入力情報として複数音声モーフィングを利用した自分の音声を好みの音声へと変換する手法

石黒　徹　　高橋　沙知　　宮崎　敬仁　　濱川　礼
中京大学　情報理工学部　情報システム工学科

1. はじめに

本論文では，入力情報として複数音声モーフィングを用いて自分の音声を好みの音声へと変換する手法について述べる．本論文での好みの音声とは，音声変換を行う際にユーザが期待する変換後の音声のことを指す．

近年インターネット上では動画共有サイトに注目が集まっている．投稿動画の中にはボイスチェンジャを用いてエンタテインメント性を高めている動画もあり人気を集めている．しかし，既存のボイスチェンジャでは好みの音声へ直感的に音声変換を行うことが困難である．そこで，本手法を考案しシステムに実装した．

2. 特徴

本手法では，複数人の音声を混ぜ合せオリジナルの音声を作成し，その音声に利用者の音声を似せるように音声変換を行う．既存のボイスチェンジャでは，2次元の座標から数値を取得し，その数値を用いて音声変換を行う手法が採られている[1]．そのため，変換後の音声を直感的に想像することが困難である．本手法では，変換に用いる数値を手動では無く，音声データの比較から自動的に取得しているため，既存の手法より直感的に音声変換を行うことが可能になる．

3. システム構成

実装したシステムは，「音声分析」，「音声混合」，「音声比較」，「音声変換」から構成されている[図 1]．「音声分析」と「音声変換」の機能は『音声分析変換合成法 STRAIGHT』[2]を用いて実装している．

Method for converting user's voice into his/her favorite voices using voices morphing mechanism
Toru Ishikuro,Sachi Takahashi,Takahito Miyazaki and Rei Hamakawa
Chukyo University Department of Information

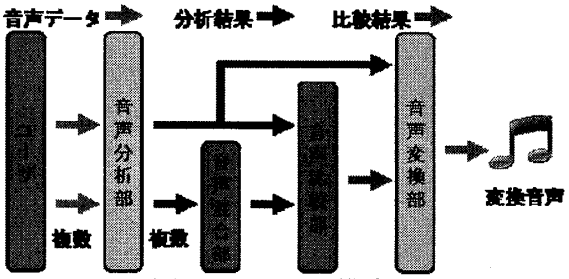

図1　システム構成図

4. 音声混合部

音声混合部では，音声モーフィングの手法を利用し，複数の音声の混合音声を生成している．音声モーフィングとは，ある話者の音声を徐々に他の話者の音声へと変化させることである．[2]音声モーフィングは1対1での音声変換である．これを3つ以上の音声に拡張し，より自分好みの音声が生成出来るようにした．

混合された音声の中にどの音声の特徴がどれだけ混ざっているかの比率を表す混合率を設定する．

$$\sum_{i=0}^{N-1} r(i) = 1.0$$

r(i)：混合率，N:音声数

また，混合音声を設定した混合率で生成するために以下の式を用いる．m(x)はモーフィング率を表す．モーフィング率とは，ユーザが最初に設定した混合率を崩さず，1対1でのモーフィングで使えるように正規化を行った値である．

最初の音声モーフィング

$$m(1) = \sum_{i=1}^{2}(r(i)/\sum_{j=1}^{2} r(j))$$

2回目以降の音声モーフィング

$$m(x) = \sum_{i=1}^{n}(r(i)+r(n-1))/\sum_{j=1}^{n+1} r(j)$$

m(x)：モーフィング率，n:モーフィング数

音声モーフィングでは，2 つの音声の基本周波数とスペクトルから中間音声を生成する．

中間音声の基本周波数とスペクトルはモーフィング率を基に、2 つの音声を単位時間ごとに区切って算出する．計算式は以下の通りとなっている．

基本周波数 = $f(1)^{m(1)} \times f(2)^{m(2)}$

スペクトル = $\sum_{k=0}^{n} (s(1)^{m(1)} \times s(2)^{m(2)})$

f(1)，f(2)：2 つの音声の基本周波数
s(1)，s(2)：2 つの音声のスペクトルの内容
m(1)，m(2)：モーフィング率
n：中間音声の周波数軸のデータ数

最終的に出力される音声が，複数の音声の混合音声である．

5. 音声比較部

音声比較部では，音声混合部から出力された混合音声とユーザが入力したユーザ音声との比較を行い，それぞれの基本周波数とスペクトルの変換倍率を求める．音声変換部では音声比較部で作成した変換倍率を受け取ることで，ユーザ音声を混合音声に近付けることが出来る．

音声を分析部で単位時間ごとに区切り分析を行い，区間ごとのデータから基本周波数の抽出を行う．抽出した複数の基本周波数の中から，最も出現頻度の高い周波数の値を音声の基本周波数として決定する．混合音声の基本周波数とユーザ音声の基本周波数を決定する．混合音声の基本周波数をユーザの基本周波数で除算することで基本周波数の変換倍率を算出する．

同様にスペクトルの変換倍率も，単位時間毎に分けられたデータから複数のスペクトルの抽出を行い，単位時間のスペクトルの平均値を算出する．求めた平均値から単位時間に存在する混合音声のスペクトル m(n)とユーザ音声のスペクトル u(n)から，以下の式で各値の倍率 A(n)を求める．

$$A(n) = \frac{1}{N} \sum_{n=0}^{N-1} \{ \frac{m(n)}{u(n)} \} \ (0 \le n \le N-1)$$

$m(n)$：混合音声のスペクトル
$u(n)$：ユーザのスペクトル
N：単位時間毎に存在するサンプル数

上記の式から求めた単位時間毎の倍率の平均を求めることでスペクトルの変換倍率を取得する．

このように求めた基本周波数とスペクトルの変換倍率を STRAIGHT の音声変換部に出力することにより，ユーザ音声を混合音声に近付けた変換音声を作成する．

6. ユーザインタフェース

[図 2]に本システムのユーザインタフェースを示す．

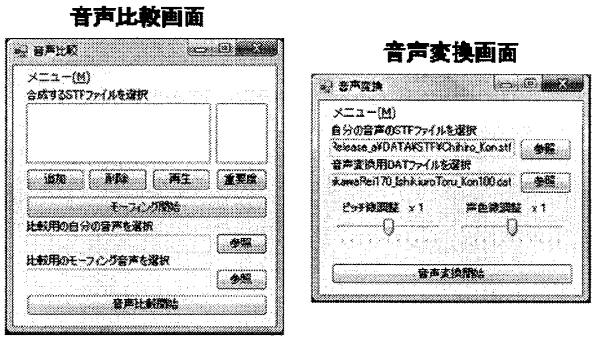

図 2　音声比較画面・音声変換画面

音声比較画面で複数の音声ファイルを混ぜ，ユーザの音声と比較し，音声変換画面で比較結果を元に音声の変換を行う．

7. まとめと今後の展望

ユーザ 17 名を対象に本システムと座標を用いて音声変換するシステムで比較評価を 5 段階で行った結果，「直感的な変換が出来たか」という項目で従来システムより平均点が 1 ポイント増加した．このことから，「直感的な音声変換を可能にする」という目的を達成することが出来たといえる．

今後の展望として，リアルタイム音声変換を実装することで，ボイスチャッでの利用などが考えられる．

参考文献

[1] STRAIGHT Voice Changer
http://w3voice.jp/straight/
[2] 音声分析変換合成法 STRAIGHT
http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_j.html
[3] 阿部 匡伸 「基本周波数とスペクトルの漸次変形による音声モーフィング」