

論文 / 著書情報
Article / Book Information

論題(和文)	日本語単音節音声認識の検討
Title(English)	
著者(和文)	古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子通信学会総合全国大会, , , pp. 5-329
Citation(English)	, , , pp. 5-329
発行日 / Pub. date	1981,
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1981 Institute of Electronics, Information and Communication Engineers.

日本語単音節音声認識の検討

1351

古井 貞昭 (日本電信電話公社 武蔵野電気通信研究所)

I まえがき 音声スペクトル時系列の大局的・局所的特徴の組合せによる単音節音声認識方式と、その10数字および単音節認識への応用に関してはすでに報告した⁽¹⁾が、ここでは対象を単音節認識⁽²⁾⁽³⁾⁽⁴⁾にしぼり、認識方法をやや変えて、45単音節の認識実験を行ったので、その結果を報告する。

II 認識方法 認識方法は図1に示す通りで、音声スペクトル時系列の大局的・局所的特徴をLPCケプストラムおよびパワー時系列の低次のフーリエ展開係数で、局所的特徴を語頭部のパラメータ時系列そのものによって抽出し、両者の特徴の組合せによって認識を行う。同波数帯域を4kHz(8kHz標準化)、分析フレーム長25ms、フレーム間隔10ms、語頭時系列のマッチングのための最大シフト量を±5フレームとした点以外は、詳細は文献(1)に示した通りである。

認識実験には男性10名が45単音節(日本語全単音節から濁音等を除いた基本部分)を繰返し5回発声したものを用い、これらを学習用サンプルセット(各音節1~3サンプル)と認識用サンプルセットにわけた認識実験を行った。認識に用いるパラメータの選択およびパラメータの重みづけの係数設定のための分散分析には、5名の話者の学習サンプルのみを用いた。入力音声の平均S/Nは、20~25dBである。

III 実験結果 各話者の各単音節の学習サンプル数をとしたとき、大局的・局所的特徴のみの場合、両者の特徴を組合せたときの平均認識誤り率を図2に示す。両者の特徴を組合せたときの誤り率は6.3%で、いずれか一方の特徴のみによる場合の平均誤り率の3/5に低下している。両者の特徴を組合せたとき、オ1~5位までに正しい単音節が入る割合を図3に示す。オ2位までで正しいときの認識率は98.7%、オ5位までで正しいときの認識率は99.8%である。

両者の特徴を組合せ、オ1位のみを用いるときの認識誤りに関し、子音別に整理したときの誤認識行列を表1に、母音別に整理したときの誤り率の比較を図4に示す。全認識サンプルのうち1800単音節のうち、母音部に誤りがあったものが表1の下に示す2サンプル(0.1%)あったが、他はすべて子音部の誤りであった。表からわかるように/sV/⇔/tV/、特に/su/⇔/tu/の誤りが最も多く、次に/hi/⇔/ki/の誤りが多い。母音別には/Ci/、/Cu/の誤り率が比較的大きい。

なお各話者の各単音節の学習サンプル数を1とすると、誤り率が約2倍

に増加する。学習サンプル数を3サンプルよりさらに増加すれば、認識率は更に向上すると思われる。

IV むすび 本方式は処理が比較的簡単であるにも拘らず、高い性能を有することが確認された。

(謝辞) 日頃御指導頂く当所野田基礎研究部長・小池中四研究室長、並びに討論して頂く研究員の諸氏に感謝する。

(文献) (1)古井: 音学季刊 1-1-18 (昭55), (2)新田: 音学季刊 1350 (昭55), (3)吉田: 音学季刊 3-2-16 (昭54), (4)白根: 電子技術 (昭54-12)

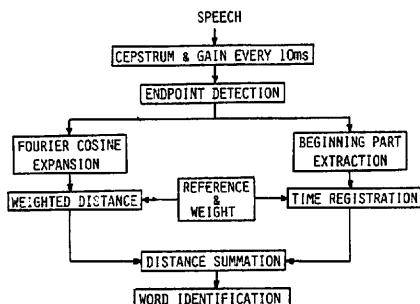


図1. 単音節認識方式のブロック図

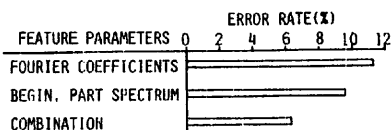


図2. 平均認識誤り率

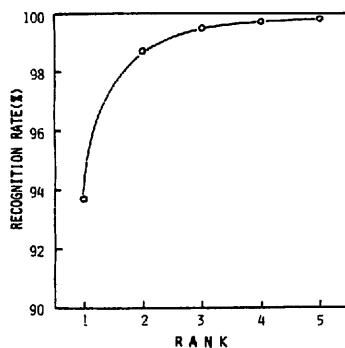


図3. オ1~5位までに正しい単音節が入る割合

表1. 誤認識行列

OUT	IN	(v)	k	s	t	n	h	m	y	r	w	N	rate
(v)	193	3						4					96.5*
k	2	189						9					94.5
s			171	29									85.5
t			1	20	178		1						89.0
n						194		1		3			97.0
h			8	12	2	1	175			2			87.5
m				1			3	196					98.0
y									120				100
r			4			3	2			191			95.5
w											40		100
N												40	100
Other errors: no→na, ni→re													average
													93.7

図4. 母音別認識誤り率