

## 調音器官表皮の動きからの日本語単音節の検出

1K-1

### — 正反射光スポット法による測定点の探索 —

竹内 伸, 村井和昌, 原田正明  
富士ゼロックス株式会社 総合研究所

#### 1. はじめに

我々は、周囲の雑音に影響されることなく発話内容を情報機器へ入力する手段として、調音器官表皮の動きから発話を検出する機械読話システムの研究を行っている。前回の報告では、正反射光スポット法を提案し、この方法を下唇と顎下の2点に同時に適用することにより、単独に発話された日本語単音節100種に対して34.5%の認識率が得られることを示した<sup>1)</sup>。今回は、調音器官表皮の測定位置と認識率との関係を把握するため、15点の測定点に対して、各点の認識率を個別に評価した。この結果をもとに、上記15点の中から認識率の高い3点を選び、この3点に正反射光スポット法を同時に適用して、日本語単音節100種に対する認識率を求め良好な結果を得た。

#### 2. 正反射光スポット法

発話に伴い調音器官の形状は変化する。その動きは、下顎と唇を除けば微少なものであるが、微少な動きの中にも発話の特徴が含まれていると考えられる。正反射光スポット法は、このような微少な動きの特徴を検出するために発案された測定法である。この測定法の概略を図1に示す。

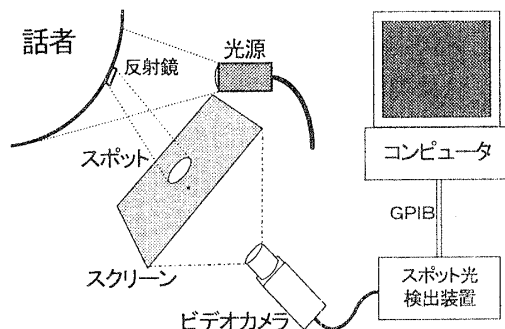


図1 正反射光スポット法の概略

#### 3. 測定点の選定

今回の実験では、日本語単音節100種の単独発話において特徴的な動きがみられる各領域から合計15点(唇:5点, 頬:3点, 顎:4点, 顎下:3点)の測定点を設定した。各測定点の位置と番号を下図に示す。以降、測定点の番号は図2の位置に対応している。

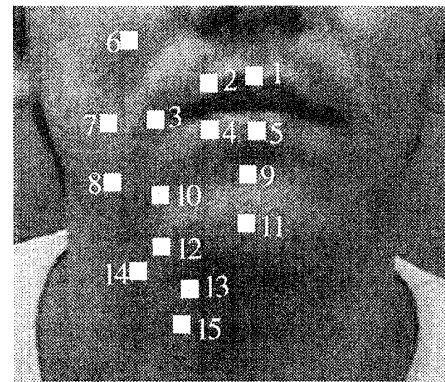


図2 各測定点の位置

各々の測定点に正反射光スポット法を適用し、被験者1名、繰り返し数4回で日本語単音節100種の発話に対応したスポット光の時間パターンを測定した。得られた測定データにDPマッチングによる認識処理を行い、各測定点における母音成分、子音成分、そして単音節の認識率を求めた。その結果を表1および図3に示す。

表1 各測定点の認識率

番号	測定点		認識率 (%)		
	位置	領域	母音成分	子音成分	単音節
1	上唇正中	唇	27.5	19.5	8.3
2	上唇横		33.5	26.5	10.0
3	口角		41.3	19.8	10.8
4	下唇横		32.3	21.0	8.0
5	下唇正中		32.3	20.0	5.5
6	頬上	頬	49.0	23.0	12.0
7	頬中央		57.3	31.3	19.5
8	頬下		32.0	26.8	10.0
9	顎正中	顎	44.0	16.8	8.8
10	顎横		33.5	20.8	8.5
11	顎正中		49.3	24.3	13.5
12	顎横		43.8	31.8	14.8
13	顎下前	顎下	53.0	41.5	24.0
14	顎下横		56.0	35.0	22.5
15	顎下後		48.5	37.0	22.0

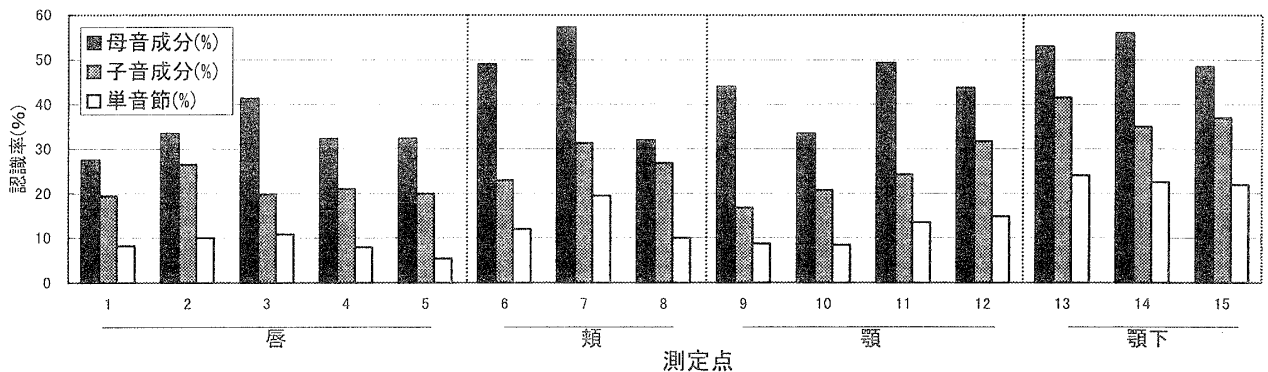


図3 各測定点における日本語単音節100種の認識結果

これらより、認識率が最も高かった点は、母音成分では頬中央7、子音成分および単音節では顎下前13であった。逆に、認識率が最も低かった点は、母音成分で上唇正中1、子音成分で顎正中9、単音節では下唇正中5であり、いずれも正中上の点であった。

領域毎に見ると、唇領域では、母音成分および単音節の認識率は、口角3に近いほど高くなっていた。また、子音成分では上唇横2の認識率が最も高かった。頬領域では、いずれの認識率も頬中央7が最も高かった。顎領域では、母音成分では顎正中11、子音成分と単音節で顎横12の認識率が最も高く、また、いずれの認識率も測定点の位置が顎先に近い方が高くなっていた。顎下領域では母音成分で顎下横14、子音成分と単音節で顎下前13の認識率が最も高かった。また、顎下領域のいずれの測定点も、頬中央7の母音成分を除けば他の全ての測定点より認識率が高く、顎下領域が発話認識に適した領域であることが確認された。

調音器官の形態情報から発話認識を行う従来の研究では、画像計測の容易さから唇の形や動きの情報を主に利用していた。しかし今回の実験結果から、頬中央7や顎横12、顎下前13などの点の方が唇領域のいずれの点よりも認識率が高いことが示された。これは頬や顎下の領域に、唇領域よりも発話認識に適した測定点があるということを示唆しており興味深い結果である。

#### 4. 三点同時測定による日本語単音節認識

前節で用いた測定点から、単音節の認識率が高く領域が重複しない3点(頬中央7、顎横12、顎下前13)を選定した。この3点に正反射光スポット法を同時に適用して、被験者2名、繰り返し10回で日本語単音節100種に対する認識実験を行った。その結果を図4に示す。

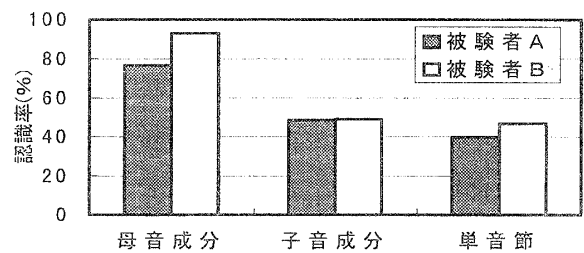


図4 三点同時測定による認識結果

この図から、両被験者の認識結果が類似していることが分かる。両者の平均認識率は、母音成分で 85.0%、子音成分で 48.8%、単音節では 43.5%であった。

#### 5. まとめ

調音器官表皮の15点に正反射光スポット法を適用し、日本語単音節100種に対する認識実験を行った。その結果、従来のほとんどの研究で用いられている唇領域よりも、頬や顎下の領域の方が発話の認識に適していることが示された。

また、これらの結果から、測定点を3点選び、正反射光スポット法に適用して日本語単音節100種に対する認識実験を行った。その結果、平均 43.5%の認識率を得た。この認識率は、わずか3点の計測点から得られており、正反射光スポット法による計測が発話の検出に優れていることを示している。

今後は、認識率の向上のため、より精細な測定点の選定や、実用化に向けての装置の小型化、そして話者の負担をより軽減する計測法の研究を進めていきたい。

#### 参考文献

- 1) 原田, 竹内, 村井, 「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」, 情処学会第58回全国大会4E-02(平成11年3月)