統計的手法に基づく声質分析・変換・制御技術とその応用*

○戸田智基 (奈良先端大・情報)

1　はじめに

音声は言語情報のみでなく，パラ言語情報や非言語情報も同時に伝達できる．様々な情報が空気振動という一次元の時系列信号に混在するわけであるが，人間はその中から個々の情報を容易に分離・抽出することができる．一方で，計算機上において，このメカニズムを実現するのは容易ではない．多種多様な情報を表す「声質」という特徴の解明が必要不可欠となる．

声質と音声特徴量の関連性について，様々な観点から研究がなされている [1, 2]．例えば，下咽頭腔形状の個人差の影響を受ける高周波数帯域のスペクトル包絡成分に個人性が現れること [3] や，音源特徴量である声門体積流波形の変化により異なる声質が得られること [4] が報告されている．知覚される声質と物理現象の関係を明らかにすることは，声質を理解する上で重要であり，さらなる研究成果が期待される．

声質が表す情報の内，個人性に限定しても，発音の癖などのように，音韻に応じて多様に変化する要因が存在する．そのため，声質と音声特徴量の関係を明らかにするためには，音韻性と声質を分離する処理が必要となる．近年の計算機資源の拡大に伴い，大量の音声データを用いて統計的に音声特徴量をモデル化する技術が発展し，音韻性と声質の分離処理を確率的に行う枠組みが提案されている．その中の一つとして，本稿では，統計的手法に基づく声質分析・変換・制御技術について概説し，その応用例を紹介する．

2　統計的手法に基づく声質モデリング

テキストから音声信号を合成するテキスト音声合成処理や，音声信号を変形して言語情報を保持したまま所望の声質のみを変換する声質変換処理において，声質のモデル化は重要な技術課題である．80 年代後半から 90 年代にかけて，事前に収録された音声データに基づき合成・変換処理を行うコーパスベース方式 [5, 6] が提案され，合成・変換処理を数理的に記述することが可能となった．本方式は日々着実な進歩を遂げており，近年では，確率モデルに基づく音声合成・変換処理が主流として盛んに研究されている．

テキスト音声合成処理は，与えられる言語情報 $\boldsymbol{l}$ に対して，音声特徴量 $\boldsymbol{x}$ の確率密度関数 $P(\boldsymbol{x}|\boldsymbol{l})$ を推定する問題としてみなせる．代表的な手法は，隠れマルコフモデル（hidden Markov model: HMM）を用いた手法 [7] である．言語情報の利用により，分節的特徴のみでなく，韻律的特徴も上手くモデル化できる．一方で，声質変換処理は，与えられる元音声の音声特徴量 $\boldsymbol{x}$ に対して，目標音声の音声特徴量 $\boldsymbol{y}$ の確率密度関数 $P(\boldsymbol{y}|\boldsymbol{x})$ を推定する問題としてみなせる．代表的な手法は，混合正規分布モデル（Gaussian mixture model: GMM）を用いた手法 [8] である．言語情報を一切必要としない変換処理が可能であり，フレーム毎の変換処理も実現できる．韻律的特徴についてはモデル化の困難性が増すが，分節的特徴は比較的上手くモデル化できる．なお，合成・変換音声を得るためには，推定された確率密度関数から音声特徴量を生成する必要がある．品質の高い音声を得るためには，時系列データの特徴を効果的に捉える動的特徴量 [9] や系列内変動 [10] などを考慮した生成法が有効である．

これらの統計的手法において，合成される音声の声質は，音声特徴量の確率密度関数を学習するために用いる音声データに依存する．所望の声質をモデル化し制御するためには，音韻性と声質を分離する枠組みを導入する必要がある．関連する技術として，複数話者による同一音韻の音声特徴量に対して補間処理を行うことで，目標話者の声質を持つ音声特徴量を生成する話者補間 [11] がある．固有声技術 [12] は，この処理を特徴量空間ではなくモデルパラメータ空間に導入したものである．確率密度関数のパラメータを声質依存部と声質非依存部に分解し，モデル化対象とする声質を幅広くカバーする音声データを用いて，個々のパラメータを学習する．これにより，声質依存パラメータによる確率密度関数の変形が可能となる．

3　固有声混合正規分布モデルに基づく声質分析・変換・制御

固有声変換 [13] は，固有声技術を GMM に基づく声質変換処理に導入したものである．本技術は，音韻性と声質を自動的に分離する仕組みを内包しており，**Fig. 1** に示すとおり，言語情報が不要な声質分析，声質変換，声質制御を実現できる．以下では，声質の要素として主に個人性に着目し，本技術を説明する．

3.1　参照話者に基づくパラレルデータセット

通常の声質変換では，確率密度関数を学習するために，元話者と目標話者による同一内容発声データ（パラレルデータ）を用いる．これにより，言語情報は同一で，変換対象の声質情報のみが異なる音声特徴量対が得られる．一方で，固有声変換では，参照話者と呼ばれるある特定の話者と，多数の事前収録話者間におけるパラレルデータのセットを用いる．個々の事前収録話者に対しては，必ずしも同一内容発声データを必要としないが，参照話者に対しては全ての事前収録話者と同一内容の発声データが必要となる．

3.2　結合確率密度関数のモデル化

3.2.1　固有声 GMM

フレーム t における参照話者の音声特徴量ベクトルを $\boldsymbol{x}_t = [x_t(1), \cdots, x_t(D_x)]^\top$ とし，それに対応す

*Voice quality analysis, conversion, and control techniques based on statistical approaches and their applications. by TODA, Tomoki (Nara Institute of Science and Technology)

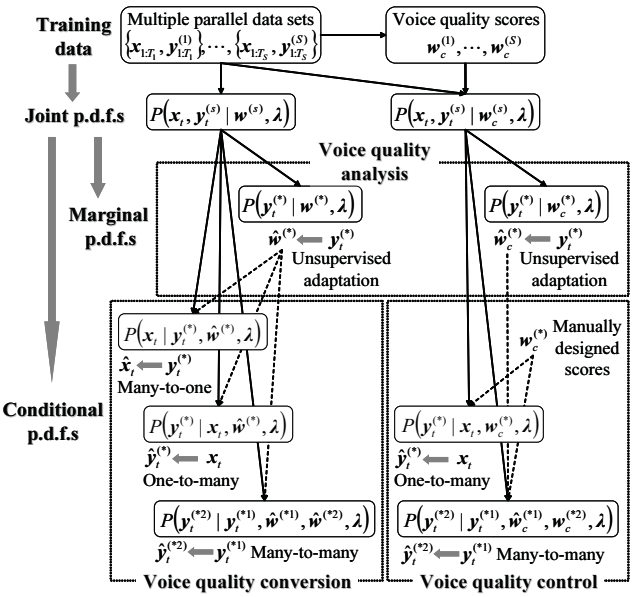

Fig. 1 Framework of voice quality analysis, conversion, and control techniques based on eigenvoices.

る事前収録話者 s の音声特徴量ベクトルを $\boldsymbol{y}_t^{(s)} = \left[y_t^{(s)}(1), \cdots, y_t^{(s)}(D_y)\right]^\top$ とする．ここで，$^\top$ は転置を表す．これらの音声特徴量ベクトルの結合確率密度関数を，M 個の多次元正規分布（次元数は $D_x + D_y$）からなる GMM でモデル化する．

$$
\begin{aligned}
&P\left(\boldsymbol{x}_t, \boldsymbol{y}_t^{(s)} | \boldsymbol{w}^{(s)}, \boldsymbol{\lambda}\right) \\
&= \sum_{m=1}^{M} P\left(m|\boldsymbol{\lambda}\right) P\left(\boldsymbol{x}_t, \boldsymbol{y}_t^{(s)} | m, \boldsymbol{w}^{(s)}, \boldsymbol{\lambda}\right) \\
&= \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\begin{bmatrix}\boldsymbol{x}_t \\ \boldsymbol{y}_t^{(s)}\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_m^{(x)} \\ \boldsymbol{\mu}_m^{(y,s)}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_m^{(xx)} & \boldsymbol{\Sigma}_m^{(xy)} \\ \boldsymbol{\Sigma}_m^{(yx)} & \boldsymbol{\Sigma}_m^{(yy)}\end{bmatrix}\right) \quad (1)
\end{aligned}
$$

ここで，α_m は m 番目の分布の混合重みであり，$\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ は平均ベクトル $\boldsymbol{\mu}$，共分散行列 $\boldsymbol{\Sigma}$ の多次元正規分布を表す．また，m 番目の分布における事前収録話者 s に対する平均ベクトル $\boldsymbol{\mu}_m^{(y,s)}$ は，次式で与えられる．

$$
\boldsymbol{\mu}_m^{(y,s)} = \boldsymbol{B}_m^{(y)} \boldsymbol{w}^{(s)} + \boldsymbol{b}_{m,0}^{(y)} \quad (2)
$$

ここで，$\boldsymbol{B}_m^{(y)} = \left[\boldsymbol{b}_{m,1}^{(y)}, \cdots, \boldsymbol{b}_{m,J}^{(y)}\right]$ 及び $\boldsymbol{b}_{m,0}^{(y)}$ は m 番目の分布の基底ベクトルセット及びバイアスベクトルであり，$\boldsymbol{w}^{(s)}$ は事前収録話者 s に対する J 次元の重みベクトルである．重みベクトルは個々の事前収録話者に依存するパラメータであり，全分布間で共有される．一方で，$\boldsymbol{\lambda}$ は全事前収録話者間で共有される分布依存パラメータセットであり，各分布における混合重み，参照話者に対する平均ベクトル，基底ベクトルセット，バイアスベクトル，共分散行列から成る．

各分布の平均ベクトル $\boldsymbol{\mu}_m^{(y,s)}$ は，基底ベクトルで張られる部分空間上で表される．話者依存パラメータである重みベクトルを変化させることで，個々の分布の平均ベクトルがシフトし，参照話者と様々な話者間における結合確率密度関数が得られる．

3.2.2 固有声 GMM の学習法

パラレルデータセットを用いて，話者適応学習 [14] に基づき，固有声 GMM のパラメータを最適化する．分布依存パラメータセット $\boldsymbol{\lambda}$ および個々の事前収録話者（話者数は S）に対する重みベクトルのセット $\boldsymbol{w}^{(1:S)} = \left\{\boldsymbol{w}^{(1)}, \cdots, \boldsymbol{w}^{(S)}\right\}$ を次式にて最尤推定する．

$$
\left\{\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{w}}^{(1:S)}\right\} = \underset{\{\boldsymbol{\lambda}, \boldsymbol{w}^{(1:S)}\}}{\operatorname{argmax}} \prod_{s=1}^{S} \prod_{t=1}^{T_s} P\left(\boldsymbol{x}_t, \boldsymbol{y}_t^{(s)} | \boldsymbol{w}^{(s)}, \boldsymbol{\lambda}\right) \quad (3)
$$

参照話者と各事前収録話者のパラレルデータ（$\left\{\boldsymbol{x}_1, \boldsymbol{y}_1^{(s)}\right\}, \cdots, \left\{\boldsymbol{x}_{T_s}, \boldsymbol{y}_{T_s}^{(s)}\right\}$）に対して，重みベクトルが適応された固有声 GMM を用いて，尤度計算が行われる．全パラレルデータに対する尤度最大化に基づき，各パラメータは最適化される．

固有声 GMM でモデル化される結合確率密度関数において，参照話者に対する周辺確率密度関数 $P(\boldsymbol{x}_t|\boldsymbol{\lambda})$ は，事前収録話者に依らず一定であるため，個々の分布がモデル化する参照話者の音韻空間は固定される．また，パラレルデータの利用により，参照話者と各事前収録話者の音声特徴量対 $\left\{\boldsymbol{x}_t, \boldsymbol{y}_t^{(s)}\right\}$ は同一の音韻性を持つため，個々の分布がモデル化する音韻空間は，全事前収録話者に対しても固定される．すなわち，参照話者の音声特徴量がアンカーの役割を果たすことで，全事前収録話者間において個々の分布と音韻空間の対応付けの統一化が図られる．結果，固有声 GMM において，音韻性は個々の分布でモデル化され，個人性は重みベクトルでモデル化されることで，音韻性と個人性の分離が行われる．

重みベクトルに対して，直感的に理解しやすい意味を持たせることも可能である [15]．HMM 音声合成における声質制御法 [16] と同様に，声質表現語 [17] を用いて，各事前収録話者に対して，声質表現語スコアを人手で付与する．得られた声質表現語スコアを要素として，各事前収録話者に対する重みベクトル $\boldsymbol{w}_c^{(s)}$ を構成する．そして，全パラレルデータに対する尤度に基づき，共有パラメータのみを最適化する．

$$
\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}} \prod_{s=1}^{S} \prod_{t=1}^{T_s} P\left(\boldsymbol{x}_t, \boldsymbol{y}_t^{(s)} | \boldsymbol{w}_c^{(s)}, \boldsymbol{\lambda}\right) \quad (4)
$$

各声質表現語に対応する基底ベクトルにより，声質表現語スコアという知覚尺度に対応した部分空間が構成される．これにより，声質表現語で表される声質要因と音声特徴量の関係がモデル化される．

3.3 声質分析

固有声 GMM を用いて，与えられた音声データに対して，個人性を表す重みベクトルを推定することで，声質分析処理を実現できる．式 (1) で表される結合確率密度関数に対して，参照話者の音声特徴量 $\boldsymbol{x}_t$ の周辺化を行うことで，次式に示す周辺確率密度関数が得られる．

$$
P\left(\boldsymbol{y}_t^{(*)} | \boldsymbol{w}^{(*)}, \boldsymbol{\lambda}\right) = \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\boldsymbol{y}_t^{(*)}; \boldsymbol{\mu}_m^{(y,*)}, \boldsymbol{\Sigma}_m^{(yy)}\right) \quad (5)
$$

分析対象音声の音声特徴量を $\boldsymbol{y}_1^{(*)}, \cdots, \boldsymbol{y}_T^{(*)}$ とすると，次式のとおり，周辺確率密度関数の尤度最大化に基づき，重みベクトルを推定することができる．

$$\hat{\boldsymbol{w}}^{(*)} = \arg\max_{\boldsymbol{w}^{(*)}} \prod_{t=1}^{T} P\left(\boldsymbol{y}_t^{(*)} | \boldsymbol{w}^{(*)}, \boldsymbol{\lambda}\right) \tag{6}$$

本推定処理は，言語情報を一切必要としないため，完全な教師無し推定処理となる．また，重みベクトルは分布間で共有されており，その次元数も小さいので（事前収録話者数未満であり，大幅に削減可能），一発話程度といった極少量の音声データのみを用いても，十分な推定精度が得られる．さらに，重みベクトルに対する事前分布を用意して，最大事後確率推定を行うことで，一単語程度の音声データに対しても，頑健な推定処理を実現できる．

推定された重みベクトルにより声質が表現されるが，その値は直感的に理解し難い．そこで，声質表現語スコアを重みベクトルとする固有声 GMM を用いることで，声質表現語スコアの推定が可能となり，直感的に理解しやすい声質分析を実現できる．

3.4 声質変換

固有声 GMM を用いて，参照話者と任意の話者間の声質変換処理を実現できる．まず，任意の話者の音声データに対して，式 (6) に基づき重みベクトルの最尤推定値を求めることで，固有声 GMM で表される結合確率密度関数を適応する．参照話者の音声データが与えられる場合，結合確率密度関数 $P\left(\boldsymbol{x}_t, \boldsymbol{y}_t^{(*)} | \hat{\boldsymbol{w}}^{(*)}, \boldsymbol{\lambda}\right)$ と周辺確率密度関数 $P\left(\boldsymbol{x}_t | \boldsymbol{\lambda}\right)$ から，次式の条件付確率密度関数が得られる．

$$\begin{aligned} &P\left(\boldsymbol{y}_t^{(*)} | \boldsymbol{x}_t, \hat{\boldsymbol{w}}^{(*)}, \boldsymbol{\lambda}\right) \\ &= \sum_{m=1}^{M} P\left(m | \boldsymbol{x}_t, \boldsymbol{\lambda}\right) P\left(\boldsymbol{y}_t^{(*)} | \boldsymbol{x}_t, m, \hat{\boldsymbol{w}}^{(*)}, \boldsymbol{\lambda}\right) \\ &= \sum_{m=1}^{M} \gamma_{m,t}^{(x)} \mathcal{N}\left(\boldsymbol{y}_t; \boldsymbol{\mu}_{m,t}^{(y,*|x)}, \boldsymbol{\Sigma}_m^{(y|x)}\right) \end{aligned} \tag{7}$$

ここで，

$$\gamma_{m,t}^{(x)} = \frac{\alpha_m \mathcal{N}\left(\boldsymbol{x}_t; \boldsymbol{\mu}_m^{(x)}, \boldsymbol{\Sigma}_m^{(xx)}\right)}{\sum_{n=1}^{M} \alpha_n \mathcal{N}\left(\boldsymbol{x}_t; \boldsymbol{\mu}_n^{(x)}, \boldsymbol{\Sigma}_n^{(xx)}\right)} \tag{8}$$

$$\boldsymbol{\mu}_{m,t}^{(y,*|x)} = \boldsymbol{\Sigma}_m^{(yx)} {\boldsymbol{\Sigma}_m^{(xx)}}^{-1} \left(\boldsymbol{x}_t - \boldsymbol{\mu}_m^{(x)}\right) + \boldsymbol{\mu}_m^{(y,*)} \tag{9}$$

$$\boldsymbol{\Sigma}_m^{(y|x)} = \boldsymbol{\Sigma}_m^{(yy)} - \boldsymbol{\Sigma}_m^{(yx)} {\boldsymbol{\Sigma}_m^{(xx)}}^{-1} \boldsymbol{\Sigma}_m^{(xy)} \tag{10}$$

である．この条件付確率密度関数に基づき，適応された話者の音声特徴量を推定することができる．本変換処理は，参照話者から任意の話者への変換を行うため，**一対多声質変換**と呼ばれる．

同様に，条件付確率密度関数 $P\left(\boldsymbol{x}_t | \boldsymbol{y}_t^{(*)}, \hat{\boldsymbol{w}}^{(*)}, \boldsymbol{\lambda}\right)$ に基づいて，任意の話者から参照話者への変換を行う**多対一声質変換**も実現できる．なお，多対一声質変換は一対多声質変換よりも本質的に容易な変換処理となるため，高精度な適応処理を行わなくても，比較的良好な変換性能が得られる．

さらに，任意の話者から任意の話者への変換である**多対多声質変換**も実現できる．話者 $*1$ から話者 $*2$ への変換を行う際には，まず，式 (6) により，各話者に対して独立に重みベクトルの最尤推定値 $\hat{\boldsymbol{w}}^{*1}, \hat{\boldsymbol{w}}^{*2}$ を求める．各話者に適応された結合確率密度関数に対して，次式の通り，参照話者の音声特徴量 $\boldsymbol{x}_t$ の周辺化を行うことで，話者 $*1$ の音声特徴量 $\boldsymbol{y}_t^{(*1)}$ と話者 $*2$ の音声特徴量 $\boldsymbol{y}_t^{(*2)}$ に対する結合確率密度関数が得られる．

$$\begin{aligned} &P\left(\boldsymbol{y}_t^{(*1)}, \boldsymbol{y}_t^{(*2)} | \hat{\boldsymbol{w}}^{(*1)}, \hat{\boldsymbol{w}}^{(*2)}, \boldsymbol{\lambda}\right) \\ &= \sum_{m=1}^{M} P\left(m | \boldsymbol{\lambda}\right) \int P\left(\boldsymbol{y}_t^{(*1)} | \boldsymbol{x}_t, m, \hat{\boldsymbol{w}}^{(*1)}, \boldsymbol{\lambda}\right) \\ &\qquad P\left(\boldsymbol{y}_t^{(*2)} | \boldsymbol{x}_t, m, \hat{\boldsymbol{w}}^{(*2)}, \boldsymbol{\lambda}\right) P\left(\boldsymbol{x}_t | m, \boldsymbol{\lambda}\right) d\boldsymbol{x}_t \\ &= \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\begin{bmatrix} \boldsymbol{y}_t^{(*1)} \\ \boldsymbol{y}_t^{(*2)} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_m^{(y,*1)} \\ \boldsymbol{\mu}_m^{(y,*2)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_m^{(yy)} & \boldsymbol{\Sigma}_m^{(yxy)} \\ \boldsymbol{\Sigma}_m^{(yxy)} & \boldsymbol{\Sigma}_m^{(yy)} \end{bmatrix}\right) \end{aligned} \tag{11}$$

ここで，

$$\boldsymbol{\Sigma}_m^{(yxy)} = \boldsymbol{\Sigma}_m^{(yx)} {\boldsymbol{\Sigma}_m^{(xx)}}^{-1} \boldsymbol{\Sigma}_m^{(xy)} \tag{12}$$

である．この結合確率密度関数から条件付き確率密度関数 $P\left(\boldsymbol{y}_t^{(*2)} | \boldsymbol{y}_t^{(*1)}, \boldsymbol{w}^{(*1)}, \boldsymbol{w}^{(*2)}, \boldsymbol{\lambda}\right)$ を導出することで，多対多声質変換を実現できる [18]．本処理は，多対一声質変換を行い，その際の変換誤差成分も考慮して，続けて一対多声質変換を行う処理に相当する．

3.5 声質制御

一対多声質変換において，声質表現語スコアを重みベクトルとする固有声 GMM を用いることで，声質表現語スコアの手動操作による変換音声の声質制御が可能となる．この枠組みでは，適応データを一切必要とせずに，参照話者の音声から手動設定した声質を持つ音声への変換が可能となる．

多対多声質変換において声質制御を行う際には，声質表現語スコアを重みベクトルとする固有声 GMM を用いて，式 (11) で表される周辺化を行えばよい．しかし，通常，声質表現語スコア数はその操作性の面から数個程度に抑えられるため，部分空間上で表現可能な声質は限定され，十分な適応性能が得られない可能性がある．そこで，声質表現語スコアのみでなく，適応学習によりデータから推定する重みベクトルを併用することで，部分空間を拡張する手法が提案されている [15]．これにより，声質操作性能と声質適応性能の両立が行われる．

なお，声質表現語スコアによっては，平均ベクトルとの対応を式 (2) で表される線形回帰モデルで上手く表現できない場合もある．その際には，カーネル回帰などの非線形回帰モデルを導入することで，声質操作性能を改善させることができる [19]．

4　応用例

統計的手法に基づく声質変換技術は，話者変換のみでなく，様々な信号間の変換処理に対して適用できる．特に，GMM に基づく変換法は言語依存性が低く，リアルタイム変換処理も実現できるため，音声コミュニケーションへの応用が期待される．例えば，電話音声の狭帯域音声スペクトル包絡から広帯域音声スペクトル包絡へと変換することで，電話音声の帯域拡張処理が実現できる [20]．雑音環境下における音声コミュニケーションのための，骨伝導音声を用いた音声強調処理にも適用可能である [21]．また，秘匿性に優れた音声コミュニケーションとして，非可聴つぶやきマイクロフォンを用いた肉伝導音声収録が提案されており [22]，その音質および明瞭性を改善するために，様々な発話様式の肉伝導音声に対する変換処理に適用されている [23]．この他にも，音声信号からの調音運動逆推定や，調音運動からの音声信号生成などに対しても，適用可能である [24]．適用の際には，個々の応用例に応じて，変換元となる特徴量および変換先となる特徴量を適切に選択することが重要となる．

固有声変換の特徴である教師なし適応性能と声質制御性能を活用することで，より利便性の高い応用技術が構築できる．例えば，声質を保持する他言語音声合成技術として，一対多声質変換の音声翻訳システムへの適用が提案されている [25]．音声翻訳の出力音声に対して，一対多声質変換を行うことで，入力話者の声質を持つ出力音声を生成できる．極少量の音声データを用いた教師無し適応技術により，翻訳システムに入力される様々な言語の音声データのみを用いて，固有声 GMM の適応が可能となる．なお，出力音声合成用のテキスト音声合成システムを用いることで，現存する多数話者の音声データと同一内容の合成音声を人工的に生成できるため，固有声 GMM 学習用のパラレルデータは容易に構築できる．

別の応用例として，発声障害者の音声をより自然で明瞭な音声へと変換する処理への適用が提案されている [26]．手術等で喉頭を取り除き，声を失った喉頭摘出者は，食道発声や電気式人工喉頭を用いた発声により，再び音声を発声することが可能となる．しかしながら，発声される音声の自然性は乏しく，話者性も大幅に失われる．そこで，統計的声質変換を用いることで，各種代替発声法により得られる音声を健常者の通常音声に変換する技術が提案されている．固有声変換を導入することで，手術前の自身の声が極少量でも録音されている際には，類似した声質での発声が可能となり，仮に録音データが存在しなくても，手動制御された声質での発声が可能となる．

5　おわりに

本稿では，統計的手法に基づく声質分析・変換・制御技術に関して概説し，その応用例を紹介した．大量の音声データを用いることで，音韻性と声質を確率的に分離する処理が実現できる．本技術は，言語依存性が低く，リアルタイム処理にも適していることから，音声コミュニケーションにおける様々な障壁（言語の違いや身体的障害など）を越える技術への発展が期待される．なお，統計的手法では，大量の音声データに基づいて，声質と音声特徴量の関係性を確率モデルで記述するが，声質と物理現象を結びつけるところまでには至っていない．声質の理解を深めるために，物理的な制約を統合した統計処理の実現が望まれる．

謝辞　本研究の一部は，科研費補助金若手研究（A）により実施したものである．

参考文献

[1] Kuwabara and Sagisaka, Speech Commun., **16**(2), 165–173, 1995.
[2] Kitamura and Akagi, J. Acoust. Soc. Jpn. (E), **16**(5), 283–289, 1995.
[3] 北村, 日本音響学会聴覚研資, **38**(6), 653–658, 2008.
[4] 粕谷, 楊, 音響誌, **51**(11), 869–875, 1995.
[5] Iwahashi *et al.*, IEICE Trans. Fundamentals, **E76-A**(11), 1942–1948, 1993.
[6] Abe *et al.*, J. Acoust. Soc. Jpn. (E), **11**(2), 71–76, 1990.
[7] Zen *et al.*, Speech Commun., **51**(11), 1039–1064, 2009.
[8] Stylianou *et al.*, IEEE Trans. Speech & Audio Process., **6**(2), 131–142, 1998.
[9] 徳田 他, 音響誌, **53**(3), 192–200, 1997.
[10] Toda *et al.*, IEEE Trans. Audio, Speech & Lang. Process., **15**(8), 2222–2235, 2007.
[11] Iwahashi and Sagisaka, Speech Commun., **16**(2), 139–151, 1995.
[12] Kuhn *et al.*, IEEE Trans. Speech & Audio Process., **8**(6), 695–707, 2000.
[13] 戸田, 信学技報, **SP2008-138**, 73–78, 2009.
[14] Anastasakos *et al.*, Proc. ICSLP, 1137–1140, 1996.
[15] Ohta *et al.*, Proc. INTERSPEECH, pp. 2158–2161, 2010.
[16] Nose *et al.*, IEICE Trans. Inf. & Syst., **E90-D**(9), 1406–1413, 2007.
[17] 木戸, 粕谷, 音響誌, **55**(6), 405–411, 1999.
[18] Ohtani *et al.*, Proc. INTERSPEECH, 1623–1626, 2009.
[19] 山本 他, 情報処理研報, **2011-SLP-85**(11), 1–6, 2011.
[20] Cheng *et al.*, IEEE Trans. Speech & Audio Process., **2**(4), 544–548, 1994.
[21] Subramanya *et al.*, Speech Commun., **50**(3), 228–243, 2008.
[22] 中島 他, 信学論, **J87-D-II**(9), 1757–1764, 2004.
[23] Toda *et al.*, Proc. ICASSP, 3601–3604, 2009.
[24] Toda *et al.*, Speech Commun., **50**(3), 215–227, 2008.
[25] 服部 他, 情報処理研報, **2011-SLP-85**(10), 1–6, 2011.
[26] 戸田 他, 信学技報, **SP2010-58**, 75–80, 2010.