モーラ情報およびアクセント位置をもちいた単語音声認識 *

○堀田波星夫, 村上仁一, 池原悟 (鳥取大 · 工)

1 はじめに

従来の単語音声認識においては, 主に音声の音韻的特徴が用いられてきた. しかし, 日本語では,「箸」,「橋」のような音韻的には同一だがアクセントの違いによって弁別できる単語が存在する. 過去の研究において, 韻律的特徴を用いた研究としては, 高橋ら [1] の研究があるが, 同音異義語の音声認識の研究はあまり行われていない [2] .

そこで本研究では, アクセントの情報を利用して同音異義語の認識精度を調査する. また, 特徴パラメータに一般的に使用されている MFCC は音韻的特徴しか含んでいない. そのため, 韻律的情報を含む特徴パラメータとして FBANK を用いて認識精度を調査する.

2 アクセントとモーラ情報

本研究では単語のアクセント型の情報と各モーラ位置のアクセントの高低の情報を同音異義語認識に用いる. なお, 研究において単語のアクセントは NHK 日本語発音アクセント辞典 [3] を利用する. また, モーラ数とモーラ位置をあわせたものをモーラ情報と定義する. なお, モーラ情報を用いると単語音声認識の精度が向上することが知られている [4] .

3 アクセントに対する特徴パラメータ

従来の音声認識の特徴パラメータに用いられている MFCC には韻律情報が含まれていないため, 同音異義語を認識する実験において認識精度が低いと予想される. そのため, 本研究では, 韻律情報が含まれている FBANK を特徴パラメータとして使用する. なお,FBANK を用いると MFCC より単語音声認識精度が向上することが知られている [5] .

4 評価実験

本研究では音素 HMM に単語のアクセント型と各モーラ位置のアクセントの高低の情報を加えたモデル (以下, アクセントモデル) を提案する. そしてアクセントモデルの有効性検証のために, 単語音声認識実験を行い, 同音異義語の認識率を調査する. なお, 通常の音素ラベルを用いて学習した音素 HMM を基本モデルとする.

4.1 ラベル分類

アクセントモデルでは, 母音, 撥音, 促音音素の後ろ 2 桁の数字で単語のモーラ数, さらにその後ろ 2 桁の数字でその音素のモーラ位置を表現する. 加えて, そのうしろ 2 桁がアクセントの型を表現する. そして, 最後の 1 桁はそのモーラ位置でのアクセントの高低を示し,0 か 1 とする.0 は低を 1 は高を表現する. アクセントモデルのラベルの分類例を表 1 に示す. 例のモーラ位置 1 の音素表記の説明を表 2 に示す.

表 1: 音素ラベルの分類例

単語:間 (音素列 aida アクセント辞書表記 011)

基本モデル	a	i	d	a
アクセントモデル	a0301000	i0302001	d	a0303001

表 2: ラベル表記

a	03	01	00	0
	単語のモーラ数	単語のモーラ位置	アクセント型	アクセントの高低

4.2 音素 HMM の作成

HMM は初期モデルが重要であるため, アクセントモデルの初期モデルは基本モデルから作成する. また, パラメータを同一にして実験を行うために, 半連続型 HMM を使用する [6].

実験手順を図 1 に示す. 初めに図のように基本モデルを作成する (図中 a). 次に作成された基本モデルの HMM を複製してアクセントモデルの初期モデルとする (図中 b). 最後に連結学習を行いアクセントモデルの HMM を作成する (図中 c).

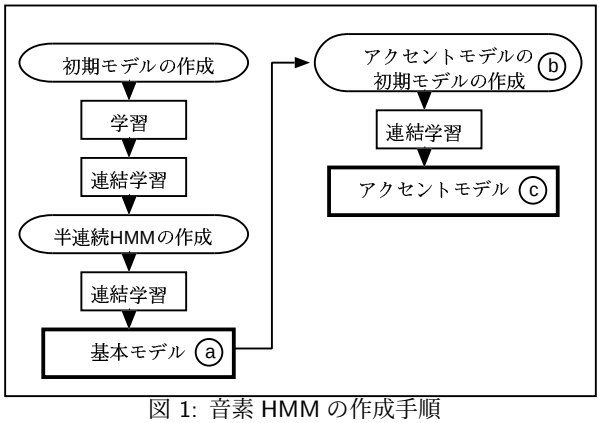

図 1: 音素 HMM の作成手順

4.3 学習データと評価データ

データベースには ATR 単語発話データベース Aset の 5240 単語を用い, 奇数番を学習データに, 偶数番を評価データに用いる. 評価データ中にはアクセントが異なる同音異義語が 11 組ある. 実験で用いられる同音異義語を表 3 に示す. 表中の括弧内の数字の 0 はアクセントの低,1 は高を意味する. 評価データ 2620 単語を音声認識し, その中の同音異義語に注目する. なお, 学習データ中にアクセント辞典から決定したアクセントと人手で聴取したアクセントが異なるデータを確認したが, 数が多いためにアクセントの訂正は行っていない. 一方, 評価データ中の同音異義語のアクセントは人手による聴取結果と一致することを確認した.

表 3: 評価データ中の同音異義語の対

1.	居る (01)	射る (10)
2.	代える (011)	返る (100)
3.	欠ける (011)	駆ける (010)
4.	機嫌 (011)	起源 (100)
5.	公開 (0111)	航海 (1000)
6.	置く (01)	億 (10)
7.	指名 (011)	氏名 (100)
8.	度 (01)	足袋 (10)
9.	徳 (01)	解く (10)
10.	付ける (010)	漬ける (011)
11.	因る (01)	夜 (10)

∗ Word Speech Recognition using Mora Infomation and Accent
By Haseo Hotta, Jin'ichi Murakami and Satoru Ikehara(Faculty of Engineering, Tottori University)

4.4 実験条件

評価実験は, 男性話者 3 名と女性話者 3 名で行う. 実験には単語音声認識ツールの HTK [7] を使用する.HMM の共分散行列には Diagonal-covariance(以下,Diagonal) と Full-covariance(以下,Full) の 2 種類を使用する. その他の実験条件を表 4 に示す。

表 4: 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 半連続分布型
stream 数	3
MFCC 特徴ベクトル	12 次 MFCC+Δ 12 次 MFCC +対数パワー+Δ対数パワー (計 26 次)
MFCC 混合分布数	MFCC 128 Δ MFCC 128 対数パワー, Δ対数パワー 16
FBANK 特徴ベクトル	24 次 FBANK+Δ 24 次 FBANK +対数パワー+Δ対数パワー (計 50 次)
FBANK 混合分布数	FBANK 128 Δ FBANK 128 対数パワー, Δ対数パワー 16

5 実験結果

5.1 同音異義語の認識精度

Diagonal で行った同音異義語認識の実験結果を表 5 に,Full で行った実験結果を表 6 に示す. 表中の括弧内の分母は評価データ中の同音異義語の数である. そして, 分子の数字は誤認識した同音異義語を示す.

表 5: Diagonal を用いた同音異義語の誤り率

話者	MFCC	FBANK
mau	18%(4/22)	5%(1/22)
mmy	14%(3/22)	18%(4/22)
mnm	14%(3/22)	9%(2/22)
faf	0%(0/22)	0%(0/22)
fms	14%(3/22)	18%(4/22)
ftk	5%(1/22)	5%(1/22)
平均	11%	9%

表 6: Full を用いた同音異義語の誤り率

話者	MFCC	FBANK
mau	9%(2/22)	5%(1/22)
mmy	9%(2/22)	0%(0/22)
mnm	5%(1/22)	9%(2/22)
faf	0%(0/22)	0%(0/22)
fms	14%(3/22)	5%(1/22)
ftk	9%(2/22)	0%(0/22)
平均	8%	3%

実験の結果,Diagonal,Full 共通して FBANK を用いた特徴パラメータの方が MFCC よりアクセントの認識精度が高かった. また比較的多くの同音異義語を誤認識していた話者 mmy,fms の認識率が FBANK の Full では,Diagonal と比較して改善されていた. 最も同音異義語を認識できた実験では平均 97%の精度が得られた.

5.2 単語音声認識精度

基本モデルとアクセントモデルの単語音声認識の実験結果を表 7 に示す. 表中の括弧内の分母は 6 話者の評価データ数である. なお,Full の FBANK の実験結果では, 学習データの不足で作成されなかった音素が存在するので分母が異なっている. 括弧内の分子の数字は誤認識した単語数を示す. なお, アクセントモデルにおいて同音異義語に誤認識している認識結果は正解として集計している.

表 7: 6 話者平均の単語音声認識の誤り率

	基本モデル	アクセントモデル
Diagonal MFCC	7.22% (1135/15720)	4.29% (675/15720)
Diagonal FBANK	10.25% (1611/15720)	7.32% (1150/15720)
Full MFCC	5.23% (822/15720)	3.38% (532/15720)
Full FBANK	5.48% (858/15666)	3.29% (515/15666)

アクセントモデルの単語音声認識精度は, 基本モデルより高かった. 最も単語音声認識精度が高かったのは,FBANK の Full でのアクセントモデルの実験で 6 話者平均 96.71% の精度が得られた. 一方, 同条件の基本モデルは平均 94.52%の精度であった.

実験結果よりアクセントモデルは単語音声認識に対しても効果があることを確認した.

6 考察

6.1 同音異義語の誤認識

全ての実験条件において, 同音異義語の誤認識としては, モーラ数 2 の高低のアクセントの同音異義語と, モーラ数 3 の低高高のアクセントの同音異義語を別の同音異義語に誤認識する例が多かった. 同音異義語の誤認識の例を表 8 に示す. なお, 表 8 , 9 の括弧内の数字の 0 はアクセントの低,1 は高を意味する.

表 8: 同音異義語の誤認識例

認識結果	正解
居る (01)	射る (10)
起源 (100)	機嫌 (011)

6.2 単語の誤認識

単語を同音異義語ではない別の単語に誤認識した認識結果は Diagonal の MFCC を用いた実験で多く見られた. なお, 誤認識の例を表 9 に示す.

表 9: 単語を同音異義語ではない単語とした誤認識例

認識結果	正解
徳 (01)	置く (01)
堪える (010)	返る (100)

7 おわりに

本研究では, アクセントを用いて同音異義語の音声認識実験を行った. 実験の結果, アクセントモデルの有効性を確認できた. また,FBANK では MFCC より同音異義語の認識精度は高いことが確認された.

参考文献

[1] 高橋, 松永, 嵯峨山:“ピッチパタン情報を用いた単語音声認識”, 日本音響学会講演論文集,1-3-20,pp.39-40(1990)

[2] 村上, 荒木, 池原:“音声におけるポーズ長およびアクセント位置の情報量の考察”, 日本音響学会講演論文集,3-3-11,pp.89-90(1988)

[3] “NHK 日本語発話アクセント辞典新版”,NHK 出版,ISBN4-14-011112-7(1998)

[4] 妹尾, 村上, 池原:“モーラ情報を用いた単語音声認識”, 日本音響学会講演論文集,1-4-8,pp.15-16(2003)

[5] 谷口, 村上, 池原:“FBANK を用いた孤立単語音声認識”, 日本音響学会講演論文集,3-Q-3,pp.157-158(2003)

[6] X.D.HUANG,Y.ARIKI,M.A.JACK: “HIDDEN MARKOV MODELS FOR SPEECH RECOGNITION”

[7] “HTK Ver3.2 reference manual”,2002 Cambridge Universit