5T-3

話者の負担を考慮した話者識別と音響モデルの検討

楠 和馬 †　　奥村 紀之 †
† 香川高等専門学校情報工学科

1　はじめに

話者識別を行うためには，話者の音声データが必要である．本研究は，話し手が切り替わる際の特徴をモデル化し，様々な特徴をもつ人物に対応する汎用性の高いシステムを目指している．

本研究では，アンケート調査を実施し，対話独特の話し手の特徴を抽出している．そして，GMM を作成し，特徴の識別結果を報告する．

2　話し手が切り替わるきっかけに関する調査

人は対話をする際，話し手は聞き手に対して配慮 (口調の速さ，イントネーション，強弱，等の変化) をする．それにより，聞き手があいづちをうったり，否定肯定をしたり，自らが話し手になるきっかけになる．この話者が切り替わるきっかけが検出できれば話者を識別する手掛かりになり，話者識別において有用なものになると考えられる．

2.1　アンケート内容について

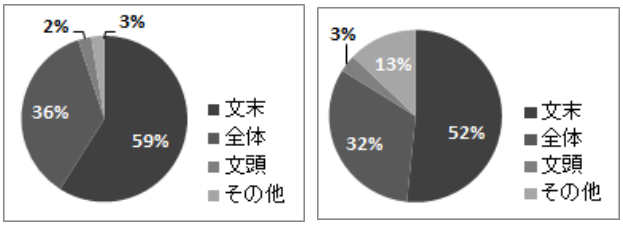
アンケート調査は男性 10 人と女性 5 人を対象に行った．また，回答者を聞き手と話し手に分けて行った．話し手と聞き手に対する質問の内容は以下に示す．

- 話し手
  - 聞き手に反応して欲しいときはどのように音声特徴に変化を加えるか
  - 音声区間 (文末，文頭，発話音声全体など) のどこを変化させるか
  - 各回答に適切な対話例はどのようなものか
- 聞き手
  - 話し手がどのように音声特徴に変化を加えた時に反応をしやすいか
  - 音声区間のどこを変化させているか
  - 各回答に適切な対話例はどのようなものか

このアンケートより得られた回答から，話し手が切り替わるきっかけの特徴を抽出する．次節にアンケートの回答から統計をとった結果を提示する

2.2　調査結果

まず，声情報を変化させる音声区間について着目した．図 1 は話し手と聞き手の回答に含まれていた音声区間を集計し，円グラフで表したものである．集計データから，対話する際に話し手が切り替わるきっかけになるのは文末に多いことが分かった．



話し手　　聞き手

図 1: 回答から集計した各音声区間の割合

次に，音声的特徴をどのように変化させるかについて着目した．表 1 は音声特徴の変化の加え方を集計し，3 個以上重複していた回答を降順で並べた結果である．話し手は「文末を長く伸ばす」,「文末の声調を上げる」が多く回答されており，聞き手は「文末を強く発音した時」,「文末の声調を上げた時」が多く回答されていた．また，話し手と聞き手が共通して「文末の調子を上げる」と多く回答していることが分かる．

表 1: 回答から集計した音声的特徴の度数

話し手

音声特徴の変化させ方	度数
文末を長く伸ばす	5
文末の声調を上げる	5
文末を弱くする	3
文末を声調を下げる	3
発話全体の声調を上げる	3

聞き手

音声特徴の変化のさせ方	度数
文末を強く発音	5
文末の声調を上げる	5
文末を弱く発音	3
発話全体の声調を上げる	3

3　音響モデルの作成と学習

この節では，前節のアンケートで得られた結果を用いて音響モデル作成・学習までの作業について報告する．

Speaker Identification with Consideration of Speaker's Burden and Examination of an Acoustic Model
†Kazuma KUSU　†Noriyuki OKUMURA
†Kagawa National College of Technology, Department of Information Engineering

対話音声から話し手がの切り替わるきっかけを検出するために，音響モデルは混合ガウス分布モデル (GMM) を用いる．音声データは，国立国語研究所・情報通信研究機構（旧通信総合研究所）・東京工業大学が共同開発し公開している日本語話し言葉コーパス (CSJ) を用いた (表 2)．CSJ 内にある対話音声データから「文末の声調が上がっている時」,「文末を長く伸ばしている時」,「文末が変化せず終わった時」の音声を切り出したものにラベリングを行い，それらを学習データとした (表 2).

表 2: 音声データ・特徴量

音声データ	CSJ(対話音声)
標本化・量子化	16kHz，16bit
分析条件	16kHz，16bit フレーム長 25msec フレーム周期 10msec
窓	ハミング窓
特徴量	25 次元 MFCC(12), Δ MFCC(12) Δ Power(1)

GMM の作成・学習には HTK(HMM Tool Kit) を用いた．分析条件，特徴量については表 2 に提示する．

4 実験

前章で作成した「文末の声調が上がっている時」,「文末を長く伸ばしている時」,「文末が変化せず終わった時」の GMM を用いて，それぞれの音声特徴を正しく識別できるか実験した．

4.1 実験概要

この実験では判別デコーダとして大語彙連続音声認識エンジン *Julius* を使用した．実験に扱うテストデータは学習データと同じ話者のデータを用いた．テストデータをあらかじめ「文末の声調が上がっている時」,「文末を長く伸ばしている時」,「文末が変化せず終わった時」で学習データと同じように音声を切り出し分類しておく．そして，分類したデータをそれぞれ *Jullius* で音声認識を行い，識別率を計算する．また，混合数別で以上を行った．

4.2 実験結果

識別結果を図 2 に示す.「文末の声調が上がっている時」,「文末を長く伸ばしている時」,「文末が変化せず終わった時」をそれぞれ，up，long，non とする．

混合数 1 から混合数 4 にかけて識別率は急激に落ちている．混合数 8 から徐々に識別率が上がり，最終的に 128 では混合数 1 の時とほぼ変わらない値となった．また，long や up に比べ non は遥かに識別率が低い結果となった．

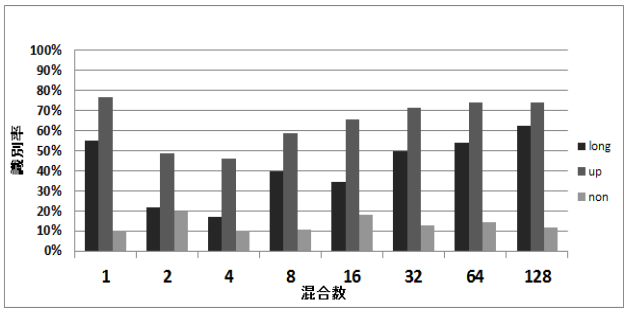


図 2: 識別結果

4.3 考察

図 2 より，今回アンケートより得られた long と up の特徴は実生活において有用なものであるため高い識別率を期待したが，それぞれ 60%台，70%台となり，望んだ結果は得られなかった．後に，アンケートの回答にもあった「語尾を強く (弱く) 発音する」といった特徴を考慮するためには，特徴量抽出の設定をより考察する必要があることも推測できる [1, 2].

また，混合数を多くしていくと，識別率が急激に落ち，徐々に識別率は再度向上した．急激に落ちた原因は，徐々に混合数の差が開くに連れて，識別率が高くなっていることから，過学習により学習データに特化してしまったことが疑われる．

5 おわりに

本稿では，アンケート調査を行い，集計した結果から話し手の文末に加える音声特徴の変化を GMM を用いてモデル化した．実験を行った結果，特徴量抽出を考察することによって識別精度を高めることができる事が分かった．

今後は，話し手 GMM の識別精度をさせていく．そして，聞き手 GMM もまた調査により作成していく．

謝辞

本研究の一部は研究費 (23720222) の助成を受けたものである。

参考文献

[1] 峯松信明，広瀬啓吉，関口真理子 (2002)．話者認識技術を利用した主観的高齢話者の同定とそれに基づく主観的年代の推定．情報処理学会論文誌，Vol.43 No.7, pp.2186-2196

[2] 水野寛之，竹内　伸一，田村　智嗣，速水　悟 (2009)．話者認識技術を利用した主観的高齢話者の同定とそれに基づく主観的年代の推定．情報科学技術フォーラム，E-047，FIT2009，pp.363-364