

論文 / 著書情報
Article / Book Information

論題(和文)	単音節入力による大語彙単語音声認識の検討
Title(English)	
著者(和文)	古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1981年春季講演論文集, , , pp. 543-544
Citation(English)	, , , pp. 543-544
発行日 / Pub. date	1981, 5

3—1—12 単音節入力による大語彙単語音声認識の検討*

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特徴と局所的特徴の組合せによる単音節認識法の有効性についてはすでに報告した⁽¹⁾⁽²⁾。ここではこの方法を用いて、音節毎に区切って発声された大語彙の単語音声と認識する可能性について検討を行った。

II 単音節認識実験 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。標本化周波数は8kHz、分析フレーム長は25ms、フレーム周期は10msで、語頭時系列のマッチングは、±1フレーム以内のシフトによる重ね合せ(時間伸縮なし)によって行った。なお比較のために、語頭部について非線形時間伸縮(DP)マッチング(端点フリー)を行う場合についても実験を行った。

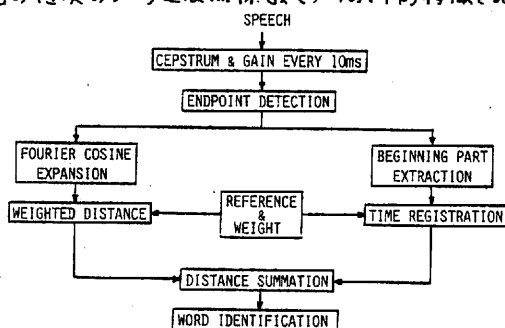


Fig.1. Principal operation of the syllable recognition system.

認識実験には男性10名が日本語68単語

節を、計算機端末室で繰返し5回発声した音声を用い、初めの3サンプルを学習用、残りを未知入力とした場合と、後の3サンプルを学習用、初めの2サンプルを未知入力とした場合の2通りの実験を行った。大局的特徴または局所的特徴のみの場合、および両者を組合せる場合の平

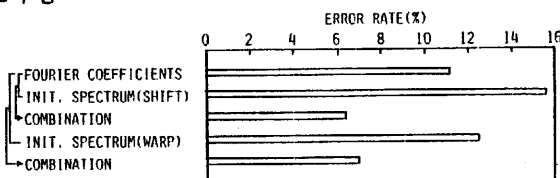


Fig.2. Syllable recognition error rate as a function of the feature parameter set.

Table 1. Confusion matrix for syllable recognition.

OUT IN	(v)	k	s	t	n	h	m	y	r	w	ŋ	g	z	d	b	p	rate
(v)	191							7									2 95.5
k		4 183		1			8										4 91.5
s			173 25											1			1 86.5
t		1	24 171					2									2 85.5
n					198			1		1							99.0
h		8 9				181							1				1 90.5
m					2		198										99.0
y								120									100
r				1 1 1					188					2 3 2 2			94.0
w										40							100
ŋ											40						100
g		1 2		1		1						188		1 3 3			94.0
z				2 1					1				192 3				1 96.0
d				1 3					3				2 111				92.5
b		1		1					4			2		2 184 6			92.0
p		3 1		3 4												189	94.5
average																	93.6

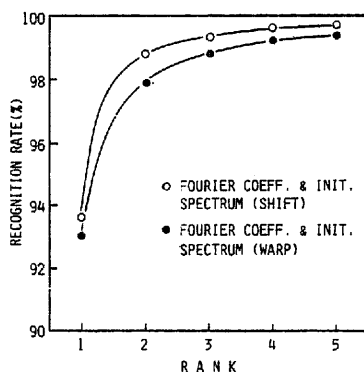


Fig.3. Accumulated syllable recognition rate.

* Automatic Recognition of Large Vocabulary Words Uttered by Isolated Syllables.
By Sadaaki FURUI (Musashino Electrical Communication Laboratory, N.T.T.)

均認識誤り率を図2に示す。語頭部のマッチングのみによる場合は、シフトによるよりも非線形伸縮による場合の方が誤りが少ないが、フーリエ係数を組合せた場合は、シフトによる方法の方が誤りが少ない。これは、シフトによる語頭マッチングを行うと、子音節は正しく母音部を誤るものが生ずるが、この誤りは大局的特徴との組合せによって完全に除去されるためである。大局的特徴と局所的特徴を組合せたとき、オ1位～オ5位に正しい音節が入る割合を図3に、シフトによる語頭マッチングと大局的特徴の組合せによるときの認識行列を表1に示す。

III. 大語彙単語音声認識 ある国語辞典に含まれている4モーラ4文字の全名詞12813単語および、そこから4つ毎に選んだ3203単語を対象とし、実際の音声には上述の実験で用いた10名の単音節音声をあてはめて認識実験を行った。これらの単語を構成する音素の出現頻度は、文献(3)に報告されている値と類似している。また距離1～4(音節単位)の単語対の割合は図4に示す通りである。単語の認識は、各音節毎の標準パターンとの距離を4音節について合計した値によって行った。

オ1位～オ5位に正しい単語が入る割合を図5に、各話者の単音節認識率と単語認識率の関係を図6に示す。図6には、

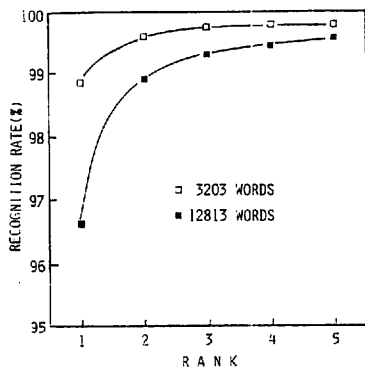


Fig.5. Accumulated word recognition rate.

図4の結果にもとづいて、簡単なモデル(文献(4), N形模型)によって計算した認識率の理論値をあわせて示した。実験値には話者によるばらつきがあるが、全体的には実験値と理論値はよく対応しており、理論式の有効性が確認された。実験結果から、単音節認識率が92%程度であれば、10,000程度の語彙を96%, 3,000程度の語彙を98%の精度で認識できることがわかる。また図5の結果から、オ2位の候補までとれば、10名の平均で、12813単語に対し99%程度の認識率を得ることができるとわかる。

誤認識の原因を調べたところ、単音節では $su \leftrightarrow tu$ の誤りが多く、これはほとんど単語の誤認識の原因とはならないことがわかった。

IV. むすび スペクトル時系列の大局的・局所的両特徴の組合せによる単音節認識方法が、単音節発声による大語彙単語の認識に極めて有効であることが確認された。

[謝辞] 日頃御指導とご所望賜る基礎部長、板倉才4研究室長に感謝する。幹事作成に協力頂いた奥倉綱吉、実験に協力頂いた実習生、加地君(久大)に感謝する。

(文献) (1) 古井: 音学精論 1-1-18 (昭55-10) (2) 同: 信学誌 46.151 (昭56-04) (3) 坂井他: 信学誌 46.11, p.1696 (昭58-4) (4) 阿部他: 信学誌 52-C, 6, p.305 (昭44-06)

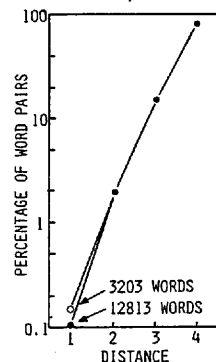


Fig.4. Statistics of word pairs.

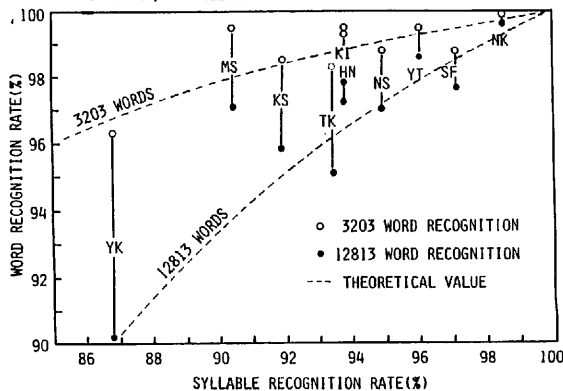


Fig.6. Theoretical & experimental word recognition rates for each speaker.