

CSJ を用いた話者交代時の発話特徴抽出

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

本研究の目的は単一マイクロフォンによって会話中の発話者の識別を行うことである。そのためには、話し手の切り替わりを検出する必要がある。そのために、本研究では話し手の発話特徴に着目し、話し手が切り替わる前の話し手の発話特徴を抽出する。

2 関連研究

西田らの研究^[1]は発話特徴については着目していない。我々は話し手が発する発話特徴が話し手の切り替わり予測には重要であると考えている。そこで、話し手の切り替わり直前に話し手が発声する発話特徴をアンケートにより取得し、モデル化を行い、各モデルの発話特徴中の識別率を評価した。^[2]アンケートにより取得した発話特徴を表1に示す。

表 1: 取得した発話特徴リスト

発話特徴の変化方法	
文末を長く伸ばす	文末の声調を上げる
文末を弱くする	文末の声調を下げる
文末を強く発音する	文末を早く喋る
全体を強く発音する	全体の声調を下げる
文末を止める	全体の声調を上げる

そして、研究成果を報告したところ、以下のような問題の指摘を受けた。

1つ目に、各発話特徴が対話中に話し手の切り替わりを検出する手がかりとしてどれほどの有用性をもっているかを調査すること、2つ目に、話し手の発話特徴が複数同時に発声されることについて考慮しているか、最後に、声調などの観測をする場合に用いられる基本周波数の利用をしていないことについてである。以上の問題点の改善方法を検討していく。

3 話者交代時の発話特徴の抽出

本節では、話者交代前に話し手の発話から観測される発話特徴を実際の対話音声データから抽出する行程について記述する。また、その対話音声を用いて各発話特徴の有用性評価実験を行う。

3.1 話者切替音声データの準備

話者交代前に実際話し手が用いる発話特徴を得るために口語調の対話音声データを準備する。話し手の切り替わりを含む音声データは日本語話し言葉コーパス (CSJ; Corpus of Spontaneous Japanese) に同梱してある対話音声データ (以下、CSJ 対話音声) を利用する。

また、切り替わりポイントの評価を行いやすいように、CSJ 対話音声を1回の話し手の切り替わりを含むように分割する。この分割作業には、CSJ 同梱の発話領域情報が記録された転記ファイルである TRN ファイルを利用した。TRN ファイルには単位情報部という転記箇所があり、単位情報部には発話 ID、発話開始・終了時間、マイクチャンネルの情報が含まれている。単位情報部のマイクチャンネルに注目し、マイクチャンネルが変化するごとに対話音声データを分割した。そのため、最低1つの話者の切り替わりが分割音声データには含まれる。

3.2 試聴による発話特徴の抽出

3.1 節で作成した分割音声データを利用し、話し手が切り替わり前に用いている発話特徴を抽出する。発話特徴の抽出には実際に分割音声データを試聴し、話し手交代前にどのような発話特徴が発声されたか抽出する。この作業は筆者を含む9名の被験者で行った。また、実際に試聴する分割音声データは全分割音声データの中からランダムで選択した100個の話し手の切り替わりポイントである。

分割音声データから発話特徴の抽出を行う際に以下のようなルールを設けている。

- 話し手は初めに喋り出した人とする。
- 最初の発言者が相槌だけでも話し手とする。
- 話し手の切り替わりが複数存在するときは1番初めの切り替わりに注目する。
- 話し手の発話中に聞き手が割り込んできた場合、割り込み前までを抽出領域とする。
- 2人の発話が重複していても可能であれば評価を行う。

また、分割音声データから抽出する発話特徴は表1に加え、被験者が新しく見つけた発話特徴についても挙げるようにしている。そして、以上の作業から得た発話特徴の情報をもとに分割音声データにラベリングを行っている。

The Utterance Feature Extraction For Detection Of Speaker-Changing Using Corpus Of Spontaneous Japanese

[†] Kazuma KUSU(c09093@sr.kagawa-nct.ac.jp)

^{††} Noriyuki OKUMURA(okumura@di.kagawa-nct.ac.jp)

Department of Information Engineering Kagawa National College of Technology ([†])

4 話者交代時の発話特徴のモデル化

本節では、3.2節で抽出した発話特徴の結果から2節で挙げた指摘点について考慮しながら、各発話特徴をモデル化する。

4.1 各発話特徴の出現確率

全試聴分割音声データ中に、各発話特徴が話し手の切り替わり前に現れた分割音声データファイル数を全試聴音声データ数で除算したものを各発話特徴の出現確率とする。図1に各発話特徴の出現確率を示す。

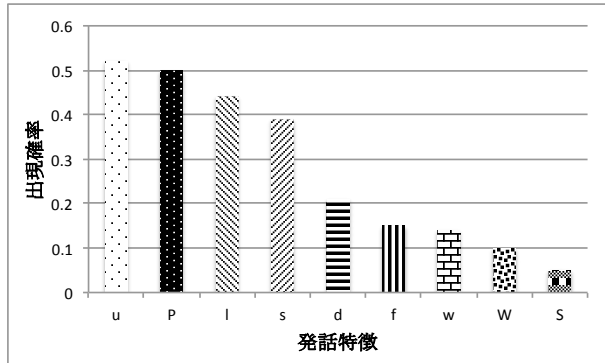


図1: 各発話特徴の出現確率

また、図中のアルファベット1文字はそれぞれ発話特徴を表しており、各文字が対応する発話特徴は表2に示す。

表2: 文字と発話特徴の対応表

文字	発話特徴	文字	発話特徴
u	声調を上げる	w	声量を小さく
d	声調を下げる	W	声が震える
l	伸ばす	f	喋りを速く
s	止める	S	喋りを遅く
P	声量を大きく		

この出現確率結果が大きい発話特徴ほど、話し手の切り替わりを予測する手がかりとして効果のある発話特徴と言える。

4.2 発話特徴の競合

3.2節でラベリングを行った結果から、複数の発話特徴同時に発声される発話特徴の組を挙げる。分割音声データ内で同時に発声された発話特徴を表3に示す。

表3: 性別組別の対話音声データ数

発話特徴の組	回数	発話特徴の組	回数
u & P	31	s & P	13
u & l	14	s & d	12
u & s	12	N & w	10
l & N	18		

4.3 基本周波数の利用

CSJは収録されてある音声データの各発話の基本周波数が記録されてある。本研究では、CSJに記録されてある基本周波数を利用する。

4.4 考察

図1に示すように、話し手の切り替わり直前によく発声されるのは「声調をあげる」である。「声調を上げる」は文中や文末に発声されることが多く、疑問や質問を聞き手に投げ掛けるときによく用いられることが対話音声を試聴して多かった。そのため、「声調を上げる」は発話の意図が疑問や質問だと予測する手がかりになる。また、「声調を上げる」は疑問や質問だけでなく、話し手の機嫌がいい時などに左右されるためか、「声量を大きく」と一緒に発声されることも多く見ることができている。

発話特徴によって、発話の意図がある程度読み取ることが出来たり、発話特徴の組み合わせによって、発話の盛り上がりなど発話の状態について得られることが期待できる。それらを定型化することにより、話し手の切り替わるかもしれないと予測できたり、聞き手の割り込みを予測することも期待できる。

5 おわりに

CSJ対話音声データを用いて、話し手が切り替る直前の話し手の発話特徴を抽出した。その結果、話し手の切り替り直前に発声されやすい発話特徴や同時に発声される発話特徴の組を示した。これらの結果と実際に対話音声を試聴し、各発話特徴や発話特徴の組み合わせには話し手の発話の情報が含まれていると考えた。そのため、各発話特徴から得られる発話の情報は話し手の切り替り予測の手がかりとして成り立つことが期待できる。

しかし、各発話特徴の有用性については出現確率だけでは明らかに出来ないため、実際にモデル化を行い、話し手の切り替りポイント予測実験等を今後行う必要があると考えられる。

参考文献

[1] 西田昌史, 緒方淳, 有木, 情報処理学会論文誌, pp.2205-2213, Vol.43, No.7, (2002) 康雄:音声認識と話者認識を統合した話者の人名付与システム

[2] 楠和馬, 奥村紀之:話者の負担を考慮した話者識別と音響モデルの検討, 情報処理学会第75回全国大会, 5T-3, (2013)