

デルタケプストラムの声道長依存性に関する実験的検討*

齋藤大輔, 朝川智, 峯松信明, 広瀬啓吉 (東大)

1 はじめに

音声認識および音声合成システムにおいては、音響特徴量としてケプストラムが広く用いられている。それに加えて動的性質を表す特徴量として、各時刻の微係数に対応する Δ ケプストラムが用いられる [1]。ケプストラムは、音声に不可避に含まれる静的な非言語性歪みの影響を直接受けるため、例えば話者の声道長差異やマイクの音響特性差異によって大きく変動する。一方 Δ ケプストラムは微係数であるため、その時刻における観測点がケプストラム空間のどこに位置しているかに依存しない。すなわちマイクの音響特性などの乗算性歪みには頑健な特徴量である。

本稿では Δ ケプストラムの声道長差異に対する頑健性について理論的かつ実験的に検討する。

声道長差異によるケプストラムの変形については、先行研究において数学的な定式化がなされている。江森らは、声道長の変化をケプストラム空間における線形写像で近似し、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 [2]。また Pitz らも声道長正規化がケプストラムの線形変換によって実装できることを言及している [3]。本稿ではこれらの数学モデルを用いて声道長の変化が Δ ケプストラムをどう変化させるのかについて検討を行う。その後、検討の妥当性を実験的に検証する。

2 声道長によるケプストラム変化の定式化

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。周波数ウォーピングにおいて、変換前後のスペクトル表現における周波数をそれぞれ ω と $\hat{\omega}$ 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。

以下、上記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。変換前後のケプストラムベクトル ($0 \sim n$

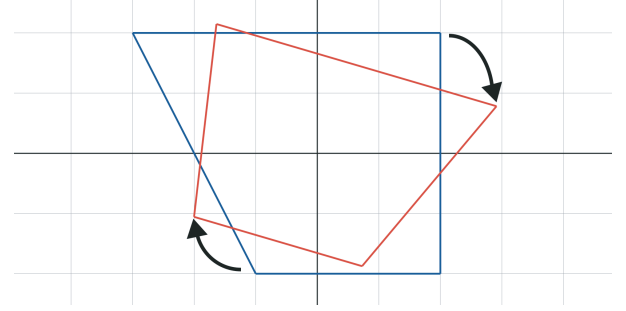


Fig. 1 2次元平面における行列演算の回転性

次) をそれぞれ $c, \hat{c}$ とする。江森らによると、式 (1) の周波数ウォーピングはケプストラム空間において以下の式で表現される [2]。

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

$$A = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \cdots \\ 0 & 1-\alpha^2 & 2\alpha-2\alpha^3 & \cdots & \cdots \\ 0 & -\alpha+\alpha^3 & 1-4\alpha^2+3\alpha^4 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (3)$$

このように声道長変換はケプストラム空間において、式 (2) のようなアフィン変換によって記述され、行列 A は式 (3) のようにウォーピングパラメータ α のみを用いて定式化される。

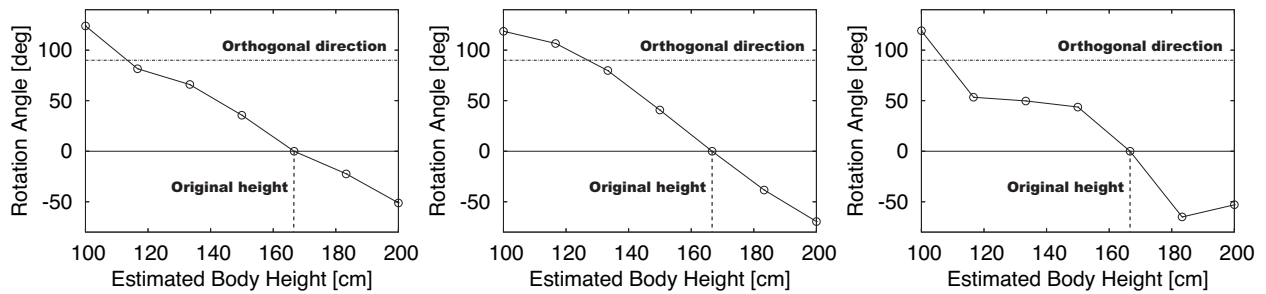
3 ケプストラム空間における回転性

式 (3) の行列演算が複数の観測点をどのように変換するのかを考える。簡単のため 2 次元のケプストラム空間を考え、パワーを考慮しないものとする。つまり式 (3) における 2,3 行と 2,3 列で構成される 2 次の正方行列を考える。このとき c から $\hat{c}$ への変換は式 (4) で表される。

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

今 $\alpha = 0.3$ として、2 次元平面における台形図形に式 (4) で表されるアフィン変換を行ったものを Fig. 1 に示す。Fig. 1 より式 (3) の行列 A がケプストラム空間において回転性の強い演算であることが示唆される。

*Experimental study on dependency of delta cepstrum coefficients on the vocal tract length.
by SAITO Daisuke, ASAKAWA Satoshi, MINEMATSU Nobuaki, and HIROSE Keikichi (The University of Tokyo)



(a): /a/ から /i/ への遷移部分 (b): /i/ から /u/ への遷移部分 (c): /u/ から /e/ への遷移部分

Fig. 2 母音遷移の中心時刻における声道長と Δ ケプストラムの回転との対応関係

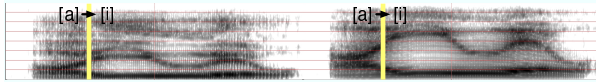


Fig. 3 ウォーピング前後の音声のスペクトログラム

4 模擬音声を用いた検証実験

実際の音声における Δ ケプストラムの声道長依存性について調べるため、STRAIGHT[4]を使って声道長特性の異なる音声試料を作成し分析を行った。

まず身長 167 cm の成人男性 1 名の日本語 5 母音の連続発声/aieuo/に対して、STRAIGHT を用いて周波数ウォーピングを行い、身長 100 cm から 200 cm に対応する音声を作成した。ウォーピング関数の設計については [5] を参考にした。ウォーピング前後の音声の例を Fig. 3 に示す。Fig. 3 において、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、これらが周波数方向にシフトしていることがわかる。作成した音声の同一時刻について Δ ケプストラムを求め、身長 167 cm の話者との比較を行った。音響分析条件を Table 1 に示す。比較については二つの Δ ケプストラムベクトルのなす角を調べた。二つのベクトル v_1, v_2 のなす角 θ は式 (5) より求められる。

$$\theta = \cos^{-1} \frac{\langle v_1, v_2 \rangle}{|v_1||v_2|} \quad (5)$$

ここで $\langle v_1, v_2 \rangle$ はベクトルの内積を、 $|v_1|, |v_2|$ はベクトルのノルムを表す。母音の遷移部分(/a/ から /i/、/i/ から /u/、/u/ から /e/) について、身長を横軸に、身長 167 cm の話者の Δ ケプストラムとのなす角を縦軸にプロットしたものを Fig. 2 に示す。Fig. 2 より、声道長の変化に対して、 Δ ケプストラムが極めて強い依存性を示し、その変化がケプストラム空間における回転性として表されていることがわかる。例えば身長が 120 cm ほどになると、成人男性話者に対して Δ ケプストラムが直交してくる。

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
フレーム長	25 ms
シフト長	5 ms
ケプストラムパラメータ	Δ MFCC 10 次元

5 おわりに

本稿では、音声の動的特徴を示すパラメータである Δ ケプストラムについて、声道長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実験的に示した。 Δ ケプストラムは相対量であり乗算性歪みに頑健である一方、線形変換歪みに極めて強く依存する。この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と見える。 Δ ケプストラムは動的特徴の表現として有用であるが、その取り扱いに注意を要するであろう。なお筆者らはこのような話者の違いに対して不変な枠組みによる音声認識システムの提案を行っており、 Δ ケプストラムを使うことの弊害についても報告している [6]。

参考文献

- [1] S. Furui, *The Journal of the Acoustical Society of America*, Vol.80, No.4, pp.1016–1025, 1986.
- [2] 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.
- [3] M. Pitz, H. Ney, *IEEE Trans. Speech and Audio Processing*, Vol.13, pp.930–944, 2005.
- [4] H. Kawahara *et al.*, *Speech Communication*, 27, 3–4, pp.187–207, 1999.
- [5] 林他, 音講論 (春), 2-Q-27, 2007.
- [6] 朝川他, 音講論 (秋), 3-Q-10, 2007 (発表予定).