメータは、１２次元ＭＦＣＣ、Δ_{C0}、１２次元ＤＭＦＣＣの計２５次元である。ＤＭＦＣＣの音響特徴パラメータは、１２次元ＤＭＦＣＣ、Δｐｏｗ、１２次元ΔＤＭＦＣＣの計２５次元である。使用した音素は、日本語分析でよく用いられる２６種類の音素である。

【００９５】

音響モデルの状態共有構造は、ＭＬ－ＳＳＳ（非特許文献１６を参照されたい）より生成した２１００状態のＨＭｎｅｔを使用した。各状態の混合数は５である。

【００９６】

学習データとして、出願人において準備した旅行会話データベースＴＲＡを用いた。このデータベースＴＲＡは、４０７名が発声した対話及び音素バランス５０３文の計３０時間である。

【００９７】

雑音適応元の音響モデルは、様々な場所で採取した１２種類の雑音を用いて生成した。ＭＦＣＣの音響モデルは、雑音とＳＮＲ毎にＰＭＣ法を用いて無雑音音声ＨＭＭを適応化することにより生成した。ＤＭＦＣＣの音響モデルは、雑音を重畳した学習データを用いて生成した。雑音重畳音声のＳＮＲは、１０ｄＢ、２０ｄＢ、３０ｄＢである。

【００９８】

ＭＦＣＣとＤＭＦＣＣの音響モデルはそれぞれ、男声女声、１２種類の雑音、及び３種類のＳＮＲとの組合せからなる、２×１２×３＝７２種類と無雑音音声モデルとの計７３種類である。

【００９９】

評価用音声データは、出願人において準備したＡＴＲ旅行会話基本表現集ＢＴＥＣ　ｔｅｓｔｓｅｔ－０１（５１０文、男性４名、女性６名、それぞれ５１文の発声データ）を使用し、１０ｄＢ、２０ｄＢ、３０ｄＢのＳＮＲで雑音を重畳した。評価用に重畳した雑音はＨＭＭの合成に用いた雑音とは異なる複数の場所で採取した雑音である。雑音ＧＭＭの混合ウェイト推定には１秒間の雑音を使用して個々の混合ウェイトの上位４つの雑音を用いて雑音重畳音声用音響モデルを生成した。

【０１００】

図２４に、３種類の評価用雑音重畳音声データに対する平均単語正解精度を示す。図中のＭＡＸは個々の音響モデル（１０ｄＢ、２０ｄＢ、３０ｄＢ、無雑音）を用いて得られた仮説を最大尤度基準で選択した場合の単語正解精度である。図２４に示すように、最大尤度基準による選択を行なうことで、実験に用いたＳＮＲ全てにおいて平均９０％以上の単語正解精度が得られた。ＤＭＦＣＣの音響モデルを用いるとＭＦＣＣの音響モデルを用いた場合よりも単語正解精度が低下している。しかしＤＭＦＣＣの無雑音音声音響モデルを用いた場合、雑音重畳音声の単語正解精度がＭＦＣＣの無雑音音響モデルよりも高い。従って、雑音の種類や雑音ＳＮＲに対する正解精度への影響がＭＦＣＣよりも小さいことがわかる。

［実験２］
さらに、言直し発話に対し頑健な音響モデルに対して雑音と発話スタイルの変動に対する単語正解精度への影響を調べるため、日本語大語彙連続音声認識実験を行なった。評価用音声として、実験１で用いた通常発声の音声と、意図的に音節ごとに区切って発声した音節強調発声の音声とを用いた。音節強調発声データは、旅行会話文、男性２名女性２名、各話者１０文の計４０文である。評価用音声には３０ｄＢ、２０ｄＢ、１０ｄＢのＳＮＲで、実験２で用いた３種類の雑音が重畳されている。

【０１０１】
言直し発話に頑健な音響モデルは、環境依存音素モデル数が通常発声モデルよりも多い。そのため探索空間が大きく広がり、通常発声音声に対して単語正解精度の低下が懸念される。そこで、上記実施の形態で説明した通り、言直し発話用音響モデルと通常発声用音響モデルとを用いて別々にデコーディングし、最大尤度基準による仮説の選択を行なった。

【０１０２】
図２５に、通常発声用音響モデルの場合、言直し発話用音響モデル単独の場合、２つの音響モデルをパラレルデコーディングした場合それぞれに対する単語正解精度を示す。図２５に示すように、言直し発話用音響モデルを単独で使用した場合、その単語正解精度は若干低下する。それに対しパラレルデコーディングを行なうことにより、通常発声の音声に対してもほぼ同等の正解精度が得られた。

【０１０３】
次に、音節強調発声の音声に対する単語正解精度を図２６に示す。図２６に示すように、言直し発話用音響モデルは、通常発声用音響モデルよりも高い単語正解精度が得られた。雑音重畳音声に対しても、実験１で得られた結果同様、１０ｄＢの音声に対しても無雑音音声や３０ｄＢの音声と同程度の単語正解精度が得られた。

【０１０４】
[実験３]
最後に、ＭＦＣＣ特徴量とＤＭＦＣＣ特徴量のデコーダから得られた仮説を統合することによる性能の改善を調べるための評価実験を行なった。予備実験から、上記実施の形態で述べたように仮説統合時における言語モデルウェイトを０．０６とした。図２７に、仮説統合を行なった場合の単語正解精度を示す。図２７に示すように、通常発声に対してはＭＦＣＣ特徴量の正解精度と同等の結果が得られた。さらに、音節強調発声に対しては、ＭＦＣＣとＤＭＦＣＣの各々の正解精度以上の性能が得られた。これは、仮説統合により、ＭＦＣＣによる仮説とＤＭＦＣＣによる仮説とが互いに相補的であったため、仮説統合によって精度が高くなったためと考えられる。

【０１０５】
以上のように本実施の形態の音声認識システム１３０では、雑音と発話スタイルの変動に頑健な音声認識を実現することを目指した。本システムでは、雑音の変動に頑健な音響特徴量としてのＤＭＦＣＣ、予め種々の雑音環境に適応化したＨＭＭを用いて雑音ＧＭＭの混合ウェイトから雑音適応ＨＭＭを高速に生成する雑音適応手法、言直し発話に頑健な音響モデル、及び複数の仮説を統合する手法を用いた。その結果、１０ｄＢから３０ｄＢのＳＮＲで雑音を重畳した通常発声の評価データに対して、平均９０％以上の単語正解精度が得られた。また、言直し発話などの発話スタイルの変動に対しても、通常発声用音響モデルのみを用いた場合よりも高い単語正解精度が得られた。

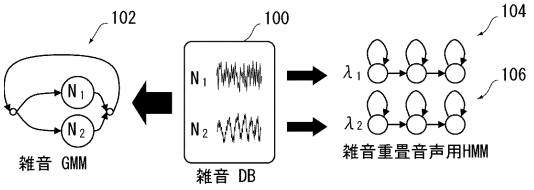
【０１０６】
今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

【図面の簡単な説明】
【０１０７】

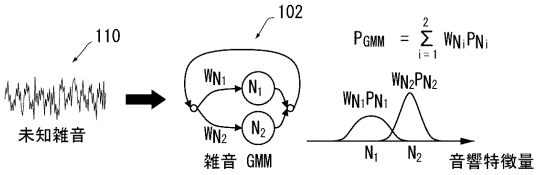
【図１】雑音ＧＭＭ及び雑音重畳音声ＨＭＭの作成を説明するための図である。
【図２】混合重みの推定を説明するための図である。
【図３】適応化ＨＭＭの生成を説明するための図である。
【図４】本発明の一実施の形態に係る音声認識システムのブロック図である。
【図５】ＨＭＭ作成部のより詳細なブロック図である。
【図６】本発明の一実施の形態における雑音重畳音声用ＭＦＣＣ・ＨＭＭ群の作成を説明するための図である。
【図７】雑音ＧＭＭの混合重み適応化において、ＰＭＣ法により準備される雑音ＨＭＭを説明するための図である。
【図８】本発明の一実施の形態において雑音重畳音声用ＤＭＦＣＣ・ＨＭＭを作成する方法を説明するための図である。
【図９】認識処理部のより詳細な構成を示すブロック図である。
【図１０】ＭＦＣＣ処理部３１０の詳細な構成を示すブロック図である。
【図１１】ＤＭＦＣＣ処理部３１２の詳細な構成を示すブロック図である。
【図１２】ＭＦＣＣ通常発声認識処理部の詳細な構成を示すブロック図である。
【図１３】ＭＦＣＣ言直し発話認識処理部の詳細な構成を示すブロック図である。
【図１４】ＭＦＣＣ女声通常発声デコーダ部３５６及び女声通常発声用適応化ＭＦＣＣ・ＨＭＭ群３５２の詳細な構成を示すブロック図である。
【図１５】本実施の形態における入力音声の発話環境から、予め準備された雑音ＨＭＭの発話環境までの距離を概念的に説明するための図である。
【図１６】入力音声の発話環境に類似した雑音を含む雑音ＨＭＭから適応化ＨＭＭを合成する概念を示す図である。
【図１７】言直し発話に頑健な音響モデルの構成を示す図である。
【図１８】認識結果統合部の詳細な構成を示すブロック図である。
【図１９】仮説統合の経過を説明するための、２つの仮説を示す図である。
【図２０】仮説統合の過程で生成される単語ラティスを示す図である。
【図２１】仮説統合の際に行なわれる最尤単語列の探索を説明するための図である。
【図２２】本発明の一実施の形態に係る音声認識システムの動作を説明するための図である。
【図２３】発話ごとの音声認識に用いられる雑音の位置を説明するための図である。
【図２４】本発明の一実施の形態を用いて行なわれた、雑音適応による頑健化の評価実験の結果を示すグラフである。
【図２５】本発明の一実施の形態の音声認識システムを用いて行なわれた、言直し発話用音響モデルによる頑健化の評価実験の結果を示すグラフである。
【図２６】音節強調発声の音声に対する単語正解精度を示すグラフである。
【図２７】本発明の一実施の形態に係る音声認識システムにおいて行なわれる仮説統語による単語正解精度を調べる実験結果を示すグラフである。
【図２８】ＰＭＣ法の概念を説明するための図である。
【符号の説明】
【０１０８】
１３０　音声認識システム、１５０　初期ＨＭＭ、１５２　雑音ＤＢ、１５３　雑音重畳学習データ、１５４　ＨＭＭ作成部、１５６　雑音重畳音声用ＭＦＣＣ・ＨＭＭ群、１５８　雑音重畳音声用ＤＭＦＣＣ・ＨＭＭ群、１４４　入力音声、１４２　認識処理部、１４６　音声認識結果、１９０　無雑音通常発声用ＭＦＣＣ・ＨＭＭ、１９２　無雑音言直し発話用ＭＦＣＣ・ＨＭＭ、２１０　男声通常発声用ＭＦＣＣ・ＨＭＭ群、２１２　男声言直し発話用ＭＦＣＣ・ＨＭＭ群、２１４　女声通常発声用ＭＦＣＣ・ＨＭＭ群、２１６　女声言直し発話用ＭＦＣＣ・ＨＭＭ群、２３０　無雑音通常発声用ＤＭＦＣＣ・ＨＭＭ、２３２　無雑音言直し発話用ＤＭＦＣＣ・ＨＭＭ、２５０　男声通常発声用ＤＭＦＣＣ・ＨＭＭ群、２５２　男声言直し発話用ＤＭＦＣＣ・ＨＭＭ群、２５４　女声通常発声用ＤＭＦＣＣ・ＨＭＭ群、２５６　女声言直し発話用ＤＭＦＣＣ・ＨＭＭ群、３１０　Ｍ

ＦＣＣ処理部、３１２　ＤＭＦＣＣ処理部、３１４　認識結果統合部、３２０　ＭＦＣＣ算出部、３２２　ＭＦＣＣ通常発声認識処理部、３２４　ＭＦＣＣ言直し発話認識処理部、３２６　最尤選択部、３３０　ＤＭＦＣＣ算出部、３３２　ＤＭＦＣＣ通常発声認識処理部、３３４　ＤＭＦＣＣ言直し発話認識処理部、３３６　最尤選択部、３５０　雑音適応化処理部、３５６　ＭＦＣＣ女声通常発声デコーダ部、３５８　ＭＦＣＣ男声通常発声デコーダ部、３７０　雑音適応化処理部、３７６　ＭＦＣＣ女声言直し発話デコーダ部、３７８　ＭＦＣＣ男声言直し発話デコーダ部、４５０　尤度正規化部、４５２　単語ラティス作成部、４５８　最尤経路探索部、４８０　単語ラティス

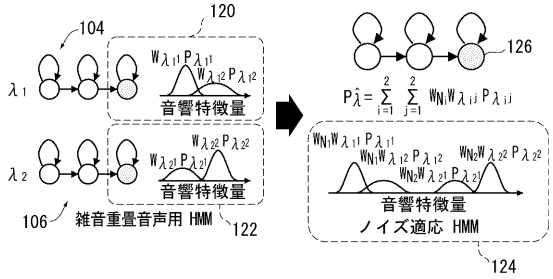
【図１】



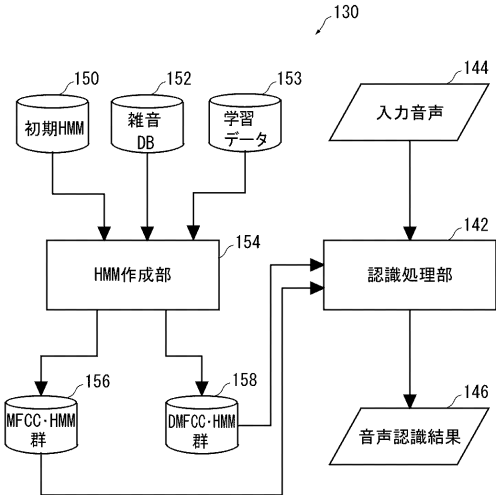
【図２】



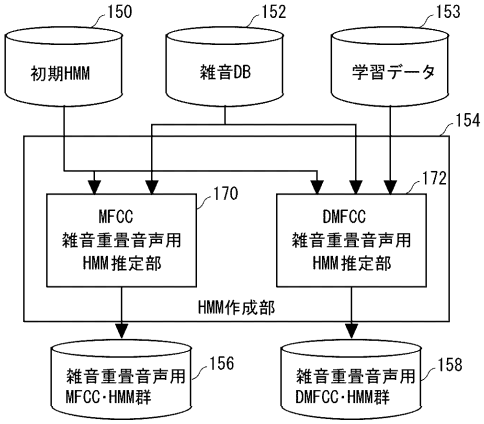
【図３】



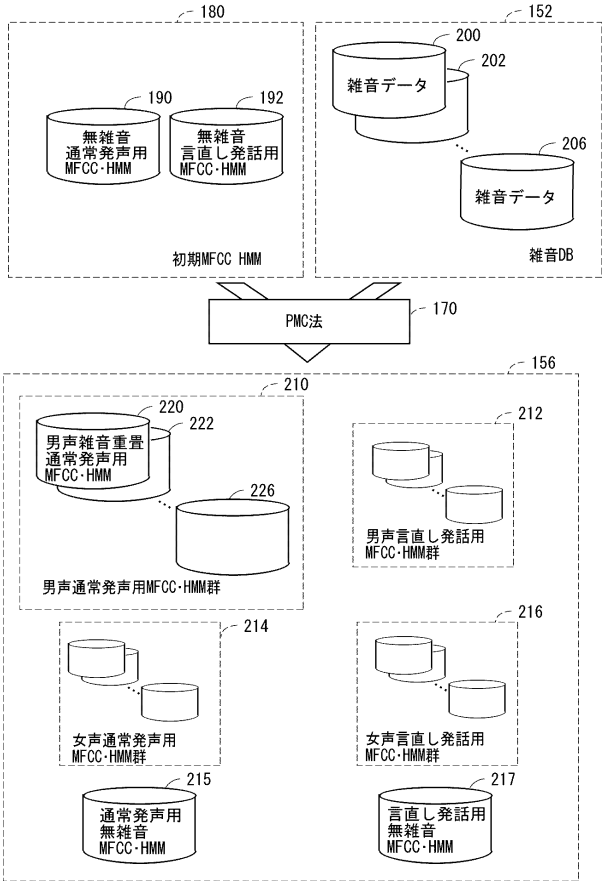
【図４】



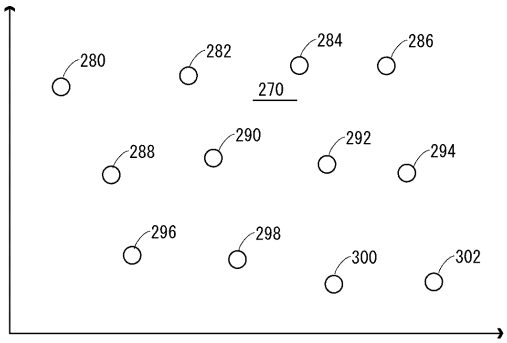
【図５】


150
初期HMM
152
雑音DB
153
学習データ
154
170
MFCC
雑音重畳音声用
HMM推定部
172
DMFCC
雑音重畳音声用
HMM推定部
HMM作成部
雑音重畳音声用
MFCC・HMM群
156
雑音重畳音声用
DMFCC・HMM群
158

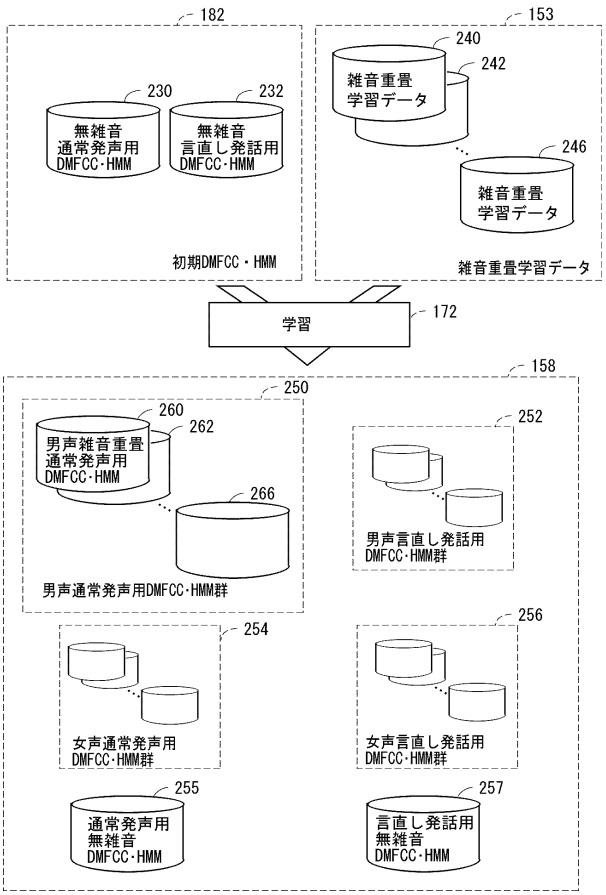
【図６】


180
190
無雑音
通常発声用
MFCC・HMM
192
無雑音
言直し発話用
MFCC・HMM
初期MFCC HMM
152
200
雑音データ
202
206
雑音データ
雑音DB
PMC法
170
156
210
220
男声雑音重畳
通常発声用
MFCC・HMM
222
226
男声通常発声用MFCC・HMM群
212
男声言直し発話用
MFCC・HMM群
214
女声通常発声用
MFCC・HMM群
216
女声言直し発話用
MFCC・HMM群
215
通常発声用
無雑音
MFCC・HMM
217
言直し発話用
無雑音
MFCC・HMM

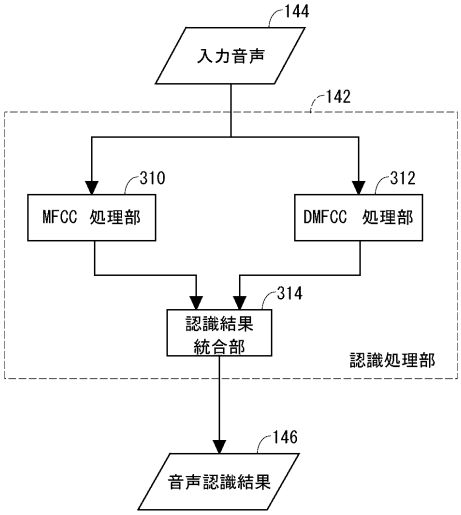
【図７】


280
282
284
286
270
288
290
292
294
296
298
300
302

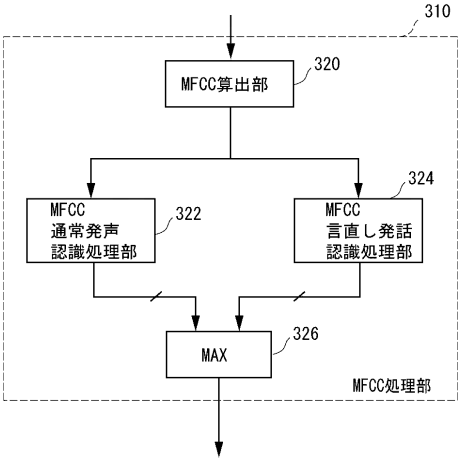
【図８】


182
230
無雑音
通常発声用
DMFCC・HMM
232
無雑音
言直し発話用
DMFCC・HMM
初期DMFCC・HMM
153
240
雑音重畳
学習データ
242
246
雑音重畳
学習データ
雑音重畳学習データ
学習
172
158
250
260
男声雑音重畳
通常発声用
DMFCC・HMM
262
266
男声通常発声用DMFCC・HMM群
252
男声言直し発話用
DMFCC・HMM群
254
女声通常発声用
DMFCC・HMM群
256
女声言直し発話用
DMFCC・HMM群
255
通常発声用
無雑音
DMFCC・HMM
257
言直し発話用
無雑音
DMFCC・HMM

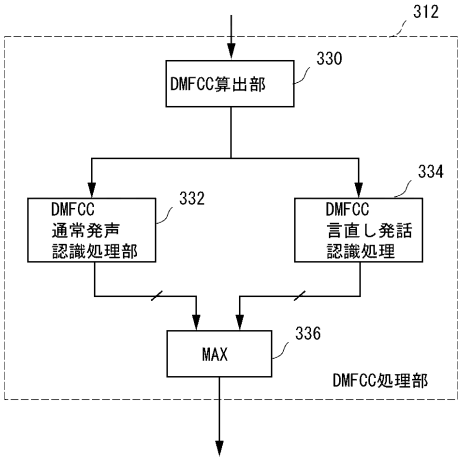
【図９】



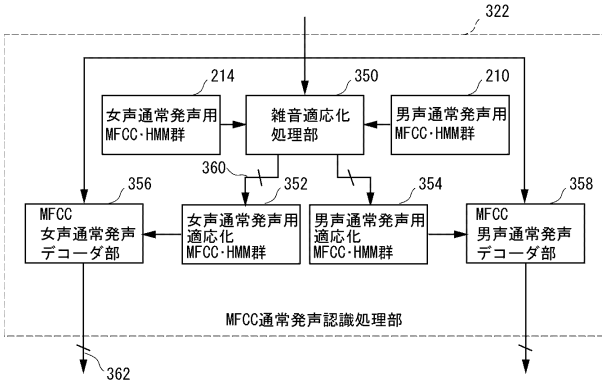
【図１０】



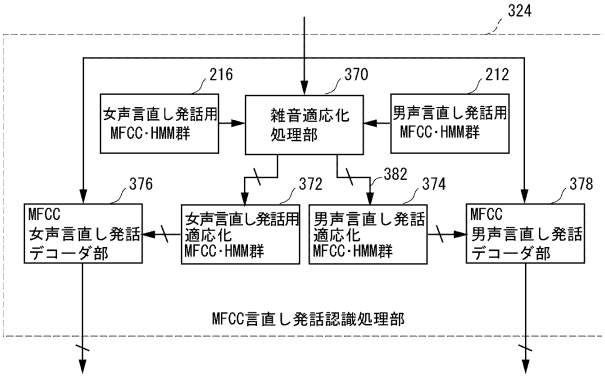
【図１１】



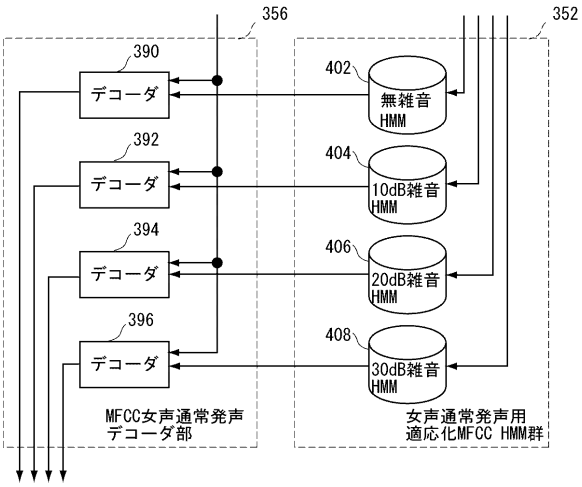
【図１２】



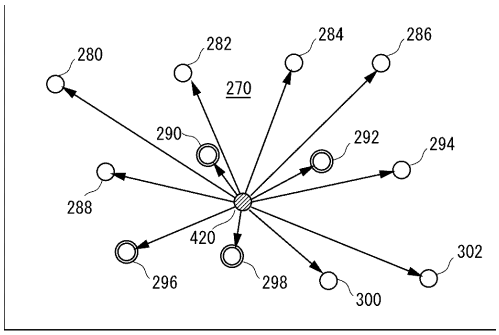
【図１３】



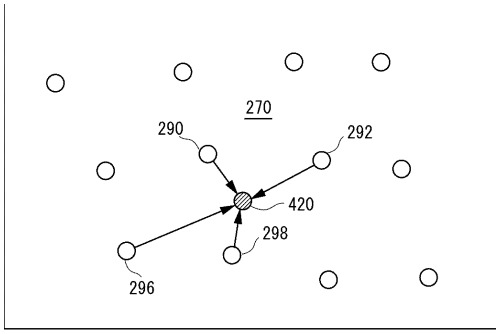
【図１４】



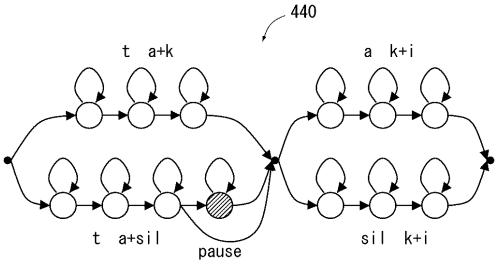
【図１５】



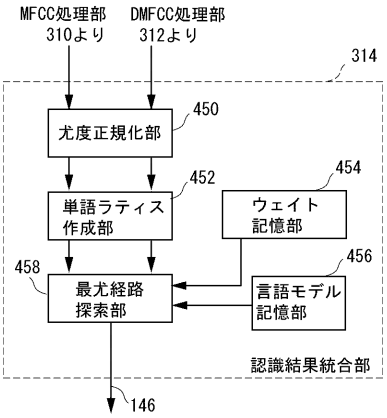
【図１６】



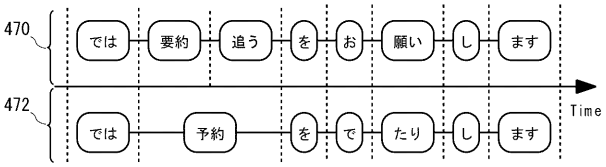
【図１７】



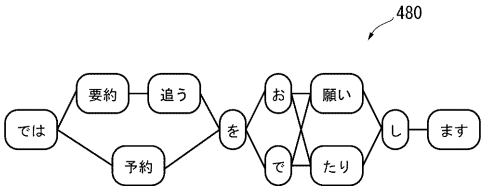
【図１８】



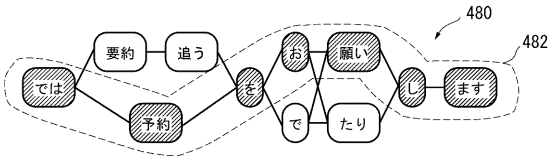
【図１９】



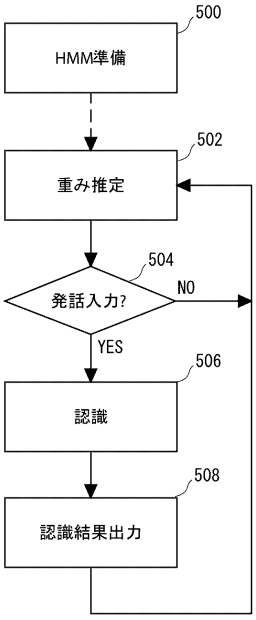
【図２０】



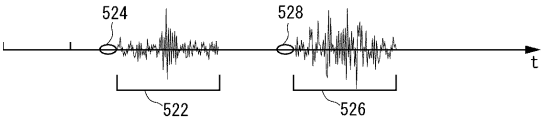
【図２１】



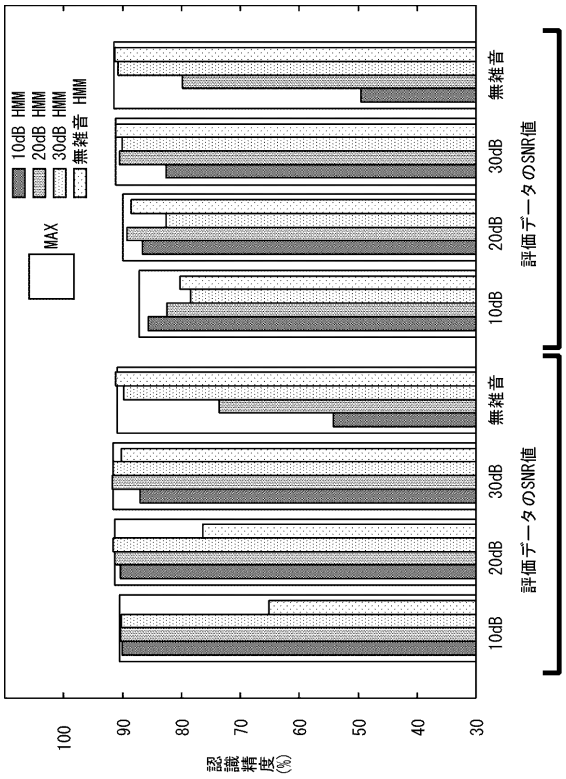
【図２２】



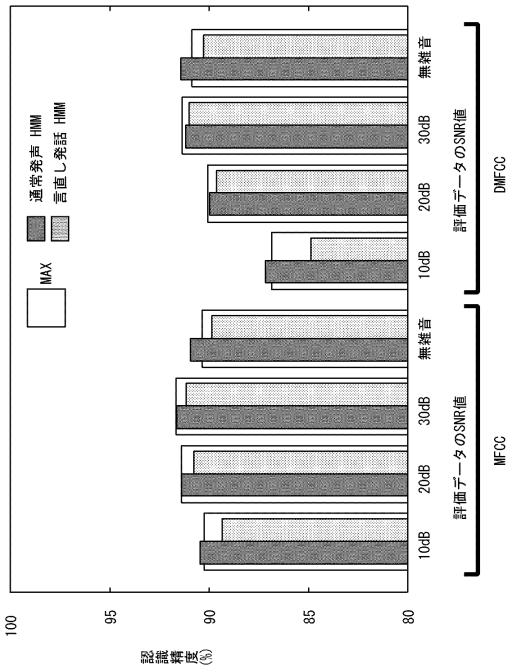
【図２３】



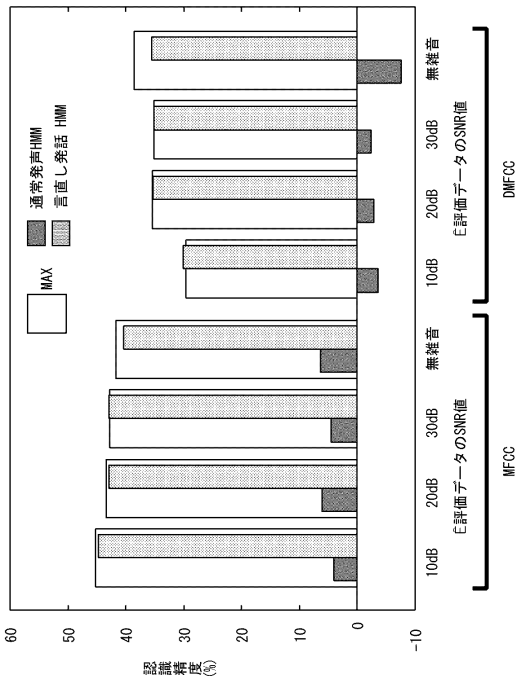
【図２４】



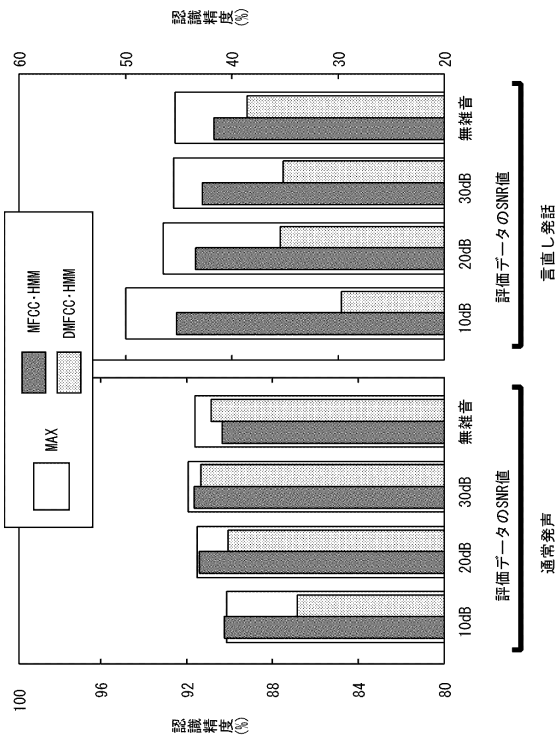
【図２５】



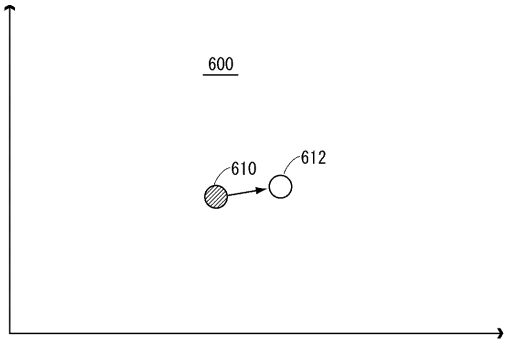
【図２６】



【図２７】



【図２８】



フロントページの続き

(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

審査官　菊池　智紀

(56)参考文献　特開２００３－１７７７８１（ＪＰ，Ａ）
特開２００２－１８９４９４（ＪＰ，Ａ）
特開２００４－００４５０９（ＪＰ，Ａ）
Konstantin MARKOV et al., "Noise and Channel Distortion Robust ASR System for DARPA SPINE2 Task", IEICE Trans. Inf. & Syst., ２００３年　3月　1日, Vol.E86-D, No.3, p.497-504
伊田政樹 他, "雑音GMMの適応化とSN比別マルチパスモデルを用いたHMM合成による高速な雑音環境適応化", 電子情報通信学会論文誌D-II, ２００３年　2月　1日, Vol.J86-D-II, No.5, p.195-203

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ１０Ｌ　１５／００－１７／００
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍＩＩ）