ウェーブレット画像処理法による音声認識

妹尾 裕○、早野 誠治、斎藤 光吉（法政大学）

Voice Cognition by Wavelets Image Processing

ISENOO,S.,HAYANO and Y.SAITO

ABSTRACT

As is well known, voice cognition system of the personal computers becomes one of the popular command input methodologies. However, it is still remaining problem, which requires a hard training in order to build up a reliable voice database. Most of the voice cognition methods are based on simple correlation between the input and database voices. This leads to a relatively low cognition rate when using a poor trained database system. To overcome this difficulty, we are now developing a new voice cognition methodology which cognizes an input voice as a solution of an inverse problem. Our inverse methodology requires each of the characteristic patterns representing essential feature of the single pronunciation. In the present paper, we describe that the characteristic patterns representing essential feature of the single pronunciation are derived by means of the simple Lissajous diagram after the threshold operations. Using the database containing the characteristic patterns, as ill-posed linear system of equations for each pronunciation input is established. Simple linear squares leads to a good approximate solution of this system. Thus, we have succeeded in cognizing the single wavelet transform. As a result, it is revealed that the Daubechies' 2nd order base function makes it possible to reduce the database into 25% quantity.

Keywords: Voice Cognition, Eigen Pattern, Method of Least Squares
2. 固有パターン法による音声認識

2.1 固有パターンの定義

音声の固有パターンとは、音声が持つ固有の不変量が生成するパターンである。音声で固有パターンを定義するのには必要である。しかししながら、計算機を用いるより大きな空間である音声に関する具体的な変化を考慮し、これを可能な限り具体化する技術が必要である。従って音声が持つ固有の特性は、音声の何がどう、すなわち音声のレベル、音声の周波数、音声の形態、音声の波形などの特性から生成されるものである。音声の固有パターンは、音声の特性を表現するものである。音声の特性を表すために、音声の固有パターンの定義を以下に示す。

具体的な音声の特性を表す音声情報処理の簡単な方法を用いる。本論文では、音声の固有パターンの定義、音声の固有パターンの特性、音声の固有パターンの生成の方法などについて述べる。

2.2 システム構成

Fig.2に示す固有パターンを生成するため、音声の固有パターンを音声信号を入力信号にし、音声の固有パターンを出力信号にし、音声の固有パターンを生成するシステムが構成される。音声の固有パターンを生成するシステムは、音声の特性を表す音声情報を基に、音声の固有パターンを生成する。音声の特性を表す音声情報を基に、音声の固有パターンを生成するシステムは、音声の特性を表す音声情報を基に、音声の固有パターンを生成するシステムが構成される。音声の特性を表す音声情報を基に、音声の固有パターンを生成するシステムは、音声の特性を表す音声情報を基に、音声の固有パターンを生成するシステムが構成される。

\[ C = \{ c_1, c_2, \ldots, c_n \} \]  

（1）

いま、任意の音声信号の固有パターンを1次元元へ並べ替え得られる入力ベクトルを \( y \) とすれば、\( y \) がベクトル放出型方程式（2）式で与えられる。

\[ Y = \bar{X} \]  

（2）

（2）式で解ベクトルの要素を

\[ X = \{ x_{1}, x_{2}, \ldots, x_{n} \} \]  

（3）

とすれば、最大値を取る要素が識別された音声を生成することができる。

（1）式の固有パターンベクトル、それぞれに対応する音声信号を \( V_1, V_2, \ldots, V_n \) とすれば、（3）式の解ベクトル \( X \) をより生成された音声信号 \( G \) は（4）式で与えられる。

\[ G = \sum_{i=1}^{n} V_i x_i \]  

（4）
ウェーブレット音響処理による音声認識

\[ X = \sum_{n} X_n \]  

(4)

2.3 最小自乗法

(2)式のシステム方程式は式の未知数に対し、
64x64点のデータであり、64x64とする。全ての式
を同時に満たす解は特別な条件を満たして存在しない。
このため、観測ベクトルのルール

\[ e = Y - CX \]  

(5)

を最小にする解ベクトル、すなわち、最小自乗法による
解ベクトルを(6)式に計算する。

\[ X = (C^T C)^{-1} C^T Y \]  

(6)

50音の中の「ア」から「ロ」の45音に対する固有バタ
ーンを生成し、(6)式の解ベクトルを計算した。その
結果、システム生成生成に使用した音声を入力とした
場合、全てを元に認識できた。Fig.3に音声「ア」と
「イ」に対する解ベクトルを示す。Fig.4の図では、
認識された音声の波形は、他はすべてゼロである。
明らかにこれらの解は(2)式を厳密に満たす特
別な解である。従来は、システム生成生成に使用し
た音声を入力とした場合、(2)式を完全に満たす解
が得られる。

2.4 ウェーブレット変換による固有パターの圧縮

Fig.2に示す音声の各固有パターを2次の元で処理し、音声から、音声の特性を強調するウェーブレット変換の母数を用いて圧縮する。2次の元で配置されるウェーブレット変換で23%のデータを圧縮する。Fig.3に示す音声の音声を圧縮する。圧縮後のデータ

\[ X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \]  

(7)

Fig.3 Subsion vectors for the input voices

Fig.4 Recovered characteristic patterns of the voice signals "a"(upper) and "i"(lower) from the compressed 25% data by means of the Daxheko's 2nd order base function.

データから音声を抽出する固有パターを用いて(6)式の最
小自乗法による解を求める。Fig.4はドミンクの2次基本
関数を用いて圧縮された音声パターから得られた音声
「ア」から「イ」に対する解ベクトルの例である。それ
ぞれのデータを取る必要が既に認識された音声に対応するか
ら、全ての音声が正確に認識できた。しかし、Fig.4に示
す解ベクトルと異なり、2式のシステム生成生成は完
全に満たない。ドミンクの2次基本関数やColmanや
Rashinなどの基本関数を用いて同様の音声を計算を行ったが、
ドミンクの2次基本関数以外は全て正確に認識でき
なかった。この理由は、Fig.2に固有パターを既知す
れば、音声パターの特徴は比較的ピーカ値のデータに
現れるため、ウェーブレット変換の母数をどのピーカ値
が特徴を示す認識率の低下を招いていると考えられる。
ドミンクの2次基本関数の変数は矩形型であるため、比較的ピーカ値の音高を表現できる。
このため、他の音声に比較して満足的な認識率が得られたと考えられる。

3.まとめ

本論文では、音声情報から音声固有の情報を抽出し、音声
情報の特性を Hardcover音声情報に提案した。抽出した音声パターを用いてシステム生成生成を行った音声認識の問題を解決し、ウェーブレット変換の圧縮法で圧縮し、小規模のデータベースから音声認識の可能性を検討した。
その結果、ドビッシーの2次基底関数は25％のデータから100％の認識率を確保できることが判明した。
本論文の主な目的は、音声データのパターンを捉え
し、音声認識の可能性を検討する点にある。この意味
で、所期の目的は達成できたと考える。

参考文献
1) 崇明充武 著、ウェーブレット変換の基礎と応用
（朝倉書店、1996年4月）
2) H. Takahashi, S. Hayano, Y. Saito. “Visualization
Of The Currents On The Printed Circuit
Boards”. IEEE Visualization 1999, Late

Fig 5 Solution vectors for the input voices
"a", "b", "u", "e" and "o" where the characteristic
patterns have compressed into 25% data
quantity by means of the Daubechies's 2nd order
base function.