

Title	ヤコビ適応法を用いた雑音・伝達特性・発声変形への同時適応
Author(s)	坂井, 伸圭
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1559
Rights	
Description	Supervisor: 嵯峨山 茂樹, 情報科学研究科, 修士

ヤコビ適応法を用いた 雑音・伝達特性・発声変形への同時適応

坂井 伸圭 (010047)

北陸先端科学技術大学院大学 情報科学研究科

2002年2月15日

キーワード: 音声認識, ヤコビ適応, 発声変形, Lombard 効果.

1 はじめに

音声認識技術の向上により現在の音声認識システムは, 雑音の無い理想的な環境では高い認識性能を示す. しかし, 雑音や伝達特性が存在する実環境においては, 認識率が大きく低下する. これは, 作成した音響モデルと発話環境における音響モデルとのミスマッチが原因である. ミスマッチの原因としては, 雑音や伝達特性の重畳による問題と, 背景雑音の存在による発話者の発声変形 (Lombard 効果) の問題とがある. これらの変動要因に対して, 音響モデルを適応させていく必要がある. これら個々の要因に対する適応法としては従来様々な研究がなされてきた. しかし, これらの要因に対して同時に適応する手法に関しては複雑な非線型問題となるため, あまり研究されていない. そこで, 複雑な非線型問題を, Taylor 展開の1次項を用いて近傍を線形近似して解く, という特徴を持つヤコビ適応法を用いて, これらの要因に対して音響モデルの同時適応を行う手法を検討する. 既にヤコビ適応法を用いた雑音・伝達特性に対する適応については報告されている.

本研究では, 発声変形におけるホルマント周波数の移動をスペクトル領域における周波数軸伸縮を用いてモデル化し, ヤコビ適応を行う手法につ

いて提案する．さらに，雑音・伝達特性・発声変形，これらの変動要因に対して同時に適応を行い，実験を通してその効果を確認する．

2 発声変形に対するヤコビ適応法

発声変形により，ホルマント周波数は，1.5 kHz 付近を境にして高周波数領域のものは低域へ，低周波数領域のものは高域へ移動することが報告されている．このホルマント周波数の移動をスペクトル領域における周波数軸伸縮を用いて以下のようにモデル化する．

通常音声のスペクトル $S(f)$ に対し，周波数軸伸縮後の周波数 f' を周波数軸伸縮係数 λ ，定数 f_α を用いて次式で表す．

$$f' = \lambda f + (1 - \lambda)f_\alpha \quad (1)$$

ここで， $\lambda > 1$ は周波数 f_α に向かう収縮を意味し， $\lambda < 1$ の場合は逆に伸長する．音声スペクトル $S(f)$ のケプストラムを C ， $S(f')$ のケプストラムを C_L とすると，それらは次式の関係にある．

$$C_L = F^{-1} F_L C \quad (2)$$

ここで， F^{-1} は離散コサイン変換行列 (F) の逆行列， F_L は前述の周波数軸伸縮を伴う離散コサイン変換行列で，その (i, j) 成分は次式で与えられる．

$$(F_L)_{ij} = \cos \frac{i(\lambda(j + 0.5) + (1 - \lambda)\omega_\alpha)\pi}{N} \quad (3)$$

N は周波数の標本化点数， ω_α は f_α に対応する周波数軸定数である．発声変形によって，伸縮係数 λ が微少量 $\Delta\lambda$ だけ変化したとすると，ケプストラム C_L の変動量 ΔC_L は，Taylor 展開による線形近似によって次式で与えられる．

$$\Delta C_L = \frac{\partial C_L}{\partial \lambda} \Delta \lambda \quad (4)$$

したがって，発声変形ヤコビ行列 $J_L \equiv \frac{\partial \mathbf{C}_L}{\partial \lambda}$ は， $\lambda \approx 1$ のとき，

$$\mathbf{J}_L = \mathbf{F}^{-1} \frac{\partial \mathbf{F}_L}{\partial \lambda} \mathbf{C} \approx \mathbf{F}^{-1} \mathbf{G} \mathbf{C} \quad (5)$$

となり， $\mathbf{G}$ の (i, j) 成分は

$$(\mathbf{G})_{ij} = \frac{-i(j + 0.5 - \omega_\alpha)\pi}{N} \sin\left(\frac{i(j + 0.5)\pi}{N}\right) \quad (6)$$

である．この $\mathbf{J}_L$ は想定する初期環境で事前に求まる．この際，周波数軸伸縮係数の変動分 $\Delta\lambda$ は，発声サンプルに Viterbi Time Alignment を行つて得られた変化分を用いて，(4) 式から最小二乗法により計算した．

3 実験結果

通常音声で作成した音響モデルから Lombard 音声への発声変形に対するヤコビ適応を行い，孤立単語音声認識実験によりその効果を確認した．発声変形モデルを用いたヤコビ適応では，最適な ω_α を用いた場合でも，認識率の改善は 2% 程度であったが，乗法性・加法性に対する適応を加えた同時適応では，15% 程度の認識率向上が見られた．発声変形には加法性・乗法性による変形要素が大きいことが確認できた．また，雑音・伝達特性を重畳した Lombard 音声に対して，雑音・伝達特性・発声変形に対する同時適応を行い，16-19% の認識率の向上を確認できた．