T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	話し言葉音声における音素の静的特徴量の分散と動的特徴量に関する分析
Title(English)	
著者(和文)	中村 匡伸, 岩野 公司, 古井 貞熙
Authors(English)	Masanobu Nakamura, Koji Iwano, SADAOKI FURUI
出典(和文)	日本音響学会講演論文集, Vol. , No. , pp. 217-218
Citation(English)	, Vol. , No. , pp. 217-218
発行日 / Pub. date	2006, 9

Powered by T2R2 (Science Tokyo Research Repository)

話し言葉音声における音素の静的特徴量の分散と動的特徴量に関する分析*

◎中村匡伸, 岩野公司, 古井貞熙 (東工大)

1 はじめに

話し言葉音声の音響的特徴の分析は，話し言葉音声の認識性能の向上や，音声合成の品質向上に役立つと考えられ，非常に重要である．我々はすでに，日本語話し言葉コーパス (以下 CSJ と呼ぶ) に収録されている同一話者の発声した話し言葉音声 (学会講演音声，模擬講演音声，対話音声) と読み上げ音声 (学会講演音声の再読み上げ) において各音素のケプストラム特徴量に関する比較を行った．その結果，話し言葉音声では読み上げ音声に比べて音素ケプストラム空間が縮小する傾向が明らかになった [1]．さらに全音素間のマハラノビス距離を分析することにより，音素間のユークリッド距離と，各音素のケプストラムの分散の双方が話し言葉音声の認識性能の低下に強い影響を及ぼしていることが明らかになった [2]．しかし，各音素のケプストラムの分散に関する定量的な分析は未だ行われていない．

これまではケプストラム特徴量として静的特徴量のみを分析の対象としてきたが，音声認識に通常用いられる動的特徴量に関しても分析を行うことにより，話し言葉音声の認識性能向上のための新たな知見が得られる可能性がある．そのため本稿では，読み上げ音声と話し言葉音声の音響的特徴の違いに関して，静的特徴量である MFCC の分散ベクトルの違い，および動的特徴量である ΔMFCC の平均ベクトルの違いに着目して分析を行う．

2 音声データ

分析には CSJ に含まれる，発話スタイルの異なる読み上げ音声，学会講演音声，模擬講演音声，対話音声を用いる．音声データは 16 kHz でサンプリングされている．実験に際して，まず人手で作成された時間ラベル付きの書き起こしに基づいて音声データを 400 ms 以上の無音区間で区切り，区切られた区間を「発話単位」として定義した．発話単位が 1 秒未満の場合には，後続する発話単位と接続し，1 つの発話単位とみなした．

本分析において分析対象とする音素は，表 1 のリストにある 31 種（母音 10 種・子音 21 種）とした．分析対象データの話者は男女各 5 名であり，各話者は 4 種類の異なる発話スタイルの音声 (読み上げ音声，学会講演音声，模擬講演音声，対話音声) を発声しており，それら全てを分析に用いる．表 2 に，分析対象データにおける発話スタイルごとの発話時間と音素サンプル数を示す．R, AP, EP, D は，それぞれ読み上げ音声，学会講演音声，模擬講演音声，対話音声を表す．

3 音響特徴量の抽出

本分析を行うにあたり，各発話スタイルにおける各音素の静的特徴量 MFCC と動的特徴量 ΔMFCC を抽出する．

Table 1 分析対象とする音素のリスト

母音	/a,i,u,e,o,a:,i:,u:,e:,o:/
子音	/w,y,r,p,t,k,b,d,g,j,ts,ch, z,s,sh,h,f,N,N:,m,n/

Table 2 音声データの発話時間および音素サンプル数

発話スタイル	発話時間 (分)	音素サンプル数
R	175	119,354
AP	160	119,330
EP	112	79,863
D	296	195,441

分析対象となる MFCC の分散ベクトル，および ΔMFCC の平均ベクトルは，以下のようにして抽出される．

1. 音声データから MFCC 12 次元とその一次微分，二次微分成分，対数パワーの一次微分，二次微分成分の計 38 次元の音響パラメータを抽出する．分析周期は 10ms, 分析窓幅は 25ms とし，発話単位ごとに CMS 処理を行っている．
2. 各発話スタイルごとに，分析対象データを用いて 1 混合 monophone HMM を学習する．全ての音素モデルは，3 状態の left-to-right 型 HMM とする．
3. 出来上がった monophone HMM のうち，分析対象音素の HMM の第 2 状態から MFCC の分散ベクトルを取り出し，第 1, 2, 3 状態から ΔMFCC の平均ベクトルを取り出す．

4 静的特徴量に関する分析

読み上げ音声と話し言葉音声における静的特徴量の分散ベクトルに関する比較を行う．発話スタイル X の音素 i における HMM の MFCC ベクトルの分散の拡大率 $ext_i(X)$ を次のように定義する．ただし X は R, AP, EP, D (それぞれ読み上げ音声，学会講演音声，模擬講演音声，対話音声) とする．

$$ext_i(X) = \frac{\sum_{k=1}^{K} \sigma_{ik}^2(X)}{\sum_{k=1}^{K} \sigma_{ik}^2(R)}$$

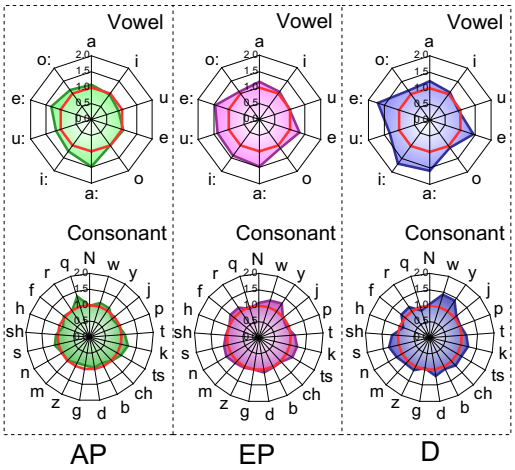

Fig. 1 各発話スタイルにおける母音・子音の分散の拡大率 (上段が母音，下段が子音の拡大率を表し，左から学会講演音声 (AP)，模擬講演音声 (EP)，対話音声 (D))

*Analysis on the variances of static features and the mean values of dynamic features in spontaneous speech by NAKAMURA Masanobu, IWANO Koji, and FURUI Sadaoki(Tokyo Institute of Technology)

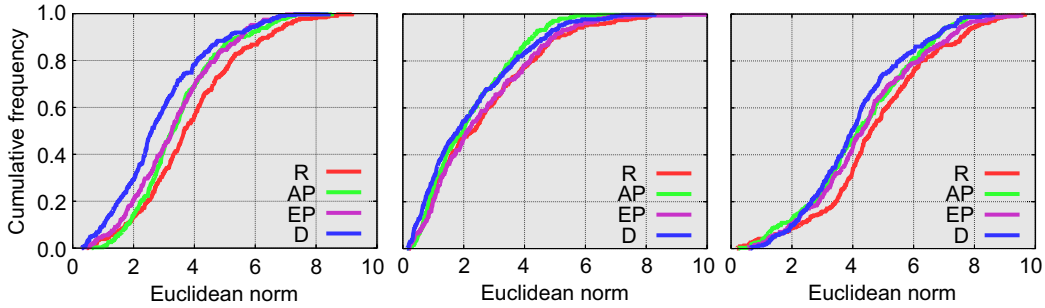

Fig. 2　各発話スタイルにおける ΔMFCC のユークリッドノルムの相対累積度数 (左から第 1, 2, 3 状態)

K は MFCC ベクトルの次元数 ($K = 12$) である．σ_{ik}^2 は，音素 i の第 2 状態から抽出された MFCC ベクトルの k 次成分の分散である．話者ごとに $ext_i(X)$ を求め，その話者平均値を $\overline{ext_i(X)}$ とする．

図 1 に，母音と子音の MFCC ベクトルの分散の拡大率 $\overline{ext_i(X)}$ を発話スタイルごとに示す．上段が母音，下段が子音を表している．$\overline{ext_i(X)} = 1$ を太線で表記する．

図 1 の左側に，読み上げ音声に対する学会講演音声の MFCC ベクトルの分散の拡大率 $\overline{ext_i(AP)}$ を表す．これより MFCC ベクトルの分散の拡大率は，短母音においては大半の音素が 1 に近く，長母音では全ての音素が 1 よりも大きいという特徴が見られる．逆に，子音における分散の拡大率はほとんど 1 に近いことが分かる．図 1 の中央に，読み上げ音声に対する模擬講演音声の MFCC ベクトルの分散の拡大率 $\overline{ext_i(EP)}$ を表す．母音・子音ともに学会講演音声での拡大率と同様の特徴が見られる．図 1 の右側に，読み上げ音声に対する対話音声の MFCC ベクトルの分散の拡大率 $\overline{ext_i(D)}$ を表す．対話音声の母音・子音における拡大率は，学会講演音声・模擬講演音声と比較すると，全ての音素に対して大きくなっていることが分かる．

また，学会講演音声，模擬講演音声，対話音声における母音の $\overline{ext_i(X)}$ の平均は，それぞれ 1.14, 1.24, 1.32 となり，子音ではそれぞれ 1.03, 1.09, 1.13 となった．発話スタイル間の母平均の差の検定を行ったところ，有意水準 5 % で MFCC ベクトルの分散の拡大率に差があることが示された．

5　動的特徴量に関する分析

読み上げ音声と話し言葉音声における動的特徴量の平均ベクトルに関する比較を行った．発話スタイルごとに ΔMFCC の平均ベクトルのユークリッドノルムを測定し，比較した．ΔMFCC ベクトルは，学習した HMM の第 1, 2, 3 状態からそれぞれ抽出した．

図 2 に，各発話スタイルにおける ΔMFCC ベクトルのユークリッドノルムの相対累積度数を，HMM の状態ごとに示す．左からそれぞれ，第 1, 2, 3 状態を表す．横軸をユークリッド距離，縦軸を相対累積度数とする．

図 2 より，第 1, 3 状態においては，自発度が高くなるにつれて ΔMFCC ベクトルのユークリッドノルムが小さくなっている傾向がある．これはすなわち，話し言葉音声では音素と音素のわたりの部分 (音素境界付近) において，MFCC の変化量が小さくなっていることを意味しており，我々の先行研究によって明らかになった音素ケプストラム空間の縮小による影響が表れていると考えられる．一方，第 2 状態においては，上記のような傾向は見られない．これはすなわち，音素の安定状態における MFCC の変化量は，発話スタイルに依存しないことを意味している．表 3 に，読み上げ音声 (R)，学会講演音声 (AP)，模擬講演音声 (EP)，対話音声 (D) の ΔMFCC のユークリッドノルムの分布の平均値を，状態ごとに示す．発話スタイル間の母平均の差の検定を行ったところ，第 1, 3 状態では，読み上げ音声と話し言葉音声 (学会講演音声，模擬講演音声，対話音声) の間に有意水準 5 % で母平均に差が見られたが，第 2 状態では有意差は見られなかった．

Table 3　各状態における ΔMFCC のユークリッドノルムの平均値

	R	AP	EP	D
1st state	3.88	3.48	3.34	2.88
2nd state	2.60	2.20	2.57	2.22
3rd state	4.77	4.25	4.43	4.11

6　まとめ

本稿では CSJ に収録された，同一話者が 4 種類の異なる発話スタイルで発話を行った音声データを用いて，静的特徴量の分散ベクトル，および動的特徴量の平均ベクトルに関する分析を行った．その結果，発話の自発性が高くなると各音素の MFCC ベクトルの分散が大きくなるという傾向が分かった．また，読み上げ音声に対して話し言葉音声では，音素と音素のわたりの部分において ΔMFCC ベクトルのユークリッドノルムが小さくなる傾向があることが分かった．この現象は，音素ケプストラム空間の縮小によるものであると考えられる．

本分析では，話者性の違いを考慮に入れないために，「同一話者による読み上げ音声と話し言葉音声」を対象とした．そのため，不特定話者においても同様の結果が得られるかどうか調査する必要がある．また，今回得られた知見を，話し言葉音声の認識性能の向上に役立てることが出来るかどうか，検討する必要がある．

謝辞

本研究は文部科学省 21 世紀 COE プログラム「大規模知識資源の体系化と活用基盤構築」の一環として実施されました．

参考文献

[1] 中村, 岩野, 古井, “日本語話し言葉コーパスを用いた話し言葉音声の音響的特徴の分析,” 情報処理学会研究報告, 2004-SLP-53 (2004-10).

[2] 中村, 岩野, 古井, “マハラノビス距離を用いた日本語話し言葉音声の音響的特徴の分析,” 日本音響学会 2005 年春季講演論文集, 2-1-4, pp.230-231 (2005-3).