

## 平成19年度 メディア科学専攻修士論文要旨

大西 研究室	氏 名	久 富 慎 二
修士論文題目	事前学習を用いないオンラインでの話者識別	

人間は、誰が何を話しているかということ、聴覚系において、「話者認識」と「音声認識」という処理を行うことで、高い精度で認識することができる。このような人間の能力を、計算機を用いて実現しようと、多くの研究がなされている。しかし、話者識別における従来研究では、話者の特徴を事前に学習し、学習によって得られた知識を基に、話者を識別するという研究が大半である。そこで本研究では、オンライン処理により同一話者の発話をグルーピングすることを目的とする。

本研究では、「現在の発話が、過去に発話した話者と同一か否か」という問題を扱う。ここで、問題の簡単化のために、以下のことを仮定する。発話区間において話者は1人とする。話者は必ず非音声部分をはさんで交替する。短い発話は無視する。

話者識別法の処理の流れは次の通りである。まず、入力音声に対して、対数パワー  $lp$  とゼロクロス数  $Z_{cr}$  により、フレーム毎に音声/非音声を判定し、発話区間の検出を行う。音声/非音声の判定は、 $(lp \geq Th_{lp}$  または  $Z_{cr} \geq Th_{Z_{cr}}$ ) という閾値処理により行う。次に、各フレームの特徴量として、対数パワーとMFCC12次元を求める。さらに、特徴量の階層的クラスタリングを行うことで、発話クラスタ(発話区間)毎に特徴クラスタ  $FC$  を生成する。最後に、話者識別は、発話クラスタ  $UC$  間距離でなされる。発話クラスタ間距離は、各発話クラスタに属する特徴クラスタの組に関する特徴クラスタ間距離の平均について、特徴クラスタの全組み合わせに対する最小値とする。そして、この最小値が閾値以下ならば、2つの発話に同じラベルをつける。

以上の手法を計算機に実装し、2人の話者による対話形式の音声(90)を用い、話者数は未知として識別実験を行った。評価に際して、出力(話者インデクス)に対する、平均クラスタ純度(average cluster purity:  $acp$ )と、平均話者純度(average speaker purity:  $asp$ )という尺度を用いた。平均クラスタ純度  $acp$  は、ある発話クラスタがどれくらい一人の話者に限定されているかを表す尺度を与える。また、平均話者純度  $asp$  は、ある話者がどれくらい一つの話者クラスタに限定されているかを表す。ただし、 $acp$  と  $asp$  はトレードオフの関係にあるため、全体の評価基準として、 $acp$  と  $asp$  の積の平方根をとった値  $K$  を識別率とした。識別結果は下の表1のようである。全体での平均クラスタ純度  $acp = 0.97$ 、平均話者純度  $asp = 0.79$ 、識別率  $K = 0.87$  であった。

表1：識別結果

	$acp$	$asp$	$K$	クラスタ数
同性話者	0.94 (0.10)	0.79 (0.17)	0.86 (0.12)	4.62
異性話者	1.00 (0.00)	0.78 (0.15)	0.88 (0.08)	5.44
全体	0.97 (0.07)	0.79 (0.16)	0.87 (0.10)	5.06

同性話者：同じ性別の話者の組

異性話者：異なる性別の話者の組