1本のマイクロフォンを利用した同時発話音声認識*

○岡本 一輝，村上仁一，池原悟 (鳥取大)

1 はじめに

複数の話者が話したときに，各話者が話した音声の認識を行う場合，複数のマイクロフォンを用いる手法が一般的である [1]．しかし，人間では1つの耳だけで複数の音声を聞き分けることが出来る．このような単1のマイクロフォンで音声認識を行う研究例は少ない．

そこで本研究では，男女 2 話者が同時に発話した場合に，単 1 のマイクロフォンを使用した状況を想定し，認識率の調査を行った．まず男女個別のモデルを利用して，単純な方法で認識実験を行った．また，雑音環境における認識の手法である，Parallel Model Conbination 法を用いて認識実験を行った．

2 単純法

単純法では，男女 2 話者が同時に発話した音声に対し，男性話者，女性話者の HMM をそれぞれ利用して認識を行う．各々の認識結果に対して男性話者と女性話者が同時に認識できた場合の認識率を調査する．

3 Parallel Model Combination 法

Parallel Model Combination 法 [2](以下 PMC 法) とは，雑音が重畳した音声を認識する一般的な方法である．無雑音音声の HMM と雑音の HMM から目的の雑音環境の音声 HMM を合成し，認識を行う．図 1 に PMC 法のモデルを示す．

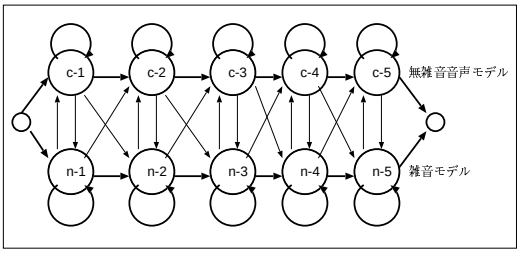

Fig. 1 PMC 法のモデル

4 評価実験

4.1 実験データ

本研究では，実際に男女 2 話者が同時に発話した音声を使用せず，各々の話者の音声を重畳した音声を利用する．具体的には，ATR 単語発話データベース Aset の男性話者 mau，mms 及び女性話者 ftk，fyn の男女各 2 名を使用する．偶数番号の音声の中から 4 モーラで発話時間がほぼ同じ語を，ランダムに 10 単語ずつ抽出する．それぞれを重畳した音声 (以下クロストーク音声) を作成する．1 セットにつき 100 単語のクロストーク音声を 4 セット作成し，テストデータとして利用する．奇数番号の音声は HMM の学習データとして使用する．表 1 に実験に使用した単語を示す．また，図 2 にクロストーク音声認識の手順を示す．

Table 1 実験に使用した単語

	男性話者	女性話者
1	悪質 (akushitsu)	足元 (ashimoto)
2	聞こえる (kikoeru)	可愛い (kawaii)
3	加える (kuwaeru)	勤勉 (kiNbeN)
4	失恋 (shitsureN)	細々 (komagoma)
5	優れる (sugureru)	すまない (sumanai)
6	そのうち (sonouchi)	対策 (taisaku)
7	中毒 (chuudoku)	手拭い (tenugui)
8	内容 (naiyou)	天才 (teNsai)
9	暴力 (bouryoku)	滅ぼす (horobosu)
10	わざわざ (wazawaza)	欲張る (yokubaru)

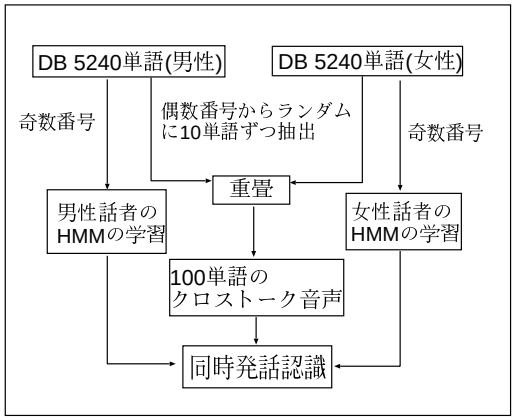

Fig. 2 クロストーク音声認識の手順

4.2 本研究における PMC 法

本研究では，HTK[3] を使用して実験を行う．しかし HTK で PMC 法を簡単に利用するのは困難である．そこで本研究では Parallel Model を構築する際，音声を音素単位で考え，各モーラごとに子音と母音に分け，子音は子音同士で母音は母音同士で相互にパスを持った PMC 法のモデルを 100 個構築する．認識実験においては各クロストーク音声に対し各 PMC 法のモデルの尤度を求め，最も尤度が高かった PMC 法のモデルを認識結果とする．図 3 に本研究での PMC 法のモデルを示す．

*Simultaneous speech recognition using single microphone. by OKAMOTO Kazuki, MURAKAMI Jin'ichi and IKEHARA Satoru(Tottori University)

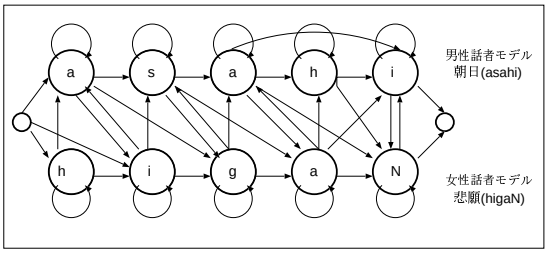

Fig. 3　本研究での PMC 法のモデル

図 3 は男性話者の音声に「朝日 (asahi)」，女性話者の音声に「悲願 (higaN)」を使用した場合の PMC 法のモデルである．

4.3　実験条件

実験環境を表 2 にまとめる．本研究では音声を加算性ノイズと仮定し，加算に有理な特徴パラメータである MELSPEC を使用する．また音素 HMM の共分散行列には Diagonal-covariance を使用し，認識率を調査する．

Table 2　実験条件

基本周波数	16kHz
分析窓	Hamming 窓
分析窓長	25ms
フレーム周期	10ms
音響モデル	3 ループ 4 状態・連続分布型
stream 数	2
混合分布数	母音・撥音・無音 4mixture 子音 2mixture
特徴 パラメータ	MELSPEC24 次+対数パワー (計 25 次)

4.4　実験結果

表 3 に各話者で組み合わせたクロストーク音声の単純法及び PMC 法の認識結果を示す．

Table 3　実験結果

話者	mau+ftk	mau+fyn	mms+ftk	mms+fyn	平均
単純法	8% (8/100)	8% (8/100)	8% (8/100)	17% (17/100)	10% (41/400)
PMC 法	30% (30/100)	45% (45/100)	24% (24/100)	32% (32/100)	33% (131/400)

PMC 法の方が単純法より認識率が高く，有効性が確認できる．また，ランダムサンプルでは認識率が 1%であることを考えると，33%の認識率は高いと考えている．

5　考察

5.1　人間による聴覚実験

実験に使用したクロストーク音声に対して，人間による聴覚実験を行い，認識率を求めた．なお，被験者は男性 3 名，女性 1 名である．表 4 に認識結果を示す．

Table 4　実験結果 (人間による聴覚実験)

話者	mau+ftk	mau+fyn	mms+ftk	mms+fyn	平均
認識率	77% (77/100)	87% (87/100)	74% (74/100)	70% (70/100)	77% (308/400)

表 3 と表 4 を比較すると，計算機による認識率は人間にはるかに及ばないことがわかる．また，計算機の方が話者ごとの認識率のばらつきが大きい．

5.2　特徴パラメータ

特徴パラメータに MFCC 及び FBANK を使用して，単純法における認識実験を行った．HMM には半連続分布型を使い，混合分布数は 128mixture とした．HMM の共分散行列には Diagonal-covariance 及び Full-covariance の 2 種類を使用した．表 5 に認識結果を示す．

Table 5　実験結果 (単純法)

話者	mau+ftk	mau+fyn	mms+ftk	mms+fyn	平均
MFCC Diagonal	41% (41/100)	49% (49/100)	50% (50/100)	44% (44/100)	46% (184/400)
MFCC Full	50% (50/100)	49% (49/100)	50% (50/100)	54% (54/100)	51% (203/400)
FBANK Diagonal	27% (27/100)	32% (32/100)	36% (36/100)	30% (30/100)	31% (125/400)
FBANK Full	63% (63/100)	49% (49/100)	56% (56/100)	47% (47/100)	54% (215/400)

この結果, FBANK Full-covariance が最も認識率が高く，認識率は 54%であった．今後は，FBANK Full-covariance に対して PMC 法を利用した認識実験を行う予定である．

6　おわりに

男女 2 話者が別々の音声単語を同時に発声する状況を想定し，単 1 のマイクロフォンを利用した音声認識における PMC 法の有効性を確認した．また，単純な手法を用いた場合の認識率も調査した．その結果，人間による聴覚実験と比べると認識率は非常に低いが，ある程度の認識率が得られることがわかった．今後は，様々な特徴パラメータで PMC 法を利用して認識率を調査する．

参考文献

[1] Panikos Heracleous, Satoshi Nakamura, Takeshi Yamada(Univ. of Tsukuba), Kiyohiro Shikano(NAIST)A Microphone Array-Based 3-D N-Best Search Method for Recognizing Multiple Sound Sources IEICE Transactions on Information and Systems Vol.E85-D, No.6, pp.994-1002, 2002

[2] M.J. Gales and S.J. Young, Robust Continuous Speech Recognition Using Parallel Model Combination, IEEE Transactions on Speech and Audio Processing, Vol. 4, No. 5, pp. 352-359, September 1996.

[3] Steve Young, Phil Woodland and Gunnar Evermann, HTK Book, Cambridge University Engineering Department 2002.