2016-7-A3-2

音素構造音響モデルを用いた非並列学習による声質変換

野 本 知 宏[†1]　小 坂 哲 夫[†1]

音響モデルに音素構造 GMM を用いた非並列声質変換手法を検討する．非並列声質変換とは，複数の話者が全く同じ内容を話した音声（並列発話）をモデルの学習に使用しない声質変換手法のことである．本研究の新規性は，非並列声質変換において，分布を音素毎に推定する音素構造 GMM を導入したことである．音素構造 GMM を用いることで，音素固有の情報を活かした変換が可能になるため，変換音声の音質が向上することが期待される．本稿では，声質変換に通常の GMM を用いた場合と音素構造 GMM を用いた場合の客観的・主観的評価について調査した結果を示す．

Non-parallel voice conversion using phonetically structured acoustic models

Tomohiro Nomoto[†1] and Tetsuo Kosaka[†1]

We propose a non-parallel voice conversion method using phonetically structured GMMs as acoustic models. Non-parallel voice conversion is a voice conversion method that does not use a parallel data set consisting of utterance pairs for model training. The novelty of this study is to introduce a phonetically structured GMM, where parameters of a Gaussian *pdf* are estimated separately depending on the type of phoneme, in non-parallel voice conversion. By using the phonetically structured GMM, phoneme specific information can be used for voice conversion. Therefore, it is expected to improve sound quality of converted voices. In this paper, we investigate on the objective and subjective evaluations of voice conversion with conventional GMMs and phonetically structured GMMs.

†1 山形大学
Yamagata University

1. はじめに

合成音声の話者性を制御する技術の一つに，声質変換がある．声質変換とは，ある話者が話した音声を別の話者の声に変換する技術である．声質変換の代表的な手法としてはコードブックマッピング法[1] やファジーベクトル量子化を用いた手法[2] が挙げられるが，現在ではこれらの手法と比較して連続的かつ高精度な変換が行える，混合正規分布モデル (GMM:Gaussian Mixture Model) を用いた手法[3] が一般的となっている．

GMM を用いた声質変換では，音声特徴量を正規分布の線形重ね合わせによって機械学習を行い変換を行う．GMM の変換手法は学習データによって複数の種類が存在し，大きく分けて並列学習と非並列学習に分かれる．並列学習は変換の元話者と対象話者の同一の発話 (並列発話) を用いて GMM を学習する手法である．一般的に並列発話を用いた声質変換は音質の良い音声ができやすいと言われている．しかし，GMM の学習のために大量の並列発話を用意する必要があるため，モデルの作成が困難だという問題がある．

これに対し，非並列学習では並列発話を用いずに GMM の学習を行う．非並列学習による変換音声は並列学習と比較して音質が悪くなりやすいが，学習データの準備が容易という利点がある．過去幾つかの非並列学習法が提案されているが，特に文献[4] の手法は, 比較的少数の発話の学習で並列学習に近い性能が得られるとされている．

また，従来の GMM による声質変換の問題点として，各音素がどの混合分布にクラスタリングされているかが分からないということが挙げられる．文献[5] では，複数の音素が同じ混合分布に学習されてしまい，平均的な音に変換されている可能性が挙げられていた．これに対し，文献[6] では音素ごとに学習した GMM を用いて変換を行い，男女間の声質変換で音質が向上したと報告されている．文献[6] の変換手法は並列学習に基づくものであったが，非並列学習においても，音素ごとに GMM を学習することで似た音素についても個別に変換を行うことができ，より精度の高い変換を行えると考えられる．

そこで，本研究では，音素固有の情報を生かした GMM を非並列学習により作成し，声質変換の精度を向上させることを目的とする．そのために，音素情報を用いた GMM の学習条件についての検討を行う．また，音素情報を用いずに学習した GMM による声質変換との比較実験も行う．

2. 非並列学習による声質変換の手法

本研究では，非並列学習による声質変換の手法として，文献[4] の手法を用いる．変換の際

は，図 1 に示す手順により GMM を学習した後，学習した GMM を用いて，図 2 の手順で声質変換を行う．

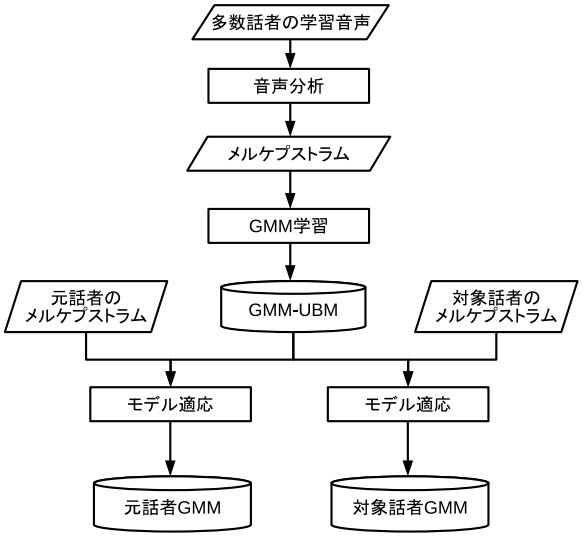

図 1 GMM の作成手順
Fig. 1 Procedure for creating GMMs

2.1 GMM の作成

2.1.1 GMM-UBM の学習

混合正規分布モデル (GMM:Gaussian Mixture Model) とは，正規分布の線形重ね合わせにより表されるモデルである．また，GMM-UBM(GMM-Universal Background Model) は，不特定多数の話者の大量の発話から学習された GMM のことである．本研究では，この GMM を用いて音響特徴をモデル化し，GMM-UBM を後述する MAP 適応により話者適応し，変換に用いる．

ある変量 $\boldsymbol{x}= [x_1, x_2, \cdots, x_d]^T$ が与えられたとき，M 個の正規分布それぞれの混合重みを α_i，$\boldsymbol{x}$ の平均ベクトル，共分散行列をそれぞれ μ_i，Σ_i とすると，変量 $\boldsymbol{x}$ の確率分布は混合分布として次式で表される．

$$p(\boldsymbol{x}) = \sum_{i=1}^{M} \alpha_i N(\boldsymbol{x}; \mu_i, \Sigma_i) \tag{1}$$

ここで，各混合重み α_i は以下の式を満たす．

$$\sum_{i=1}^{M} \alpha_i = 1, \qquad \alpha_i \geq 0 \tag{2}$$

また，$N(\boldsymbol{x}; \mu_i, \Sigma_i)$ は以下で示すような，平均 μ_i，分散共分散行列 Σ_i の多次元正規分布である．

$$N(\boldsymbol{x}; \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma_i|}} \exp(-\frac{1}{2}(\boldsymbol{x} - \mu_i)^T \Sigma_i^{-1} (\boldsymbol{x} - \mu_i)) \tag{3}$$

2.1.2 GMM-UBM のモデル適応

声質変換には，変換の元話者の GMM と，対象話者の GMM が必要である．また，後述するガウス正規化による声質変換を行うためには，元話者と対象話者の GMM の構造が同じである必要がある．そこで，GMM-UBM を元話者や対象話者の音声で適応することで，同一の構造を持つ元話者の GMM，対象話者の GMM を作成する．本研究では，適応法に MAP 適応[7] を用いる．MAP 適応は，事後確率最大化法 (MAP 推定) を用いてパラメータを更新する手法であり，以下の式で平均ベクトルを更新する．

$$\mu_i^{'} = \frac{\tau}{N_i + \tau} \mu_i + \frac{N_i}{N_i + \tau} e_i \tag{4}$$

ここで，μ_i は GMM-UBM のインデックス i の分布の平均を表している．また，N_i，e_i は，インデックス i の分布に推定された適応データのサンプル数と平均である．τ は事前分布の信頼度を表すパラメータで，定数で与えられる．

2.2 声質変換処理

声質変換処理の流れを，図 2 に示す．声質変換では，まず入力音声に対して音声分析を行い，メルケプストラムと f0 を抽出する．その後，前小節の方法で作成された元話者と対象話者の GMM を用いてメルケプストラムを変換する．また，元話者と対象話者の基本周波

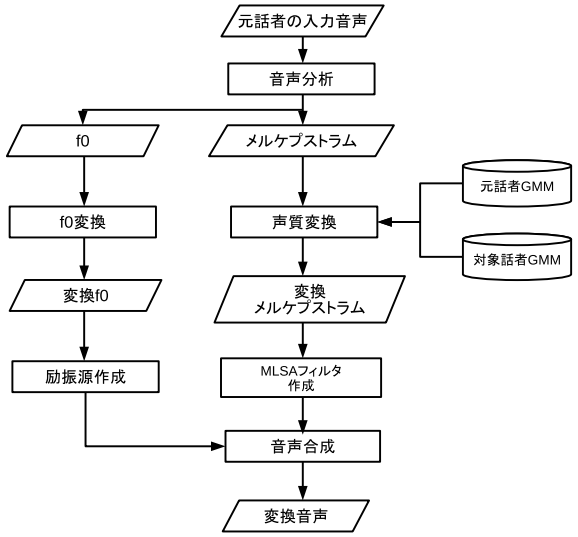

図 2 声質変換の手順
Fig. 2 Procedure of voice conversion

数に大きく差がある場合，元話者の f0 をそのまま使用すると変換音声の品質が劣化する恐れがあるため，f0 については元話者と対象話者の適応データの f0 の対数平均を用いて変換を行う．このようにして変換したメルケプストラムから MLSA フィルタ[8] を作成し，変換された f0 から作成した励振源を通過させることで音声を合成する．

以下では，メルケプストラムと f0 の変換処理について述べる．

2.2.1 ガウス正規化による声質変換

ガウス正規化による声質変換は，文献[4] で用いられている変換手法の一つである．あるフレーム t における入力音声の特徴量 $\boldsymbol{x}_t$ と元話者に適応されたモデルから，次式により最大事後確率を有する正規分布を選択する．

$$m = \arg\max_i p(i|\boldsymbol{x}) \qquad (i = 1, 2, \cdots, M) \tag{5}$$

ここで，i は分布のインデックスであり，M はモデルの混合数である．また，$p(i|\boldsymbol{x})$ は入力特徴量 x に対する分布 i の事後確率である．このようにして推定された正規分布に属する元話者と対象話者の特徴量について，以下の式が成り立つ．

$$\frac{x - \mu_m^x}{\sigma_m^x} = \frac{y - \mu_m^y}{\sigma_m^y} \tag{6}$$

ここで，μ, σ は平均，分散を，下付きの m は混合の番号を表し，上付きの x, y はそれぞれ元話者，対象話者のパラメータであることを表している．従って，元話者と対象話者のモデル間の，インデックス m の混合に関する変換関数は，次式で与えられる．

$$F(x) = \frac{\sigma_m^y}{\sigma_m^x} x + \mu_m^y - \frac{\sigma_m^y}{\sigma_m^x} \mu_m^x \tag{7}$$

ただし，式 (6) と式 (7) は，元話者と対象話者のモデルが同じ構造であることを前提としている．従って，この変換手法は，2.1 節の方法で同じ GMM-UBM から適応したモデル間でのみ成り立つ．

2.2.2 基本周波数の変換

本研究では時刻 t における元話者の f0 情報 $p_t^{(x)}$ を次式で線形変換し合成に用いる．

$$p_t^{(y)} = \exp(\ln p_t^{(x)} - \mu^{(\ln x)} + \mu^{(\ln y)}) \tag{8}$$

ここで，$\mu^{(\ln x)}$，$\mu^{(\ln y)}$ はそれぞれ，GMM の適応時に用いた元話者，対象話者の発話の f0 情報全体を対数変換した際の平均値を表している．

3. 音素構造を考慮した GMM

従来の GMM の問題点として，音素個々の発音の違いを表現することができないということが挙げられる．例えば，ある話者の母音/a/と他の話者の母音/o/などが混同されて学習される場合がある．これは，音素コンテクストの情報を無視して尤度計算を行っているためである．このような GMM では，1 つの混合に複数の音素の情報が内包されてしまうため，音素ごとの詳細な違いが平均化されてしまう．従って，従来の GMM では音響特徴全体を表現することはできても，音素固有の情報を表現することはできない．

この問題に対し，音素と対応関係があるガウス分布を用いた GMM により音素固有の情報を表現する方法が提案されている．例えば，文献[9] では，音素構造 GMM[10] を話者識別のアンカーモデルに用いた場合，従来の GMM よりも性能が向上したと報告している．文

献[9] は話者識別に関する研究だが，声質変換においても音素ごとに分布を推定した GMM を用いることで，音素固有の情報を表現することができ，性能が向上することが期待される．

よって，本研究では，文献[4] の非並列声質変換手法における音素構造 GMM の有用性について検討を行う．以下では，音素構造 GMM の概要について説明した後，従来の GMM と音素構造 GMM の作成アルゴリズムについて述べる．

3.1 音素構造 GMM の概要

音素構造 GMM は，アライメントにより得られたラベルデータとセグメントデータを利用して，音素ごとに学習を行う GMM である．従来の GMM と音素構造 GMM の作成法を，図 3，図 4 に示す．

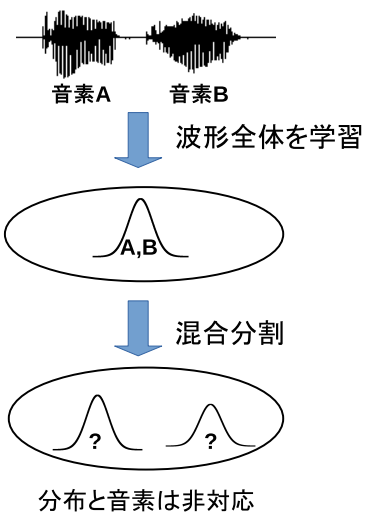

図 3 従来の GMM の作成方法
Fig. 3 Procedure for creating conventional GMM

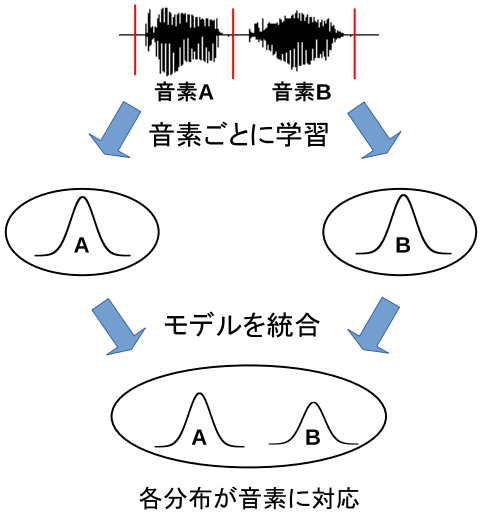

図 4 音素構造 GMM の作成方法
Fig. 4 Procedure for creating Phonetically Structured GMM

従来の GMM では，音声波形全体を同じ正規分布で学習した後，分布を分割して分布数を増加させ，音声をモデル化する．このため，必ずしも音素と正規分布が 1 対 1 に対応しない．これに対して音素構造 GMM では，音声波形を音素単位で分割し，音素ごとに分布を学習する．その後，各音素の分布を 1 つのモデルにまとめ GMM に変換する．音素構造 GMM を作成するためには音声波形を音素単位に分割するためのラベルデータとセグメントデータが必要となるが，距離の近い音素を混同することなく，音素固有の情報を活かした分布を学習することができる．

3.2 従来の GMM の作成アルゴリズム

従来の GMM は，次のような手順で作成する．

(1) 学習データ全体を 1 つの正規分布で表現
(2) 正規分布を分割
(3) 分布が増加したモデルの再学習
(4) 2，3 を目標の混合数に達するまで繰り返し

従来の GMM ではまず，全音声の学習データを 1 つの正規分布で表現する．次に，分布を適当に分割し，分布が増加したモデルを用いて再び全学習データを学習し，各分布の平均，分散，混合重みを更新する．これを目標の混合数に達するまで繰り返し，GMM を作成する．以降では，手順 1 の学習を初期学習，手順 3 の目標混合数に達するまでの再学習を途中学習，目標混合数に達した後の再学習を最終学習と記載している．

3.3 音素構造 GMM の作成アルゴリズム

音素構造 GMM は，次のような手順で作成される．

(1) 音声パラメータ，ラベルデータ，セグメントデータを使用した，隠れマルコフモデル (HMM:Hidden Markov Model) の音素境界を固定しての初期学習
(2) HMM の Viterbi アルゴリズムによる音素境界を固定しない連結学習
(3) HMM を 1 状態の混合ガウス分布に変換 (平均・分散のみ参照)
(4) GMM の重み学習

音素構造 GMM ではまず，アライメントにより作成されたラベルデータとセグメントデータを音素時間長として用いて，音素ごとに HMM を学習する．続いて，ラベルデータのみを用いて，音素境界を固定せずに連結学習を行う．ここまでで音素ごとに分布が推定されているため，HMM の各分布を持つ GMM への変換を行う．このとき，分布の平均と分散の値のみを参照し，重みについては適当に与える．最後に，GMM の各混合の重みの推定を行う．この際，分布の平均と分散については変更しない．

また，第 4 節の従来法 GMM と音素構造 GMM の比較実験では，HMM を GMM に変換し，学習データで GMM の重み学習を行った後に適応を行っている．つまり，手順 1〜4 を図 1 の「GMM 学習」の部分で行っている．これは，従来法との比較を行う都合上，適

応時の条件を揃えるためである．一方，第 5 節の混合数と状態数の検討では，図 5 のように，HMM のまま適応を行った後に GMM に変換し，適応データで重み学習を行っている．つまり，図 5 の「HMM 学習」の部分で手順 1，2 を行い,「GMM へ変換」,「重み学習」部分で手順 3，4 を行っている．これは，混合数を増加させた場合，出現頻度の低い分布が適応されにくくなると考えられるためである．HMM のまま適応を行うと，各混合の重みは GMM よりも大きくなるため，出現頻度の低い分布についても適応されやすくなると考えられる．

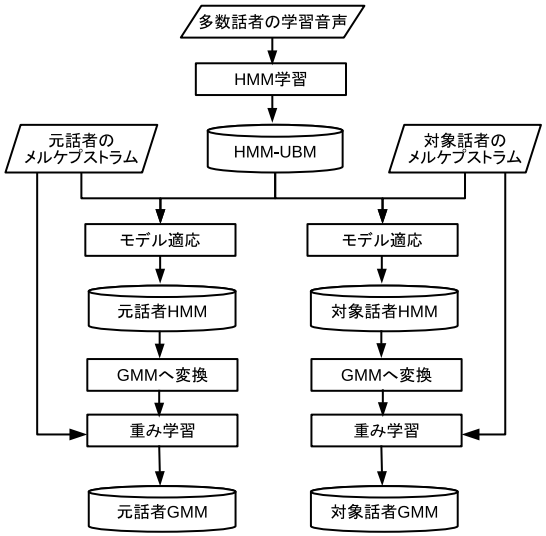

図 5 HMM の適応手順
Fig. 5 Procedure of HMM adaptation

4. 従来法 GMM と音素構造 GMM の性能比較

提案法である音素構造 GMM の有用性を検証するため，音素構造 GMM を使用した場合と，音素構造を用いずに学習した同混合数の GMM を使用した場合の変換音声の，客観評価及び主観評価による比較を行った．

4.1 実験条件

使用した音声の分析条件を表 1 に，GMM-UBM の学習に使用した学習データの詳細を表 2 に示す．また，GMM-UBM の適応条件と評価データの詳細を表 3 に示す．表 3 の話者名の M015，F015 は話者の識別番号であり，M が男性，F が女性，数字が話者の番号を表している．従来法 GMM 作成の際は，初期学習 5 回，途中学習 3 回，最終学習 5 回で学習を行った．また，混合数は，2，4，8，16，32 と 2 の倍数で増加させ，35 混合のモデルを作成した．音素構造 GMM 作成の際は，初期学習 5 回，連結学習 5 回で HMM を作成した．HMM の音素数は 35，状態数・混合数は 1 とした．また，HMM から GMM に変換した後の重み学習には学習データを使用し，5 回学習を行った．

表 1 音声分析条件
Table 1 Analysis conditions

サンプリング周波数	16KHz
分析周期	5msec
分析窓幅	25msec
窓関数	Hamming
FFT 長	1024
特徴量の種類	メルケプストラム
特徴量の次元数	19 次
周波数圧縮パラメータ	0.42

表 2 学習データの詳細
Table 2 Details of training data

学習データ	
コーパス	新聞記事読み上げ音声コーパス (JNAS)
話者	男女各 5 名，計 10 名
発話数	各話者約 150 発話
	計 1528 発話

4.2 評価方法

4.2.1 客観評価

本研究では，客観評価の指標として，メルケプストラム距離 (MCD) を用いる．MCD は，変換された音声と目標音声の誤差を表す指標であり，声質変換精度の客観的な指標としてよ

表 3 GMM-UBM の適応条件と評価データの詳細
Table 3 Adaptation conditions of GMM-UBM and details of evaluation data

適応データ	
話者	M015，F015
発話数	各話者 50 発話
適応条件	
適応法	MAP 法
事前分布パラメータ τ	10
更新パラメータ	平均ベクトル
学習回数	5
評価データ	
話者	M015，F015
発話数	各話者 103 発話 （学習データ，適応データに Open）

く用いられる．メルケプストラム距離は，以下の式で定義される．

$$MCD = \frac{10}{\log 10}\sqrt{2\sum_{d=1}^{D}(m_d^t - m_d^c)^2} \tag{9}$$

ここで，m_d^t と m_d^c は目標音声と変換音声のメルケプストラムの d 次の係数を表す．客観評価では，表 3 の評価音声 103 発話全文のメルケプストラム距離の平均により比較している．

4.2.2 主 観 評 価

主観評価として，被験者に従来法 GMM と音素構造 GMM それぞれで変換した同発話内容の音声を聴取し，1〜5 の 5 段階で評価してもらう MOS 評価を行った．評価項目は，「了解度：発音の聞き取りやすさ」，「音質：雑音や音割れの少なさ」の 2 点である．また，評価音声として，表 3 の評価音声 103 発話の内，従来法と音素構造それぞれ 40 発話，計 80 発話を使用した．また，40 発話の内訳は，男性から女性 20 発話，女性から男性 20 発話である．男性から女性の音声と女性から男性の音声は，発話内容が同じ音声を使用した．これらの条件で，変換音声に聴きなれていない 20 代の男子学生 16 名と女子学生 1 名，50 代の男性 1 名，計 18 名を対象に実験を行った．

4.3 実験結果及び考察

従来法と提案法の，評価音声 103 発話のメルケプストラム距離の平均を，表 4 に示す．参考のため，元音声と対象音声の距離と，従来法からの改善率も示してある．また，了解度と音質の平均評価スコアを図 6〜図 9 に示す．グラフ中のエラーバーは 95%信頼区間を表している．

表 4 従来法と提案法のメルケプストラム距離の平均 [dB]
Table 4 Average of mel cepstrum distance of conventional method and proposed method[dB]

	元音声	従来法 GMM	音素構造 GMM	改善率
男性から女性	0.360	0.308	0.271	12.0%
女性から男性	0.346	0.287	0.270	5.9%

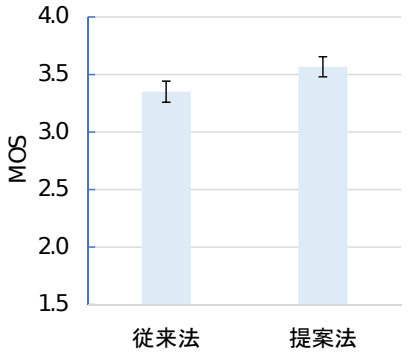

図 6 了解度の平均スコア（男性から女性）
Fig. 6 Mean opinion score of intelligibility (Male to Female)

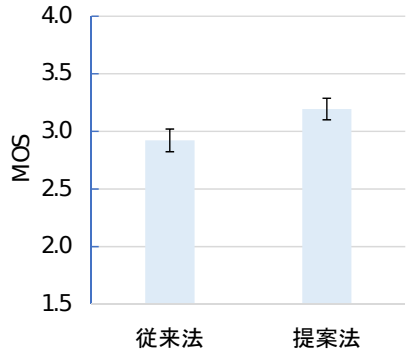

図 7 音質の平均スコア（男性から女性）
Fig. 7 Mean opinion score of sound quality (Male to Female)

表 4 を見ると，男性から女性，女性から男性共に，従来法と比較して提案法の距離が改善している．改善率では男性から女性の方が大きく，従来法と比較して 12%の改善が見られたが，女性から男性では 5.9%しか改善が見られなかった．しかし，男性から女性，女性から男性共に，音素構造 GMM を用いた方が距離が縮まる傾向が見られた．また，図 6〜図 9 を見ると，男性から女性，女性から男性共に，了解度と音質が有意に改善していることが分かる．

以上より，同混合数の GMM を用いた声質変換では，客観評価，主観評価共に音素構造

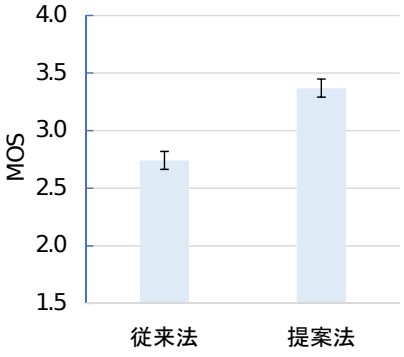

図 8 了解度の平均スコア（女性から男性）
Fig. 8 Mean opinion score of intelligibility (Female to Male)

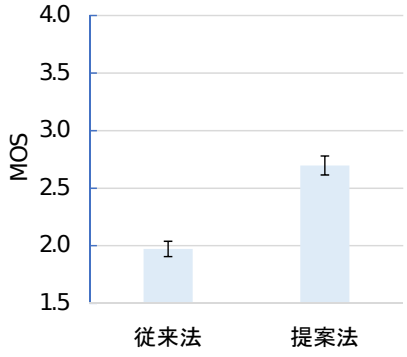

図 9 音質の平均スコア（女性から男性）
Fig. 9 Mean opinion score of sound quality (Female to Male)

を用いた方が良い結果が得られた．従って，非並列声質変換において，音素構造 GMM を用いるのは有効だと考えられる．

5. 音素構造 GMM の状態数・混合数に関する検討

前節にて，同混合数において音素構造が有用であることが確かめられたが，前節の実験は 35 混合の GMM による変換であったため，変換音声の品質は良くなかった．そこで，変換音声の品質向上のため，変換に使用する GMM の状態数や混合数を増加させて実験を行った．

5.1 実 験 条 件

本実験では，HMM の状態数を 3 状態とし，各状態の混合数を 1，2，4，8，16，32，64 と増加させた．また，重み学習は適応後に適応データを用いて行った．学習回数は，初期学習 5 回，途中学習 5 回，最終学習 5 回，重み学習 5 回とした．学習データや適応条件，評価データ等は前節の実験と同じである．

5.2 実験結果及び考察

各混合数における変換音声のメルケプストラム距離の平均を，表 5 に示す．

表 5 を見ると，男女間，女男間共に，混合数を増加させると距離が悪化した．また，混合数を増加させると，明らかに変換音声の声質が元話者に近づく現象が発生した．このような結果になった原因として，UBM の分布が話者毎に分かれてしまい，適応がうまく行われなかったことが考えられる．例えば，UBM の分布が男女で別れた場合，男性話者 A で適応を行った場合は UBM の男性に相当する分布が適応されるが，女性に相当する分布は適応されないと考えられる．同様に，女性話者 B で適応した場合は男性の分布が動かないと考えられる．このようなモデルで男性から女性への変換を行った場合，音声は A のモデルの男性に相当する分布から，B のモデルの適応されていない男性の分布に正規化されるため，ほぼ声質は変化しなくなる．このような現象が，実際のモデルでも発生していた可能性がある．

表 5 各混合数における変換音声のメルケプストラム距離の平均 [dB]
Table 5 Average of mel cepstrum distance for each number of mixture components[dB]

	1 混合	2 混合	4 混合	8 混合	16 混合	32 混合	64 混合
男性から女性	0.267	0.305	0.319	0.325	0.333	0.340	0.340
女性から男性	0.273	0.284	0.297	0.306	0.310	0.311	0.312

6. 結論と今後の課題

現在一般的に用いられている GMM に基づく声質変換の問題点として，各音素がどの分布にクラスタリングされるかが分からないということが挙げられる．本研究ではこの問題へのアプローチとして，文献[4] の非並列声質変換手法において音素構造 GMM を導入し，有用性の検討を行った．その結果，音素構造 GMM を用いない場合と比較して，客観評価，主観評価の両方において改善が見られた．このことから，非並列声質変換において音素構造 GMM を用いることはある程度の有用性があることが確かめられた．しかし，混合数を増加させた場合，メルケプストラム距離が悪化し，変換音声の声質が元話者に近づく現象が発生した．このため，現在の手法では，これ以上の性能向上は見込めないと考えられる．

この問題を解決し，変換精度を向上させるのが今後の課題である．そのための方法として，SAT(Speaker Adaptive Training)[11] の導入が考えられる．SAT は話者適応手法の一つであり，話者毎に分かれた分布を一つのコンパクトなモデルに変換し適応を行うことで，全話者の分布をまとめて適応する手法である．SAT は元は音声認識における手法だが，文献[12] のように声質変換における有用性を示した論文もあり，本研究においても有効だと考えられる．

参 考 文 献

1) 中村 哲，鹿野 清宏，”セパレートベクトル量子化を用いたスペクトグラムの正規化”，日本音響学会誌 44 巻 8 号，pp.595-602，1998.

2) 中村 哲, 鹿野 清宏，”ファジィベクトル量子化を用いたスペクトログラムの正規化”，日本音響学会誌 45 巻 2 号，pp.107-114，1989.
3) 花園 正也，川波 弘道，猿渡 洋，鹿野 清宏，”GMM に基づく声質変換への尤度基準学習の適用”，IEICE technical report.Speech 103(750)，pp.43-48, 2004.
4) Peng Song，Wenming Zheng，Li Zhao，”Non-parallel training for voice conversion based on adaptation method”，ICASSP2013，IEEE，pp.6905-6909，2013.
5) 佐藤慶，加藤正治，小坂哲夫，”統計的手法による水中音声の声質変換における精度向上の検討”，2013 年度第 7 回情報処理学会東北支部研究会，13-7-B1-2，2014.
6) 佐々木 志貢，小坂 哲夫，”音素の音響的特徴を利用した水中音声の声質変換”，電子情報通信学会総合大会，ISS-SP-219，2016.
7) J.L. Gauvain and C.H. Lee，”Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains”，IEEE Transactions on，Speech and Audio Processing，vol.2，no.2，pp.291-298，1994.
8) 今井聖，住田一男，古市千枝子，”音声合成のためのメル対数スペクトル近似 (MLSA) フィルタ”，電子情報通信学会論文誌 A，Vol.J66-A，No2，pp．122-129，1983.
9) 小坂 哲夫，赤津 達也，加藤 正治，好田 正紀，”音素モデルを用いた話者ベクトルに基づく話者識別”，電子情報通信学会論文誌 Vol.J90-D No.12，pp．3201-3209，2007.
10) R.Faltlhauser，G.Ruske，”Improving Speaker Recognition Performance Using Phonetically Structured Gaussian Mixture Models”，EUROSPEECH01，pp．751-754，2001.
11) Tasos Anastasakos，John McDonough，Richard Schwartz，John Makhoul，”A Compact Model for Speaker-Adaptive Training”，Proc．ICSLP96，vol．2，pp．1137-1140，1996.
12) Yamato Ohtani，Tomoki Toda，Hiroshi Saruwatari，Kiyohiro Shikano，”Speaker Adaptive Training for One-to-Many Eigenvoice Conversion Based on Gaussian Mixture Model”，INTERSPEECH2007，pp．1981-1984，2007.