

ケプストラムの幾何学的特性に着眼した声質変換に関する実験的検討*

◎齋藤大輔, 峯松信明, 広瀬啓吉 (東大)

1 はじめに

音声に含まれる情報は多様であり、音韻のみならず、年齢や性別、声道長や音響機器などの諸要因により変化に富んでいる。音声の持つ情報は大きく分けて、言語的情報、非言語的情報、パラ言語的情報に分類する事ができる。通常、音声を用いたアプリケーションではこれらのうち一つに着眼して情報を抽出する。例えば音声認識では言語的情報を抽出し、話者認識や話者同定では、非言語的情報に着目する。このような着目する情報を取り出す際、その他の情報が多様に变化した学習データを用いて、着目する情報以外を平均化することでモデルを構築する事が広く行われている。不特定話者音響モデルや GMM による話者モデルなどがこれに当たる。しかし子供に代表されるような特異な音声に対してはこれらのモデルが有効に機能しないことがある [1]。これらはミスマッチ問題と呼ばれ、統計的手法における問題の一つと考えられる。すなわち特徴ベクトルの性質は音声の多様な情報の影響によって容易に変化し、大量データによる統計的手法が必ずしも多様な変化を適切にモデル化している訳ではないと言える。そのため音声に含まれる各種の情報が特徴ベクトルにどのような影響を与えるのかを分析する必要がある。

音声認識におけるミスマッチ問題への対応として、特に声道長正規化 (VTLN) がよく用いられる。VTLN は音声に含まれる非言語的情報のうち、特に声道長の違いを取り除くために用いられる [2]。VTLN ではケプストラム空間における変換行列を推定する事で入力話者を標準声道長の話者の特徴へと変換する。

筆者らはこの声道長変化を表す変換行列に着目し、その幾何学的性質を理論的および実験的に検討してきた [3, 4, 5]。声道長変化を表す変換行列は強い回転性を有し、ケプストラムの方向成分は声道長が変化する事でおよそ同様に回転する。さらにこの回転性を詳細に分析し、2次元部分空間における回転の組み合わせとして記述できることを示している [6, 7]。

本稿では声道長変換行列の回転を表すパラメータが音声のどのような性質に対応するかを調べるため、回転パラメータを操作して分析再合成音を作成し、その分析を行う。これはケプストラム空間における幾何学的特性を変化させる事で声質変換を実現する枠組みといえる。

2 声道長変化の定式化

話者の声道長の変化は、音声のスペクトル表現における周波数ウォーピングとして考えることができる。今、周波数ウォーピングにおける変換前後の正規化角周波数を $\omega, \hat{\omega}$ ($0 \leq \omega, \hat{\omega} \leq \pi$) とする。このとき $z = e^{j\omega}$, $\hat{z} = e^{j\hat{\omega}}$ として、周波数ウォーピングとして以下の1次全域通過関数を考える。

$$\hat{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (1)$$

このとき α は $|\alpha| < 1$ の実数であり、 $\alpha < 0$ の場合、周波数軸が低域に変換され声道長は長くなる。一方 $\alpha > 0$ の場合、周波数軸は高域に変換され声道長が短くなる。以後 α をウォーピングパラメータと呼ぶ。

江森らは上記の周波数ウォーピングをもとに、そのケプストラム空間における記述を導出している [8]。パワーを表現するケプストラムの0次項 ($c_0, \hat{c}_0$) を考慮しない場合、周波数ウォーピングは以下の式で表現される。

$$\hat{c} = A c \quad (2)$$

$$A = \begin{pmatrix} 1-\alpha^2 & 2\alpha-2\alpha^3 & \cdots & \cdots \\ -\alpha+\alpha^3 & 1-4\alpha^2+3\alpha^4 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (3)$$

さらに Pitz らによれば、式 (3) における行列 A の要素 a_{ij} はウォーピングパラメータ α を用いて以下のよう表せる [9]。

$$a_{ij} = \frac{1}{(j-1)!} \sum_{m=\max(0, j-i)}^j \binom{j}{m} \times \frac{(m+i-1)!}{(m+i-j)!} (-1)^{(m+i-j)} \alpha^{(2m+i-j)} \quad (4)$$

ただし

$$\binom{j}{m} = \begin{cases} jC_m & (j \geq m) \\ 0 & (j < m) \end{cases} \quad (5)$$

とする。

* Experimental examination of voice conversion based on geometrical properties of a cepstrum space.
by SAITO Daisuke, MINEMATSU Nobuaki, and HIROSE Keikichi (Univ. of Tokyo)

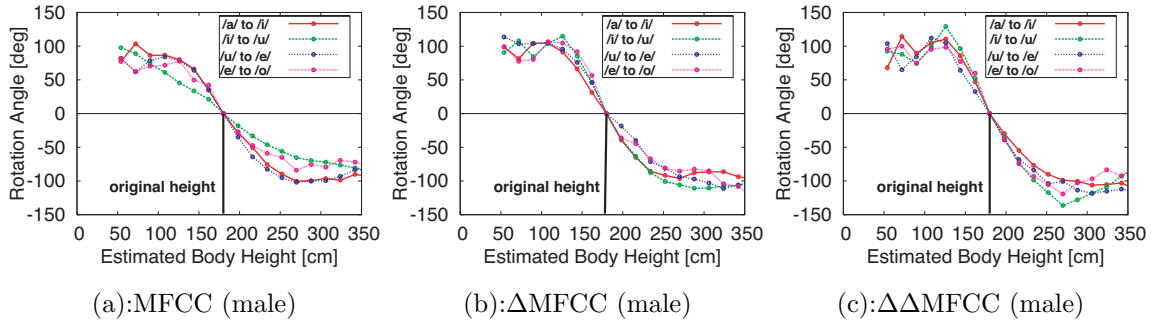


Fig. 1 男性話者（身長 180 cm）における分析再合成音の推定身長と回転角との対応関係

3 ケプストラム空間における回転性

筆者らは多次元空間での幾何学を通して、式 (3) の性質を考察してきた [4, 5]. n 次元空間において一般に行列 R が以下の性質を満たすとき、 R を回転行列と呼ぶ.

$$R^t R = R R^t = I \quad (6)$$

$$\det R = +1 \quad (7)$$

先行研究では式 (3) が近似的に式 (6)(7) を満たすことを理論的に示している. 加えて日本語 5 母音発声 /aieuo/ について STRAIGHT[10] を用いて人工的に声道長の長さを変化させた再合成音を用いて、その音声の MFCC パラメータの回転性を実験的に示している. Fig. 1 は二つのベクトルの角度を式 (8) で定義し、母音遷移の中心部分における声道長変換前後の角度と声道長（身長）との対応関係を示したものである.

$$\theta = \arccos \frac{a \cdot b}{|a||b|} \quad (8)$$

Fig. 1 より音声のパラメータベクトルの方向が声道長変化に依存して回転している事が分かる.

4 固有値による回転性の記述

4.1 回転行列の対角化

行列 A の回転性を詳細に記述するため、行列の固有値に基づく解析を考える. 以下 n 次元の回転行列 R_n を用いて議論を進める. R_n は n 次のユニタリ行列 U_n を用いて複素数を要素に持つ対角行列 D_n の形に対角化できる.

$$R_n = U_n D_n U_n^\dagger \quad (9)$$

一方 2 次元の回転行列 R_2 は一つの回転パラメータ θ を用いて以下のように表される.

$$R_2(\theta) = U_2 D_2(\theta) U_2^\dagger \quad (10)$$

ここで

$$R_2(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (11)$$

$$D_2(\theta) = \begin{pmatrix} e^{j\theta} & 0 \\ 0 & e^{-j\theta} \end{pmatrix} \quad (12)$$

R_n の固有方程式は $n=2m$ ($m \in \mathcal{N}$) の時、絶対値 1 の m 組の複素共役解を持ち、 $n=2m+1$ の時、1 と m 組の複素共役解を持つ. 故に式 (9) の D_n は m 次の回転パラメータベクトル $\Theta = (\theta_i | i=1, 2, \dots, m)$ を用いて以下のように表現できる.

$$D_n(\Theta) = \begin{cases} \begin{pmatrix} D_2(\theta_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_2(\theta_m) \end{pmatrix} & (n : \text{even}) \\ \begin{pmatrix} 1 & \cdots & 0 \\ & D_2(\theta_1) & \\ \vdots & & \ddots \\ 0 & \cdots & D_2(\theta_m) \end{pmatrix} & (n : \text{odd}) \end{cases} \quad (13)$$

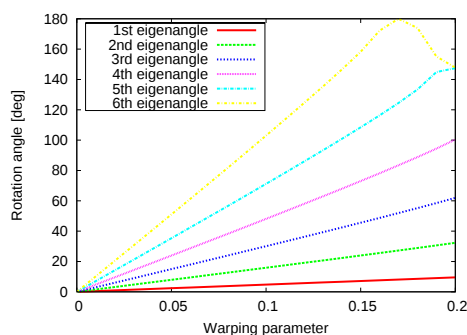
さらに式 (10) を用いることで R_n は実数の回転行列のみを用いて以下のように表現することができる [11].

$$R_n(\Theta) = R_n'' R_n'(\Theta) R_n'^t \quad (14)$$

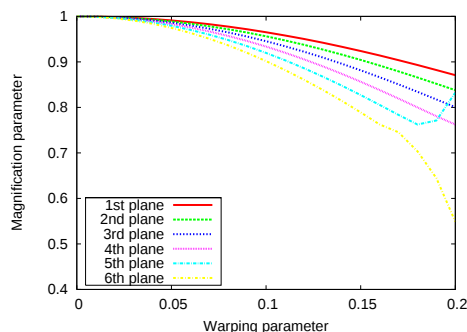
ここで、 $n=2m+1$ のとき、 $R_n'(\Theta)$ は以下のようになり、 $n=2m$ では式 (15) の 1 行目、1 列目を削った形となる.

$$R_n'(\Theta) = \begin{pmatrix} 1 & \cdots & 0 \\ & R_2(\theta_1) & \\ \vdots & & \ddots \\ 0 & \cdots & R_2(\theta_m) \end{pmatrix} \quad (15)$$

このとき R_n'' は式 (9) におけるユニタリ行列と対応する実回転行列であり、回転行列 R_n の固有ベクトルと



(a):固有回転角と α の対応関係



(b):縮小率と α の対応関係

Fig. 2 各回転パラメータと α の関係

等価の情報を持つ。これは回転行列 $R_n(\theta)$ が持つ固有姿勢へ補正する回転行列と考える事ができる。 R_n^t はその逆操作である。

さらに回転性に拡大縮小を伴う場合を考える。このような変換において軌跡は渦を描く。このとき回転角 θ_i に対応する拡大縮小成分を r_i として式 (15) における $R_2(\theta_i)$ を $r_i R_2(\theta_i)$ に置き換えればよい。

4.2 声道長変換行列の分解

前節における議論から、 n 次元の声道長変換行列は以下の 3 つの要素に分解して考える事ができる。ここで $n = 2m$ とする。

1. 回転行列の固有姿勢を表す行列 R_n^t
2. 回転行列の固有値に基づく回転パラメータベクトル $\Theta = (\theta_i | i = 1, 2, \dots, m)$
3. 各回転角 θ_i に対応する拡大縮小パラメータベクトル $\mathbf{r} = (r_i | i = 1, 2, \dots, m)$

これらの要素と声道長変換行列 A のウォーピングパラメータ α との対応について考える。先行研究において声道長変換行列の固有姿勢は α に対する依存性が低い事を実験的に示している [6]。ここでは回転パラメータベクトルおよび拡大縮小パラメータベクトルと α との対応関係について示す。

今、固有値分析との対応から θ_i を回転角の小さいものから順に第 1 固有角、第 2 固有角と呼ぶ事にす

る。さらにこれらの固有角に対応する 2 次元の部分空間を第 1 回転平面、第 2 回転平面と呼ぶ。固有姿勢の α に対する依存性が低いことから、回転平面は α によらず一定であると仮定する。各回転平面における固有角と拡大縮小性について、 α との対応を示したものが Fig. 2 である。Fig. 2 によれば、 α の変化に応じて各回転平面上でケプストラムベクトルが連続的に回転し、さらにその絶対値成分が縮小していることがわかる。一方、実際の音声の変化は、これらのパラメータを回転平面毎に操作する事でより多様な声質変換のバリエーションを作成できると考えられる。現在検討している回転性は声道長変化に着眼しており、音韻への影響は少ないと考えられる。

5 実験

5.1 実験方法

実際に回転行列の各要素を変化させた場合に、音声に与える影響について調べるため、実験を行った。実験音声として成人男女 2 名ずつの日本語 5 母音の連続発声 /aiueo/ を収録した。このときサンプリング周波数 16 kHz、量子化は 16 bit とした。次にこれらの音声について STRAIGHT[10] を用いて、その基本周波数、スペクトルについて分析した。得られたスペクトルについて、ケプストラムを求め、その低次の 12 次元の係数を抽出した。一方で $\alpha = 0.05, 0.1, 0.15$ の場合について式 (3) によって 12 次元の声道長変換行列 A_α を求めた。さらにそれぞれの固有姿勢、固有回転角、拡大縮小パラメータについて算出した。

4.2 で示した 3 つの要素がそれぞれどのように音声を変化させるかを調べるために、上記の声道長変換行列を基に以下のような操作を行い、新たに変換行列を作成した。

固有姿勢の変化 それぞれの A_α における固有姿勢を、 α の異なる A_α における姿勢、および単位行列に置き換えて行列を作成した。本稿では各 α について 3 種類ずつ検討した。固有姿勢を単位行列とした場合、元のケプストラム空間を低次から 2 次元ずつの部分空間に分割して回転させることに相当する。

拡大縮小パラメータの正規化 Fig. 2 において A における拡大縮小パラメータは 1 より小さくなっている。それぞれの A_α についてこれらの値を全て 1 に置き換えて行列を作成した。

固有回転角および拡大縮小パラメータ それぞれの A_α における固有回転角および拡大縮小パラメータについて、各々の回転平面について α の異な

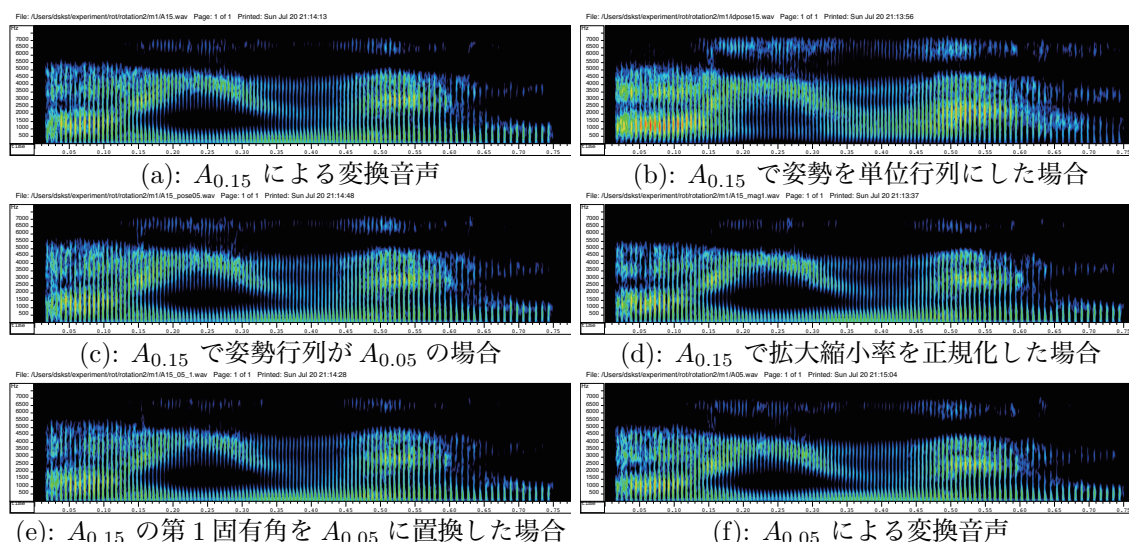


Fig. 3 合成音声の一例

る A_α で算出した固有回転角に置き換えて行列を作成した。本稿では一つだけ固有回転角を置換し、各 α について12種類を検討した。

上述の操作に基づいて各 α 毎に17種類、計51種類の行列を作成し実際に変換を行い、STRAIGHTを用いて合成音を作成した。この際、F0についてはすべて原音声の1.2倍とした。

5.2 実験結果

成人男性話者1名について、合成音声の結果の一例をFig. 3に示す。Fig. 3において(a)から(c)はその姿勢行列が異なっている。(a)と(b)を比較すると、低域のスペクトルに大きな変化があり、聴感上も変化が見られた。一方(a)と(c)ではスペクトル上および聴感上でも大きな変化はなかった。これは固有姿勢の α に対する依存性が低いことを示している。

(a)と(d)では、拡大縮小パラメータが異なっている。スペクトル上はあまり明確な差が見られないが、聴感上は(d)の方がより明瞭度が高い傾向があった。

一方(a)と(e)は、第1固有角およびその拡大縮小パラメータを変化させたものである。(e)はスペクトル上では、あまり明確な差は見られなかった。一方聴感上では、(a)と(f)の中間的な音声として聴取された。これは音声パラメータの個々の回転平面への射影量が大きく寄与していると考えられる[6]。すなわち音声に含まれる第1回転平面の成分が大きく、その変化が、声道長の変化を表す特徴として大きく寄与する事を示唆しているといえる。また全ての音声において音韻性は保存されていた。このことは本稿で提案する行列の分解が、声道長の変化を適切に捉えている事を示唆している。

6 おわりに

本稿では、ケプストラムの幾何学的特性に着眼した声質変換について検討を行った。筆者らはこれまでに声道長変換を表す行列がケプストラム空間において強い回転性を有する事に言及し、固有値分析を通してその詳細な回転性について分析してきた。今回、分解された個々の要素を変化させ、分析再合成を行う事で、幾何特性に着眼した声質変換を検討した。今回は、比較的固定されたパラメータによって実験を行ったが、今後、分解した要素を高い自由度で変化させた場合の影響の調査が必要になってくる。また大規模な聴取実験についても検討していく。回転パラメータに基づく分解は、球面線形補間などで柔軟な合成を実現できる可能性がある[11]。ケプストラムの幾何学的性質をうまく捉える事で話者情報を表現する手法を検討していく予定である。

参考文献

- [1] M. Russel, S. D'Arcy, SLATE2007, 2007.
- [2] E. Eide, H. Gish, ICASSP'96, vol. 1, pp. 346–348, 1996.
- [3] 齋藤他, 音講論 (秋), 3-3-13, 2007.
- [4] 齋藤他, 信学技報, SP2007-128, pp.189–194, 2007.
- [5] D. Saito et al., ICASSP2008, pp.4485–4488, 2008.
- [6] 齋藤他, 音講論 (春), 1-Q-19, 2008.
- [7] D. Saito et al., INTERSPEECH2008, 2008 (to appear).
- [8] 江森, 篠田, 電子情報通信学会論文誌 D-II, vol.J83-D-II, No.11, pp.2108–2117, 2000.
- [9] M. Pitz, H. Ney, IEEE Trans. Speech and Audio Processing, vol.13, pp.930–944, 2005.
- [10] H. Kawahara et al., Speech Communication, 27, 3-4, pp.187–207, 1999.
- [11] 高橋他, “高次元回転行列の補間とその応用” Visual Computing / グラフィックスと CAD 合同シンポジウム, 2007.