論文 / 著書情報
Article / Book Information

論題(和文)	音源・声道特徴の組合せによる話者認識
Title(English)	
著者(和文)	古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1990年春季講演論文集, , , pp. 57-58
Citation(English)	, , , pp. 57-58
発行日 / Pub. date	1990, 3

2—3—4

音源・声道特徴の組合せによる話者認識*

⊙ 松井知子　　古井貞熙
(NTT ヒューマンインタフェース研究所)

1　まえがき

従来の話者認識実験 [1] - [4] では、声道に関する特徴量が用いられることが多く、それと音源に関する特徴量を組み合わせた例は少ない。本稿では、ベクトル量子化による話者認識において、声道に関する特徴量、音源に関する特徴量を、テキスト独立に、かつ時期的な変動に対して頑健に組み合わせる方法について検討した。

2　時期的な変動を考慮した話者認識法

本稿では、各話者の標準パターンは1時期の学習データのみから作成する。以下の検討を通して、異なる時期のテストデータに対しても有効な話者認識法を考える。

2.1　話者認識システムの構成

一般にベクトル量子化による話者認識では、入力音声から抽出した特徴量（テストデータ）と、登録している各話者の特徴量の標準パターン（符号帳）との類似度（平均量子化歪）を計算し、最も歪が小さい話者を選び出して、その人の音声と判定する。[3][4]

ここでは、声道に関する特徴量としてケプストラム、Δケプストラム、音源に関する特徴量としてピッチ、Δピッチを扱う。複数の特徴量に注目することにより、時期的な変動の吸収を試みる。ピッチ特徴量は、無声音区間などでは抽出できないので、そのような区間の処理が問題となる。ここでは、ピッチが抽出できる区間とできない区間とで、別々に符号帳を作成することで対処した。話者認識システムの構成を図1に示す。

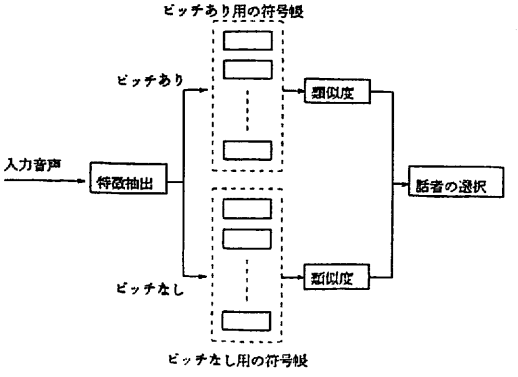


図 1: 話者認識システムの構成

2.2　特徴量正規化法

特徴量は、一つのベクトル $\vec{x} = (x_1, x_2, \ldots, x_i, \ldots)$ に統合して表す。複数の特徴量を扱う場合、特徴量間で分布が異なるので、平均量子化歪を計算する時に、各特徴量を正規化する必要がある。従来の話者認識では、簡易マハラノビス法が用いられることが多い。この方法は、各特徴量を話者内分散で割るという方法で、各特徴量の話者内の分布のばらつきを考慮している。

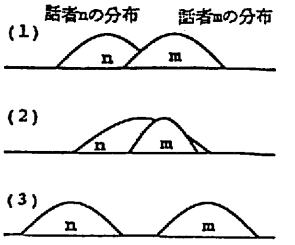
ここでは、各特徴量の話者間の分布のばらつきや、時期的な変動も考慮した正規化法を提案する。この方法では、正規化特徴量 $\vec{y}$ を、次に示す x_i に対する重み w_i を用いて、$y_i = w_i \cdot x_i$ で与える。

$$w_i = \frac{1}{T_i} \sum_{n=1}^{N} \sum_{m=1, m \neq n}^{N} \frac{\sigma_{ni}}{L_{mni}^2}$$

T_i は特徴量 x_i の時期的な変動量を表す尺度で、各話者毎に異なる時期のデータを用いて求めた標準偏差を、全話者について平均したものである。L_{mni} は話者 m と話者 n の特徴量 x_i に関する分布の重なりを示す尺度である。$1/T_i$ で時期差を吸収し、σ_{ni}/L_{mni}^2 で話者間のばらつきを強調し、かつ話者内のばらつきを吸収する。μ_{ni} と σ_{ni} は、各話者の1時期の学習データから計算する。

$$L_{mni} = \begin{cases} 3(\sigma_{mi} + \sigma_{ni}) - |\mu_{mi} - \mu_{ni}| & \text{if (1)} \\ 6 \cdot min(\sigma_{mi}, \sigma_{ni}) & \text{if (2)} \\ 1/|\mu_{mi} - \mu_{ni}| & \text{if (3)} \end{cases}$$

N	:	話者数
m, n	:	話者の識別子
μ_{ni}	:	話者 n の平均値
σ_{ni}	:	話者 n の標準偏差
(1), (2), (3)	:	下図に示す各状態に対応



*Speaker recognition based on the combination of pitch and vocal tract features.
By Tomoko Matsui and Sadaoki Furui (NTT Human Interface Laboratories)

2.3 歪・重なり尺度

時期差を含んだ、テキスト独立な話者認識では、テストデータの分布の中で、符号帳の分布から外れた部分が生ずることがある。この部分は、時期的な変動を受け易い部分、もしくは符号帳の作成には出現しなかった音韻が含まれている部分である可能性が大きい。

ここでは、類似度を計算する方法として、テストデータの分布と符号帳の分布の重なり部に含まれるテストデータのサンプル数、及びその平均量子化歪の両者を考慮する「歪・重なり尺度」を提案する。この方法では、正規化したテストデータ系列 $\{\vec{y}_j\}$ (サンプル数：J) と符号帳 $\{\vec{c}_k\}$ (符号帳サイズ：K) の (広義の) 距離を次式で表す。

$$\frac{\sum_{j=1}^{J} d_j + d_{out}(U_{Max} - U)}{U_{Max}}$$

$$d_j = \begin{cases} \min\limits_{k \in \Lambda_j} \| \vec{y}_j - \vec{c}_k \|^2 & \text{if } \Lambda_j \neq \phi \quad (4) \\ 0 & \text{otherwise} \quad (5) \end{cases}$$

$$\Lambda_j = \{k \mid 1 \leq k \leq K \text{ and } \| \vec{y}_j - \vec{c}_k \| \leq l_k\}$$

U : (4) に該当するサンプル数
U_{Max} : 全ての符号帳における (4) に該当するサンプル数の最大値
d_{out} : (5) に該当するサンプルのための固定距離
l_k : k番目のコードワードの覆う範囲
$\| \cdot \|$: ユークリッド距離

$\sum_{j=1}^{J} d_j$ は、テストデータの分布と符号帳の分布の重なり部の量子化歪の和を表す。重なり部は、符号帳に含まれる各コードワードに対して、それぞれの覆う範囲 l_k を定めることによって表す。サンプルがコードワードの覆う範囲に入れば、(4) のように歪を計算する。入らないならば、d_{out} とする。$d_{out}(U_{Max} - U)$ は、重なり部に含まれないサンプル数に比例している。重なり部のサンプル数が多い場合、「歪・重なり尺度」は小さくなる。

3 実験

3.1 概要

男性9名が約3年間に渡る4つの時期に、繰り返し発声した5母音、5単語、3文章 (平均時間長：A. 約5秒、B. 約12秒、C. 約30秒) を対象とする。ケプストラムは標本化周波数 12kHz、フレーム長 32ms、フレーム周期 8ms、LPC分析次数 15で抽出した。Δケプストラムはケプストラムの短区間 (88ms) の回帰係数である。ピッチは標本化周波数 12kHz、フレーム長 32ms、フレーム周期 8ms、LPC分析次数 15で、変形相関法により抽出した。Δピッチはピッチの短区間 (150ms) の回帰係数である。それらの特徴量を単独に、あるいは組み合わせて、一つの特徴ベクトルとする。

符号帳は、5母音、5単語、2文章 (A、B) を用い、LBGアルゴリズムで各話者、各時期毎に作成する。テストには、符号帳の作成に使った文章とは別の1文章 (C) を用いる (「テキスト独立」)。各話者、各時期毎にその1文章を、3つの異なる時期の符号帳に対してテストする。結果は全時期における、異なる時期の符号帳に対する平均識別誤り率で評価する。

3.2 認識結果

認識結果を表 1に示す。声道に関する物理的特徴量ケプストラムだけを用いるより、その動的な特徴量Δケプストラムを組み合わせて用いた方が平均誤り率が2.89 %低下する。更にそれらに、音源に関する物理的特徴量ピッチ、その動的な特徴量Δピッチを組み合わせた方が4.17 %低下する。「歪・重なり尺度」を適応すると、更に1.85 %低下することが分かった。

表 1: 認識結果

特徴量 (方法)	平均誤り率 %
ケプストラム	9.43
Δケプストラム	11.81
ピッチ	74.42
Δピッチ	85.88
ケプストラム, Δケプストラム (特徴量正規化)	6.54
ケプストラム, Δケプストラム ピッチ, Δピッチ (特徴量正規化)	2.37
ケプストラム, Δケプストラム ピッチ, Δピッチ (特徴量正規化＋歪・重なり尺度)	0.52

4 むすび

テキスト独立な話者認識において、声道に関する特徴量と音源に関する特徴量の組み合わせ、及び「歪・重なり尺度」が有効であり、時期的な変動がある場合でも、高い話者認識精度が得られることがわかった。

本研究に対して、御討論頂いた音声部、基礎研誉田グループの皆様に深く感謝いたします。

参考文献

[1] 古井貞熙, "音声波に含まれる個人性情報の研究," 東京大学学位論文 *(1978)*

[2] S.Furui, "Comparison of speaker recognition methods using statistical features and dynamic features," *IEEE Trans. ASSP,pp.342-350(1981)*

[3] F.K.Soong, et al., "A vector quantization approach to speaker recognition," *Proc. ICASSP,pp.387-390(1985)*

[4] F.K.Soong, et al., "On the use of instantaneous and transitional spectral information in speaker recognition," *Proc. ICASSP,pp.877-880(1986)*