

ケプストラムの声道長依存性に関する幾何学的考察

齋藤 大輔[†] 松浦 良[†] 朝川 智[†] 峯松 信明[†] 広瀬 啓吉^{††}

[†] 東京大学大学院新領域創成科学研究科 〒 277-0882 千葉県柏市柏の葉 5-1-5

^{††} 東京大学大学院情報理工学系研究科 〒 113-0033 東京都文京区本郷 7-3-1

E-mail: {dsk_saito,matsuura,asakawa,mine,hirose}@gavo.t.u-tokyo.ac.jp

あらまし 本報では、ケプストラムベクトルの方向成分が声道長の変化に対して強く依存していることを理論的、実験的に示す。さらにこの依存性が n 次元のケプストラム空間における回転として表出されることを示す。音声認識の研究においては、年齢や性別の違いといった歪みを取り除くため、声道長正規化 (VTLN) とよばれる技術が広く用いられている。VTLN はスペクトルドメインにおける周波数ウォーピングによって実現されるが、ケプストラムドメインでは線形変換 $\hat{c} = Ac$ として表現する事ができる。しかしこの変換行列 A の幾何学的な性質に関しては今まで十分に議論されてこなかった。本研究では n 次元空間における幾何学を通して、これらの変換が全てのケプストラムベクトルをおよそ等しく回転させる事を示す。さらに分析再合成音を用いて、実際にケプストラムベクトルが回転している事を実験的に確認した。身長 180 cm の話者と身長 120 cm の話者を比較した場合、そのケプストラムベクトルがおおよそ直交していることがわかった。本報の結果から従来の音声認識システムが子供の声のような特異音声を苦手とする一因を定量的に示す事ができた。

キーワード 周波数ウォーピング, ケプストラム, 幾何学的性質, 回転行列, 声道長

A study of geometric dependency of cepstrum on vocal tract length

Daisuke SAITO[†], Ryo MATSUURA[†], Satoshi ASAKAWA[†], Nobuaki MINEMATSU[†], and
Keikichi HIROSE^{††}

[†] Graduate School of Frontier Sciences, The University of Tokyo
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-0882, Japan

^{††} Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

E-mail: {dsk_saito,matsuura,asakawa,mine,hirose}@gavo.t.u-tokyo.ac.jp

Abstract In this paper, we theoretically and experimentally prove that the direction of cepstrum vectors strongly depends on vocal tract length and that this dependency is represented as rotation in the n dimensional cepstrum space. In speech recognition studies, vocal tract length normalization (VTLN) techniques are widely used to cancel age- and gender-differences. In VTLN, a frequency warping is often carried out and it can be implemented as a linear transformation in a cepstrum space; $\hat{c} = Ac$. However, the geometric properties of this transformation matrix A have not been well discussed. In this study, its properties are made clear using n dimensional geometry and it is shown that the matrix rotates any cepstrum vector similarly and apparently. Experimental results using resynthesized speech demonstrate that cepstrum vectors extracted from a speaker of 180 [cm] in height and those from another speaker of 120 [cm] in height are reasonably orthogonal. This result clarifies one of the reasons why children's speech is very difficult for conventional speech recognizers to deal with adequately.

Key words frequency warping, cepstrum, geometric property, rotation matrix, vocal tract length

1. はじめに

音声は年齢、性別、声道長や音響機器などの非言語的特徴に

よって不可避に歪む。これらの歪みは音声認識システムの性能に大きな影響を与えている。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に対して、話者正規化の技術が広く用いられている．話者正規化技術は、差分処理や微分演算に基づくものと変換に基づくものの二つに大別する事ができる．前者の例としてはケプストラム平均正規化 (CMN) や Δ ケプストラムの利用、後者として声道長正規化 (VTLN) がある．

CMN では、ケプストラム特徴量の長時間平均をもとの特徴量から削除する [2]．CMN は相対特徴量を導出することから、音響機器の伝送特性に対応するスペクトルの乗算性歪みに対して効果がある． Δ ケプストラムも直前のフレームとの差を捉えており、乗算性歪みに頑健で音声の動的特徴を表現することから広く用いられる．

VTLN は声道長の違いを取り除くために用いられる [3]．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あらかじめ想定した話者の特徴へと変換する．一方この変換行列のケプストラム空間における幾何学的な性質は今まであまり議論されてこなかった．本稿ではこの声道長変化を表す変換行列の幾何学的性質について、理論的および実験的に検討する．また Δ パラメータに対して変換行列が与える影響についても言及する．

2. 声道長変化の定式化

2.1 周波数ウォーピング

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．図 1 はウォーピングパラメータを変化させた場合の式 (1) の様子を示している．

2.2 ケプストラム空間における定式化

以下、前述のスペクトルドメインにおける周波数ウォーピングをケプストラム空間における記述に置き換える．江森らは声道長の変化をケプストラム空間で記述し、これらのパラメータ推定に基づく声道長正規化を行っている [4], [5]．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．

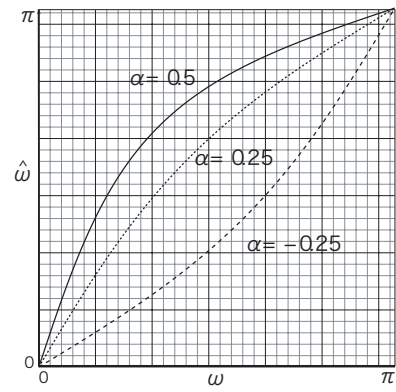


図 1 周波数ウォーピング関数

Fig. 1 Examples of frequency warping functions for different values of α .

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

$$\hat{c} = (\hat{c}_1 \ \hat{c}_2 \ \hat{c}_3 \ \hat{c}_4 \cdots)^t$$

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

$$c = (c_1 \ c_2 \ c_3 \ c_4 \cdots)^t$$

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 [6]．

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

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

3.1 2次元ケプストラム空間における回転性

本節では、式 (3) の性質について幾何学的に考察する．簡単のためまず、ケプストラムの 1 次および 2 次の項のみに着目し、2 次元のケプストラム空間を考える．次に、 n 次元のケプストラム空間へと議論を拡張する．

2 次元空間において、式 (2) における c から $\hat{c}$ への変換は以下のように記述する事ができる．

$$\begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \end{pmatrix} = \begin{pmatrix} 1-\alpha^2 & 2\alpha-2\alpha^3 \\ -\alpha+\alpha^3 & 1-4\alpha^2+3\alpha^4 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \quad (6)$$

今、式 (6) の変換行列を行列 T と呼ぶ事にする．このとき T は以下のように分解する事ができる．

$$T = R + O \quad (7)$$

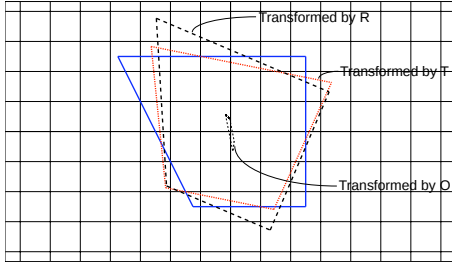


図 2 $\alpha = 0.2$ における行列 T, R, O による変換の様子

Fig. 2 Effects of transformations of T, R , and O for $\alpha = 0.2$.

ただし

$$R = \begin{pmatrix} 1-2\alpha^2 & 2\alpha(1-\frac{1}{2}\alpha^2) \\ -2\alpha(1-\frac{1}{2}\alpha^2) & 1-2\alpha^2 \end{pmatrix} \quad (8)$$

$$O = \begin{pmatrix} \alpha^2 & -\alpha^3 \\ -\alpha & -2\alpha^2+3\alpha^4 \end{pmatrix} \quad (9)$$

とする．行列 R について着目すると，一次近似である $(1+t)^k \simeq 1+kt$ を用いる事で次式のように変形できる．

$$R \simeq \begin{pmatrix} 1-2\alpha^2 & 2\alpha\sqrt{1-\alpha^2} \\ -2\alpha\sqrt{1-\alpha^2} & 1-2\alpha^2 \end{pmatrix} \quad (10)$$

$$= \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{pmatrix} (\alpha = \sin \theta) \quad (11)$$

式 (11) は行列 R が回転行列であることを示しており， R による変換は原点を中心として，全てのベクトルを時計まわりに 2θ 回転させる事になる．

一方行列 O は，行列 T による変換において，その影響が非常に小さいといえる．なぜならば $|\alpha| < 1$ であり， O のうち 3 つの要素が 2 次以上の高次項で構成されているためである．故に 2 次元平面における変換行列 T の性質は行列 R の性質およそ記述でき，すなわち T による変換は高い回転性を示す事になる．図 2 は $\alpha = 0.2$ において，2 次元平面における台形が行列 T, R, O によってどのように変換されるかを示している．原点付近に存在する小さな四角形が O によって変換されたものである．図 2 により，明らかに T によって変換された台形が時計回りに回転しており，その様子が R による変換と非常に近い事がわかる．一方 O における変換は全ての点を原点付近へ圧縮しているといえ，その性質は零行列に近いため， T における O の影響は非常に小さい．

さらに別の側面から行列 T の性質を記述する事を考える．図 3 は以下のベクトル関数による 2 次元平面におけるベクトル場を表している．

$$y = (T - I)c = \hat{c} - c \quad (12)$$

ここで I は 2 次元における単位行列を表している．行列 $(T - I)$ は変換前後の差を表現しており，ベクトル y は 2 次元平面の各点における変換行列 T の影響を表している．図 3 を見ると，式 (12) によるベクトル場は渦を描いている．このことは行列 T による変換が強い回転性を持つことを示唆している．

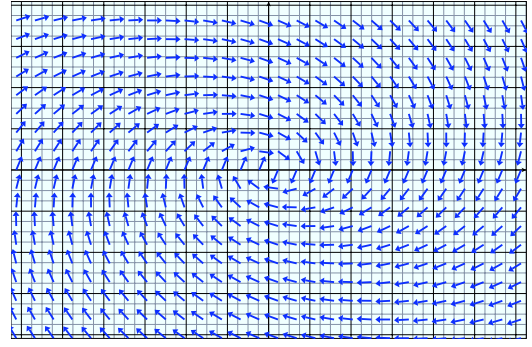


図 3 式 (12) のベクトル関数のベクトル場表示

Fig. 3 Vector field given by Equation (12) for $\alpha = 0.2$.

3.2 n 次元ケプストラム空間における回転性

n 次元空間については，前述の 2 次元空間のように変換行列からその回転性のみを抽出してくることは難しい．故に本節では n 次元における回転行列の一般的な定義に基づき，式 (3) の行列 A の幾何学的性質を記述する．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．

$$R^t R = R R^t = I \quad (13)$$

$$\det R = +1 \quad (14)$$

以下，式 (13) および式 (14) が式 (3) の A に関して近似的に成立することを示す． $|\alpha| \ll 1$ が成立すると仮定すると， A は 2 次以上の高次項を無視して以下のように近似する事ができる．

$$A_n = \begin{pmatrix} 1 & 2\alpha & 0 & \cdots & \cdots \\ -\alpha & 1 & 3\alpha & 0 & \cdots \\ 0 & -2\alpha & 1 & 4\alpha & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (15)$$

このとき行列 A_n の要素 a_{ij} は以下のように表される．

$$a_{ij} = \begin{cases} 1 & (i = j) \\ \text{sgn}(j-i) * j\alpha & (|i-j| = 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (16)$$

ここで $\text{sgn}(j-i)$ は $j-i > 0$ のとき $+1$ を， $j-i < 0$ のとき -1 を返す関数である．このとき $A_n^t A_n$ および $A_n A_n^t$ がおよそ単位行列となることを示す．

$$A_n^t A_n = \begin{pmatrix} 1+\alpha^2 & \alpha & -3\alpha^2 & 0 & \cdots \\ \alpha & 1+8\alpha^2 & \alpha & -8\alpha^2 & \cdots \\ -3\alpha^2 & \alpha & 1+18\alpha^2 & \alpha & \cdots \\ 0 & -8\alpha^2 & \alpha & 1+32\alpha^2 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (17)$$

式 (17) の対角成分は $k \in \mathcal{R}$ として $1+k\alpha^2$ と表される．さらに $|i-j| = 1$ の要素は α ， $|i-j| = 2$ の要素は $m \in \mathcal{R}$ として $m\alpha^2$ ，それ以外の要素は 0 となる．一方 $A_n A_n^t$ は以下で表される．

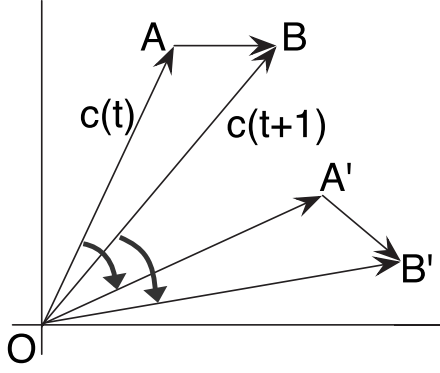


図4 ケプストラムベクトルの回転とそのデルタベクトル

Fig. 4 Rotation of two cepstrum vectors and their Δ vector.

$$A_n A_n^t = \begin{pmatrix} 1+4\alpha^2 & \alpha & -4\alpha^2 & 0 & \cdots \\ \alpha & 1+10\alpha^2 & \alpha & -9\alpha^2 & \cdots \\ -4\alpha^2 & \alpha & 1+20\alpha^2 & \alpha & \cdots \\ 0 & -9\alpha^2 & \alpha & 1+34\alpha^2 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (18)$$

式 (17)(18) とともに, $|\alpha| \ll 1$ の仮定のもと α^2 の項を無視する事ができる. この時, 二つの行列積は対角成分が 1, $|i-j|=1$ の成分が α となる三重対角行列と考える事ができる. このことから $A_n^t A_n$ および $A_n A_n^t$ は厳密には単位行列にはならないものの, 高い直交性を持っているといえる. そのため近似的に式 (13) を満たしているといえる.

一方 A_n は三重対角行列であるため, その行列式を求める事ができる [7]. n 次の三重対角行列の行列式は再帰的に以下の形で求められる.

$$\det A_n = a_{nn} \det A_{n-1} - a_{n(n-1)} a_{(n-1)n} \det A_{n-2}. \quad (19)$$

式 (15) より, $a_{nn} = 1$ および $a_{n(n-1)}, a_{(n-1)n} \sim \alpha^2$ となる. ここで $|\alpha| \ll 1$ の仮定を用いれば, $\det A_n \approx \det A_{n-1} \approx \cdots \approx \det A_1 \approx 1$ と再帰的に行列式を求める事ができる. よって A_n は近似的に式 (14) を満たす.

上記の議論から, 式 (3) の行列 A による変換は n 次元空間において回転性を有しているといえる. しかし, 上記の議論にはいくつかの近似を導入しているため, 実験的に A の回転性を確かめる必要がある.

いま 2 次元空間で得られた図 3 が n 次元空間においても観察されると仮定する. このとき行列 A による変換の性質が予測できる. 図 3 は行列 T による変換が全ての点を, およそ等しい回転角で回転させている事を示している. さらにこの回転角はおよそ α のみに依存している. このことからケプストラムベクトルは音素や性別に依存せずウォーピングパラメータのみに依存して近い回転角で変化することが示唆される.

一方, Δ パラメータに関する性質も予測ができる. 図 4 は時刻 t および $t+1$ におけるケプストラムを c_t, c_{t+1} として, c_t, c_{t+1} および $\Delta c = c_{t+1} - c_t$ の回転の様子を示している. 空間中の各ベクトルがおよそ等しい角度で回る場合, これらの Δ

表 1 音響分析条件

Table 1 Acoustic conditions.

サンプリング条件	16 kHz / 16 bit
フレーム窓	Hamming window
フレーム長	25 ms
シフト長	5 ms
ケプストラムパラメータ	MFCC (1-12)

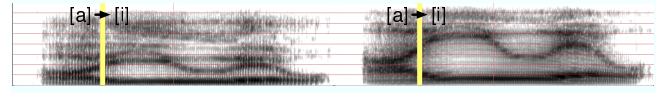


図5 ウォーピング前後の音声のスペクトログラム

Fig. 5 Spectrogram of the original speech (left) and its warped version (right).

パラメータも同様に回転することになる. これらは二つの Δ ベクトルに対して $\Delta\Delta$ ベクトルを考えた場合も同様である. 即ち, 空間中の各ベクトルがおよそ等しく回転する事はこれらの高次の Δ ベクトルも同様に回転する事を示している. 1. で述べたように, Δ パラメータの導入は微分演算に相当し, 音声認識における学習データと評価データのミスマッチ問題に対してある程度の効果を示す事が知られている. しかし, 本節における考察により, 空間における回転に相当する変換に対しては, Δ パラメータの導入がおよそ効果がない事が示唆される.

4. 実験

4.1 実験条件

前節で議論した声道長変化に伴うケプストラムの回転性が実際の音声でどのように表出するかを調べるため再合成音による評価実験を行った.

本実験では成人男女 1 名ずつの日本語 5 母音の連続発声 /aiueo/ の音声を用いた. これらに対して STRAIGHT [8] を用いて周波数ウォーピングを行った. これらのウォーピング後の音声が声道長の異なる話者に対応する. これらの音声について表 1 に示す音響分析条件で分析を行った. 加えて Δ MFCC, $\Delta\Delta$ MFCC についても抽出した. 作成した音声資料の同一時刻についてウォーピング前後のパラメータの比較を行った. 比較については二つのパラメータベクトルのなす角を用いた. 二つのベクトル a, b のなす角 θ については式 (20) で求められる.

$$\theta = \arccos \frac{a \cdot b}{|a||b|} \quad (20)$$

ここで $a \cdot b$ はベクトルの内積を, $|a|, |b|$ はベクトルのノルムを表す. 声道長変化との対応関係を明確にするため, 周波数ウォーピングに際して式 (1) および式 (3) を直接的に用いず, 下記の折れ線関数を用いて行った.

$$\hat{\omega} = \begin{cases} \frac{1}{m}\omega & (0 \leq \omega < \frac{m}{1+m}\pi) \\ m(\omega - \pi) + \pi & (\frac{m}{1+m}\pi \leq \omega \leq \pi) \end{cases} \quad (21)$$

上記の折れ線関数は図 1 を低域, 高域ごとに二つの直線で近似したものである. 式 (21) における m はウォーピング前後の

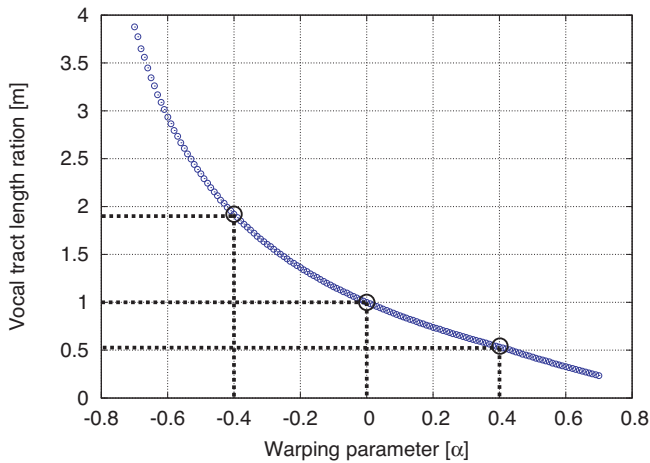


図6 ウォーピングパラメータ α と声道長比 m との対応関係

Fig.6 Relation between warping parameter and vocal tract length ratio.

音声の声道長の比を表している．式 (21) によってウォーピングを施した音声の例を図 5 に示す．図 5 において，左がウォーピング前，右がウォーピング後の音声である．色の濃い部分がフォルマントを表しているが，ウォーピングに伴い，これらが高周波数方向にシフトしていることがわかる．

式 (21) における m と α は 1 対 1 に対応させる事ができる．それぞれの α について，図 1 におけるウォーピング関数を式 (21) で近似した場合の 2 乗誤差が最小となる m を求めた．これらの α と m との対応関係を図 6 に示す．図 6 によれば身長 180 cm の話者について $\alpha = 0.4$ のウォーピングは身長約 90 cm の話者に対応し， $\alpha = -0.4$ のウォーピングは身長約 360 cm の話者に対応する．

4.2 実験結果

図 7 は母音の遷移部分 ($/a/$ から $/i/$ ， $/i/$ から $/u/$ ， $/u/$ から $/e/$ ， $/e/$ から $/o/$) について，身長を横軸に，ウォーピング前のパラメータベクトルを基準とした式 (20) の角度を縦軸にプロットしたものである．(a) から (c) まだが男性話者の，(d) から (f) まだが女性話者のデータである．さらに図 7 の左から順に MFCC， Δ MFCC， $\Delta\Delta$ MFCC に対応している．図 7 によれば，性別，音韻にあまり依存しない回転性を示している事がわかる．加えて MFCC， Δ MFCC， $\Delta\Delta$ MFCC ともに強い回転性を示しており，微分演算の回数を増やす事でむしろ回転角が大きくなっている傾向がみられる．

5. 考察

例えば図 7(b) に着目すると，身長 180 cm の話者とウォーピング後の身長 120 cm に対応する話者について， Δ MFCC のベクトルがおおよそ直交してくることになる．すなわち成長に伴って，話者の音声のパラメータベクトルが回転し，いずれベクトルとして直交性を持つ事になる．このことからパラメータベクトルの回転度という指標が性別や年齢，身長といった声道長に依存する話者情報の表現として有用であり，話者の声道長の情報を推定する有効なパラメータである事が示唆される．一方，

α の絶対値が小さい場合にはおおよそ等しい回転傾向がみられるが， α の絶対値が大きく，すなわち話者の声道長の違いが大きくなる場合には音韻や性別によって回転角が異なる様子が観察される．これらは 3.2 における議論において α の絶対値に基づく近似を多く含んでいるため， α の絶対値が大きい場合の回転が異なる様子を示している事が考えられる．しかしこれらの異なる回転性によって音韻や性別を表現できる可能性も示唆しており，今後詳細な検討を行っていく必要がある．

一方，これらの結果が音声認識システムに与えている影響について考える．前述の通り，同一の音韻を表すパラメータベクトルが，身長 180 cm の話者と身長 120 cm の話者とではおおよそ直交している．すなわちこれは成人で構築した音響モデルによって子供の音声の認識させるタスクに対応しており，今回の実験結果は従来の音声認識システムが子供の音声のような特異音声の認識を苦手とすることの一因を定量的な形で示したものである．

1. で述べたように，音声の動的特徴を表現する場合に Δ ケプストラムは広く用いられている． Δ パラメータの導入は微分演算に相当し，マイクの伝送特性に代表されるような，スペクトルに対する乗算性歪みに起因するミスマッチ問題に対してある程度の効果を示す事が知られている．しかし本報で議論したような線形変換性歪みによるミスマッチ問題に対してはその効果が十分でない事が示唆される．動的特徴の表現として Δ ケプストラムのようなベクトル表現を用いる限り，本質的にこのような問題は不可避であると考えられる [9]．一方筆者らはすでにスカラー特徴量に基づいた動的特徴のみに着目し，乗算性歪み，線形変換性歪みのいずれに対しても不変な音声表象を提案している [10]．朝川らは，この表象を用いて音声認識システムを構築し，ごく少量の学習データによって従来の大規模学習データを用いた音声認識システムと同等の性能を示している [11], [12]．

6. おわりに

本報ではケプストラム特徴量について，その声道長依存性を幾何学的に検証した．声道長の変化をケプストラム空間における線形変換として表現し，この変換行列の性質を数学的に議論する事で，その変化がケプストラム空間における回転性として表出することを示した．加えてこれらの回転性が音韻や性別， Δ パラメータの次数におおよそ依存しない事を確認した．またこれらの幾何学的性質を確認するため，周波数ウォーピングを施した分析再合成音を用いて，実際に声道長の変化が回転性として表出することを示した．これらの性質は定性的には先行研究で指摘されているが [9]，本研究ではこれを理論的，実験的，そして定量的に検討した．

今回の実験結果はケプストラム空間における声道長変化に起因する幾何学的特徴が，身長や性別といった話者情報の表現として有用であることを示唆している．一方，頑健な音声認識の観点から今回の結果を考えたとき，従来の音声認識システムが，声道長の異なる話者から抽出した，時に直交性を示す特徴量の一つのカテゴリとして扱い音響モデルを構築しているといえる．これらのベクトル特徴量の幾何学的な性質をよく考慮した上で

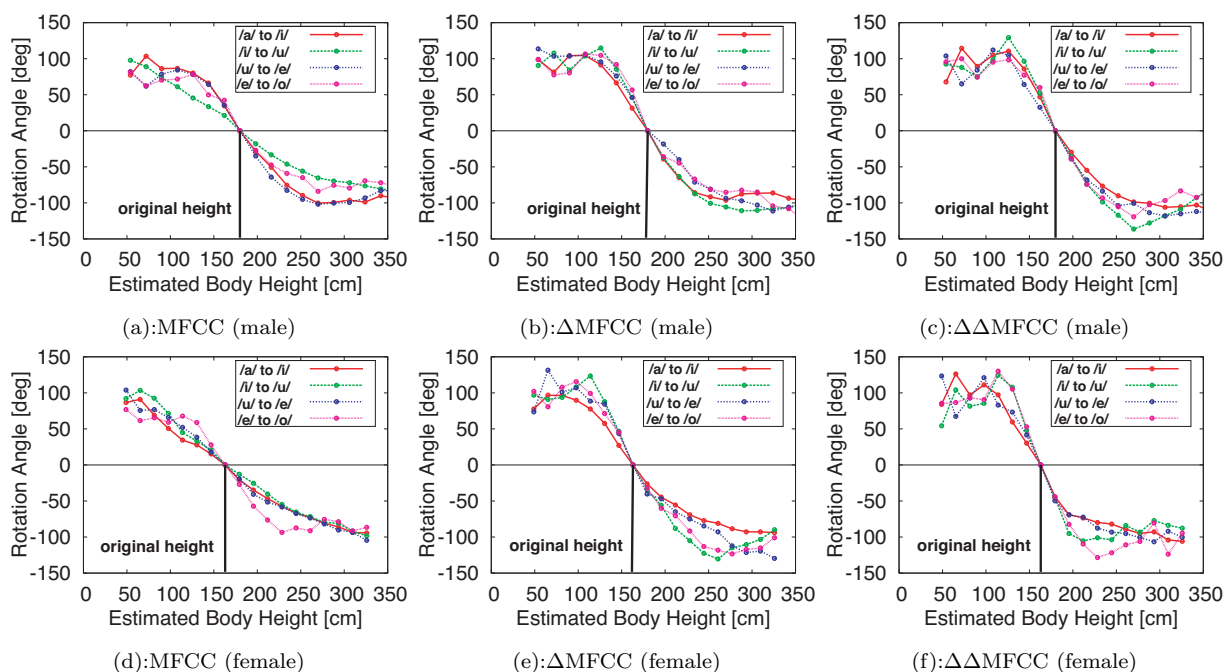


図 7 身長と回転角との対応関係: (a)–(c): 身長 180 cm の男性話者のデータ; (d)–(f): 身長 163 cm の女性話者のデータ

Fig. 7 Relation between the rotation angle and the estimated body height. (a) to (c) are from a male speaker of 180 cm in height and (d) to (f) are from a female speaker of 163 cm in height.

音響モデルの構築を考えなければならないだろう。

声道長変化による回転性はおよそ等しく作用しているが、音韻や性別、 Δ パラメータの次数による影響も多少見られた。一方、高次の回転行列については対角化などの操作によりその様子を詳細に記述することが可能である [13]。今後はこれらの実験結果をふまえ、回転行列の性質をより詳細に分析することで、ケプストラム空間における幾何学的特徴が話者情報の表現としてどのように利用できるかを検討していく予定である。

文 献

- [1] M. Russel and S. D’Arcy: “Challenges for computer recognition of children’s speech,” *CD-ROM of SLaTE2007*, 2007.
- [2] B. Atal: “Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification,” *J. Acoust. Soc. America*, vol. 55, pp. 1304–1312, 1974.
- [3] E. Eide and H. Gish: “A parametric approach to vocal tract length normalization,” *ICASSP96*, vol. 1, pp. 346–348, 1996.
- [4] 江森正, 篠田浩一: “音声認識のための高速最尤推定を用いた声道長正規化,” *電子情報通信学会論文誌 D-II*, vol. J83-D-II, no. 11, pp. 2108–2117, 2000.
- [5] T. Emori and K. Shinoda: “Rapid Vocal Tract Length Normalization using Maximum Likelihood Estimation,” *Eurospeech2001*, pp. 1649–1652, 2001.
- [6] M. Pitz and H. Ney: “Vocal tract normalization equals linear transformation in cepstral space,” *IEEE Trans. Speech and Audio Processing*, vol. 13, pp. 930–944, 2005.
- [7] R.A. Horn and C.R. Johnson: “Matrix Analysis,” Cambridge University Press, 1985.
- [8] H. Kawahara *et al.*: “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 Communica-*

tion, vol. 27, pp. 187–207, 1999.

- [9] 峯松信明, 松井健, 広瀬啓吉: “声に内在する音響的普遍構造とそれに基づく音声コミュニケーション,” 第三回「話し言葉の科学と工学」ワークショップ論文集, pp. 143–150, 2004.
- [10] N. Minematsu: “Mathematical evidence of the acoustic universal structure in speech,” *ICASSP2005*, pp. 889–892, 2005.
- [11] S. Asakawa *et al.*: “Automatic Recognition of Connected Vowels Only Using Speaker-invariant Representation of Speech Dynamics,” *INTERSPEECH2007*, pp. 890–893, 2007.
- [12] 朝川他: “音声の構造的表象を用いた音声認識における特徴量空間分割とその効果,” *日本音響学会秋季講演論文集*, 3-Q-10, 2007.
- [13] 高橋友和: “高次元回転行列の補完とその応用,” *Visual Computing / グラフィックスと CAD 合同シンポジウム*, 2007.