サウンドスペクトログラムにおける連鎖パタンマイニング—個人識別への応用—

サウンドスペクトログラムにおける連鎖パタンマイニング―個人識別への応用―

サウンドスペクトログラムにおける連鎖パタンマイニング
-個人識別への応用-

○三澤慶太　後藤廉太　岡田吉史（室蘭工業大学）

概要　我々は以前、複数の系列データに跨って繰り返される頻出パタンである連鎖パタンを抽出する方法を提案した。現在我々は、この手法を母音のサウンド スペクトログラムに適用し個人識別を試みている。本発表では、個人間の判別に寄与する周波数点を特定し、それを用いた場合の個人識別結果について 報告する。
キーワード: 話者認識，連鎖パタンマイニング，頻出パタン,マイニング,スペクトログラム

1. はじめに

話者認識技術は，ニュースや対談における内容の書き起こし，対話を行うロボット，個人認証などへの応用が期待され，盛んに研究されている[1].

本研究では，母音のサウンドスペクトログラム（以下，スペクトログラム）に含まれる時系列情報を用いた話者認識手法を提案する．既存のスペクトログラムを対象とした話者認識手法として，スペクトログラムの画像解析を用いた方法がある[2]．これは，発話者の特徴量としてスペクトログラムの画素値を用いている．一方，本研究で提案する手法は，連鎖パタンマイニング[3,4]と呼ばれる方法を用いて，個々人のスペクトログラムの時系列パタンを特徴量として話者認識を行う．連鎖パタンマイニングにより，異なる周波数点間に跨って繰り返されるスペクトルパワーの時間的推移パタン（連鎖パタン）を発見できる．

本稿では，母音「あ」のスペクトログラムから連鎖パタンを抽出し，これを特徴量とした話者認識の結果を示す．

2. 方法

Fig1 にスペクトログラムからの連鎖パタン抽出とユーザプロファイルデータベース（UPD）の構築方法を，Fig2 に話者認識の概要を示す．

2.1 ユーザプロファイルデータベース（UPD）の構築

2.1.1　スペクトログラムの作成

スペクトログラムは，縦軸と横軸にそれぞれ周波数と時間をとり，色の濃さでスペクトルパワーを表現した3次元グラフである．スペクトログラムは，音声データに対して窓処理を行い，切り出された分析区間に周波数分析を繰り返すことにより生成される．

2.2.2　周波数選択

スペクトログラムから個人の識別に寄与すると思われる周波数点のみを選択する．事前に各被験者から音声サンプルを収集し，それらのスペクトログラムを生成しておく．続いて，周波数点ごとにすべての被験者ペアでクラスカル・ウォリス検定を行い（有意水準α=0.05），有意差があったペアの個数をカウントする．なお，ここでのp値としてボンフェローニ法による補正後の確率が用いられた．最終的に，カウント数が全ペア数の60%以下の周波数点をスペクトログラムから削除し，それ以外を被験者間に差別化に有効な周波数点として使用することとする．

2.2.3　スペクトログラムからの連鎖パタン抽出

スペクトログラムからの連鎖パタンマイニングは次の手順で行われる．まず，周波数点ごとに，繰り返し現れるスペクトルパワーの時系列パタン（頻出パタン）を抽出する．次に，異なる周波数点に跨って同時刻帯に繰り返し現れる頻出パタンの集合を連鎖パタンとして抽出する．Fig1では連鎖パタン{(f1,A),(f2,B),(f4,C)}などが抽出されている．ここで，例えば(f1, A)は，周波数点f1で頻出パタンAを持つことを意味している．

2.2.4　連鎖パタンへのスコア付け

本研究では，種類数と出現頻度の高い頻出パタンを多く含む連鎖パタンほど，被験者を特徴付ける情報であるとみなす．そこで，スペクトログラムから抽出された連鎖パタンLのスコアを，(1)式に基づいて算出する．

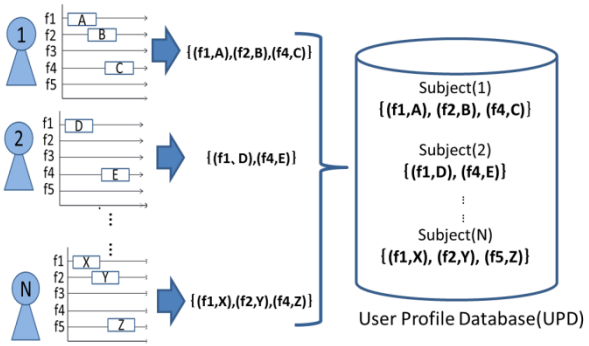

Fig 1: Linkage Pattern Mining and Creation of UPD

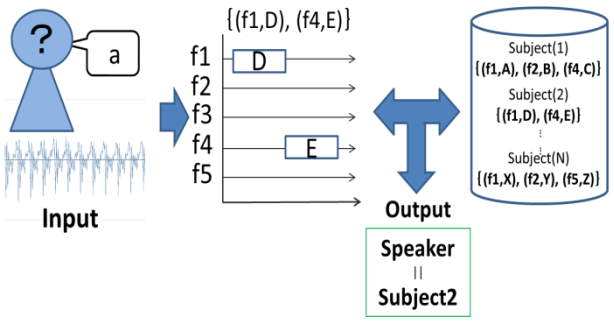

Fig 2: Speaker recognition

$$Score(L) = (no * np * r) / C \quad (1)$$

ここで, noは連鎖パタンLの出現頻度, npは連鎖パタンLに含まれる頻出パタンの数, rは連鎖パタンLと同じサイズの連鎖パタンをもつサンプルの割合を表す. また, Cは定数であり, 本研究ではC=5とした.

2.2.5 UPDへの連鎖パタンの格納

2.2.3節から2.2.4節の操作をすべての被験者の音声サンプルに対して行う. なお, 本研究では, 2.2.4節のスコアが上位10%の連鎖パタンのみをユーザプロファイルデータベース (UPD: User Profile Database) に格納した.

2.2 話者認識

まず, 認識対象となる発話者の入力音声に, 2.2.3節から2.2.4節の操作を施す. 次に, 抽出された連鎖パタンとUPDに格納された連鎖パタンを照合し, 最も類似した連鎖パタンを保持している被験者を発話者として出力する.

3. 実験

3.1 音声データの収集とスペクトログラム作成

男子大学生 5 名の母音「あ」の音声データを収集した. 母音の録音に使用した機材は OLYMPUS 製リニア OCM レコーダー(型番:LS-11)である. 各被験者には, 母音「あ」を 1 秒以上発音してもらい, 録音した音声から 1000 ミリ秒を切り出し, これを 4 分割した 250 ミリ秒の音声を 1 サンプルとした. 上記の作業は, 各被験者に対して 3 回ずつ行われた. その結果, 被験者ごとに 12 サンプル, 計 60 サンプルの音声データを収集した. 続いて, 各サンプルからスペクトログラムを作成した. なお, スペクトログラム算出のパラメータは, フレーム幅 5ms, シフト幅 0.5ms と設定した.

3.2 話者認識

本手法の話者認識精度を評価する. 実験手順は以下のとおりである. まず, 被験者 5 名の 12 サンプル (計 60 サンプル) から連鎖パタンを抽出し, スコア付けを行い、上位 10%の連鎖パタンを UPD に格納する. 次に, UPD から 1 サンプルを取り出して入力音声とし, 2.5 節の話者認識により出力結果の正誤を記録し集計する. 上記の実験を UPD 内の全てのサンプルに関して実行する.

4 結果・考察

Table1 は本実験による話者認識の結果を示している. 行が入力に用いた被験者番号, 列が出力結果の被験者番号である. 行列内の各数値は結果として出力された回数を示す. Table1 の右端の列は入力被験者ごとの正解率を示している. 全体での識別率は約36%となっており, 被験者の中では被験者1が一番高く約67%, 被験者3, 5が一番低く約17%となった. 認識結果の傾向として, 正解率の低い話者は正解率の高い話者に誤認されている.

Table1:Recognition Result

		Output					
		subject(1)	subject(2)	subject(3)	subject(4)	subject(5)	Recognition rate
Input	subject(1)	8	2	1	1	0	0.67
	subject(2)	5	3	1	3	0	0.25
	subject(3)	3	2	2	4	1	0.17
	subject(4)	2	1	1	7	1	0.58
	subject(5)	2	2	0	6	2	0.17

このような, 被験者間の識別率の差は, 抽出された連鎖パタン数や種類数の違いによるものと考えられる. 高い識別率が得られなかった理由として, 被験者間で共通の連鎖パタンが多数含まれることにより, 互いに誤認してしまう場合, 同じサンプル内で似通った連鎖パタンが残っており, その連鎖パタンがまとめて誤認されていることなどが考えられる.

5 まとめ

本稿では, 母音「あ」のスペクトログラムから得られる連鎖パタンを用いた話者認識手法を示した. 実験として, 被験者5名の音声を用いた話者認識を行った. 結果, 全体的な識別率は約36%と低く, 被験者間での識別率にもばらつきがあった.

今後は, 被験者間を差別化する連鎖パタン抽出の方法を考案する. また, 被験者間で抽出される連鎖パタン数にバラツキが出ないよう被験者ごとにパラメータ設定を変更して実験を行う. また, 類似した連鎖パタンを1つの連鎖パタンとしてまとめる処理を追加することにより, 連鎖パタンの質を上げて話者認識を試みる.

参考文献

1) 御船 正樹,鈴木 基之,人 福継ぎ,北 研二:“クラスタリングに基づく GMM 学習法による話者モデルの構築”, 電子情報通信学会技術研究報告. SP, 音声,2011
2) 和田 はるか,西 仁司:”音声スペクトログラムの画像解析による話者識別の研究”, 2012
3) 三浦貴大, 岡田吉史:”複数の系列データにまたがる頻出パタン抽出方法の提案”,第 14 回日本感性工学会,日本感性工学会,2012
4) Saerom Lee, Takahiro Miura and Yoshifumi Okada: “A New Method for Improving the Performance of Linkage Pattern Mining”, Proc. of IMECS 2014, International Association of Engineers, 36-40, 2014.