Vol.2017-MUS-115 No.27
2017/6/17

喉頭摘出者のための歌唱支援を目指した統計的電気音声変換法

森川 一穂[1,a)]　戸田 智基[2,b)]

概要：本研究では，喉頭摘出者のための歌唱支援システムの構築を目指し，電気式人工喉頭を用いて得られる音声（電気音声）をより自然な歌声に変換する手法を提案する．電気音声は，その音高が電気式人工喉頭から機械的に生成される音源により決定されるため，歌唱を行うためには事前にメロディを設定する必要があるなど，歌唱の自由度は著しく限定される．また，その声質に関しても，機械的な音源の影響により，通常の歌唱と比べて機械的に聞こえることや，また，生成された音源が外部に漏れて雑音として混入するため，音質が劣化することといった問題がある．この問題に対して，本研究では，楽器を利用した音高の入力と，雑音抑圧や声質変換といった信号処理を適用することで，電気音声の音高の制御と音質の向上を可能とする電気音声変換法を提案する．電気音声による歌唱の音高の制御を実現するために，楽器から入力される音高情報に対して，歌声特有の音響特徴を付与し，電気音声に反映させる処理を導入する．また，より自然な声質による歌唱を可能とするために，スペクトル減算法に基づく雑音抑圧処理を適用し，さらに，統計的手法による電気音声から自然な歌唱への変換処理の導入を試みる．実験的評価から，1）電気音声に対する雑音抑圧手法の導入により，歌声の人間らしさが大きく改善されること，および，2）電気音声の音高に対する歌声特有の音響特徴付与処理や声質変換処理に関しては，必ずしも十分な効果が得られるとは限らないことを示す．

1. はじめに

発声や歌唱という行為は，音声による情報伝達や歌声による情報表現において必要不可欠であり，人が人らしい生活を送る上で，重要な意味を持つ．一方で，喉頭がんなどにより，喉頭を全摘出した人（喉頭摘出者）は，発声や歌唱で必要不可欠となる声帯振動による音源の生成能力を失うため，深刻な発声・歌唱機能障害を患う．発声機能障害は，音声コミュニケーションにおける大きな障壁となるため，喉頭摘出者の生活の質を著しく低下させる．また，歌唱機能障害においても，過去に歌唱に従事していた人や，歌唱を余暇活動として楽しんでいた人にとって，歌唱を行えないことは，その人の人生そのものにかかわる極めて重大な問題であり，生活の質を著しく低下させる．そのため，喉頭摘出者に対する発声・歌唱支援技術の構築が切望されている．

†1 現在，名古屋大学 情報学研究科
Presently with Graduate School of Informatics, Nagoya University
†2 現在，名古屋大学 情報基盤センター
Presently with Information Technology Center, Nagoya University
a) morikawa.kazuho@h.mbox.nagoya-u.ac.jp
b) tomoki@icts.nagoya-u.ac.jp

喉頭摘出者が再び発声を行う方法として，電気式人工喉頭を用いた発声法がある．この発声法は，外部から機械的に生成される音源信号を用いて発声を行う方法である．他の代替発声法に比べ，習得が容易で，比較的明瞭な音声を発することができるという大きな利点が存在する．しかし，発せられる音声（電気音声）は機械的であり，人間の発話と比べて自然性は著しく低下する．また，電気式人工喉頭から音源が漏れ，雑音として音声に混入することで，音質が劣化するという欠点がある．

これまでに，電気式人工喉頭による発声法の欠点を克服するために様々な研究がなされ，多くの成果を挙げている．例えば，櫛型フィルタによる雑音抑圧 [1]，音響特徴量に対する平滑化処理 [2]，フォルマント操作 [3]，聴覚マスキングに基づく雑音抑圧 [4] など，信号処理に基づく電気音声強調法が提案されている．また，より高精度な強調処理の実現を目指し，統計的声質変換 [5] によって電気音声を自然な音声に変換する研究 [6] や，電気音声を制御して抑揚を表現することで，電気音声に自然性を持たせる研究 [7]，一対多固有声変換法 [8] を用いて喉頭摘出者の発する音声を個人性をもたせつつより自然な音声に変換する研究 [9] など，統計的手法に基づく電気音声強調法が提案されており，その高い有効性が示されている．近年，統計的手法の

© 2017 Information Processing Society of Japan

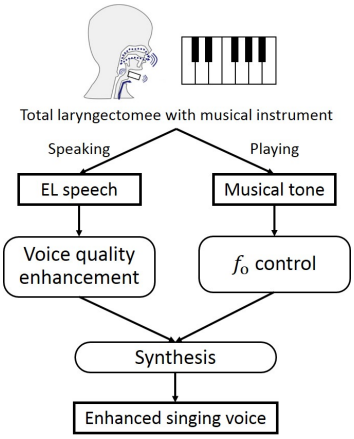


Fig 1 Process of the proposed singing aid system using electrolarynx for laryngectomee.

基盤技術は大幅な性能改善がなされているため，今後さらなる改善が期待される．などが挙げられる．

このように，喉頭摘出者が再びより自然な音声コミュニケーションを実現するための研究が多くなされている一方で，歌唱のように，より複雑な音源信号の制御を必要とするものに対しては，十分な研究がなされておらず，特に過去に歌唱に従事していた人や，歌唱を余暇活動として楽しんでいた人の生活の質は未だ低い水準にあるといえる．電気式人工喉頭を用いた歌唱支援として，例えば，事前に決められた楽曲のメロディーに対応した音高を設定しておくことで，対応する楽曲の歌唱を可能とする機能が実現されている．しかしながら，歌唱可能な楽曲は限定的であり，歌唱の自由度は極めて低い．より柔軟な歌唱支援を実現するためには，喉頭摘出者が自由に音高を調整できる機能の実現が望まれる．また，より人間らしい歌声による歌唱を実現するためには，電気音声の機械的な声質ではなく，より自然な声質による歌声を生成する仕組みが必要となる．

本研究では，喉頭摘出者のための歌唱支援技術として，楽器演奏による音高制御処理と，電気音声の声質を改善するための電気音声強調処理に基づく歌唱支援システムの構築を目指す．そのための第一歩として，本稿では，１）歌声特有の特徴付与に基づく音高制御法と，２）信号処理による電気音声の雑音抑圧ならびに統計的手法による電気音声の声質変換に基づく電気音声強調法を提案する．実験的評価結果から，提案法の有効性を調査するとともに，今後の課題を明らかにする．

2. 喉頭摘出者のための歌唱支援システム

本研究で提案する喉頭摘出者のための歌唱支援システムの概要を図 1 に示す．歌唱者として電気式人工喉頭を使用する喉頭摘出者を対象とし，電気式人工喉頭を用いた発声に合わせて楽器を演奏する状況を想定する．電気音声から韻律情報を，楽器音から音高情報を抽出し，これらをリアルタイムで制御・統合することで，電気式人工喉頭と楽器

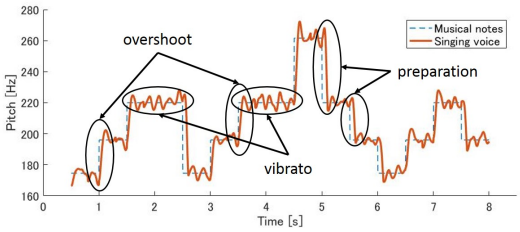


Fig 2 f_o patterns of musical notes and singing voice.

を用いたリアルタイムな歌唱を実現する．

本稿では，歌唱支援システムの基盤となる電気音声と楽器音の制御・統合の手法を提案する．より具体的には，通常音声に対して有効性が示されている，話し声を歌声に変換する手法 [10], [11] を応用することで，電気音声に歌声特有の特徴を付与する．また，従来の電気音声に対して有効性が示されている雑音抑圧手法 [12] を，電気音声による歌唱に対しても適用する．さらに，統計的声質変換 [5] を電気音声に対して適用することで，電気音声をより人間らしい声質に近づけることを目指す．そして，これらを組み合わせることで，電気音声による歌唱における音高情報の制御を可能としつつ，より人間らしい歌声への変換を目指す．

3. 関連研究

本章では，電気音声をより自然な歌声に変換するための方法として関連のある，話し声を歌声に変換する手法，電気音声の雑音抑圧の手法，および統計的声質変換の手法について述べる．

3.1 話し声を歌声に変換する手法

これまでに，通常の音声には見られない，歌声に特有の音響特徴が存在することが数多く報告されている [13]．本節で述べる手法は，歌声に特有の音響特徴として，音高（基本周波数，f_o），スペクトル包絡，音韻長の 3 つに着目している．本節では，音高に関する歌声特有の音響特徴に焦点を絞り、これらの音響特徴を話し声に付与することで話し声を歌声に変換する手法 [10], [11] について述べる．

歌声の f_o には次のような動的変動成分が存在することがわかっている．

オーバーシュート (Overshoot)
滑らかな音高の変化，およびその直後に目的音高を越える瞬時的な変動成分

ヴィブラート (Vibrato)
同一音高区間で観測される 4 ～ 8 Hz の準周期的な変動成分

プレパレーション (Preparation)
音高変化直前に変化とは逆方向に振れる瞬時的な変動成分

微細変動 (Fine fluctuation)

© 2017 Information Processing Society of Japan

発声区間全体に観測される不規則で細かい変動成分

この手法は，歌詞の朗読音声とそれに対応する楽譜情報を入力とし，これらに対して上記の歌声特有の音響特徴を付与することによって合成歌声を得るものである．図 2 に譜面上の f_{o} パターンと歌声の f_{o} パターンの例を示す．本手法では，譜面中の音符をステップ関数で記述し，メロディの遷移の概形を生成する．次に，歌声特有の音響特徴の付与を行うために，入力音声を音声分析することによって f_{o} パターンを得る．これに上記の f_{o} 動的変動成分を付与することで歌声の f_{o} パターンを生成する．オーバーシュート，ヴィブラート，プレパレーションはメロディの遷移の概形に対して，以下の伝達関数を適用することで得られる．

$$H(s) = \frac{k}{s^2 + 2\zeta\omega s + \omega^2} \tag{1}$$

ここで，ω は固有角周波数，ζ は減衰項，k は振幅項を表す．オーバーシュートとプレパレーションは減衰振動モデル ($0 < |\zeta| < 1$)，ヴィブラートは定常振動モデル ($|\zeta| = 0$) で記述される．微細変動は白色雑音をカットオフ周波数 10 Hz の低域通過フィルタに通した後，最大振幅が 5 Hz になるように正規化することで生成し，f_{o} パターン全体に付与する．

3.2 電気音声の雑音抑圧

電気式人工喉頭から得られる電気音声には，電気式人工喉頭自体から漏れ出した音源が雑音として含まれ，明瞭性や自然性を欠いてしまうという問題がある．本節では，電気式人工喉頭の雑音を除去する手法として，スペクトル減算法 [14] を挙げ，電気音声に適応する方法 [12] について述べる．

3.2.1 スペクトル減算法の電気音声への適用

スペクトル減算法を電気音声に適用する場合の処理の過程を図 3 に示す．電気音声 (EL Speech) に含まれる雑音信号 (EL noise) は，電気式人工喉頭から生成される音源信号であるため，事前にその振幅特性を求めることが可能である．雑音信号は，電気式人工喉頭を喉元に押し当てて音源信号を生成し，口元のマイクで収録する．このとき，音源信号のみを収録するために，口を閉じて，電気音声が発声されないようにする．得られた雑音信号から振幅スペクトルの期待値を計算し，雑音振幅スペクトルのプロトタイプ $|\hat{L}(\omega)|$ を求める．観測信号の振幅スペクトルを $|Y(\omega,t)|$ とすると，雑音が抑圧された電気音声の振幅スペクトル $|\hat{S}(\omega,t)|$ は次式によって求められる．

$$|\hat{S}(\omega,t)| = \begin{cases} |Y(\omega,t)| - 2|\hat{L}(\omega)|, & (|Y(\omega,t)| > 2|\hat{L}(\omega)|) \\ 0, & (\text{otherwise}) \end{cases} \tag{2}$$

最後に，雑音成分が減算された振幅スペクトルと電気音声の位相スペクトルを用いて逆フーリエ変換を行うことで雑音が抑圧された電気音声を得る．

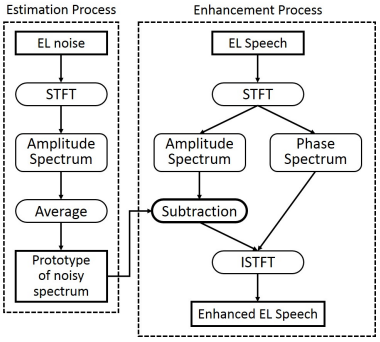


Fig 3 Process of spectral subtraction for electrolaryngeal (EL) speech.

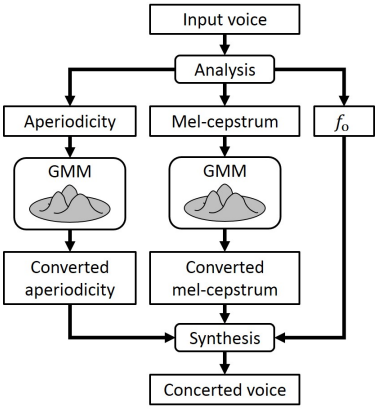


Fig 4 Process of statistical voice conversion.

3.3 統計的声質変換

電気音声をより人間らしい歌声に変換するためには，歌声の声質を人間らしいものに変換する必要がある．本研究では，電気音声の声質を人間らしい声質に変換するために統計的声質変換の技術を用いる．

統計的声質変換 (VC: statistical Voice Conversion)[5] は，ある人が発した音声の特徴量を統計的手法によって変換することによって，別の人が発したような音声に変換する技術のことである．この手法は学習処理と変換処理によって構成される．

学習処理は，入力音声と目標音声の特徴量の関係を表すために，入力音声と目標音声のパラレルデータ（同一内容の発話）を用いて，各特徴量の結合確率密度関数を混合ガウス分布モデル (GMM: Gaussian Mixture Model) によってモデル化する．時刻 t における入力音声の静的・動的特徴ベクトルを $\mathbf{X}_t = [\mathbf{x}_t^{\mathrm{T}}, \mathbf{\Delta x}_t^{\mathrm{T}}]^{\mathrm{T}}$，目標音声の静的・動的特徴ベクトルを $\mathbf{Y}_t = [\mathbf{y}_t^{\mathrm{T}}, \mathbf{\Delta y}_t^{\mathrm{T}}]^{\mathrm{T}}$ とすると，GMM は以下の式で表される．

© 2017 Information Processing Society of Japan

$$P(\mathbf{X}_t, \mathbf{Y}_t|\lambda) = \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\begin{bmatrix} \mathbf{X}_t \\ \mathbf{Y}_t \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_m^{(X)} \\ \boldsymbol{\mu}_m^{(Y)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_m^{(XX)} & \boldsymbol{\Sigma}_m^{(XY)} \\ \boldsymbol{\Sigma}_m^{(YX)} & \boldsymbol{\Sigma}_m^{(YY)} \end{bmatrix} \right) \quad (3)$$

ここで，$\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ は平均ベクトル $\boldsymbol{\mu}$，共分散行列 $\boldsymbol{\Sigma}$ をもつ正規分布を，M は GMM の混合数を，$m(1 \leq m \leq M)$ は分布番号を，α_m は各分布の混合重みを表す．また，$\boldsymbol{\lambda} = \{\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m, \alpha_m | m = 1, \ldots, M\}$ である．動的特徴量を考慮することによって，フレーム間の相関を考慮してパラメータ推定を行うことができる．また，一般には，入力音声と目標音声の特徴量の系列長は異なるため，動的時間伸縮 (DTW: Dynamic Time Warping) を行い，系列同士を対応させる必要がある．

変換処理の過程を図 4 に示す．変換処理では，動的特徴量を考慮した最尤系列変換法 [5] によって，入力音声の特徴量を目標音声の特徴量に変換する．入力音声と目標音声の特徴量の系列ベクトルをそれぞれ $\mathbf{X} = [\mathbf{X}_1^{\mathrm{T}}, \ldots, \mathbf{X}_T^{\mathrm{T}}]$, $\mathbf{Y} = [\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_T^{\mathrm{T}}]$ とすると，変換される静的特徴量系列 $\hat{\mathbf{y}} = [\hat{\mathbf{y}}_1^{\mathrm{T}}, \ldots, \hat{\mathbf{y}}_T^{\mathrm{T}}]$ は次式で表される．

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} P(\mathbf{Y}|\mathbf{X}, \boldsymbol{\lambda}) \text{ subject to} \mathbf{Y} = \mathbf{W}\mathbf{y} \quad (4)$$

ここで，$\mathbf{W}$ は静的特徴量ベクトルを静的・動的特徴量ベクトルに拡張する行列である．なお，過剰な平滑化による変換歌声の音質劣化を緩和するため，系列内変動 (GV: Global Variance)[5] を考慮する．

4. 電気音声変換法

本章では，電気音声の音高の制御と声質強調によって，電気音声をより人間らしい歌声に変換する手法を提案し，各手法の内容について述べる．

4.1 音高制御処理

電気音声をより人間らしく歌っているような歌声に変換するために、電気音声に対して歌声特有の音響特徴を付与する．はじめに，楽器音から f_{o} を抽出し，(1) 式で表される歌声特有の特徴を付与するフィルタを抽出した f_{o} の系列に適用することで，歌声特有の f_{o} パターンを得る．次に，歌声特有の f_{o} パターンと声質強調処理において得られるスペクトル包絡成分と非周期性指標をパラメータとして合成処理を行うことで，歌声特有の音響特徴を付与した歌声が得られる．

4.2 声質制御処理

本研究では，電気音声の声質制御処理として，スペクトル減算法と統計的声質変換の 2 種類の手法を個別に適用する．

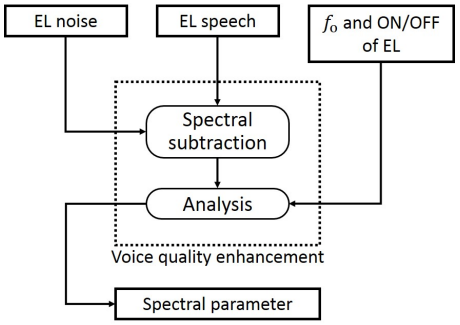


Fig 5 Process of voice quality enhancement based on spectral subtraction.

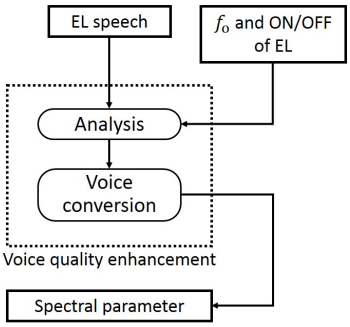


Fig 6 Process of voice quality enhancement based on voice conversion.

4.2.1 スペクトル減算法に基づく声質制御

図 5 にスペクトル減算法に基づく声質制御の処理の過程を示す．この手法では，電気音声に含まれる電気式人工喉頭から漏れ出した雑音信号を抑圧することで，声質をより人間に近いものにすることを目指す．雑音信号はあらかじめ電気式人工喉頭を用いて収録する必要がある．雑音抑圧を行った後，電気音声に対して音声分析処理を適用することで，電気音声のスペクトル包絡成分と非周期性指標を得る．しかし，一般的な音声分析の手法では，電気音声を所望のパラメータに分析できない場合がある．その理由は，電気音声の基本周波数推定および有声無声判定が正しく行われないためである．しかしながら，電気音声の f_{o} はほぼ一定であると仮定すれば，あらかじめ与えることが可能である．また，電気音声の有声無声判定情報は電気式人工喉頭の ON/OFF 情報を用いることで対処することができる．

4.2.2 統計的声質変換に基づく声質制御

図 6 に統計的声質変換に基づく声質制御の処理の過程を示す．この手法では，統計的声質変換によって電気音声の声質を通常の発声の声質に変換することで，声質を寄り人間に近いものにすることを目指す．統計的声質変換を行うために，電気音声と通常の発声による音声のパラレルデータを用意し，GMM の学習を行う必要がある．GMM の学習後，電気音声に対して音声分析を適用し，得られた特徴量を学習した GMM を用いて変換を行う．電気音声に対する音声分析の問題点は先と同様の方法で回避することがで

© 2017 Information Processing Society of Japan

表 1　Parameter settings of f_o control

	Duration [ms]	ω	ζ	k
Overshoot	300	34.8	0.403	34.8
Preparation	100	34.8	0.603	34.8
Vibrato	-	34.8	0	0.0164

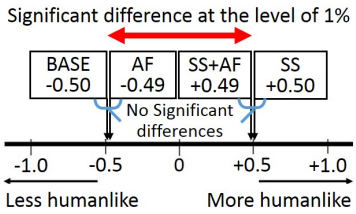


Fig 7　Result of subjective evaluation to investigate the effectiveness of the proposed singing aid system using spectral subtraction.

きる．

5.　実験的評価

前章で提案した各変換手法に対する歌声の人間らしさの差異を評価するために主観評価実験を行った．本章では，実験内容および実験結果について述べる．

5.1　スペクトル減算法に基づく電気音声変換法の評価

提案法の有効性を調査するために，健常者による電気音声による歌唱と MIDI による音高制御を用いた模擬システムを用いて，シェッフェの一対比較法による評価を行った．20 代男性 10 人を被験者とし，日本語歌唱 6 フレーズに対する電気音声を以下の 4 種の手法によって変換した歌声を実験刺激とした．(1) 式で表される歌声特有の特徴を付与するフィルタのパラメータは表 1 のように設定した．スペクトル減算法を適用した後の分析合成処理には WORLD[15] を使用し，スペクトル包絡成分を表す特徴量は，FFT 点数 512 のスペクトル包絡とした．

BASE：　音高情報のみ付与
AF：　歌声特有の特徴の付与（提案法）
SS：　雑音抑圧，音高情報の付与（提案法）
SS+AF：　雑音抑圧，歌声特有の特徴の付与（提案法）

図 7 に実験結果を示す．実験結果から，1）電気音声に対する雑音抑圧手法の導入により，歌声の人間らしさが大きく改善されること，2）歌声特有の特徴付与手法に関しては，話し声からの変換とは大きく異なり，電気音声においては歌声の人間らしさの改善が得られないことがわかった．この結果から，電気音声による歌唱においては，通常音声とは異なる変換技術が必要であることがわかった．

5.2　統計的声質変換に基づく電気音声変換法の評価

提案法の有効性を調査するために，先と同様の実験を行った．ただし，被験者は 20 代男女 12 人とし日本語歌唱 3 フレーズに対する電気音声を以下の 4 種の手法によって

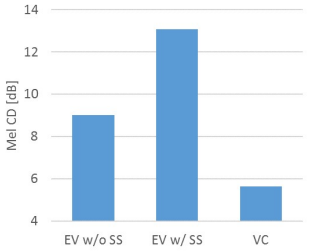


Fig 8　Comparison of mel-cepstrum distortion.

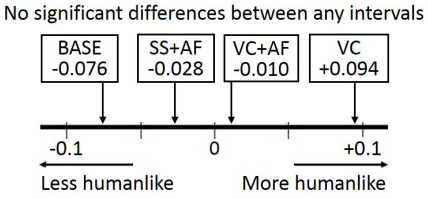


Fig 9　Result of subjective evaluation to investigate the effectiveness of the proposed singing aid system using statistical voice conversion.

変換した歌声を実験刺激とした．(1) 式で表される歌声特有の特徴を付与するフィルタのパラメータは先と同様に設定した．統計的声質変換に関しては，学習用のパラレルデータとして，日本の童謡 16 曲を健常者による電気式人工喉頭を用いた発声法と通常の発声法によって収録したものを使用し，音声分析合成系として STRAIGHT[16] を使用した．また，スペクトル包絡成分を示す特徴量として 24 次のメルケプストラムを使用し，声質変換のために使用する GMM の混合数は，メルケプストラムに対しては 64，非周期性指標に対しては 16 とした．

BASE：　音高情報のみ付与
SS+AF：　雑音抑圧，歌声特有の特徴の付与（提案法）
VC：　統計的声質変換，音高情報の付与（提案法）
VC+AF：　統計的声質変換，歌声特有の特徴の付与（提案法）

また，電気音声を通常の歌声の声質に変換するときの精度を評価するために，以下で計算されるメルケプストラム歪みを用いた．

$$\text{Mel CD [dB]} = \frac{10}{\ln 10}\sqrt{2\sum_{p=1}^{P}(m_p^{(x)} - m_p^{(y)})^2} \qquad (5)$$

ここで，$m_p^{(x)}, m_p^{(y)}$ は比較対象となる音声の p 次のメルケプストラム係数を示す．メルケプストラム歪みの値が小さいほど，比較対象の声質はより近いものになる．

図 8 にメルケプストラム歪みによる客観評価の結果を示す．左側から順にそれぞれ通常の歌声と電気音声との間のメルケプストラ歪みの値，スペクトル減算法を適用した電気音声との間のメルケプストラ歪みの値，声質変換を行った歌声との間のメルケプストラムひずみの値を表す．客観評価の結果から，変換元である電気音声と変換音声におけ

© 2017 Information Processing Society of Japan

るメルケプストラ歪みの値の差が約 3 dB となり，電気音声の声質が通常の歌声のものに近づいたことがわかった．

図 9 に主観評価実験の結果を示す．実験結果から，各手法間で有意差が確認されず，声質変換によって必ずしも歌声の人間らしさが向上するとは限らないということがわかった．

5.3 提案法の課題点

以上の実験結果から，歌声特有の音響特徴の付与が効果的でないということがわかった．本実験では，歌声特有の音響特徴の付与に関するパラメータを恣意的に設定したため，統計的手法の導入などによって，データから最適な f_0 パターンの制御法を求める必要がある．

声質変換に関しては，電気音声の有声・無声子音の誤判別によって，不自然な音声に変換されてしまうため，電気音声の音源と変換スペクトル間において，有声無声情報の不一致が生じるため，不自然な音声に変換される傾向にある．そのため，有声無声情報の取り扱いを考慮した電気音声変換法が必要となる．また，本実験で用いた統計的声質変換では，メルケプストラムによるスペクトル表現や，ボコーダに基づく音源信号生成を用いているため，リフタリングや音源モデリングにおける近似の影響により，通常の歌声よりも音質が劣化するという問題がある．音質の劣化を防ぐためには，差分スペクトル補正に基づく声質変換法 [17] や，より高精度な音源モデルの導入が必要である．

6. おわりに

本稿では，喉頭摘出者のための歌唱支援技術の構築を目指し，電気音声をより人間らしい歌声に変換するための手法を提案し，実験的評価を通して比較を行った．実験的評価の結果から，電気音声に対する雑音抑圧手法の導入により，歌声の人間らしさが大きく改善されたが，電気音声に対する歌声特有の音響特徴の付与や声質変換による人間らしさの向上は必ずしも起こりえないということを確認した．今後は，歌声の人間らしさに関する調査をさらに進め，実際に利用することのできる歌唱支援システムの構築を行う予定である．

謝辞 本研究の一部は，JST，さきがけ JPMJPR1657 により実施したものである．

参考文献

[1] A. Hisada, H. Sawada, “Real-time clarification of esophageal speech using a comb filter,” *Proc. ICD-VRAT*, pp. 39–46, 2002

[2] K. Matsui, N. Hara, N. Kobayashi, H. Hirose, “Enhancement of esophageal speech using formant synthesis,” *Proc. ICASSP*, pp. 1831–1834, 1999.

[3] H. R. Sharifzadeh, I. V. McLoughlin, F. Ahmadi, “Reconstruction of normal sounding speech for laryngectomy patients through a modified CELP codec,” *IEEE Trans. Biomedical Engineering*, Vol. 57, No. 10, pp. 2448–2458, 2010.

[4] H. Liu, Q. Zhao, M. Wan, S. Wang, “Enhancement of electrolarynx speech based on auditory masking,” *IEEE Trans. Biomedical Engineering*, Vol. 53, No. 5, pp. 865–874, 2006.

[5] T. Toda, A. W. Black, and K. Tokuda, “Voice conversion based on maximum likelihood estimation of spectral parameter trajectory,” *IEEE Trans. Audio, Speech, Language Processing*, Vol. 15, No. 8, pp. 2222–2235, 2007.

[6] K. Nakamura, T. Toda, H. Saruwatari, K. Shikano, “Speaking-aid systems using GMM-based voice conversion for electrolaryngeal speech,” *Speech Communication*, Vol. 54, No. 1, pp. 134–146, 2012.

[7] K. Tanaka, T. Toda, G. Neubig, S. Sakti, S. Nakamura, “Direct F0 Control of an Electrolarynx based on Statistical Excitation Feature Prediction and its Evaluation through Simulation,” *INTERSPEECH*, pp.31–35, 2014.

[8] T. Toda, Y. Ohtani, K. Shikano, “One-to-many and many-to-one voice conversion based on eigenvoices.” *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on* Vol. 4, IEEE, 2007.

[9] H. Doi, T. Toda, K. Nakamura, H. Saruwatari, K. Shikano, “Alaryngeal Speech Enhancement Based on One-to-Many Eigenvoice Conversion,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 1, pp. 172–183, 2014.

[10] T. Saitou, M. Unoki, M. Akagi, “Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis,” *Speech Communication*, Vol.46, N0.3 pp.405–417, 2005.

[11] T. Saitou, M. Goto, M. Unoki, M. Akagi, “Speech-to-Singing Synthesis System: Vocal Conversion from Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices,” *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp.215–218, 2007.

[12] K. Tanaka, T. Toda, G. Neubig, S. Sakti, S. Nakamura, “A Hybrid Approach to Electrolaryngeal Speach Enhancement Based on Noise Reduction and Sattistical Excitation Generation,” *IEICE Transactions on Information and Systems*, Vol.97, No.6, pp.1429–1437, 2014.

[13] J. スンドベリ, “歌声の科学,” 東京電機大学出版局, pp.117–119, 2007

[14] S.F.Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” *IEEE Trans. Acoust, Speech, Signal Processing*, Vol.27, No.2, pp.113–120, 1979.

[15] M. Morise, F. Yokomori, and K. Ozawa, “WORLD: a vocoder-based high-quality speech synthesis system for real-time applications,” *IEICE transactions on information and systems*, Vol. E99-D, No. 7, pp. 1877–1884, 2016.

[16] H. Kawahara, I. Masuda-Katsuse, A. Cheveigné, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds” *Speech Communication*, Vol.27, No.3, pp.187–207, 1999.

[17] K. Kobayashi, T. Toda, G. Neubig, S. Sakti, S. Nakamura, “Statistical singing voice conversion with direct waveform modification based on the spectrum differential,” *INTERSPEECH*, pp. 2514–2518, 2014

© 2017 Information Processing Society of Japan