音声言語情報処理 39－6
(2001. 12. 20)

発声重複区間を含む対話音声の話者区分化に関する検討

小林 雅史† 北村 達也‡ 北澤 茂良‡

†静岡大学情報学研究科 ‡静岡大学情報学部
〒432-8011 静岡県浜松市城北 3-5-1

E-mail: {cs6036,kitamura,kitazawa}@cs.inf.shizuoka.ac.jp

あらまし 談話分析作業を支援するツールの開発を目的として，HMM を用いた発声重複区間を含む対話音声の話者区分化に関する検討を行った．発声重複区間の HMM を，（1）対象話者の同一発声の単語音声データを加算したデータを用いて学習する方法と，（2）対象話者の単語音声データをランダムに加算したデータを用いて学習する方法でそれぞれ構築した．これらの方法で発声重複区間の HMM を構築し，発声重複区間を含むように作成した評価用データに対して話者区分実験を行った．その結果，正しく話者区分化を行えた区間の割合が前者より後者の方が良い結果を得た．また，認識に用いる HMM ネットワークに関して不要な状態遷移を削除し，それぞれの状態遷移に遷移確率を付加することで，エラー数を減らすことができた．

キーワード 複数話者，話者認識，話者区分化，対話音声

A study of speaker segmentation of dialogue speech with speech overlapped section

Masahumi KOBAYASHI†, Tatsuya KITAMURA‡, and Shigeyoshi KITAZAWA‡

†Graduate School of Informatics ‡Faculty of Information, Shizuoka University
Johoku 3-5-1, Hamamatsu, Shizuoka, 432-8011 Japan

E-mail: {cs6036,kitamura,kitazawa}@cs.inf.shizuoka.ac.jp

Abstract In this paper, speaker segmentation of dialogue speech was studied to develop a support system for dialogue analysis. The dialogue speech addressed in this study was spoken by two speakers, and had speech overlapped section. To train HMM for the overlapped section, two different speech databases were constructed: (1) same words spoken by the two speakers were added; (2) pairs of different words spoken by the two speakers were added. These two training approaches were compared using dialogue speech data. The experimental result showed the latter have better performance. In addition, removing redundant arcs on the HMM network was improved the performance.

Key words multispeaker, speaker recognition, speaker segmentation, dialogue speech

1 はじめに

本研究の目標は，談話分析作業を支援するツールの開発である．本研究が対象としている談話分析作業が対象としている音声は，複数話者による対話を単一マイクで収録したものである．このような複数話者の対話音声に対して，各話者の発声区間を検出することができれば，談話分析作業を容易に進めることが可能になる．これを実現するには，話者区分化の技術を用いることが考えられる．話者区分化を行うことにより，話者交代のタイミングや，各話者の発声区間の長さなどを自動的に検出することも可能になる．このようなことから，話者区分化の技術は対象としている談話分析作業だけでなく，対談番組やニュース音声などにも用いることができ，各発声話者の発言に関するデータベースの作成などに応用できる．

これまでに，話者区分化に対して ergodic HMM を用いる研究[1]や，部分空間法を用いた研究[2]，話者交替の検出に対して GMM を用いる研究[3][4][5][6]や VQ 歪みを用いる研究[7][8]，などが行われている．しかし，談話分析が対象としている音声など複数話者による対話音声では，ある話者の発声が終わる前に別の話者が発声を始めたり，ある話者の発声の最中に別の話者が発声をするなど，発声重複が起こる場面は多いと考えられる．例えば，日本語地図課題対話において発声重複の占める割合が全発話中の 45%にも昇るという報告もされている[9]．実際にこのような発声重複を考慮に入れた研究[10][11][12]も行われており，話者区分化を行う際にはそのような発声重複区間が存在する状況を考慮する必要がある．このことを踏まえ，本研究は発声話者が既知という前提で，HMM を用いた発声重複区間を含む対話音声の話者区分化に関する検討を行っている．

本研究では発声重複区間の HMM を，同一単語加算とランダム単語加算の二つ方法を用いて構築した[13]．前者は，対象話者の同一発声の単語音声データを加算したデータを用いて学習する方法であり，後者は，対象話者の単語音声データをランダムに加算したデータを用いて学習する方法である．本稿では，これらの方法で構築した HMM による実験結果の比較を行った．

また，認識に用いる HMM ネットワークの変更を行った．全ての状態への遷移を許すものから不要な遷移を削除し，それぞれの状態遷移に遷移確率を付加することで，話者区分化の精度向上を試みた．

2 話者区分化

話者 A と話者 B の対話を例に話者区分化について説明をする．2 人の対話音声から，話者 A と話者 B の発声区間，無音区間をそれぞれ検出する作業を話者区分化と言う．また，本研究では発声重複区間を含む対話音声を対象としており，その区間に関しては発声重複区間として検出する．本研究における話者区分化の例を図 1 に示す．

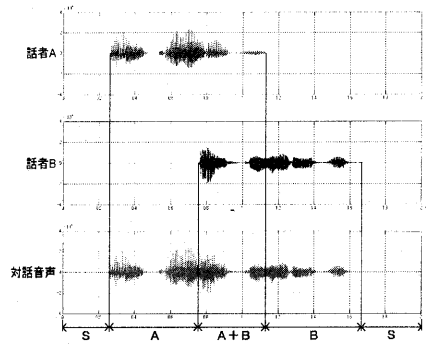

図 1. 話者区分化の例
A:話者 A，**B**:話者 B，**A+B**:発声重複，**S**:無音

3 予備実験

3.1 実験条件

予備実験として，発声重複区間を含まないデータを対象に話者区分化実験を行う．対象とする話者は男性 1 名と女性 1 名であり，抽出する区間は 2 話者の発声区間と無音区間である．予備実験の実験条件を表 1 に示す．

表 1. 実験条件

音声データ	ATR 単語データベース
話者	MHT，FAF
抽出区間	MHT，FAF，無音
標本化周波数	12 kHz
フレーム長	25 msec
フレーム周期	10 msec
特徴量	MFCC(12)+Power(1) 計 13 次
高域強調	$1-0.97z^{-1}$
HMM	1 状態モデル 4 混合
学習データ	音韻バランス語 各 216 発声
評価用データ	重要語 各 5240 発声
認識エンジン	HTK ver.3.0 [14]

3.2 評価方法

無音区間を含む評価用データの全区間に対して，正しく抽出された区間の割合を区間抽出率として求めることで評価を行う．区間抽出率の定義を以下に示す．

区間抽出率

$$= \frac{\text{正しく抽出した区間（時間）}}{\text{全区間（時間）}} \times 100(\%)$$

実験の結果，93.1%の区間抽出率を得た．

4 発声重複区間を含む音声を対象とした実験

4.1 実験条件

発声重複区間を含む音声を対象に，話者区分化実験を行う．予備実験と同様に男女各1名を対象話者とするが，抽出する区間は2話者の発声区間，発声重複区間，無音区間となる．実験条件に関して予備実験と異なる項目を表2に示す．ここでは，HMMの混合数の違いに関しても比較をする．

表 2. 予備実験からの変更項目

抽出区間	MHT，FAF，発声重複，無音
HMM	1状態モデル 1，4，8混合
評価用データ	重要語 5240 発声

4.2 発声重複区間の HMM の構築方法

同一単語加算とランダム単語加算の2つの方法を用いて，発声重複区間の HMM を構築する．それぞれの学習方法を以下に示す．

同一単語加算
対象話者の同一発声の単語音声データを加算したデータを用いて学習

ランダム単語加算
対象話者の単語音声データをランダムに加算したデータを用いて学習

また，どちらも学習方法も学習データの音声区間を振幅比1対1で加算する．

4.3 評価用データと評価方法

対象とする男女の単語音声データを加算することにより，発声重複区間を含む評価用データを作成する．

具体的には，予備実験で評価用データとして用いた重要語（各5240発声）の中から同一発声の単語音声データを振幅比1対1で加算する．ただし，全ての評価用データに対して，男性の発声区間，発声重複区間，女性の発声区間の順にそれぞれの区間が存在するようにするように，女性の単語音声データに160msec の遅延を与えたものと男性の単語音声データを加算する．

また，予備実験と同様に区間抽出率を求めることで実験結果の評価を行う．さらに，各区間を対象とした区間抽出率を求めることで，より詳細な評価を行う．例えば男性の発声区間の場合，以下のような式で定義される．

男性の発声区間における区間抽出率

$$= \frac{\text{男性の発声区間で正しく抽出した区間（時間）}}{\text{男性の発声区間（時間）}}$$

$\times$ 100 (%)

4.4 実験結果と考察

実験結果を図2に示す．

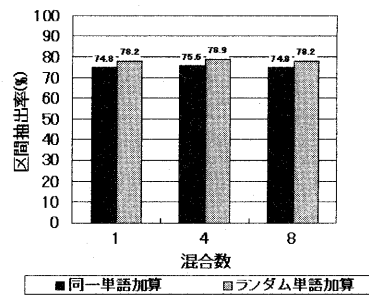

図 2. 実験結果

実験結果から，以下のようなことが分かる．

- どの混合数でも，同一単語加算よりランダム単語加算で発声重複区間の HMM を構築した方が，良い区間抽出率を得た．
- どちらの構築方法でも，混合数が4のときに最も良い識別率を得た．

また，各区間を対象とした区間抽出率を用いた検討を以下に示す．

発声重複区間の HMM の構築方法の比較

ここでは，発声重複区間の HMM の構築方法による比較として混合数 4 の結果を用いる．各区間を対象とした区間抽出率を図 3 に示す．

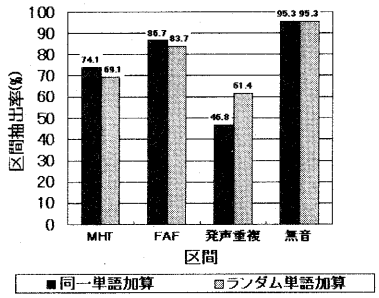

図 3. 発声重複区間の HMM の構築方法による比較

この結果から，他の区間に比べて発声重複区間における区間抽出率の差が大きいことが分かる．この区間の差が全区間における区間抽出率の差に現れていると考えられる．

混合数による比較

ここでは，混合数による比較としてランダム単語加算の結果を用いる．各区間を対象とした区間抽出率を図 4 に示す．

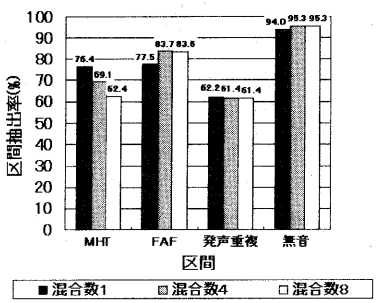

図 4. 混合数による比較

この結果から，混合数の変化は各話者の発声区間における区間抽出率のみに影響を与えることが分かる．また，混合数が増えても男女の発声区間で区間抽出率が増減しているため，適切な混合数を選ぶ必要があると考えられる．

5 HMM ネットワークの変更

5.1 HMM ネットワークについて [14][15]

HMM を用いた話者区分化は，話者の発声区間や発声重複区間など各区間ごとに学習した HMM を連結することで行われる．HMM の連結規則の表現方法や記述にはいくつか方法があり，HTK[14] ではネットワーク形式で表現することができる言語が提供されている．

5.2 HMM ネットワークの変更

まず，第 4 節までの実験で用いた HMM ネットワークを図 5 に示す．ここで，発声重複区間を含む音声を対象とする場合，以下のようなことが考えられる．

- 発声重複区間の後に無音区間は存在しない
- 無音区間や発声重複区間を介さずに話者交代は起こらない

これらのことを踏まえると，図 5 の HMM ネットワークには不要な遷移が存在する．そこで，そのような遷移を削除することで，より話者区分化に適した HMM ネットワークを作成する．新たに作成した HMM ネットワークを図 6 に示す．ここで，矢印の隣の数字は遷移確率を表す．

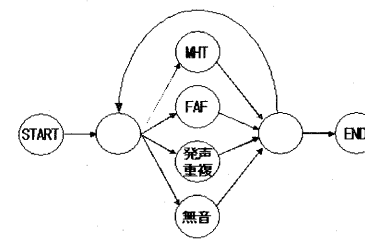

図 5. 等確率エルゴートネットワーク

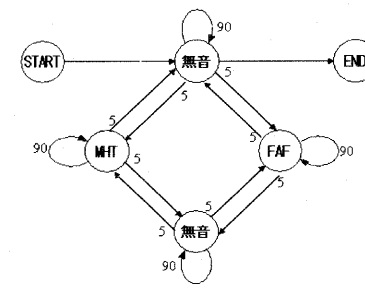

図 6. 条件付きエルゴートネットワーク

5.3 実験結果と考察

新たに作成したHMMネットワークを用いて，話者区分実験を行う．実験条件は基本的に第4節のものと同じである．ただし，HMMの混合数は4，発声重複区間のHMMの構築方法はランダム単語加算を用いる．HMMネットワークの比較として，以下に示すHMMネットワークを用いて実験を行う．

ネットワーク1
図5で示した等確率エルゴートネットワーク

ネットワーク2
図6で示した条件付きエルゴートネットワーク

ネットワーク3
図6で示したHMMネットワークで遷移確率を付加しないエルゴートネットワーク
（条件付き等確率エルゴートネットワーク）

実験結果を表3に示す．また，各区間を対象とした区間抽出率の結果を図7に示す．

表 3. 実験結果

	区間抽出率（%）
ネットワーク1	78.9
ネットワーク2	80.8
ネットワーク3	80.0

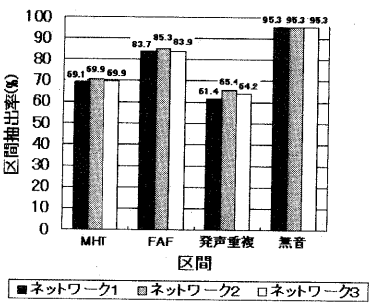

図 7. 各区間における区間抽出率の結果

全区間を対象とした区間抽出率は，高い順にネットワーク2，ネットワーク3，ネットワーク1という結果になった．また，各区間を対象とした区間抽出率においても，無音区間を除く区間全てが同様の結果となった．これらのことから，話者区分化により適したHMMネットワークを用いることと，状態遷移に遷移確率を付加することによる優位性が見られる．

6 同性話者同士を対象とした実験

6.1 実験条件

同性話者同士を対象とした話者区分化実験を行う．第4節と，第5節の実験と同様に発声重複区間を含む音声を対象とし，実験条件は基本的に第4節のものと同じである．ただし，HMMの混合数は4，発声重複区間のHMMの構築方法はランダム単語加算を用いる．また，HMMネットワークは第5.3節のネットワーク2を用いる．対象とする話者の組み合わせを表4に示す．

表 4. 対象話者の組み合わせ

	対象話者
男性	MHT，MMS
女性	FAF，FFS

6.2 実験結果と考察

実験結果を表5に，各区間を対象とした区間抽出率の結果を図8と，図9に示す．

表 5. 実験結果

	区間抽出率（%）
男性話者	72.0
女性話者	69.6

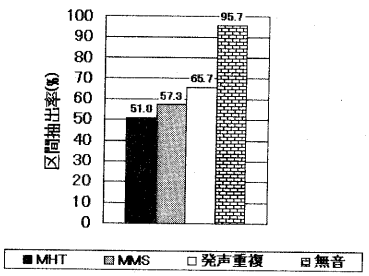

図 8. 男性話者同士の音声に対する各区間の区間抽出率

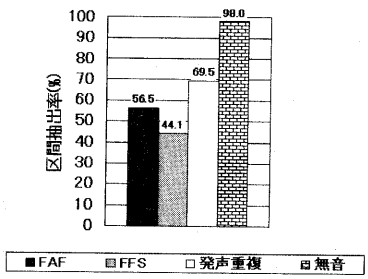

図 9. 女性話者同士の音声に対する各区間の区間抽出率

対象話者が男女のとき（第5節のネットワーク3の結果）と比較して，区間抽出率はどちらも低下する結果となった．このことから，同性話者同士の対話音声を対象とした方が話者区分化は難しいことが分かる．

また，各区間を対象とした区間抽出率の場合，発声重複区間と無音区間では対象話者の組み合わせに関わらず大差は無い．しかし，各話者の発声区間では区間抽出率が大幅に低下する結果となった．この差が全区間を対象とした区間抽出率において，対象話者が男女の結果との差になっていると考えられる．

7 おわりに

HMM を用いた話者区分化に関する各種検討を行った．具体的には以下に示す実験をし，結果の比較を行った．

- 同一単語加算とランダム単語加算による，発声重複区間の HMM の構築と混合数の比較
- より話者区分化に適した HMM ネットワークの作成と遷移確率の付加
- 同性話者同士を対象とした話者区分化実験

これらの実験を行った結果，以下のようなことが判明した．

- 同一単語加算より，ランダム単語加算で発声重複区間の HMM を構築した方が良い区間抽出率を得る．
- HMM の混合数が4のときに最も区間抽出率が良くなったが，混合数が変化することで区間ごとの区間抽出率が増減することがある．
- 話者区分化に適した HMM ネットワークを用いることと，遷移確率を付加することによる優位性が見られる．
- 同性話者同士を対象とした場合，各話者の発声区間の区間抽出率が低下する．

今後はこれらのことを踏まえ，以下に示すことを行う．

- 混合数の変化に伴う区間抽出率の変化と，エラーに関する調査
- HMM ネットワークに付加する適切な遷移確率を求める検討
- 同性話者同士を対象としたときの各話者の区間抽出率向上に関する検討

また，録音データなどより実音場の音声に近いデータを対象とした話者区分化に関する検討も行う．

謝辞 本研究の一部は（財）浜松科学技術研究振興会，および（財）スズキ財団の支援により行われたものである．

参考文献

[1] 村上仁一，杉山雅英，渡辺秀行，“HMM を用いた未知，複数信号クラスタリング問題の検討”，信学技報 SP92-74(1992).

[2] 西田昌史，有木康雄，“自動学習による話者セグメンテーション”，信学技報 SP97-57(1997).

[3] 中川聖一，岩井直美，山本一公，“話者の同定を組み込んだニュース音声の認識”，信学技報 SP99-33(1999).

[4] 村井則之，小林哲則，“統計的発話交代・話者モデルを用いた複数話者対話音声の認識”，信学技報 SP99-68(1999).

[5] 村井則之，小林哲則，“話者性と発話交代を考慮に入れた複数話者対話音声の認識”，信学論 D-Ⅱ，Vol.J83-D-Ⅱ，No.11(2000).

[6] 張子鵬，古井貞熙，“話者交代検出を含むオンライン話者適応の検討”，音響講論 (秋)，1-1-23(1999).

[7] 森一将，山本一公，中川聖一，“発話間の VQ 歪みを用いたオンライン話者交替識別と話者クラスタリング”，信学技報 SP2000-18(2000).

[8] 森一将，中川聖一，“ニュース音声における VQ 歪み尺度を用いた話者交替検出と話者クラスタリングの評価”，音響講論 1-5-5(2000).

[9] 榎本美香，土屋俊 “オーバーラップ発話の評定方法とその基礎統計 –日本語地図課題対話を通して–”，信学技報，SP99-117(1999).

[10] 滝口哲也，西村雅史，“HMM 合成法を用いた混合音声の認識”，音響講論 (秋)，2-Q-12(2000).

[11] 滝口哲也，西村雅史，“HMM 合成と遅延アレーの統合による混合音声の認識”，音響講論 (春)，3-P-9(2001).

[12] 斉藤洋平，古井貞熙，“対談音声を対象とした音声認識の検討”，音響講論 (春)，1-3-17(2001).

[13] 小林雅史，北岡教英，北村達也，北澤茂良，“HMM を用いた対話音声の話者区部化に関する検討”，音響講論 (春)，1-3-2(2001).

[14] http://htk.eng.cam.ac.uk/

[15] 鹿野清宏，伊東克亘，河原達也，武田一哉，山本幹雄，“音声認識システム”，オーム社 (2001).