(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6234060号
(P6234060)
(45)発行日　平成29年11月22日(2017.11.22)　(24)登録日　平成29年11月2日(2017.11.2)

(51)Int.Cl.　　F I
G10L 15/06 (2013.01)　G10L 15/06　300Z
G10L 15/20 (2006.01)　G10L 15/20　153
G10L 15/14 (2006.01)　G10L 15/14　200Z

請求項の数 12　(全 20 頁)

(21)出願番号　特願2013-99645(P2013-99645)
(22)出願日　平成25年5月9日(2013.5.9)
(65)公開番号　特開2014-219605(P2014-219605A)
(43)公開日　平成26年11月20日(2014.11.20)
審査請求日　平成27年11月27日(2015.11.27)

前置審査

(73)特許権者　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
INTERNATIONAL BUSINESS MACHINES CORPORATION
アメリカ合衆国10504　ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
New Orchard Road, Armonk, New York 10504, United States of America
(74)代理人　100108501
弁理士　上野　剛史

最終頁に続く

(54)【発明の名称】ターゲットドメインの学習用音声データの生成方法、生成装置、および生成プログラム

(57)【特許請求の範囲】
【請求項１】
コンピュータの演算処理により、ソースドメインの音声データを利用してターゲットドメインの学習用の音声データを生成する方法であって、
ターゲットドメインのクリーンな音声データを用いて学習された前記ターゲットドメインの混合ガウスモデル(Gaussian mixture model :GMM)を読み出すステップと、
前記ターゲットドメインのGMMを参照して、入力として受け取ったソースドメインの音声データの補正量を求めることにより、前記ソースドメインの音声データを、前記ターゲットドメインの音声データに、該ターゲットドメインの音声データのチャネル特性に基づきマッピングするステップと、
マッピングした前記ソースドメインの音声データに、前記ターゲットドメインのノイズを加えて、擬似的なターゲットドメインの音声データを出力するステップとを含み、
前記マッピングするステップは、前記ソースドメインの音声データを前記ターゲットドメインの音声データにチャネルマッピング・パラメータによってマッピングする生成モデル式を決定するステップと、前記生成モデル式を参照して前記ターゲットドメインのGMMから変換したソースドメインのGMMに含まれるチャネルマッピング・パラメータを、Expectation Maximization(EM)アルゴリズムを用いて推定するステップとを含む、
学習データ生成方法。
【請求項２】
前記チャネルマッピング・パラメータは、前記擬似的なターゲットドメインの音声デー

タを求めるために前記ソースドメインの音声データから差し引くチャネルバイアスを含む、請求項１に記載の学習データ生成方法。
【請求項３】
前記チャネルマッピング・パラメータは、前記擬似的なターゲットドメインの音声データを求めるために前記ソースドメインの音声データに乗算するチャネル振幅を更に含む、請求項２に記載の学習データ生成方法。
【請求項４】
前記EMアルゴリズムを用いて推定するステップは、変換した前記ソースドメインのGMMにソースドメインの観測値を入力して得られる音響尤度を求めるステップと、求めた音響尤度に基づいた目的関数を最小化する前記チャネルマッピング・パラメータを求めるステップとを交互に繰り返すステップを含む、請求項１に記載の学習データ生成方法。
【請求項５】
前記音響尤度を求めるステップにおいて、現在の推定されている前記チャネルマッピング・パラメータに基づき算出した雑音成分を参照する、請求項４に記載の学習データ生成方法。
【請求項６】
前記生成モデル式を参照して前記ターゲットドメインのGMMから変換したソースドメインのGMMを求める際にVector Taylor Series(VTS)近似を用いる、請求項５に記載の学習データ生成方法。
【請求項７】
学習された前記ターゲットドメインの混合ガウスモデル（GMM）は話者の性別ごと用意されており、入力として受け取ったソースドメインの音声データの１発話ごとに男性らしさ及び女性らしさを求めながら、前記ターゲットドメインの音声データにマッピングする、請求項１に記載の学習データ生成方法。
【請求項８】
前記マッピングするステップは、ソースドメインの各音声データについて、前記ターゲットドメインのGMMのガウス分布コンポーネントのうち、該音声データとの音響空間における距離が近いガウス分布コンポーネントの平均との差を求め、該差を、ガウス分布コンポーネントごとの尤度で重みづけした上で、時間方向の平均として求めて前記ソースドメインの音声データに足し合わせるステップを含む、請求項１に記載の学習データ生成方法。
【請求項９】
コンピュータに、請求項１乃至８のいずれか一項に記載の学習データ生成方法の各ステップを実行させるための学習データ生成プログラム。
【請求項１０】
請求項１乃至８のいずれか一項に記載の学習データ生成方法の各ステップを実行するように適合された手段を備える、学習データ生成システム。
【請求項１１】
ソースドメインの音声データを利用してターゲットドメインの学習用の音声データを生成するシステムであって、
ターゲットドメインの少量のクリーンな音声データを用いて学習された前記ターゲットドメインの混合ガウスモデル(Gaussian mixture model :GMM)を格納する混合ガウスモデル格納部と、
前記ターゲットドメインのGMMを参照して、入力として受け取ったソースドメインの音声データの補正量を求めることにより、前記ソースドメインの音声データを、前記ターゲットドメインの音声データに、該ターゲットドメインの音声データのチャネル特性に基づきマッピングするチャネルマッピング部と、
マッピングした前記ソースドメインの音声データに、前記ターゲットドメインのノイズを加えて、擬似的なターゲットドメインの音声データを出力するノイズ付加部と、
を含み、

前記チャネルマッピング部は、前記ソースドメインの音声データを前記ターゲットドメインの音声データにチャネルマッピング・パラメータによってマッピングする生成モデル式を決定し、該生成モデル式を参照して前記ターゲットドメインのGMMから変換したソースドメインのGMMに含まれるチャネルマッピング・パラメータを、Expectation Maximization(EM)アルゴリズムを用いて推定する、

学習データ生成システム。

【請求項１２】

前記チャネルマッピング・パラメータは、前記擬似的なターゲットドメインの音声データを求めるために前記ソースドメインの音声データから差し引くチャネルバイアスと、前記ソースドメインの音声データに乗算するチャネル振幅を含む、請求項１１に記載の学習データ生成システム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、ソースドメインの音声データを用いてターゲットドメインの音声データを生成する技術に関し、より詳細には、ソースドメインの音声データをターゲットドメインの音声データのチャネル特性に基づいてマッピングする技術に関する。

【背景技術】

【０００２】

音声認識の性能はターゲットドメインの音響環境に大きく依存する。即ち、音響モデルを学習する環境と、音声を評価する環境の間に音響的ミスマッチがあると、音声認識システムの性能は多くの場合低下する。環境間のミスマッチは、背景雑音、録音機器の音響特性、チャネル歪みなどの様々な原因によって大きくなる。そこで従来は、ターゲットドメインの音響モデルを構築することで環境間のミスマッチを避けるため、特定の環境における音声データを録音することに非常に多くの時間と労力とが費やされた。

【０００３】

これとは対照的に、近年のスマートフォン等のハンドヘルド・デバイスを用いたインターネット・サービス（例えば、音声検索やボイスメールなど）によって、大量の生きた音声データを低コストで取得することが可能となった。そこでこのような種々の音響環境における豊富な音声データを再利用することが望まれる。

【０００４】

音声認識におけるクロスドメインの問題は、伝統的に以下の４つのアプローチに大別される。

１．再利用法

２．モデル適応法

３．特徴量変換法

４．正規化法

１．の再利用法は、ターゲットドメインの音響モデルを構築するために、ソースドメインの音声データを用いてターゲットドメインの音声データをシミュレートする手法である（例えば、非特許文献１～２を参照）。

２．のモデル適応法は、ソースドメインの音響モデルのパラメータを変更してテスト音声に合わせる手法であり、最大事後確率推定法（MaximumA Posteriori Estimation : MAP）や、最尤線形回帰法（Maximum LikelihoodLinear Regression : MLLR）がこれに該当する（例えば、特許文献１、非特許文献３～非特許文献５を参照）。なお、該手法とは異なるが、同様にモデルを適応させる技術として特許文献２～３、非特許文献６が存在する。

３．の特徴量変換法は、デコード時にテスト音声の特徴量をソースドメインの音響モデルに合わせるように変換する手法であり、特徴量空間最尤線形回帰法（Feature space Maximum Likelihood Linear Regression : fMLLR）や、特徴量空間相互情報量最小化法（Feature speace Minimum Mutual Information : fMMI)がこれに該当する（例えば、非特許文献３～５、非特許文献７を参照）。

４．の正規化法は、テスト音声の特徴量の分布を正規化してソースドメインの音響モデルに合わせる手法であり、ケプストラム平均正規化法（Cepstral MeanNormalization: CMN）、分散平均正規化法(Mean and Variance Normalization: MVN)がこの手法に該当する（例えば、非特許文献８を参照）。

【０００５】

これら１．～４．の各手法は組み合わせて用いることも可能である。また、２．～４．の各手法は既に確立した技術である。一方で、１．の手法は全プロセスの出発点として重要な技術であるが、該手法に属する既存技術は、上述したインターネットを介して収集される音声データに対して適用することはできない。

【０００６】

非特許文献１は、ソースドメインのクリーン音声を入力として、まずターゲットドメインにおけるインパルス応答を畳み込み、その後雑音を加えて、ターゲットドメインの音声をシミュレートする手法を開示する（図２Ａを参照）。該手法は、チャネルと雑音の特性を補償する最も直接的な手法であるが、インターネット上の音声データをソースデータとする場合には本手法は適さない。なぜならば、ソースデータはクリーン音声とはいえず、また、入力データのチャネル特性が単一のインパルス応答に対しては多様すぎるためである。

【０００７】

非特許文献２は、ステレオデータを用いたマッピング手法を開示する。即ち、非特許文献２の技術は、ソースドメインの音声データとターゲットドメインの音声データを同時に記録することを必要とする。ソースとなる音声データをインターネット上のライブデータとする場合、ステレオデータを用意するのは困難であるため、本手法を利用することはできない。

【０００８】

また、１．の手法とは異なるが、目的タスクに合致した音声データを既存の音声コーパスから選択することで目的タスクの音声コーパスを構築する技術を開示するものもある（非特許文献９を参照）。

【０００９】

なお、以下の非特許文献１０～１１は、事前に用意したクリーン音声の混合ガウスモデル(Gaussian mixture model :GMM)と、クリーン音声と観測音声の関係式から、Vector Taylor Series (VTS)近似より観測音声のGMMを算出する技術を示す背景技術として列挙するものである。

【先行技術文献】

【特許文献】

【００１０】

【特許文献１】特開２０１２－４２９５７号公報
【特許文献２】特表２００２－５２９８００号公報
【特許文献３】特開平１０－１４９１９１号公報

【非特許文献】

【００１１】

【非特許文献１】V. Stahl、A. Fischer、R. Bippus、"AcousticSynthesis of Training Data for Speech Recognition in LivingRoom Environments," Proc. of ICASSP、Vol. 1、pp. 285-288、2001.
【非特許文献２】J. Droppo、L. Deng、A. Acero、"Evaluationof the SPLICE Algorithm on the Aurora2 Database"、Proc. of Eurospeech、pp. 217-220, 2001.
【非特許文献３】P. J. Moreno、B. Raj、R. M. Stern、"A vector Taylor series approach forenvironment-independent speech recognition"、Proc. of ICASSP、Vol. 2、 1996.
【非特許文献４】M. L. Seltzer、A. Acero、"Factored Adaptation for SeparableCompensation of Speaker and Environmental Variability", Proc. of ASRU、pp. 146-151、

2011.
【非特許文献５】M. J. F. Gales、"Maximum likelihoodlinear transformations for HMM based speech recognition、"Computer Speech and Language、Vol. 12、pp.75-98、1998.
【非特許文献６】M. J. F. Gales、S. J. Young、"Robust continuous speech recognition using parallel model combination"、IEEE Trans. on Sp.and Audio Proc.、Vol. 4, pp. 352-359、1996.
【非特許文献７】B.Kingsbury、L. Mangu、G. Saon、H. Soltau、G. Zweig、 ”fMPE: Discriminatively Trained Features for SpeechRecognition “ 、Proc. ICASSP、Vol.1、pp.961-964、2005
【非特許文献８】小川　厚徳、高橋　敏、「ケプストラム正規化の実行単位に関する実験的検証」、電子情報通信学会論文誌D、 Vol. J90-D No.9、pp.2648-2651
【非特許文献９】T. Cincarek、T. Toda、H. Saruwatari、K. Shikano、“ Utterance-based Selective Training for the AutomaticCreation of Task-Dependent Acoustic Models, ” EICETRANSACTIONS on Information and SystemsVol.E89-DNo.3pp.962-969
【非特許文献１０】B. Raj、E. Gouvea、R. M. Stern、"Cepstral compensationusing statistical linearization" 、Proc. of the ETRW、1997.
【非特許文献１１】D. Y. Kim、C. K. Un、N. S. Kim、"Speech recognition in noisy environments using first-order vector Taylor series"、 Speech Communication、 Vol. 24、pp. 39-49、1998.

【発明の概要】
【発明が解決しようとする課題】
【００１２】
本発明は、上記従来技術における問題点に鑑みてなされたものであり、インターネット上の音声データのように豊富に存在する音響環境の異なる音声データを再利用してターゲットドメインの音声データをシミュレートする方法、装置、およびプログラムを提供することを目的とする。
【課題を解決するための手段】
【００１３】
本願発明は、上記従来技術の課題を解決するために以下の特徴を有する音声データを生成する方法を提供する。本願発明の音声データを生成する方法は、ターゲットドメインのクリーンな音声データを用いて学習された前記ターゲットドメインの混合ガウスモデル(Gaussian mixture model :GMM)を読み出すステップと、前記ターゲットドメインのGMMを参照して、入力として受け取ったソースドメインの音声データを、前記ターゲットドメインの音声データに、該ターゲットドメインの音声データのチャネル特性に基づきマッピングするステップと、マッピングした前記ソースドメインの音声データに、前記ターゲットドメインのノイズを加えて、擬似的なターゲットドメインの音声データを出力するステップとを含む。ここで、ターゲットドメインのGMMの学習に使用されるターゲットドメインのクリーンな音声データは少量であってよい。
【００１４】
好ましくは、前記マッピングするステップは、チャネルマッピング・パラメータによって前記ソースドメインの音声データを前記ターゲットドメインの音声データにマッピングする生成モデル式を決定するステップと、前記生成モデル式を参照して前記ターゲットドメインのGMMから変換したソースドメインのGMMに含まれるチャネルマッピング・パラメータを、Expectation Maximization(EM)アルゴリズムを用いて推定するステップとを含む。
【００１５】
好ましくは、前記チャネルマッピング・パラメータは、前記擬似的なターゲットドメインの音声データを求めるために前記ソースドメインの音声データから差し引くチャネルバイアスを含む。
【００１６】

より好ましくは、前記チャネルマッピング・パラメータは、前記擬似的なターゲットドメインの音声データを求めるために前記ソースドメインの音声データに乗算するチャネル振幅を更に含む。

【００１７】

好ましくは、前記EMアルゴリズムを用いて推定するステップは、変換した前記ソースドメインのGMMにソースドメインの観測値を入力して得られる音響尤度を求めるステップと、求めた音響尤度に基づいた目的関数を最小化する前記チャネルマッピング・パラメータを求めるステップとを交互に繰り返すステップを含む。

【００１８】

より好ましくは、前記音響尤度を求めるステップにおいて、現在の推定されている前記チャネルマッピング・パラメータに基づき算出した雑音成分を参照する。

【００１９】

更に好ましくは、前記生成モデル式を参照して前記ターゲットドメインのGMMから変換したソースドメインのGMMを求める際にVector Taylor Series（VTS)近似を用いる。

【００２０】

好ましくは、学習された前記ターゲットドメインの混合ガウスモデル（GMM）は話者の性別ごと用意されており、入力として受け取ったソースドメインの音声データの１発話ごとに男性らしさ及び女性らしさを求めながら、前記ターゲットドメインの音声データにマッピングする。

【００２１】

また好ましくは、前記マッピングするステップは、ソースドメインの各音声データについて、前記ターゲットドメインのGMMのガウス分布コンポーネントのうち、該音声データとの音響空間における距離が近いガウス分布コンポーネントの平均との差を求め、該差を、ガウス分布コンポーネントごとの尤度で重みづけした上で、時間方向の平均として求めて前記ソースドメインの音声データに足し合わせるステップを含む。

【００２２】

以上、ターゲットドメインの学習用の音声データを生成する方法として本発明を説明した。しかし本発明は、そのような学習用音声データ生成方法の各ステップをコンピュータに実行させる学習用音声データ生成プログラム、及び該学習用音声データ生成プログラムを１以上のコンピュータにインストールして実現される学習用音声データ生成システムとして把握することもできる。

【発明の効果】

【００２３】

本願発明によれば、クリーンな音声データを用いて学習されたターゲットドメインのGMMを参照して、入力として受け取ったソースドメインの音声データを、ターゲットドメインの音声データに、該ターゲットドメインの音声データのチャネル特性に基づいてマッピングして、ターゲットドメインの音声データをシミュレートする。このため、両方の音声について書き起こしデータがある必要も音声認識する必要もない。そして、インターネット上の音声データのように豊富に存在する音響環境の異なる音声データを再利用してターゲットドメインの音声データをシミュレートすることが可能となる。本発明のその他の効果については、各実施の形態の記載から理解される。

【図面の簡単な説明】

【００２４】

【図１】本願発明の実施形態による学習用音声データ生成システムを実現するのに好適な情報処理装置のハードウェア構成の一例を示す。

【図２Ａ】従来の学習用音声データをシミュレートするシステムの機能ブロック図である。

【図２Ｂ】本発明の実施形態による学習用音声データ生成システムの機能ブロック図である。

【図３Ａ】正規化された自動車のノイズを示すグラフである。

【図３Ｂ】正規化された自動車のノイズを示すグラフである。
【図４】本発明の実施形態による学習用音声データ生成処理の全体の流れの一例を示すフローチャートある。
【図５】本発明の実施形態によるマッピング処理の流れの一例を示すフローチャートある。
【図６】本発明を適用して生成した学習用音声データから構築された音響モデルを用いて行った音声認識の実験結果を示す図である。
【発明を実施するための形態】
【００２５】
以下、本願発明を実施するための形態を図面に基づいて詳細に説明するが、以下の実施形態は特許請求の範囲にかかる発明を限定するものではなく、また実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。なお、実施の形態の説明の全体を通じて同じ要素には同じ番号を付している。
【００２６】
図１は、本発明を実施するためのコンピュータ１００の例示的なハードウェア構成を示す。外部記憶装置１１４やＲＯＭ１０６は、オペレーティング・システムと協働してＣＰＵ１０２に命令を与え、本発明を実施するための複数のコンピュータ・プログラムのコードや各種データを記録することができる。そして外部記憶装置１１４やＲＯＭ１０６に格納された複数のコンピュータ・プログラムは各々ＲＡＭ１０４にロードされることによってＣＰＵ１０２により実行される。なお、外部記憶装置１１４は、ＳＣＳＩコントローラなどのコントローラ（図示しない）を経由してバス１０８へ接続されている。また、複数のコンピュータ・プログラムのコードには、本発明の実施形態に係る学習用音声データ生成プログラムが含まれる。また、各種データには、本発明においてソースドメインの音声データとするインターネット上の様々な音響環境の音声データや、ターゲットドメインの少量のクリーンな音声データを用いて学習されたターゲットドメインのGMMが含まれる。
【００２７】
コンピュータ・プログラムは圧縮し、また複数に分割して複数の媒体に記録することもできる。なお、ＣＰＵ１０２が、外部記憶装置１１４から渡されるデジタル信号に対して、学習用音声データ生成プログラムにより行う処理の詳細は後述する。
【００２８】
コンピュータ１００はまた、視覚データをユーザに提示するための表示装置１１６を含む。表示装置１１６は、グラフィックスコントローラ（図示しない）を経由してバス１０８へ接続されている。コンピュータ１００は、通信インタフェース１１８を介してネットワークに接続し、他のコンピュータ等と通信を行うことが可能である。
【００２９】
以上の説明により、コンピュータ１００は、通常のパーソナルコンピュータ、ワークステーション、メインフレームなどの情報処理装置、又は、これらの組み合わせによって実現されることが容易に理解されるであろう。なお、上記説明した構成要素は例示であり、そのすべての構成要素が本発明の必須構成要素となるわけではない。同様に本発明を実施するためのコンピュータ１００は、キーボードやマウスのような入力デバイス、スピーカー等の他の構成要素を含むことも可能であることは言うまでもない。
【００３０】
図２Ｂは、本発明の実施形態による学習用音声データ生成システム２２０の機能ブロック図を示す。学習用音声データ生成システム２２０は、ソースドメインの音声データをターゲットドメインの音声データにマッピングしてターゲットドメインの音声データを擬似的に生成するべく、ソースドメイン・データ格納部２２２と、混合ガウスモデル格納部２２４と、チャネルマッピング部２２６と、ノイズデータ格納部２２８と、ノイズ付加部２３０と、ターゲットドメイン・データ格納部２３２とを備える。
【００３１】
ソースドメイン・データ格納部２２２は、インターネット上で公開されている、あるい

はインターネットを使った音声検索サービスなどのサービスにより提供される、種々の音響環境下における豊富な音声データをソースドメインの音声データとして格納する。したがって、ソースドメインの音声データには若干のノイズが重畳されている可能性が高い。このようなインターネットを使った音声データの収集は、例えば、携帯端末に音声認識を行うアプリケーションプログラムを実装し、その使用者の発声がインターネットを通じて送信され、サーバーに蓄積されることで行われる。

【００３２】

混合ガウスモデル格納部２２４は、ターゲットドメインのクリーンな音声データを用いて学習されたターゲットドメインのGMMを格納する。即ち、ターゲットドメインのクリーンな音声データは、ｋ混合のGMMとしてモデル化される。ターゲットドメインのGMM学習は、EMアルゴリズムを用いて行われる。なお、学習に用いられるターゲットドメインのクリーンな音声は少量であってよい（例えば、数名から数十名の男女話者によるトータル数時間の音声、精度に応じてより少ない時間の音声でもよい）。

【００３３】

チャネルマッピング部２２６は、ソースドメイン・データ格納部２２２から入力として読み出したソースドメインの音声データの１発話ごと、その補正量を混合ガウスモデル格納部２２４から読み出したターゲットドメインのGMMを参照して求め、ターゲットドメインの音声データにマッピングする。より具体的には、チャネルマッピング部２２６は、ソースドメインの各音声データ（１発話）について、ターゲットドメインのGMMのガウス分布コンポーネントのうち、該音声データとの音響空間における距離が近いガウス分布コンポーネントの平均との差を求め、該差を、ガウス分布コンポーネントごとの尤度で重みづけした上で、時間方向の平均として求める。続いてチャネルマッピング部２２６は、時間方向の平均として求めた量をチャネル補正量としてソースドメインの音声データに足し合わせる。チャネルマッピングの処理の更なる詳細は後述する。

【００３４】

ノイズデータ格納部２２８は、ターゲットドメインにおけるノイズであって、マイクロフォンに音声以外の音源から混入し、音声データの波形に加法的な変形を加えるノイズを格納する。なお、ターゲットドメインにおけるノイズは、混合ガウスモデル格納部２２４に格納されるターゲットドメインのGMMの学習時と同じ条件（同一の音響環境）で記録する必要がある。

【００３５】

ノイズ付加部２３０は、チャネルマッピング部２２６から出力されるチャネルマッピング後の音声データに、ノイズデータ格納部２２８からランダムに読み出したノイズを付加してターゲットドメインの音声データをシミュレートする。シミュレートされた擬似的なターゲットドメインの音声データはその後ターゲットドメイン・データ格納部２３２に格納される。なお、付加するノイズの割合は、最終的にターゲットドメインにおけるＳＮ比の統計的分布に一致するよう１発話ごとに調整される。

【００３６】

このように、音声データをターゲットドメインの音声データにマッピングするにあたり、チャネル特性及びノイズ特性は主要な要素である。そして、上述した学習用音声データ生成システム２２０において、ノイズ付加の前にチャネル補償を行うことが重要である。これを例証するために、２種類の音声データに同一の自動車のノイズを加えた。図３Ａ、図３Ｂは、それぞれ、１発話ごとにCMN処理してチャネル歪みによる影響を軽減した後のノイズの分布を示す。なお、縦軸はｃ２（２番目のケプストラム）を示し、横軸はｃ１（１番目のケプストラム）を示す（ただし、ケプストラムの最初の項は０番である）。

【００３７】

図３Ａは、停止した車中で録音された音声の特徴量ベクトルの平均値を０に正規化した後のノイズの分布を示す。図３Ｂは、携帯型機器に録音された音声の特徴ベクトルの平均値を０に正規化した後のノイズの分布を示す。同一のノイズを付加したにもかかわらず、ランタイム時のチャネル正規化（CMN）処理により結果の信号は全く異なるものとなって

いる。この結果からも、デコーダにおいて類似した信号を取得するためには、最初にチャネル補償を行う必要があることが分かる。

【００３８】

次に、チャネルマッピング部２２６による処理の詳細を説明する。上述したように本発明ではターゲットドメインのチャネル特性の事前知識として、ターゲットドメインの少量のクリーン音声データを用いて学習されたターゲットドメインのGMMを用いる。そしてターゲットドメインのGMMに似せるためのチャネル補正量（以下、チャネルマッピング・パラメータという）を大量の入力音声における1発話ごとに求めて、データのマッピングを行う。

【００３９】

ターゲットドメインのGMMは多少の話者バリエーションを含んでおり、これによって入力に話者依存要素が含まれることが許容される。入力は一切のノイズが重畳されていないクリーン音声が好適とされるが、実際の音声データには若干のノイズが含まれる。そこで、本発明では、VTS近似とEMアルゴリズムを用いてチャネルマッピング・パラメータを推定する。以下、チャネルマッピング・パラメータとして、特徴量のシフト量（以下、チャネルバイアスという）のみを考慮する場合と、特徴量のシフト量に加えて、特徴量の大きさを変換する特徴量の係数（以下、チャネル振幅という）をも考慮する場合の２つについて順に説明する。

【００４０】

＜チャネルバイアスのみの場合＞

ここではマッピング・パラメータによってソースドメインの音声データをターゲットドメインの音声データにマッピングする生成モデル式において、チャネルバイアスのみを考慮する場合について説明する。時間領域におけるソースドメインの観測音声の特徴量ベクトルY_sは、チャネルの特徴量ベクトルH_s、クリーン音声の特徴量ベクトルX、ノイズの特徴量ベクトルNを用いて下記式（１）のように表すことができる（添え字ｓはソースドメインであることを示す）。なおNは、マイクロフォンに音声以外の音源から混入し波形に加法的な変形を与える雑音であり、Hは伝送系により加えられる乗法性の歪みである。

【数１】

$$Y_s = H_s \cdot X + N - (1)$$

同様に、時間領域におけるターゲットドメインの観測音声の特徴量ベクトルY_tは、チャネルの特徴量ベクトルH_t、クリーン音声の特徴量ベクトルXを用いて下記式（２）のように表すことができる（添え字ｔはターゲットドメインであることを示す）。

【数２】

$$Y_t = H_t \cdot X - (2)$$

【００４１】

上記式（１）をケプストラム領域に書き直すと下記式（３）、（４）になる。

【数３】

$$y_s = h_s + x + G(x + h_s, n) - (3)$$

$$\text{但し、} G(x,n) = C\log\left(1 + \exp\left(C^{-1}(n - x)\right)\right) - (4)$$

同様に上記式（２）をケプストラム領域に書き直すと下記式（５）になり、これを更に上記式（３）を用いて変形すると、最終的に下記式（８）、（９）が得られる。ここで式（９）より定義されるｃが求めるべきチャネルバイアスである。ここで、行列Cは、Discrete Cosine Transform(DCT)行列を、C-1はその逆行列を表す。

【数４】

$$y_t = h_t + x - (5)$$

$$= h_t + \left(y_s - h_s - G\left(x + h_s, n\right)\right) - (6)$$

$$= \left(h_t - h_s\right) + y_s - G\left(x + h_s, n\right) - (7)$$

$$= y_s - c - G\left(y_t + c, n\right) - (8)$$

$$\text{但し、} c = h_s - h_t - (9)$$

【００４２】

上記式（８）において、ターゲットドメインの観測音声の特徴量ベクトルｙ$_t$を

$$\hat{y}$$

、及びソースドメインの観測音声の特徴量ベクトルｙ$_s$をｙと書き換えると、最終的に下記式（１０）で表される生成モデル式が得られる。

【数５】

$$y = \hat{y} + c + G\left(\hat{y} + c, n\right) - (10)$$

上述したように、本発明では、チャネルバイアスｃをVTS近似とEMアルゴリズムを用いて推定する。具体的には、上記式（１０）で表される生成モデル式を参照し、混合ガウスモデル格納部２２４に事前に用意したターゲットドメインのGMMから変換したソースドメインのGMMに含まれるチャネルバイアスｃを、EMアルゴリズムを用いて推定する。即ち、ターゲットドメインのGMMから変換したソースドメインのGMMにソースドメインの観測音声の特徴ベクトルｙを入力して得られる音響尤度を求めるステップと、求めた音響尤度に基づいた目的関数を最小化するチャネルバイアスｃを求めるステップとを交互に繰り返す。以下、数式を用いてこの繰り返しステップを説明する。

【００４３】

まず、ソースドメインのGMMにソースドメインの観測音声の特徴量ベクトルｙを入力して得られる音響尤度ｐ（ｙ）の対数をとったものにマイナスを掛けたものを考え、下記式（１１）を得る。ここでγ$_k$、μ$_{y,k}$、Σ$_{y,k}$は順に、ソースドメインのGMMのｋ番目の正規分布の事前確率、平均ベクトル、共分散行列を示す。式（１１）から式（１２）への変形は、イェンゼンの不等式を用いたものであり、更に変形を行って最終的に式（１４）を得る。なお下記数式においてｄは、ソースドメインの観測音声の特徴量ベクトルｙのｄ番目の成分を示し、Ｄはその次元数を示す。

【数６】

$$-\log(p(y)) = -\log\left(\sum_k \gamma_k \cdot \mathrm{N}(y;\mu_{y,k},\Sigma_{y,k})\right) - (11)$$
$$\leq -\sum_k \gamma_k \cdot \log(\mathrm{N}(y;\mu_{y,k},\Sigma_{y,k})) - (12)$$
$$= -\sum_k \gamma_k \cdot \log\left(\sqrt{2\pi}^{-D}\sqrt{|\Sigma_{y,k}|}^{-1}\exp\left(-\sum_d^D (y_d-\mu_{y,k,d},)^2/\Sigma_{y,k,d}\right)\right) - (13)$$
$$= const + \sum_k \gamma_k \left\{\sum_d (y_d-\mu_{y,k,d})^2/\Sigma_{y,k,d} + \log|\Sigma_{y,k}|^{\frac{1}{2}}\right\} - (14)$$

【００４４】

チャネルバイアスｃは音響尤度ｐ（ｙ）が最大になるように推定されるが、これは上記式（１４）の右辺第２項が最小になるようにチャネルバイアスｃを推定することに等しい。そこで、下記式（１５）で表される目的関数Φを新たに導入する。

【数７】

$$\Phi = E\left[\sum_k^K \rho_k(y)\cdot\left\{\sum_d^D (y_d-\mu_{y,k,d})^2/\Sigma_{y,k,d} + \log|\Sigma_{y,k}|^{\frac{1}{2}}\right\}\right] - (15)$$

ここで、事後確率ρは下記式（１６）より定義される。

【数８】

$$\rho_k(y) = \gamma_k \cdot \mathrm{N}(y;\mu_{y,k},\Sigma_{y,k}) \Big/ \sum_{k'} \gamma_{k'} \cdot \mathrm{N}(y;\mu_{y,k'},\Sigma_{y,k'}) - (16)$$

【００４５】

チャネルバイアスｃは、上記式（１５）で表される目的関数Φを最小にするように推定される。そこで、事前に用意していたターゲットドメインのGMMと、上記式（１０）の生成モデル式とから、共分散対角行列を仮定してVTS近似よりソースドメインのGMMを算出し、下記式（１７）、（１８）を得る。μ_n、$\Sigma_{n,d}$は、それぞれノイズの平均ベクトル、共分散行列を示し、δはクロネッカーのデルタを示す。

【数９】

$$\mu_{y,k} \cong \mu_{\hat{y},k} + c + G(\mu_{\hat{y},k} + c, \mu_n) - (17)$$

【数１０】

$$\Sigma_{y,k,d} \cong \sum_l \left(\delta_{d,l} - F(\mu_{\hat{y},k}+c,\mu_n)_{d,l}\right)^2 \cdot \Sigma_{\hat{y},k,l} + \sum_l F(\mu_{\hat{y},k}+c,\mu_n)_{d,l}^{\ 2} \cdot \Sigma_{n,l} - (18)$$

なお、式（１８）に表される共分散対角行列は、さらなる近似を用いて簡略化してもかまわない。例えば、ターゲットドメインの共分散行列と同一にしても、精度の劣化は少量に

とどまる。逆に、対角近似の条件を外して精密な共分散行列としても良い。上記式（１７）に現れるベクトルＧ、及び、上記式（１８）に現れるヤコビ行列Ｆは、それぞれ下記式（１９）、（２０）により定式化される雑音成分である。ここで、行列Ｃは、Discrete CosineTransform(DCT)行列を、C^{-1}はその逆行列を表す。

【数１１】

$$G(x,n)=C\log\left(1+\exp\left(C^{-1}(n-x)\right)\right)-(19)$$

【数１２】

$$F(x,n)_{i,j}=\sum_k C_{i,k}\cdot\left(C^{-1}_{k,j}\right)\cdot\exp\left(\sum_l C_{k,l}{}^{-1}(n_l-x_l)\right)\Big/\left\{1+\exp\left(\sum_l C_{k,l}{}^{-1}(n_l-x_l)\right)\right\}-(20)$$

【００４６】

上記式（１９）及び式（２０）によりそれぞれ表される雑音成分のベクトルＧとヤコビ行列Ｆは、ケプストラム領域において実装することが好ましいが、対数スペクトル領域において表すと、下記式（２１）、（２２）のようにそれぞれ表される。

【数１３】

$$F(x,n)=\{1+\exp(x-n)\}^{-1}-(21)$$

【数１４】

$$G(x,n)=\log(1+\exp(n-x))-(22)$$

【００４７】

次に上記式（１７）～（２０）を参照して、上記式（１５）、（１６）により表される目的関数Φをチャネルバイアスc_dに関して微分したものを０に設定することで、チャネルバイアスｃの現在の推定値を得る。また、チャネルバイアスｃの現在の推定値を用いて雑音成分のベクトルＧ及びヤコビ行列Ｆを更新し、目的関数Φを求める。そして、求まった目的関数Φを最小にするように再びチャネルバイアスｃを推定する。この２ステップからなる処理をチャネルバイアスｃが収束するまで繰り返すことで、チャネルバイアスｃの最終的な推定値が得られる。

【００４８】

収束したチャネルバイアスｃの推定値を生成モデル式の式（１０）に代入し、更に式（１０）の右辺第３項をＭＭＳＥ推定により近似することで、最終的に下記式（２３）が得られる。式（２３）によりソースドメインの音声データからマップされたターゲットドメインのクリーン音声を得ることができる。

【数１５】

$$\hat{y}=y-c-\sum_k^K \rho_k(y)\cdot G\left(\mu_{\hat{y},k}+c,\mu_n\right)-(23)$$

なお、式（２３）では、ソースドメインのノイズ除去とチャネル特性の補正を同時に行っている。ソースドメインのノイズを無視できる場合には、下記式（２４）のように、ノイ

ズ除去を省略しても良い。

【数１６】

$$\hat{y} = y - c - (24)$$

【００４９】

＜チャネル振幅を考慮する場合＞

ここではマッピング・パラメータによってソースドメインの音声データをターゲットドメインの音声データにマッピングする生成モデル式において、チャネルバイアスｃに加えて、チャネル振幅ａを新たに導入する場合について説明する。この場合、上記式（１０）の生成モデル式は、下記式（２５）のように拡張される。式中記号＊は、ベクトル要素ごとの内積を表す。

【数１７】

$$y = a * \hat{y} + c + G(a * \hat{y} + c, n) - (25)$$

上述したように、本発明では、チャネルバイアスｃとチャネル振幅ａをVTS近似とEMアルゴリズムを用いて推定する。具体的には、上記式（２５）で表される生成モデル式を参照し、混合ガウスモデル格納部２２４に事前に用意したターゲットドメインのGMMから変換したソースドメインのGMMに含まれるチャネルバイアスｃとチャネル振幅ａを、EMアルゴリズムを用いて推定する。即ち、ターゲットドメインのGMMから変換したソースドメインのGMMにソースドメインの観測音声の特徴ベクトルｙを入力して得られる音響尤度を求めるステップと、求めた音響尤度に基づいた目的関数を最小化するチャネルバイアスｃとチャネル振幅ａを求めるステップとを交互に繰り返す。以下、数式を用いて説明する。

【００５０】

上記式（２５）の生成モデル式を参照して、事前に用意していたターゲットドメインのGMMから、共分散対角行列を仮定してVTS近似より得られるソースドメインのGMMの平均ベクトル$\mu_{y,k}$と共分散行列$\Sigma_{y,k,d}$は、それぞれ下記式（２６）、（２７）のようになる。μ_n、$\Sigma_{n,d}$は、それぞれノイズの平均ベクトル、共分散行列を示し、δはクロネッカーのデルタを示す。また式（２６）、（２７）中にそれぞれ現れるＧとＦは、上記式（１９）及び式（２０）によりそれぞれ表される雑音成分のベクトルＧとヤコビ行列Ｆである。

【数１８】

$$\mu_{y,k} \cong a * \mu_{\hat{y},k} + c + G(a * \mu_{\hat{y},k} + c, \mu_n) - (26)$$

【数１９】

$$\Sigma_{y,k,d} \cong \sum_l a_l^2 \cdot \left(\delta_{d,l} - F(a * \mu_{\hat{y},k} + c, \mu_n)_{d,l}\right)^2 \cdot \Sigma_{\hat{y},k,l} + \sum_l F(a * \mu_{\hat{y},k} + c, \mu_n)_{d,l}^2 \cdot \Sigma_{n,l} - (27)$$

【００５１】

なお、目的関数Φはチャネルバイアスｃのみを考慮する場合と同じであり上記式（１５）、（１６）により表される。そこで上記式（１９）～（２０）、（２６）～（２７）を参照して、上記式（１５）、（１６）により表される目的関数Φをチャネルバイアスc_dに関して微分したものを０に設定することで、チャネルバイアスｃの現在の推定値を得る。同様に上記式（１９）～（２０）、（２６）～（２７）を参照して、上記式（１５）、（１６）により表される目的関数Φをチャネル振幅a_dに関して微分したものを０に設定することで、チャネル振幅ａの現在の推定値を得る。続いて、チャネルバイアスｃとチャ

ネル振幅ａのそれぞれの現在の推定値を用いて雑音成分のベクトルG及びヤコビ行列Fを更新し、目的関数Φを求める。そして、求まった目的関数Φを最小にするように再びチャネルバイアスｃとチャネル振幅ａとを推定する。この２ステップからなる処理をチャネルバイアスｃとチャネル振幅ａとが収束するまで繰り返すことで、チャネルバイアスｃとチャネル振幅ａの最終的な推定値が得られる。なお、チャネルバイアスｃとチャネル振幅ａの初期値はそれぞれ順に値０と値１としてよい。また、チャネルバイアスｃについては、通常その全ての成分を更新するが、チャネル振幅ａについては低次の成分、例えば０番目と２番目の成分のみを更新してもよい。これはケプストラムの低次の成分がチャネル特性について支配的であるためである。

【００５２】

収束したチャネルバイアスｃ及びチャネル振幅ａのそれぞれの推定値を生成モデル式の式（２５）に代入し、更に式（２５）の右辺第３項をＭＭＳＥ推定により近似して、最終的に下記式（２８）を得る。式（２８）によりソースドメインの音声データからマップされたターゲットドメインのクリーン音声を得ることができる。

【数２０】

$$\hat{y} = a^{-1} * \left(y - c - \sum_{k}^{K} \rho_k(y) \cdot G(a * \mu_{\hat{y},k} + c, \mu_n) \right) - (28)$$

式（２８）では、ソースドメインのノイズ除去とチャネル特性の補正を同時に行っている。ソースドメインのノイズを無視できる場合には、下記式（２９）のように、ノイズ除去を省略しても良い。

【数２１】

$$\hat{y} = a^{-1} * (y - c) - (29)$$

【００５３】

＜話者の性別ごと用意されたGMMを用いる場合＞

上記説明した２つのケースでは、混合ガウスモデル格納部２２４に事前に用意したターゲットドメインのGMMは１つであり、性別の区別なく多少の話者バリエーションを含んでいた。ここでは、ターゲットドメインのGMMは、話者の性別ごとに用意するものとする。そして、入力として受け取るソースドメインの音声データの１発話ごとに男性らしさ及び女性らしさを求めながら、ターゲットドメインの音声データにマッピングするものとする。すると目的関数Φは下記式（３０）のように表される。

【数２２】

$$\Phi = E\left[\sum_{d}^{D} \sum_{g=f,m} \lambda_g \sum_{k}^{K} \rho_{g,k}(y) \cdot \left(y - \mu_{y,g,k} \right)_d^{\,2} / \left(\Sigma_{y,g,k} \right)_d \right] - (30)$$

ここでｇはgenderインデックスであり、女性(female)又は男性(male)のいずれかを示す。また、事後gender確率λ$_g$は、ガウス分布の尤度を事後確率とみたてて合計を１になるように正規化したものである。また、ρは上記式（１６）により定義される事後確率である。

【００５４】

ここで上記式（１０）の生成モデル式と上記式（３０）の目的関数Φが与えられたとする。すると、上記式（１７）～（２０）を参照して、上記式（３０）により表される目的関数Φをチャネルバイアスｃ$_d$に関して微分したものを０に設定することで、チャネルバ

イアスｃの現在の推定値を得る。また、チャネルバイアスｃの現在の推定値を用いて雑音成分のベクトルＧ及びヤコビ行列Ｆを更新し、目的関数Φを求める。そして、求まった目的関数Φを最小にするように再びチャネルバイアスｃを推定する。この２ステップからなる処理をチャネルバイアスｃが収束するまで繰り返すことで、チャネルバイアスｃの最終的な推定値が得られる。

【００５５】

事後gender確率λ_gもまた、チャネルバイアスｃを繰り返し更新する間、ターゲットドメインの性別ごとのGMMの事後確率に基づいて、下記式（３１）、（３２）に示すように更新する。

【数２３】

$$\lambda'_g = E\left[\sum_k \gamma_{g,k} \cdot \mathrm{N}\left(y; \mu_{y,g,k}, \Sigma_{y,g,k}\right)\right] - (31)$$

4/5

【数２４】

$$\lambda''_g = \lambda'_g \Big/ \sum_{g'} \lambda'_{g'} - (32)$$

オプションとして、下記式（３３）に示す事後gender確率λ_gを用いてもよい。

【数２５】

$$\lambda_g = \exp\left(\beta \cdot \lambda''_g\right) \Big/ \sum_{g'} \exp\left(\beta \cdot \lambda''_{g'}\right) - (33)$$

但しβの値は定数である。なお、ここまで、ケプストラム領域を前提に定式化をしてきたが、よく行われるように対数メルスペクトル領域や対数スペクトル領域でも同様に定式化できることは当業者に明らかである（ケプストラム領域とは、対数メルスペクトル領域を離散コサイン変換したものである。）。

【００５６】

次に図４及び図５を参照して、本発明の実施形態による学習用音声データ生成処理の流れを説明する。図４は、学習用音声データ生成処理の全体の流れの一例を示すフローチャートある。図５は、図４に示すフローチャートのステップ４０４のマッピング処理の流れの一例を示すフローチャートある。

【００５７】

図４に示すフローチャートは、ステップ４００で開始し、チャネルマッピング部２２６は、ソースドメイン・データ格納部２２２からソースドメインの音声データを入力として取得する。続いてチャネルマッピング部２２６は、混合ガウスモデル格納部２２４からターゲットドメインのGMMを読み出す（ステップ４０２）。続いてチャネルマッピング部２２６は、ターゲットドメインのGMMを参照して、ソースドメインの音声データを、ターゲットドメインの音声データに、ターゲットドメインの音声データのチャネル特性に基づいてマッピングする（ステップ４０４）。マッピング処理の詳細は図５を参照して後述する。

【００５８】

続いて、ノイズ付加部２３０は、ノイズデータ格納部２２８からランダムに読み出したノイズをマッピング後の音声データに付加してターゲットドメインの擬似音声データを生成し（ステップ４０６）、これをターゲットドメイン・データ格納部に２３２へ出力する

（ステップ４０８）。そして処理は終了する。
【００５９】
図５に示すフローチャートは、ステップ５００で開始し、チャネルマッピング部２２６はソースドメインの観測音声の特徴量ベクトルとターゲットドメインの観測音声の特徴量ベクトルとの関係を示す生成モデル式を決定する。上述したように利用可能な生成モデル式には、チャネルマッピング・パラメータとしてチャネルバイアスｃのみを考慮する生成モデル式と、チャネルマッピング・パラメータとしてチャネルバイアスｃとチャネル振幅とを考慮する生成モデル式の２つがある。
【００６０】
続いてチャネルマッピング部２２６は、ターゲットドメインのGMMから変換したソースドメインのGMMに含まれるチャネルマッピング・パラメータを、EMアルゴリズムを用いて推定する（ステップ５０２）。ここでソースドメインのGMMは、上述したように、
事前に用意したターゲットドメインのGMMとステップ５００で決定した生成モデル式とから、VTS近似により算出される。
【００６１】
続いてチャネルマッピング部２２６は、ステップ５０２で求まったチャネルマッピング・パラメータを用いて、ステップ５００で決定した生成モデル式に従いソースドメインの音声データをターゲットドメインの音声データにマッピングする（ステップ５０４）。そして処理は終了する。
【００６２】
以上、実施形態を用いて本発明の説明をしたが、本発明の技術範囲は上記実施形態に記載の範囲には限定されない。上記の実施形態に、種々の変更又は改良を加えることが可能であることが当業者に明らかである。従って、そのような変更又は改良を加えた形態も当然に本発明の技術的範囲に含まれる。
【００６３】
上述した実施形態の各機能は、Ｃ、Ｃ＋＋、Ｃ＃、Ｊａｖａ（登録商標）などのオブジェクト指向プログラミング言語などで記述された装置実行可能なプログラムにより実現でき、本実施形態のプログラムは、ハードディスク装置、ＣＤ－ＲＯＭ、ＭＯ、ＤＶＤ、フレキシブルディスク、ＥＥＰＲＯＭ、ＥＰＲＯＭなどの装置可読な記録媒体に格納して頒布することができ、また他装置が可能な形式でネットワークを介して伝送することができる。
【実施例】
【００６４】
以下、本発明について、実施例を用いてより具体的に説明を行なうが、本発明は、後述する実施例に限定されるものではない。
【００６５】
先に開示した本発明の方法をコンピュータに実行させるためのコンピュータ・プログラムを作成し、一般社団法人情報処理学会（ＩＰＳＪ）が提供する自動車内音声認識の評価用フレームワークを使用して、各コンピュータ・プログラムの性能を評価した。
【００６６】
＜実験条件＞
本実験は、自動車内音声認識の評価用データベースCENSREC-3を用いて行った。評価条件は、学習データと評価データの双方に遠隔マイクロフォンを使用するミスマッチのない場合（Ａ）と、学習データに接話マイクロフォンを使用し、評価データに遠隔マイクロフォンを使用するミスマッチのある場合（Ｂ～Ｅ）とした。更に、ミスマッチのある場合（Ｂ～Ｅ）は、本発明の適用の無い場合（Ｂ）と、本発明の適用のある場合（Ｃ～Ｅ）とし、本発明の適用のある場合は以下の３つの条件について検討した。学習データに対しチャネルバイアスのみを考慮したチャネルマッピングを適用した場合（Ｃ）。学習データに対し、チャネルバイアスのみを考慮し、話者の性別を区別したチャネルマッピングを適用した場合（Ｄ）。学習データに対し、チャネルバイアスとチャネル振幅を考慮し、話者の性

別を区別したチャネルマッピングを適用した場合（Ｅ）。

【００６７】

本発明を適用する場合、学習データとして、駐車した車内で記録された男性２０２人、女性９１人の計２９３人のドライバーによる３６０８の発話を用いた。また、事前に用意するターゲットドメインのクリーン音声のGMMは、混合数は２５６とし、遠隔マイクロフォンで収録しランダムに選択した５００の発話データで学習した。

【００６８】

一方、評価データとしては、駐車した車内で記録された男性８人、女性１０人の計１８人のドライバーによる８９８の発話を用いた。

【００６９】

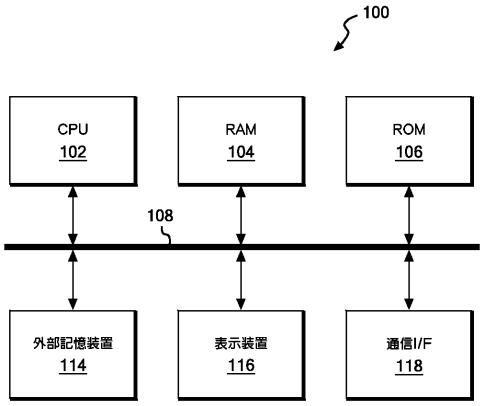
また、実験に必要な様々な特徴量を出力するフロントエンドを用意し、学習データと評価データの双方に適用した。特徴量は、MFCC 12次元 + ΔMFCC 12次元 + ΔΔMFCC12次元 + 対数パワー の39次元で、発話単位のCMNを適用した場合と適用しない場合の両方の値を得た。音響モデルの作り方などバックエンドの構成は無変更とした(Category0)。

【００７０】

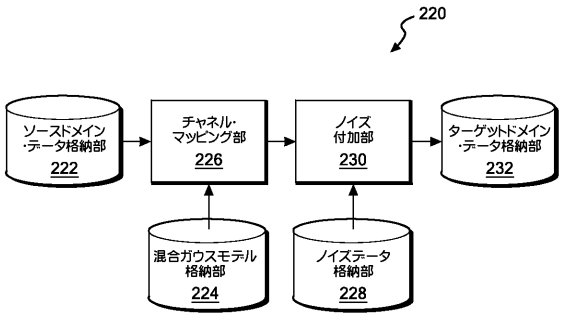
＜実験結果＞

図６は、上述したＡ～Ｅの６つの条件下それぞれでのCMNオンとCMNオフでの単語正解精度（％）を示す。条件Ａはミスマッチの無い場合の単語正解精度であるため、その値は上限を示す。条件Ｂはミスマッチのある場合であって、本発明によるチャネルマッピングがなされていないため、その値はベースラインとなる。条件Ｃ～条件Ｅは、ミスマッチのある場合であって、本発明によるチャネルマッピングが適用されている場合である。いずれの場合も単語正解精度はベースラインよりも高くなっている。また、条件Ｄでの単語正解精度の値は条件Ｃのそれよりも改善されているため、本発明によるチャネルマッピング手法は、チャネル特性と話者特性の両方に有効であるといえる。また、条件Ｃ～条件Ｅのすべてにおいて CMNオンの場合の単語正解精度の値はベースラインのそれよりも高くなっているため、本発明によるチャネルマッピング手法はCMNと相性がよいといえる。

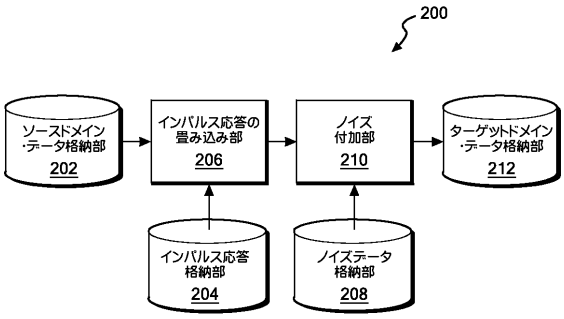
【図１】



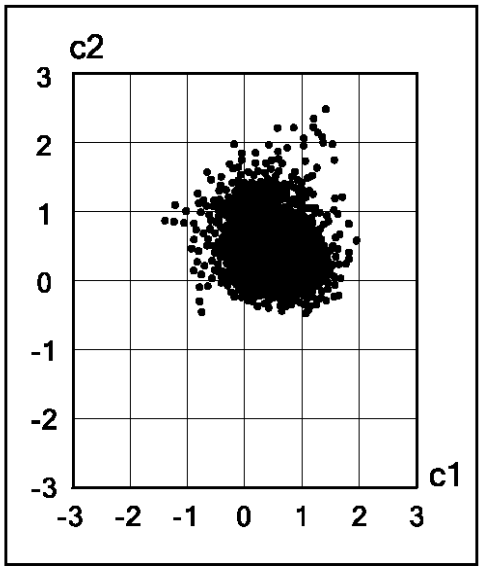
【図２Ｂ】



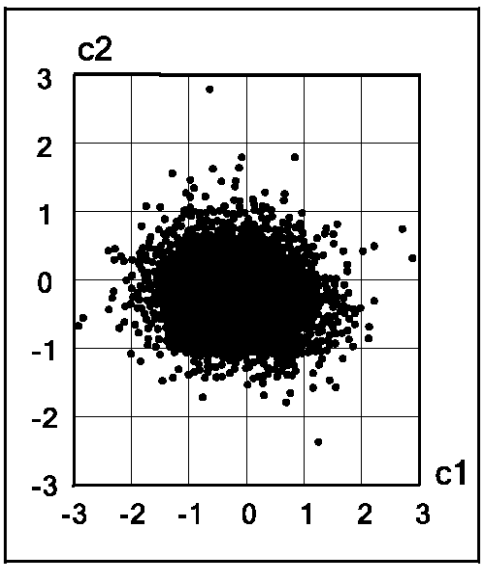
【図２Ａ】



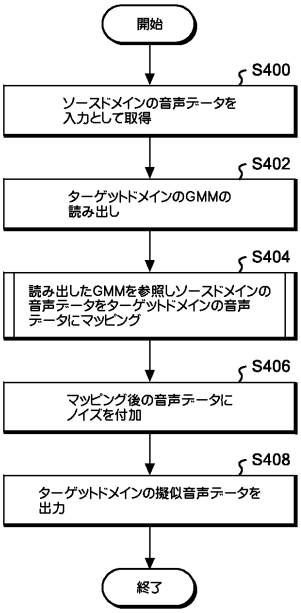
【図３Ａ】



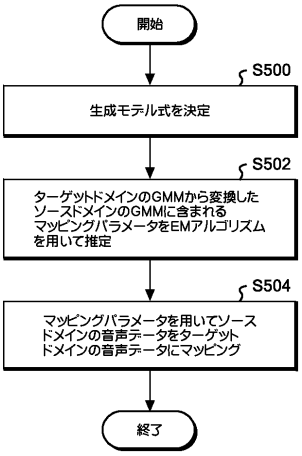
【図３Ｂ】



【図４】



【図５】



【図６】

	学習データ	評価データ	CMN off	CMN on
A	遠隔マイク、アイドリング	遠隔マイク、アイドリング（ファン・オフ、窓閉）	99.7	99.7
B	接話マイク、アイドリング	遠隔マイク、アイドリング（ファン・オフ、窓閉）	92.3	95.2
C	接話マイク、アイドリング、チャネルマッピング（バイアスのみ）あり	遠隔マイク、アイドリング（ファン・オフ、窓閉）	93.1	96.0
D	接話マイク、アイドリング、チャネルマッピング（性別依存、バイアスのみ）あり	遠隔マイク、アイドリング（ファン・オフ、窓閉）	94.5	96.8
E	接話マイク、アイドリング、チャネルマッピング（性別依存、バイアス及び振幅）あり	遠隔マイク、アイドリング（ファン・オフ、窓閉）	96.4	98.0

フロントページの続き

(74)代理人 100112690
弁理士 太佐 種一
(72)発明者 市川 治
東京都江東区豊洲五丁目６番５２号 ＮＢＦ豊洲キャナルフロント 日本アイ・ビー・エム株式会社 東京基礎研究所内
(72)発明者 スティーブン・ジェイ・レニー
アメリカ合衆国１０５９８ニューヨーク州ヨークタウン・ハイツルート１３４ キチャワン・ロード１１０１番地

審査官 鈴木 圭一郎

(56)参考文献 特開２００８－０２６４８９（ＪＰ，Ａ）
特開平０５－０７３０８８（ＪＰ，Ａ）
特開２００５－１９６０２０（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １５／００