構音障害者の音素認識誤りの傾向*

☆吉岡利也，高島遼一，滝口哲也，有木康雄 (神戸大)，李義昭 (追手門学院大)

1　はじめに

近年，情報技術の向上に伴い，福祉分野への情報技術の適用が行われている．例えば，画像認識技術を用いた手話認識 [1] や，文書内の文字の音声化などが行われている [2]．また，音声合成を用いて，発話障害者支援のための音声合成器の作成なども行われている [3].

音声認識技術は，これまでの多くは成人を対象としたものであったが，現在では子供や高齢者などの発話スタイルが成人と異なる人物を対象とした場合や，車内や会議室といった様々な実環境下での利用を目的とした場合など多様化しており，使用される機会が増加している．文献 [4] では，言語障害者を対象とした特徴量抽出や音響モデル適応と構築を行っているが，言語障害に関する研究はまだまだ少ない．発話が不安定で手足も不自由な場合など，音声に頼るしかない状況が考えられ，言語障害者を対象とした音声認識の実現が望まれている．

言語障害の原因の一つとして，脳性マヒが考えられる．脳性マヒの原因は，中枢神経系の損傷によるものであり，それに伴う運動障害であると考えられている．発症時期として出生前（胎内感染，母体の中毒，栄養欠損など），分娩時（脳外傷，脳出血，脳の無酸素状態，胎児黄疸，仮死状態，未熟での出産など），出生後（脳内出血，脳炎など）の 3 つの要因が考えられている．脳性マヒの程度や症状は，人によってそれぞれである．分類としては，1) 痙直型，2) アテトーゼ型，3) 失調型，4) 緊張低下型，5) 固縮型，6) 混合型に分類することが出来る．

本稿では，アテトーゼ型の脳性マヒによる構音障害者を対象としている．アテトーゼ型とは脳性マヒ患者の約 10〜15%に発生する症状であり，大脳基底核と呼ばれる視床下部，脳幹，小脳と関連を持ち随意運動，姿勢，筋緊張を調節する働きをしている部位に損傷をうけたため，アテトーゼと呼ばれる不随意運動が伴う型である．アテトーゼは緊張時や意図的な動作を行う際に出現しやすい．症状は軽度から重度まで様々であり，知能障害を合併していないケースや比較的知能障害の程度が軽いケースも多いことが特徴である．そこで本稿では，知能障害を合併していないアテトーゼ型に着目した．

近年，音声認識システムにおいて，音声特徴量として MFCC (Mel-Frequency Cepstrum Coefficient) が広く使われている．これは，対数メルフィルタバンク出力に対して DCT (Discrete Cosine Transform，離散コサイン変換) を行うことにより得られる特徴量であり，正規化手法や特徴量の線形回帰である ΔMFCC や ΔΔMFCC と組み合わせることで，音声認識において高い認識率を示している．しかし，アテトーゼ型の構音障害者の発話スタイルは，筋肉の緊張のため健常者と大きく異なり不安定になりやすい．そこで，我々は離散コサイン変換の代わりに，第 2 発話以降のより安定したデータを利用した，PCA (Principal Component Analysis) による発話スタイル変動にロバストな特徴量抽出法を提案してきた [5]．また文献 [6] では，非負値行列因子分解 (NMF:Non-negative Matrix Factorization) を使ったスパース表現に基づく特徴量抽出法を提案している．これらの手法により，認識率の改善が見られたが，健常者と比べると以前として低い．

これまで構音障害者の音声認識は，健常者の音素体系を基に行われてきた．しかし，両者の発声方法は大きく異なり，音素体系は一致しないと考えられる．そこで本稿では，構音障害者の音声に対して音素認識を行い，その誤り傾向から音素体系の解析を試みる．

2　構音障害者音声データ

実験用データとして，構音障害者 1 名のデータを収録した．発話内容として ATR 音素バランス単語 (216 単語) から 210 単語を無作為に選択した．アテトーゼ型脳性マヒの構音障害者は，アテトーゼの影響により，発話毎に発話スタイルが変動しやすい．そこで本稿では，収録方法として，同一単語の複数回連続発話を行った (Fig. 1)．実験では各単語は連続で 5 回発話されており，合計 1,050 単語を使用する．

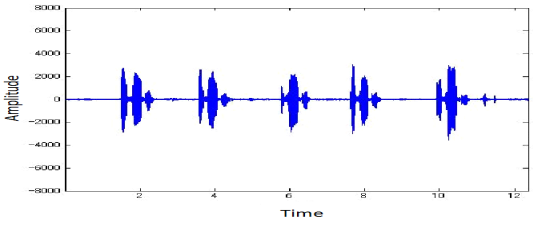

Fig. 1　Example of recorded speech data

*Tendencies of phone recognition errors due to articulation disorders, by Toshiya Yoshioka, Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Ariki (Kobe Univ.), Li Ichao (Otemon Gakuin Univ.)

3　音素認識実験

3.1　実験条件

構音障害者の音素誤り傾向を調べるために音素認識実験を行った．被験者は男性の構音障害者 3 名 (障害者 A，障害者 B，障害者 C とする) とし，それぞれ同条件で複数回連続発話データの収録を行った．

音声の標本化周波数は 16kHz，語長 16bit であり，音響分析には Hamming 窓を使用した．フレーム幅，フレームシフト幅はそれぞれ 32ms，16ms である．

構音障害者は発話スタイルが健常者と大きく異なるため，従来の不特定話者モデルでは認識が困難である．そこで，本稿では構音障害者の特定話者モデルを作成して認識実験を行った．実験では，学習データ，評価データともに収録した 1,050 単語すべてを用いた (closed 条件)．音響モデルは音素 HMM で，各 HMM の状態数は 5，状態あたりの混合分布数は 14 である．音声特徴量には，12 次元 MFCC+ΔMFCC の 24 次元を用いた．音響モデルの学習と音素認識には HTK [7] を用いた．

3.2　単語ネットワークファイル

音素認識で使用する認識文法ファイルは，単語ネットワークファイルから HTK の HParse で作成される．この単語ネットワークファイルには，「入力音声が全 53 音素のいずれかで構成され，その前後に無音 (pau) がつく」ことが記述されている．そして，単語ネットワークファイルをフォーマット変換したものが認識文法ファイルとなる．通常，各単語毎に一つの単語ネットワークファイルを用意する．しかし，本稿ではより詳しい解析を行うために，入力音声を構成する音素毎に単語ネットワークファイルを用意する．

例えば，単語/i e d e/に対しては以下の 4 つの単語ネットワークファイルを用意する．

$$\$all\_phonemes = a \mid i \mid u \mid \cdots \mid z \mid pau;$$

1. ((pau)($all_phonemes)(e)(d)(e)(pau))
2. ((pau)(i)($all_phonemes)(d)(e)(pau))
3. ((pau)(i)(e)($all_phonemes)(e)(pau))
4. ((pau)(i)(e)(d)($all_phonemes)(pau))

3.3　音素正解率

まず，構音障害者 3 名に対して音素認識実験を行い，音素誤り傾向を，各音素毎の正解率 (Phone Correct) から評価した．ただし，各音素の主なルールは Table 1 のようになっている．構音障害者 3 名の各音素毎の正解率を Table 2 に示す．

Table 1　phoneme rules

長母音	a, a → aa
二重母音	a, i → ai
母音+撥音 (ん)	a, N → aN
幼音 (ゃ，ゅ，ょ)	sh, a → sh, a^{-} (母音に-が付く)
無声化母音	f, u, k → f, u^{+}, k (母音に+が付く)
促音	k, k, u → Q, k, u (‘っ’ は Q に変換)

Table 2 より，3 名の構音障害者において，正解率が悪くなっている音素が，母音，子音ともに類似していることが分かる．特に，母音 /a/,/i/,/u/,/e/,/o/ や無声化母音 /u^{+}/ や長母音 /uu/,/ii/，また二重母音 /ou/ などの正解率が低い．また，無声子音 /h/,/k/ や子音 /g/,/r/や鼻音 /n/,/m/，母音+撥音/aN/,/oN/ などの正解率も低いことが分かる．

Table 2　Phone correct rate (%)

サンプル数	音素	構音障害者 A	構音障害者 B	構音障害者 C	平均正解率
300	r	76	82	91.67	83
365	u	79.73	80.82	89.32	83.29
515	a	78.64	85.44	93.01	85.70
345	i	85.80	85.22	92.75	87.92
380	o	85.79	87.63	90.79	88.07
325	e	83.38	90.77	92.62	88.92
35	iN	80	85.71	100	89
190	uu	88.42	86.32	95.26	90.00
145	n	91.03	88.28	91.03	90.11
170	u^{+}	90.59	86.47	95.88	90.98
135	g	87.41	92.59	93.33	91.11
60	u^{-}	93.33	86.67	96.67	92.22
85	h	89.41	94.12	95.29	92.94
245	ou	91.02	91.02	97.14	93.06
45	ii	86.67	95.56	97.78	93.33
180	m	94.44	92.22	95	93.89
85	y	90.59	95.29	97.65	94.51
380	k	92.63	96.32	95.79	94.91
70	Q	95.71	94.29	95.71	95.24
115	j	90.43	99.13	97.39	95.65
155	s	90.97	98.06	98.06	95.70
55	oN	94.55	98.18	94.55	95.76
75	o^{-}	94.67	94.67	98.67	96.00
55	i^{+}	98.18	94.55	96.36	96.36
85	z	98.82	96.47	95.29	96.86
50	N	96	96	100	97
155	t	94.84	98.06	98.71	97.20
135	b	94.81	97.78	99.26	97.28
115	sh	94.78	99.13	98.26	97.39
95	d	96.84	97.89	97.89	97.54
75	aN	96	98.67	98.67	98
50	my	100	94	100	98
50	eN	94	100	100	98
70	p	100	95.71	98.57	98
35	gy	97.14	97.14	100	98.10
40	ts	97.50	100	97.50	98.33
110	ch	98.18	98.18	99.09	98.48
45	ei	95.56	100	100	98.52
50	hy	98	100	98	99
25	ee	100	100	96	99
50	ry	98	98	100	99
40	aa	100	100	97.50	99
65	ky	100	100	98.46	99
75	ai	98.67	98.67	100	99.11
40	py	97.50	100	100	99.17
115	a^{-}	98.26	100	100	99.42
65	w	98.46	100	100	99.49
45	by	100	100	100	100
20	ui	100	100	100	100
15	ao	100	100	100	100
20	f	100	100	100	100
50	ny	100	100	100	100

3.4 音素の誤認識解析

Table 3 Tendencies of phone recognition errors (articulation disorder A)

音素	正解率	誤認識音素 (/音素/誤り率)		
a	78.64%	/e/4.01%	/i/2.14%	/r/2.14%
i	85.8%	/e/4.35%	/r/1.74%	/a/1.45%
u	79.73%	/o/4.38%	/a/3.01%	/uu/2.47%
e	83.38%	/i/5.54%	/a/3.69%	/r/2.46%
o	85.79%	/a/3.16%	/u/3.16%	/ou/1.58%
u$^+$	90.59%	/pau/2.94%	/u/2.35%	/uu/1.18%
uu	88.42%	/u/1.58%	/r/1.58%	/a/1.58%
ii	86.67%	/i/4.44%	/e/4.44%	/r/2.22%
ou	91.02%	/uu/2.86%	/u/1.63%	/o/1.22%
h	89.41%	/o/2.35%	/r/2.35%	/j/1.18%
r	76%	/o/4%	/a/3.33%	/i/2%
k	92.63%	/g/1.58%	/ky/0.79%	/pau/0.79%
g	87.41%	/o/2.96%	/u/2.96%	/r/1.48%
m	94.44%	/u/1.67%	/ou/1.67%	/uu/0.56%
n	91.03%	/a/1.38%	/u/1.38%	/e/1.38%
aN	96%	/a/4%		
oN	94.55%	/o/3.64%	/u/1.82%	

Table 4 Tendencies of phone recognition errors (articulation disorder B)

音素	正解率	誤認識音素 (/音素/誤り率)		
a	85.44%	/e/3.11%	/o/2.33%	/u/1.36%
i	85.22%	/r/2.32%	/n/1.45%	/uu/1.45%
u	80.82%	/r/3.56%	/i/2.19%	/uu/1.64%
e	90.77%	/r/2.16%	/i/1.54%	/a/0.92%
o	87.63%	/ou/4.21%	/u/2.11%	/a/1.84%
u$^+$	86.47%	/pau/3.53%	/u/2.35%	/s/1.76%
uu	86.32%	/r/2.63%	/u/2.11%	/i/1.58%
ii	95.56%	/pau/4.44%		
ou	91.02%	/u/3.27%	/uu/1.22%	/o/1.22%
h	94.12%	/pau/2.35%	/k/1.18%	/o/1.18%
r	82%	/u/2.33%	/e/2.33%	/n/2%
k	96.32%	/t/1.05%	/ky/0.79%	/pau/0.79%
g	92.59%	/k/2.22%	/r/2.22%	/i/0.74%
m	92.22%	/u/1.67%	/b/1.11%	/ou/0.56%
n	88.28%	/r/4.83%	/i/3.45%	/m/1.38%
aN	98.67%	/a/1.33%		
oN	98.18%	/pau/1.82%		

Table 2 より，3 名の構音障害者に共通して正解率が低下している音素について，その誤り傾向を調べる．Table 3，Table 4，Table 5 は，それぞれ構音障害者 A，構音障害者 B, 構音障害者 C の音素認識結果から，3 名に共通して正解率の低下が見られた音素について，上位 3 個の誤認識結果を調べたものである．

これらの表から，音素誤り傾向として以下のものが考えられる．

1. 周波数スペクトルが近似している母音の誤り
   (例えば，/a/と/e/，/o/と/u/の誤り等)
2. 無声化母音と/pau/の誤り
   (例えば，/u$^+$/と/pau/の誤り等)
3. 長母音と短母音の誤り
   (例えば，/uu/と/u/，/ii/と/i/の誤り等)

Table 5 Tendencies of phone recognition errors (articulation disorder C)

音素	正解率	誤認識音素 (/音素/誤り率)		
a	93.01%	/o/2.14%	/a-/1.96%	/aN/0.78%
i	92.75%	/e/1.16%	/pau/1.16%	/ii/0.87%
u	89.32%	/i/2.80%	/o/1.64%	/u+/1.64%
e	92.62%	/i/3.69%	/r/1.85%	/a-/0.31%
o	90.79%	/u/3.68%	/a/1.05%	/ou/0.79%
u$^+$	95.88%	/u/1.18%	/ts/1.18%	/pau/0.59%
uu	95.26%	/u+/1.58%	/u/0.53%	/i/0.53%
ii	97.78%	/i/2.22%		
ou	97.14%	/o/1.63%	/u/0.41%	/uu/0.41%
h	95.29%	/o/1.18%	/n/1.18%	/r/1.18%
r	91.67%	/i/1.33%	/b/1.33%	/e/1%
k	95.79%	/b/0.79%	/j/0.79%	/g/0.53%
g	93.33%	/k/1.48%	/i/1.48%	/j/0.74%
m	95%	/o/2.22%	/n/1.11%	/b/0.56%
n	91.03%	/pau/2.07%	/u/1.38%	/m/0.69%
aN	98.67%	/a/1.33%		
oN	94.55%	/u+/3.64%	/u/1.82%	

4. 二重母音と短母音の誤り
   (例えば，/ou/と/o/,/u/の誤り等)
5. 子音/r/と母音の誤り
   (例えば，/r/と/o/,/u/の誤り等)
6. 鼻音と母音の誤り
   (例えば，/m/と/u/の誤り等)
7. 母音+撥音の音素における撥音の欠落
   (例えば，/aN/と/a/の誤り等)

同じ母音が連続して出現するような箇所で/uu/が/u/，/ii/が/i/と誤認識するような間違いが多かった．これより，構音障害者は母音の連続発話が不明瞭になる場合があり，後ろの母音が欠落しやすいと考えられる．

また，/r/が母音と誤認識するような間違いが見られた．/r/は舌の動きだけで表す子音のため，舌の動きが上手くコントロールできない構音障害者では，/ro/や/ru/などの/r/+母音の音素と区別できない場合があると考えられる．

母音+撥音の音素/aN/において撥音の欠落による誤認識が見られた．Fig. 2 に健常者の/k aN by ou/のスペクトログラム，Fig. 3 に構音障害者 A の/k aN by ou/のスペクトログラムを示す．Fig. 3 より，構音障害者のスペクトログラム上では/N/の音素がほとんど見られず，大部分が/a/の音素になっていることが分かる．

また，音素認識実験の結果から，音素が削除されている可能性のあるデータが見られた．例えば，/n e a g e/という音素系列が/n e e g e/という音素系列として誤認識された．このとき，/a/の音素が直前と同じ/e/と認識されており，/a/の音素が消えている

と考えられる．Fig. 4 に健常者の/n e a g e/のスペクトログラム，Fig. 5 に構音障害者 A の/n e a g e/のスペクトログラムを示す．Fig. 4 , 5 より，健常者のスペクトログラム上では/e/と/a/の境目がはっきりと見て取れるが，構音障害者のスペクトログラム上では/e/の音素が長くなっており，/a/の音素が消えていると判断できる．音素が無いと判断出来れば，学習時のラベルからその音素を削除してモデルを学習することになる．また，評価ラベルとしても，その音素があり，無し両方の音素系列を辞書として登録しておくことになる．

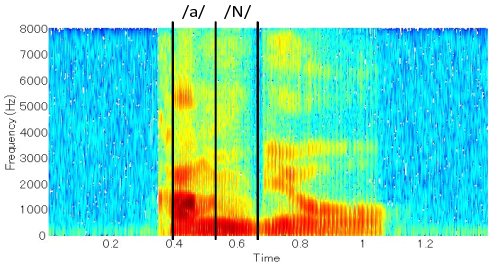

Fig. 2　Example of a spectrogram spoken by a physically unimpaired person /k aN by ou/

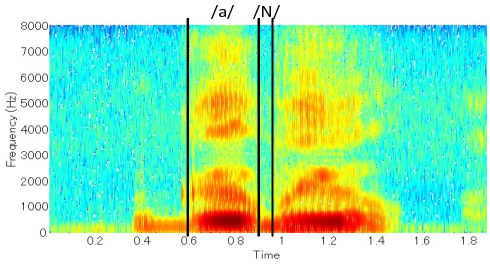

Fig. 3　Example of a spectrogram spoken by a person with articulation disorders /k aN by ou/

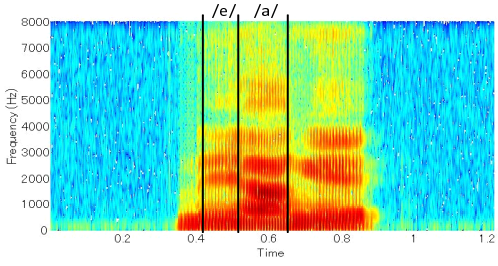

Fig. 4　Example of a spectrogram spoken by a physically unimpaired person /n e a g e/

4　おわりに

本稿では，構音障害者の音素体系に注目し，音素認識実験を行いその誤り傾向について検討を行った．

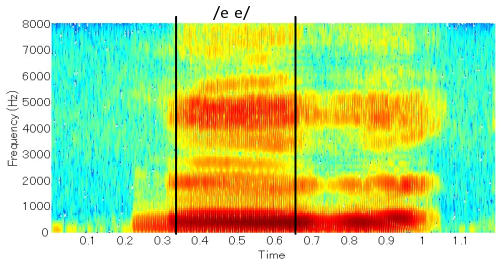

Fig. 5　Example of a spectrogram spoken by a person with articulation disorders /n e a g e/

構音障害者 3 名を対象とした音素認識実験により，正解率が低下している音素が，母音，子音ともに類似していることが確認できた．また，正解率が低下している音素において，いくつかの誤り傾向が見られた．

今後はデータ数を増やすなどして，より詳しい解析を行うと共に，その誤り傾向を基に音素系列の書き換えを行うなどして，音声認識精度の改善に取り組んでいく予定である．

参考文献

[1] 佐川浩彦，酒匂裕，大平栄二，崎山朝子，阿部正博，“圧縮連続 DP 照合を用いた手話認識方式，” 電子情報通信学会論文誌，Vol.J77-D-II，No.4，pp. 753-763，1994.

[2] 鈴木悠司，平岩裕康，竹内義則，松本哲也，工藤博章，大西昇，“視覚障害者のための環境内の文字情報抽出システム，” 電子情報通信学会技術研究報告，WIT2003-314，pp. 13-18, 2003.

[3] 藪謙一郎，伊福部達，青村茂，“発話障害者支援のための音声合成器の基礎的設計，” 電子情報通信学会技術研究報告， SP2006-321, pp. 59-64, 2006.

[4] 中村圭吾，田村直良，鹿野清宏，“発話障害者音声を対象にした健常者音響モデルの適応と検証，” 日本音響学会講演論文集，3-7-4，pp. 109-110，2005.

[5] H. Matsumasa, T. Takiguchi, Y. Ariki, I. LI and T. Nakabayashi, “PCA-Based Feature Extraction for Fluctuation in Speaking Style of Articulation Disorders,” INTERSPEECH，pp. 1150-1153，2007.

[6] 吉岡 利也, 高島 遼一, 滝口 哲也, 有木 康雄，“スパース表現に基づく構音障害者の発話スタイル変動にロバストな特徴量抽出，” 日本音響学会研究発表会講演論文集，1-P-4，pp.127-128，2012.

[7] “HTK (Hidden Markov Model Toolkit),” http://htk.eng.cam.ac.uk/.