

ケプストラムの声道長依存性に対する定量的解析とその応用*

齋藤大輔[†], 松浦良[†], 鎌田敏明^{†‡}, 朝川智[†], 峯松信明[†], 広瀬啓吉[†] ([†]東大, [‡]科警研)

1 はじめに

音声は音韻等の言語的特徴に加えて、年齢、性別、声道長や音響機器などの非言語的特徴を含んでいる。言語的特徴のみを精確に捉えることを音声認識システムの命題とした時、これらの非言語的特徴は音声に歪みを与えていることになる。このような音声の多様な歪みに対応するため、大規模な学習話者による学習データを用いて、個々の音韻に対応する音響モデルを構築することが一般に広く行われている。これらは話者非依存 (Speaker Independent: SI) の音声認識システムと呼ばれている。一方、子供の音声に代表されるような特異な音声に対しては、話者非依存のシステムは時に全く性能を発揮できないことがある [1]。このような問題はミスマッチ問題と呼ばれ、この意味において話者非依存のシステムは“真に話者非依存”ではないといえる。

上記のような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性歪みのうち、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およそ同様に回転する。しかしながら先行研究の結果では、少なからず音韻や性別によって回転の様子が異なっていた。本稿では声道長変化を表す変換行列の固有値に着目し、その幾何学的性質を定量的かつ詳細に記述する。加えてこの幾何学的性質の応用可能性について検討する。

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 [6]。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に表せる [7]。

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

ただし

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

とする。

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。

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

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

* Quantitative analysis of dependency of cepstrum on vocal tract length and its application.
by SAITO Daisuke, MATSUURA Ryo (Univ. of Tokyo), KAMADA Toshiaki (Univ. of Tokyo / NRIPS),
ASAKAWA Satoshi, MINEMATSU Nobuaki, and HIROSE Keikichi (Univ. of Tokyo)

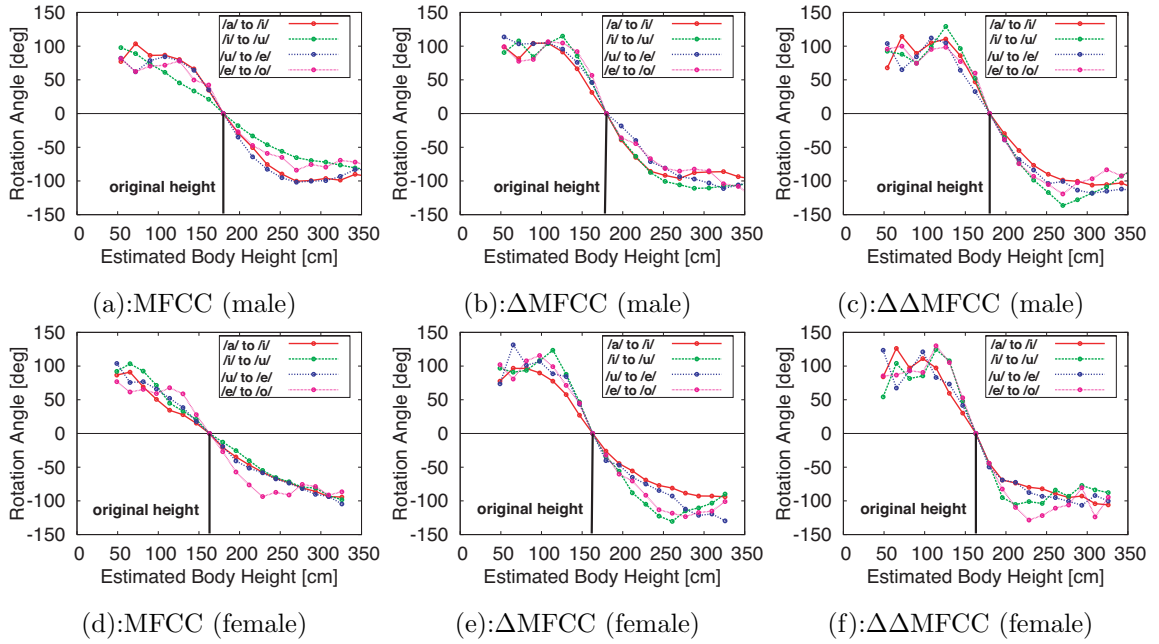


Fig. 1 身長と回転角との対応関係: (a)–(c): 男性話者 (身長 180 cm) ;(d)–(f): 女性話者 (身長 163 cm)

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 /aiueo/ について STRAIGHT[8] 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．一方で，性別や音韻の違いによって回転角の大きさに違いがみられる．すなわち式 (8) の定義だけでは式 (3) の行列 A の回転性を完全には表現できないことを意味している．

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 [9]．

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

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．

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

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 はその逆操作である．すなわち回転の様子はこれらの姿勢補正行列と Θ で記述できる．

4.2 内積による角度との対応

式 (8) のようなベクトルの内積によって定義される角度と Θ の対応関係について考える．今，簡単のため式 (14) に基づく姿勢補正によりベクトルが固有姿勢に変換されている場合を考える． $n=2m$ のときベクトル $X = (x_i | i = 1, 2, \dots, n)$ を $R'_n(\Theta)$ で変換したベクトルを $Y = R'_n(\Theta)X$ とすると

$$Y = (R_2(\theta_i)v_i | i = 1, 2, \dots, m) \quad (16)$$

となる．ただし $v_i = (x_{2i-1}, x_{2i})^t$ とする．よって X と Y の内積によってされる角度を θ' とすると θ' は以下のように表される．

$$\begin{aligned} \cos \theta' &= \frac{X \cdot Y}{|X||Y|} \\ &= \frac{1}{|X|^2} \sum_{i=1}^m v_i^t R_2(\theta_i) v_i \\ &= \frac{1}{|X|^2} \sum_{i=1}^m |v_i|^2 \cos \theta_i \end{aligned} \quad (17)$$

今， v_i はベクトル X の $2i-1, 2i$ 次元目の成分で構成される 2 次元ベクトルである．これは回転角 θ_i による回転に対応する部分空間へのベクトル X の射影成分である．以下この部分空間を回転平面と呼ぶ．すなわち式 (17) より内積による角度表現は変換前のベクトル X の回転平面への射影成分の大きさ ($|v_i|^2$) を重みとする，回転パラメータ Θ の各値に対応する回転角の余弦の重み付き線形和となる． $n = 2m + 1$ の場合は実固有値に対応する回転角について， $\cos \theta_i = 1$ として同様の議論が可能である．

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)(16) における $R_2(\theta_i)$ を $r_i R_2(\theta_i)$ に置き換えればよい．よって拡大縮小を伴う回転において式 (17) は以下のように変形できる．

$$\cos \theta' = \frac{\sum_{i=1}^m r_i |v_i|^2 \cos \theta_i}{\sqrt{(\sum_{i=1}^m |v_i|^2) (\sum_{i=1}^m r_i^2 |v_i|^2)}} \quad (18)$$

Table 1 音響分析条件

サンプリング条件	16 kHz / 16 bit
フレーム窓	Hamming window
フレーム	25 ms length / 5 ms shift
ケプストラム	FFT Cepstrum (1-12)

4.3 声道長変換行列に基づくパラメータ空間

声道長変換行列 A について，その回転パラメータ Θ がウォーピングパラメータ α のみに依存し，式 (14) における姿勢補正の行列が α に独立であると仮定する．このとき回転平面は α に依存せず常に一定になる．その場合，式 (18) において， $|v_i|$ のみによって，ある α における θ を求めることができる．すなわち $|v_i|$ は上記の仮定のもと， α によるウォーピングのような「規則的な声道長変化」に頑健なパラメータとなる．一方調音の違い，声道形状の違いといった「不規則な声道変化」はパラメータ値の違いとして表出することになる．このようなパラメータ値の違いの結果として Fig. 1 の回転性に差異があらわれたものと考えられる．

5 実験

5.1 実験方法

前節で述べた声道長変換行列 A の固有値に基づく分析の有効性を確かめるため実験を行った．まず先行研究と同様，成人男女 1 名ずつの日本語 5 母音の連続発声/aieuo/の音声を用い，STRAIGHT を用いて周波数ウォーピングを行った．これらの音声について Table 1 に示す音響分析条件で分析を行い，式 (8) によって変換前後のケプストラムのなす角 θ を調べた．

一方， θ が式 (18) によって記述可能であることを示すため， $\alpha = -0.2, -0.1, 0.1, 0.2$ の場合の行列 A について固有ベクトルによって張られる基底を導出し，変換前のケプストラムのこれらの基底への射影を求めることで，パラメータ $|v_i|$ を導出した．この値と α を変化させて得た回転パラメータ Θ および拡大縮小パラメータ r_i を用いて式 (18) から変換前後の角度 θ' を得た．声道長（身長）と角度の関係について式 (8) による θ と式 (18) による θ' を比較した．

5.2 実験結果

母音の 3 カ所の遷移部分 (/a/ から /i/，/i/ から /u/，/u/ から /e/) の Δ ケプストラムにおける結果を Fig. 2 に示す．Fig. 2 において点を含む赤線が式 (8) に基づく観測された回転角，その他の線が各基底について式 (18) によって導出した回転角である．身長 100 cm から 250 cm の範囲においては式 (18) による回転角

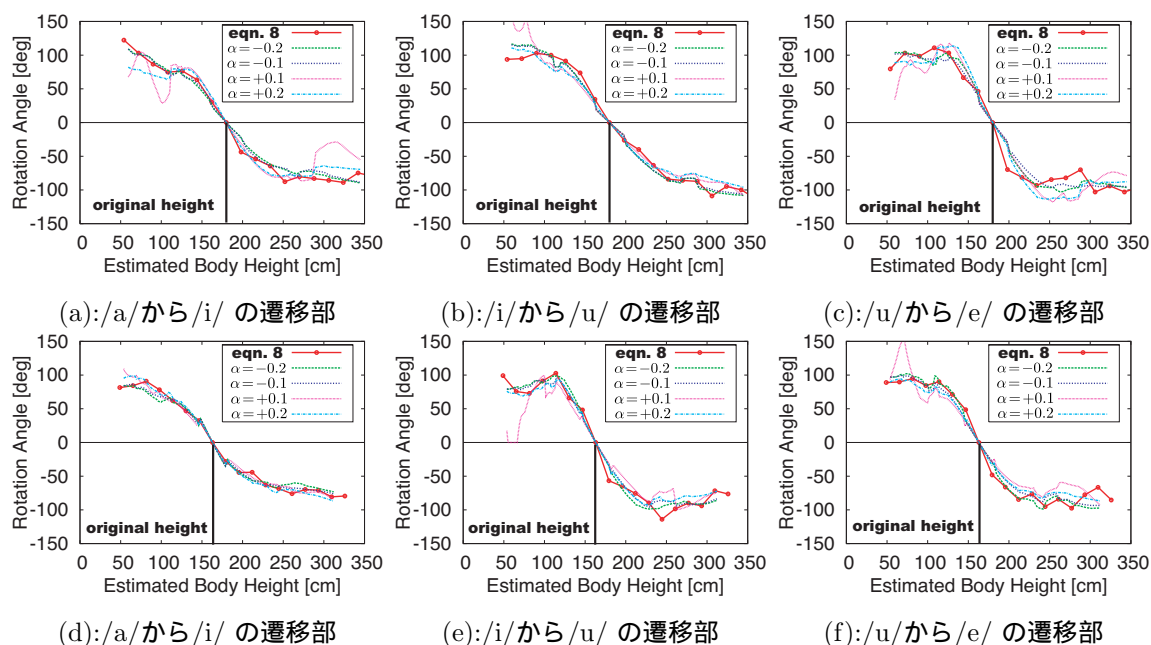


Fig. 2 固有回転による回転角の導出: (a)–(c): 男性話者 (身長 180 cm); (d)–(f): 女性話者 (身長 163 cm)

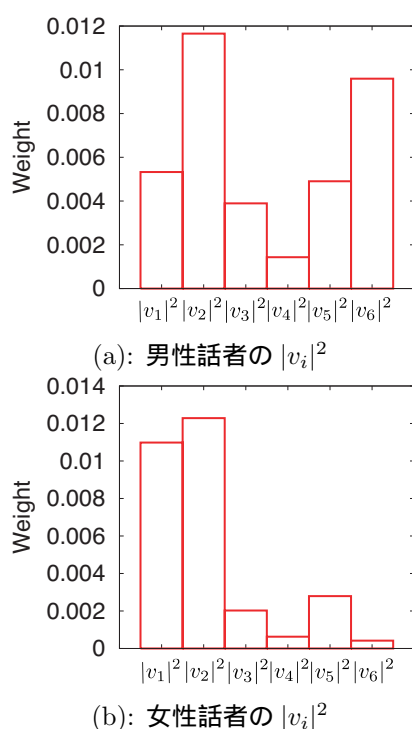


Fig. 3 回転平面への音響特徴量の射影量

がおよそ観測角と同様の傾向を示していることが分かる．また基底を決定する α が変化してもおよそ影響がないことから， α が回転行列の固有姿勢に影響しておらず，回転角が α によって決定できる事が分かる．一方，/a/ から /i/ の遷移部分について， θ_i に対応する回転平面への射影量を表す $|v_i|^2$ の値を表示したものを Fig. 3 に示す．Fig. 3 から，性別の差が回転平面への射影量として現れており，この事が回転性の差に反映されていることがわかる．

6 おわりに

本稿ではケプストラム特徴量について，その声道長依存性を定量的に解析した．筆者らはす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てきたが [3, 4, 5]，今回この変換行列の固有値に着目し，変換行列が持つ固有の回転パラメータを導出した．さらにこれらの回転パラメータから従来の実験結果を説明した．また声道長変換行列の回転平面について，ウォーピングパラメータへの依存性が低いことが分かった．

今後は回転平面への射影量を表す $|v_i|^2$ の利用法について，検討を行う必要がある．このパラメータは声道長の影響を除いた，その他の音韻，性別の情報を表現しうる．また回転行列の個々の回転平面は独立であるため，構造的表象に基づく音声処理における次元分割処理の枠組みへの応用も可能と考えられる [10]．

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, 2008 (to appear).
- [6] 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.
- [7] M. Pitz, H. Ney, IEEE Trans. Speech and Audio Processing, Vol.13, pp.930–944, 2005.
- [8] H. Kawahara et al., Speech Communication, 27, 3-4, pp.187–207, 1999.
- [9] 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.
- [10] S. Asakawa et al., ICASSP2008, 2008 (to appear).