

調音器官の形態情報から日本語拍を検出するための一試案

4E-2

原田 正明、 竹内 伸、 村井 和昌  
富士ゼロックス(株) 総合技術研究所

1. はじめに

我々は、音声情報を周囲の雑音に影響されることなく情報機器へ入力する手段として、調音器官の形態情報を利用する機械的な読話システムの研究を行っている。このため、調音器官の微妙な動きを精度良く測定する方法として正反射光スポット法を提案した。この方法を用い音声情報を採取する測定点を口唇部と顎下部に配置することによって日本語 100 拍に対する認識実験を行った。また、口唇部と顎下部によって得られたそれぞれの結果を拍の母音成分および子音成分に分解し認識率として比較した。

2. 正反射光スポット法と測定点

発話に伴い調音器官の形状は微妙に変化する。その中には位置の動きは小さいが、その角度変化に特徴を有する部分も多い。正反射光スポット法とは、このような測定点の動きを増幅して検出できるように、話者の調音器官または、その周辺の外皮に小さな反射鏡（～5mm角）を配置し、さらにその反射鏡に光線を照射することにより、そこから正反射された光スポットを位置検出センサ上に投影する方法である（図1参照）。

今回の実験では位置検出センサとして、CCD カメラ前面にスクリーンを配置した装置を2セット使用した。スクリーン上に照射した光スポットは CCD カメラで撮影され、出力される走査線の情報から光スポット位置の座標を算出している。

実験に用いた測定点は、下唇の口角寄りの1点と顎下部の1点とした（図2参照）。それぞれの点は、事前に母音を入力した実験で認識率が高かった点の中から選択したものである。

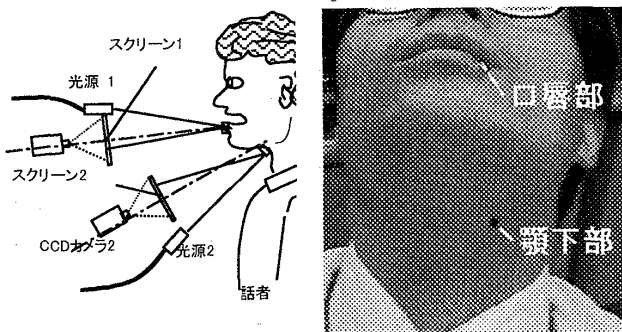


図.1 正反射光スポット法 図.2 測定点

On detecting Japanese morae by movement of vocal organ.  
M. Harada, S. Takeuchi, K. Murai  
Fuji Xerox Co., LTD. Corporate Research Labs.  
430, Sakai, nakaimachi, ashigarakami-gun, kanagawa,  
259-0157, Japan

3. 日本語 100 拍の認識実験

実験においては、日本語の 100 拍を単音で区切りながら入力した。話者は特定話者とし、40 代の男性 1 名を被験者とした。1 拍の発話につき、画像 200 フレーム分（入力時間 3.3 秒）と各フレーム中の光スポット 2 点の X-Y 座標から 800 個のデータが得られる。入力の順番は前後の拍の影響を受けないようにランダム化した。日本語 100 拍の入力の繰り返し数は 4 回であり、この内の 1 回目を標準パターンとし 2 回目をテストデータとして DP マッチングにより 1 組目の拍の認識率を求めた。同様に 3 回目と 4 回目の繰り返しデータから 2 組目の認識率を求め、これらの認識率の平均値を算出した。口唇部および顎下部の各測定点での認識結果と両方の測定点を組み合わせた場合の認識結果を表1に示した。

表1 DP マッチングによる日本語 100 拍の認識結果

	1 組目	2 組目	平均
口唇部の認識率	15.0%	19.0%	17.0%
顎下部の認識率	12.0%	19.0%	15.5%
口唇部と顎下部の認識率	33.0%	36.0%	34.5%

これらの結果をさらに分解して、入力した拍と認識された拍の母音成分が一致する場合の認識率および入力した拍と認識された拍の子音成分が一致する

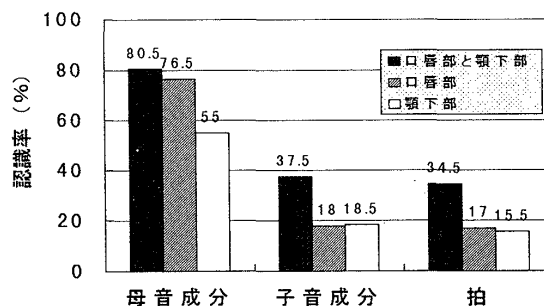


図3 拍・母音成分・子音成分の認識率

場合の認識率とを図3に示した。

どの成分も口唇部と顎下部の測定結果を併せた認識率が単独の測定結果よりも高くなっている。この中で拍および子音成分においては、口唇部による認識率と顎下部による認識率の和が、ほぼ両測定点を併せた認識率の値となっていることが分かる。このことは認識に適した拍が、口唇部と顎下部にそれぞれ

れ独立して存在することを示唆している。一方、母音成分においては、ほとんどの認識率が口唇部において達成されており、顎下部による認識率は相対的に小さくなっている。この母音成分をさらに5つの音素(/a/, /i/, /u/, /e/, /o/)に分解し、それぞれの認識率を比較したものを図4に示した。

これによると顎下部の認識結果は、どの音素においても口唇部の認識率と比べて悪く、特に/i/の認識率は低かった。また、口唇部だけの認識結果の方が口唇部と顎下部を合わせた認識結果より良いものもあった。

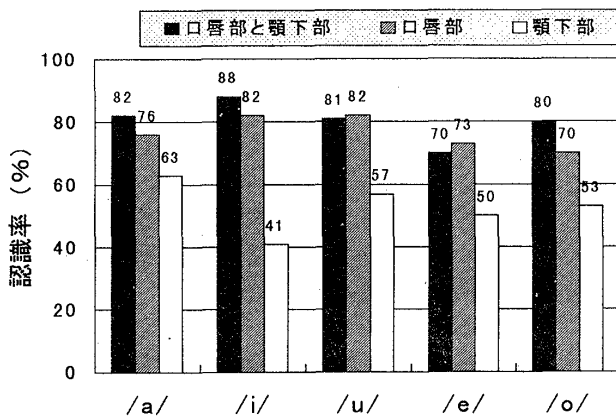


図4 各母音成分に対する認識率

#### 4. 音声学の分類による認識率

つづいて各拍を音声学の分類法<sup>1)</sup>に従いその子音成分が属する調音位置と調音方法の2通りに分類して認識率を議論する(表2参照)。

表2 音声学による代表的な子音分類

		調音位置					
		唇	歯茎	硬口蓋歯茎	硬口蓋	軟口蓋	声門
調音方法	閉鎖音	p	t		c	k	
	破裂音	b	ts	tʃ	ʃ	g	
調音方法	摩擦音	φ	dz	dʒ	ç		h
	摩擦音		s	ʃ			
調音方法	鼻音	m	n	ɲ			
	流音			ɺ			
調音方法	わたり音				j	w	

図5は調音位置に対する認識率を示し、図の横軸は左から右に向かうに従い、調音位置が唇から口腔内部に進むように配置されている。唇を調音に使用する「唇」音は口唇部の測定点による認識率が高く、反対に口腔の奥に対応する「軟口蓋」音では顎下部の測定点による認識率が高いという傾向が見られる。

図6は調音方法に対する認識率を示し、図の横軸

は左から右に向かうに従い調音時の呼気流の流れが緩やかになる様に配置されている。これらの中で「摩擦音」と「鼻音」では、2つの測定点からのデータを組み合わせることにより、その認識率が大きく向上していることが分かる。また「わたり音」は口唇部において主に認識されていることが分かる。これは、わたり音が半母音と呼ばれるように母音と近い発声法であることから母音と似た傾向が出たものと思われる。

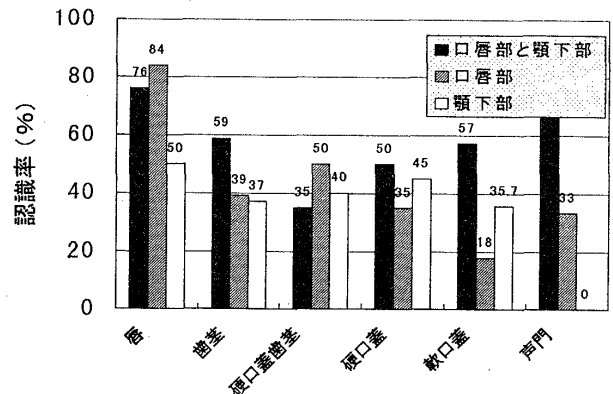


図5 調音位置による子音成分の認識率

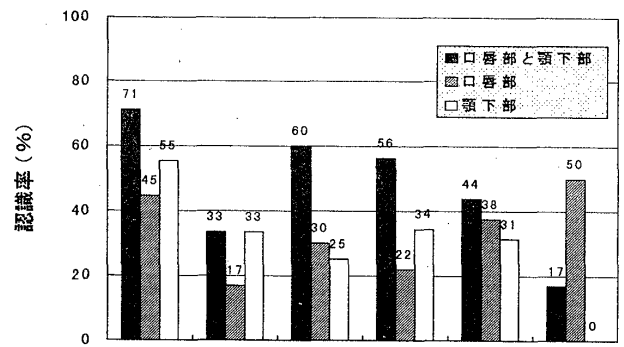


図6 調音方法による子音成分の認識率

#### 5. まとめ

正反射光スポット法を口唇部と顎下部のそれぞれ1点に適用し、日本語100拍を機械的に認識する実験を行った。その結果、平均34.5%の認識率を得ることができた。また、2つの測定点を組み合わせることにより認識率が大きく向上する子音成分があることが判明した。現状では、まだ実用的な認識率には及ばないが、測定点の数および位置の最適化と後段の処理技術とを改善することによってさらなる認識率の向上を目指していく予定である。

#### 6. 参考文献

- 1) 柴谷他、「言語の構造 音声・音韻編」第10刷、くろしお出版、1994年11月7日