

読話における情報伝達率の改善に関する基礎研究

— 語頭文字の同時提示による正答率の向上に関する検討 —

尾田 継之・奥 英久

岡山理科大学大学院工学研究科修士課程情報工学専攻

岡山理科大学工学部情報工学科

(1999年11月4日 受理)

1. 緒論

聴覚障害者は、音あるいは音声により伝達される情報を聴覚で受け取ることが困難、あるいは不可能である。このため、これらの情報を視覚など他の感覚機能による補足的、あるいは代行的に受け取る方法を用いた、補助的なコミュニケーション手段が考案され、使用されている。この代表例として、手話、筆談、読話などがある。

手話は、話し手が単語や文などを上肢（手・指・腕）の形・位置・動きと顔の表情などで二次元または三次元的な形態および動作パターンとして示し、聴覚障害者が視覚を通じて理解する補助的コミュニケーション手段である。手話は、手話通訳者の資格制度創設あるいはテレビジョン放送における手話ニュースなどにみられるように、現在では社会全般に受け入れられるようになってきている。しかし、常にコミュニケーションの相手方が手話を知っている必要があるため、何時でも何処でも使用できる補助的コミュニケーション手段とは言えない。

一方、広く知られている補助的コミュニケーション手段として、筆談がある。筆談は、伝えたい内容を文字で書き表すため、正確な伝達が可能であり、重要な情報を確認するときには欠かせない手段である。しかし、筆談は、「書く」という労力を必要とし、常に紙と鉛筆という「道具」を準備しなければならず、加えて会話よりも遙かに時間を必要とする問題もある。

これらに対し、読話は、話し手が発語する時における唇の動的情報（上唇と下唇の形、上歯や下歯の見え方、舌の先端や裏面の動き、顎の動きなど）と顔の表情から発語の内容を読みとる補助的コミュニケーション手段である。この方法から明らかなように、読話では、話し手に特別な技術を必要とせず、ゆっくりと大きく口を開けて話すなど、聞き手（聴覚障害者）へのわずかな配慮でコミュニケーションが成立する。さらに、僅かながら残存聴力のある聴覚障害者では、補聴器と併用することにより語音認識率が向上し、コミュニケーションがスムーズになることが報告されている。この状況は健聴者においても同様で、騒音下の環境あるいは声の小さい人との対話などにおいて、話し手の口元や表情から補足的に情報を得ている。一方、読話では、口型と語音の対応を覚える必要がある。しかし、後述するように、複数の語音がほぼ同じ口型となるため、一般的には発語内容を正しく受け取ることができる割合（正答率）の低いことが実験的に示されている。

筆者らは、これらの聴覚障害者の補助的コミュニケーション手段の中で読話に着目し、その情報伝達効率、すなわち認識率の向上に関する研究を行っている。情報工学的視点からは、音声認識技術による正答率 100 %の文字変換が、一つの理想的な読話形態であると考えられる。しかし、現状では、不特定話者に対する音声認識ではまだ若干の誤りが発生すること、そして変換に若干の時間を要するなどの課題があり、実用には至っていない。そこで、これらの技術が実用化されるまでの過渡的手段として、健聴者の発語時に、その先頭文字だけを取り出して、読話の補助情報として同時に提示する方式を提案した。本論文では、提案した方式と従来方式との正答率の差について、健聴者と聴覚障害者による実験を行った結果について報告する。第2章では、読話における問題点を述べ、提案した方式の可能性について論じる。次に、第3章では、提案した方式を評価するために行った実験の方法について説明する。第4章は実験結果で、これに基づいた考察を第5章に示す。

2. 読話の問題点と口型パターン

読話は、既述したように、話し手の口唇の動きと顔の表情から発語内容を読み取る方法である。しかし、読話だけで話し手の発語内容を完全に理解することは一般的に困難である。これまでの研究においては、読話による正答率が 30%程度であることが実験的に示されている¹⁾²⁾³⁾⁴⁾。読話による正答率が低い理由としては、発声発語器官の動き全てが口唇の動きとして発現しないこと、および語音の種類に比して口型の形状変化が少ないこと、などが指摘されている。特に後者では、話し手の口型が同じで発音が違う単語（同口型異音）が多いことにより、発声発語器官の動き全てを捉えることはかなりの慣れが必要と言われている。話し手との慣れの違いにより若干の差異はあるが、個々の話し手における口型が同じでないこと、および発語の際の口型変化が個人間で必ずしも同じではないことも影響を与えているものと考えられる。

日本語においては、100 個余りの語音があるのに対して、これらが発語する場合の口型は僅か 6 種類の基本口型の組み合わせで構成されている⁵⁾。これらは、図 1 に示すように、5 個の母音を示す口型と、口を閉じる形（閉唇）である。

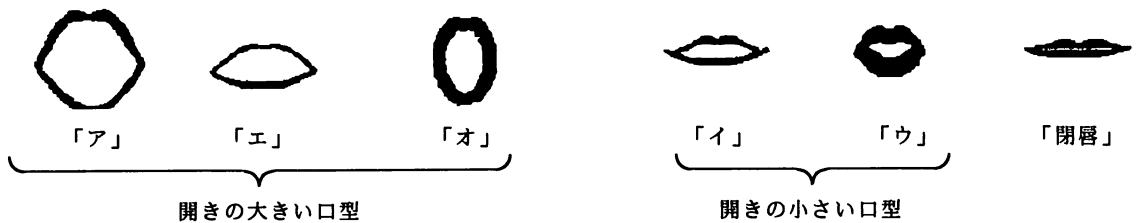


図 1 日本語の語音における 6 種類の基本口型

		音の終わりの口の形(～段の口型)					
		「ア」	「エ」	「オ」	「イ」	「ウ」	「閉唇」
音の 始めの 口型	後と同じ (なし)						
	「イ」						
	「ウ」						
	「閉唇」						
次に来る音の 始めの口型 「閉唇」							

図 2 15 通りの口型パターン

そして、これら 6 種類の口型を組み合わせる日本語の語音を発音するが、その口型パターンは図 2 に示すように 15 通りある。それぞれの口型パターンは、仮名 50 音表の各段（ア段、オ段など）の「～段」に相当する「終わりの口型」の前に、ごく短い小さな「始まりの口型」を付けた組み合わせパターンである。なお、「始まりの口型」は、カ行の K、サ行の S などに相当する部分である。これらの種類は以下

の通りである。

始まりの口型：「イ」、「ウ」、「閉唇」の3種類、

終わりの口型：「ア」、「エ」、「オ」、「イ」、「ウ」の5種類

これらの口型を組み合わせたパターンは次の通りである。

(1)ア段の口型パターン

- ・口型「ア」：あ、か、は、が
- ・口型「イ」+「ア」：さ、た、な、や、ら、ざ、だ、きゃ、、しゃ、ちゃ、にゃ、ひゃ、りゃ、...
- ・口型「ウ」+「ア」：わ
- ・口型「閉唇」+「ア」：ま、ば、ぱ、みゃ、びゃ、びゃ

(2)エ段の口型パターン

- ・口型「エ」：え、け、へ、げ
- ・口型「イ」+「エ」：せ、て、ね、へ、れ
- ・口型「閉唇」+「エ」：め、べ、ぺ

(3)オ段の口型パターン

- ・口型「オ」：お、こ、ほ、ご
- ・口型「ウ」+「オ」：そ、と、の、よ、ろ、きよ、しよ、ちよ、によ、ひよ、りよ、...
- ・口型「閉唇」+「オ」：も、ぼ、ぽ、みよ、びよ、びよ

(4)イ段の口型パターン

- ・口型「イ」：い、き、し、ち、に、ひ、り、ぎ、じ
- ・口型「閉唇」+「イ」：み、び、び

(5)ウ段の口型パターン

- ・口型「ウ」：う、く、す、つ、ぬ、ふ、る、ぐ、ず、きゅ、しゅ、ちゅ、にゅ、ひゅ、...
- ・口型「閉唇」+「ウ」：む、ぶ、ぶ、みゅ、びゅ、びゅ

他に、促・撥音節の口型がある。ア段、エ段、イ段の音に続く促・撥音節は口型「イ」と同じになり、「オ」段、「ウ」段の音に続く促・撥音節は口型「ウ」と同じになる。例外として、「ぶどう」のように延ばして発音する単語の場合には、口型「オ」のままで終わることがある。

一方、同じ口型パターンでも異なる言葉が多数存在する。例えば、「ゆか」と「くわ」などである。これを口型パターンで示すと、それぞれ口型「ウ」+「ア」、口型「ウ」+（口型「ウ」+「ア」）となる。すなわち、「くわ」と発音する場合には、最初の口型「ウ」に次の短い口型「ウ」が含まれるため、口型パターンは口型「ウ」+「ア」に見えてしまう。このように、聴覚的には異なる語音であるのに対して、視覚的には同じ口型パターンを有する語句群を同口型異音という。図3はこのような同口型異音の例で、「たいこ」と「ラジオ」の口型パターンを示している。

図3に示した「たいこ」、「ラジオ」は別の単語であるが、口型を読み取る場合、聞き手に見える口型は全く同じである。このように、口型から得られる情報は音韻レベルの一部であるため、口型を読み取るだけの読話では、音声情報のごく限られた部分しか得られないことになる。このため、語句の中には、30種類以上もの同口型異音を有する口型パターンが存在する。

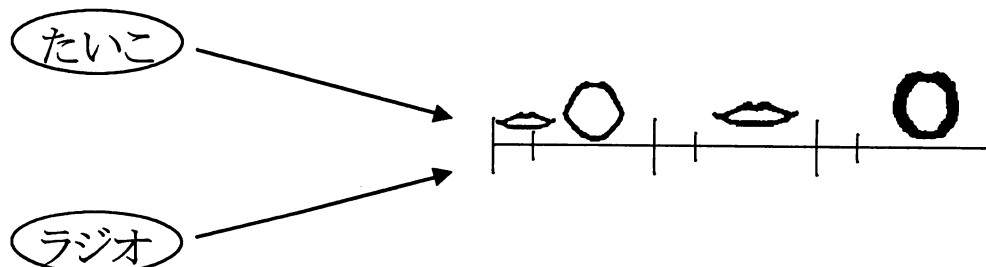


図3 同口型異音の例

しかし、残存聴力を有する人においては、補聴器により若干の音声情報を得ることによる、読話の正答率向上が報告されている。このことは、発語時にその内容に関して口型以外の情報を僅かでも与える

ことにより読話の正答率が向上することを示す一方で、残存聴力を有しない聴覚障害者における読話の正答率向上に他の情報を提供する必要があることを示している。本研究では、この観点から、他の情報として、発声した語句の先頭文字（語頭文字）の同時提示を提案している。

3. 実験方法

実験のために、発語における口唇の動きだけを見て読話を行う環境と、これに加えて発語内容の語頭文字を同時に提示して読話を行う環境を設定した。そして、被験者を用いてそれぞれの場合における読話結果をデータとして収集し、これから正答率を算出して、語頭文字付加の有無における正答率の違い等について検討した。

発語を見る環境は動画を用いるものとし、サンプルとして、(社)全日本難聴者・中途失聴者団体連合会が制作したビデオテープ・テキスト全 12 巻（「家族編」4 巻、「生活編」4 巻、「社会編」4 巻）を使用した。この中で、「家族編」は読話の基礎的な知識と技術、「生活編」では読話の知識と応用、「社会編」では読話の活用について構成されている。これらのビデオテープから、顔全体がはっきり映り比較的的口型が明確に読み取れる映像の中から、さらに 2～6 文字で構成される単語を発声している部分を実験用動画像として抽出した。最終的には 50 種類の動画部分を選定し（「家族編」第 1 巻から 16 語、「社会編」第 4 巻から 34 語）、これらを AVI ファイルとして保存した。50 種類とした理由は、ビデオテープに記録されている映像内容による制限である。AVI ファイル作成に使用したシステムは、パソコン（日本アイ・ピー・エム製 PC300GL）、ビデオデッキ（日本ビクター製 HR-D100）、ビデオキャプチャボード（アイ・オー・データ機器製 GV-VCP/PCI）を使用した。保存した各 AVI ファイルの再生時間は、発語する単語の種類により異なるが、概ね 2～3 秒である。この各 AVI ファイルにおいて、その語頭文字を平仮名で同時表示する条件の有無を設定し、再生するためのプログラムを作成した。図 4 は、単語「み・か・ん」を発声している時の再生画像に、語頭文字「み」を同時に掲示している例である。

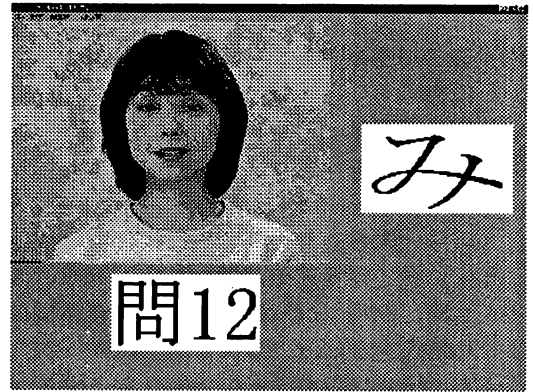


図 4 語頭文字併用方式の再生例

各被験者ごとの実験手順を以下に示す。

- 1) 50 個の動画データの中からランダムに 20 個選び、語頭音文字の提示なしで 1 問ずつ再生する。
- 2) 被験者に、各再生画像に対して、発語していると考えられる単語を筆記で回答させる。
- 3) 次の問題画像再生を被験者の指示により行わせ、20 個すべての画像について 2) を行う。
- 4) 50 個の動画データの中からランダムに 20 個選び、語頭文字の提示付きで 1 問ずつ再生する。
- 5) 被験者に、各再生画像に対して、発語していると考えられる単語を筆記で回答させる。
- 6) 次の問題画像再生を被験者の指示により行わせ、20 個すべての画像について 5) を行う。
- 7) 2)～3) および 5)～6) で使用した単語の認知度・使用頻度について回答させる。
- 8) 上記 2 種類の回答結果から、それぞれの場合における正答率を算出する。

問題に使用する 20 種類の単語をランダムに 2 回選択した理由は、学習効果を軽減するためである。

実験の被験者として、16 人の健聴者（22～25 歳／男性）と、普段健聴者とのコミュニケーションにおいて読話を用いている聴覚障害者 9 人（両耳 100dB 以上の聴覚障害／2 級、24～28 歳／男性 7 人・女性 2 人）を対象とした。

4. 結果

4.1 結果

実験結果として、健聴者グループと聴覚障害者グループそれぞれにおける正答率を図 5 と図 6 に示す。いずれのグラフにおいても、各被験者ごとに、従来の読話方式だけを用いた場合の正答率と、語頭文字併用方式による正答率の違いを各線分の傾きにより示している。次に、図 7 と図 8 は、読話方式だけの

場合と語頭文字を同時に提示した場合のそれぞれにおいて、健聴者と聴覚障害者の正答率を比較したものである。

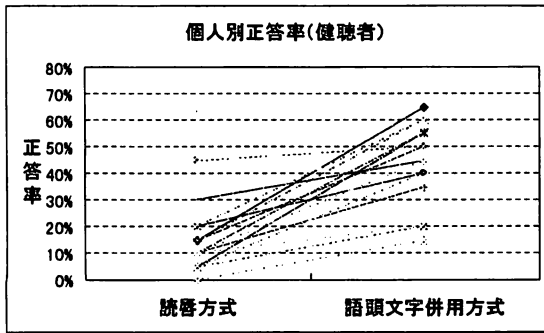


図5 健聴者における個人別正答率

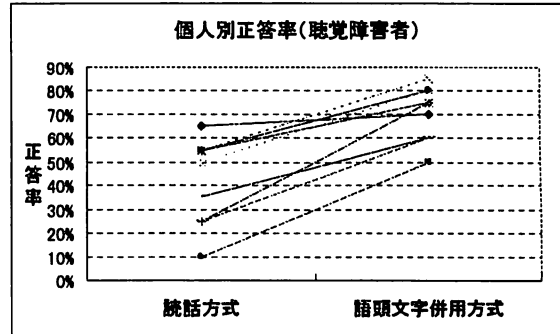


図6 聴覚障害者における個人別正答率

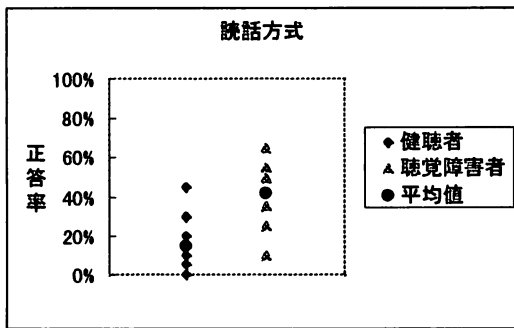


図7 読話のみにおける正答率の差異

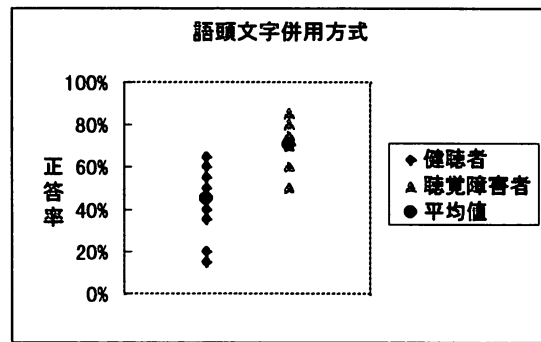


図8 語頭文字付加における正答率の差異

4.2 正答率の改善

図5と図6から明らかなように、健聴者および聴覚障害者のいずれにおいても、読話方式だけの場合よりも、語頭文字を同時に提示した場合において、正答率が有意に改善された（危険率5%）。

一方、図7と図8から明らかなように、語頭文字の同時提示の有無に関わらず、聴覚障害者グループにおける正答率が健聴者グループのそれよりも有意に高いことが示された（危険率5%）。

4.3 文字数と正答率の関係

図9は、発声した単語の文字数と正答率の関係を、健聴者グループと聴覚障害者グループのそれぞれにおいて示したものである。各文字数ごとのサンプルデータ数が少ないため統計的解析には至っていないが、読話方式だけの場合よりも、語頭文字を同時提示した場合において正答率の改善される傾向が見られた。

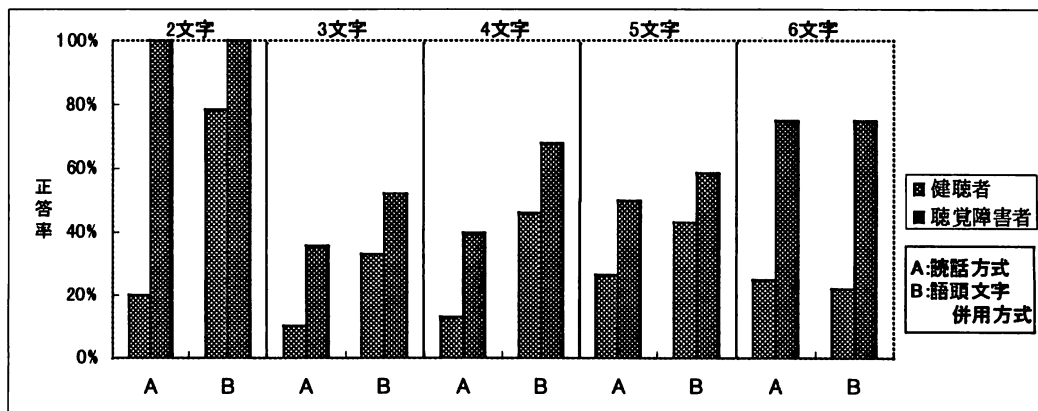


図9 単語の文字数別正答率

5. 考察

5.1 正答率の改善

既述したように、これまでの研究により、聴覚障害者を対象とした単語レベルにおける読話の正答率は30%程度で、補聴器の併用により50%程度まで改善されることが示されている。これに対して、本実験の結果は、聴覚障害者において語頭文字の同時提示を行うことにより正答率が約70%にまで向上することを示している(図10)。これにより、補聴器の併用が効果をもたらさない障害レベルの聴覚障害者においても、語頭文字の同時提示を行うことによって正答率を向上できる可能性があると考えられる。一方、健聴者においても語頭文字の同時提示により正答率が14.7%から44.7%と向上したが、日常生活において読話方式を使用していないため、聴覚障害者ほど高い正解率に至らなかったものと考えられる。

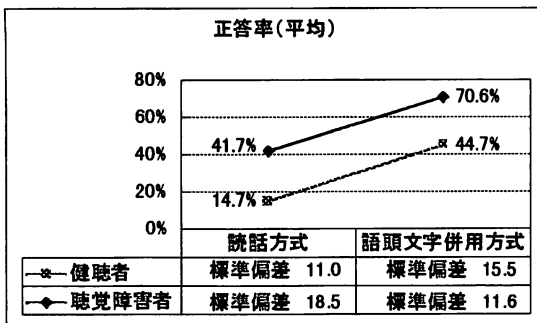


図10 個人別正答率の平均

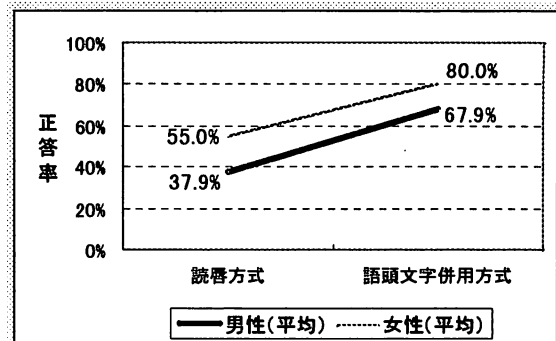


図11 聴覚障害者の男女別正答率の平均

5.2 男女差

図11は、聴覚障害者グループのデータを、更に男女別で示したものである。この結果、男性と女性のいずれも語頭文字の同時提示により正答率が有意に向上する事が示された(危険率5%)。一方、今回の実験では被験者数が少ないため統計的な評価には至っていないが、男性と女性のいずれも、正答率の向上による差は無いと考えられる。

5.3 語句の文字数と正解率

2文字の単語では、健聴者・聴覚障害者ともに、語頭文字併用方式では正答率が高いことが示された。この理由として、2文字の単語に比べて、3文字～5文字の単語の方が、単語の文字数自体を誤認しやすくなること、同口型異音の種類が多くなることなどが理由として考えられるので、正答率が低いことが考えられる。

5.4 実験に使用した単語の認知度

実験後のアンケート結果においては、単語間に認知度の相違が見られたが、有意な差は認められなかった(危険率5%)。これは、実験に使用した単語について難易度を考慮した分類を行わず、全体から無作為に抽出したためと考えられる。

一方、アンケート結果から、当該単語を知らないことによる無回答、知っていても使用頻度が低いことによる誤答、などが認められた。これらも、語頭文字を同時提示しても正答率を向上できない理由と考えられる。例えば、「まわた」は「真綿」と書くが、アンケートの結果を見ると単語自体を知らない人は健聴者で43.8%、聴覚障害者で77.8%と高い割合を示した。

一方、提示した動画において、発語開始時の口型と終了時の終わりの口型がはっきり分からないと言う意見もあり、ファイルの保存方法等について考慮する必要があると考えられる。

これらに対して、同口型異音による誤認識も見られ、特に口型パターンの中で各口型間の変化(動き)の少ない単語において、両方式で低い正解率となった。例えば、表1に示すように、読話方式において、発語が「いっぶく」という口型変化の少ない語句では、その同口型異音である「しんぶん」および「きっぶ」という回答が多くみられた。「きっぶ」の場合は、「ぶ」の後に続く口型パターンが変化しないために、読み取れなかったものと考えられる。このことから、単語レベルにおける同口型異音の問題は、語頭文字併用方式を用いても解決されない場合があるものと考えられる。

表1 読話方式での「いっぷく」の回答例

	出題数	正答数	回答例
健聴者グループ	6問	0	「きつぷ」2
聴覚障害者グループ	4問	0	「きつぷ」1、「しんぷん」2、無回答1

しかし、今回のような単語レベルではなく文章レベルの場合には、文脈などによる付加情報を利用した単語の推定が期待できるため、このような誤認識が軽減される可能性があると考えられる。一方、同口型異音の複数単語においては、各被験者が生活の中でどの単語に馴染んで多用しているかという観点からの検討も必要であり、今後の検討課題である。

6. 結論

本研究では、読話における正答率を改善する一つ的手段として、発語と同時に語頭文字を提示する方式を提案し、その効果を検証するため、健聴者および難聴者による実験を行った。この結果、語頭文字を同時提示することにより、読話による正答率が有意に改善されることを示した。一方、単語の種類によっては、同口型異音の有無あるいは周知度の相違により正答率が向上しない場合があることが明らかとなった。今回の実験で使用した単語数は高々50個であるため、日常会話で使用する単語を基本とした更なる検証が今後の課題である。一方、同口型異音の問題は、文章の読話においては文脈の流れから推定できる可能性もあり、これによる正答率の向上が期待できる。この検証についても今後の課題である。

謝辞

本研究においては、国立身体障害者リハビリテーションセンター研究所感覚機能系障害研究部聴覚・言語障害室、福田友美子室長より文献を紹介いただきました。また、実験に際しては、本学に在学する健聴者及び近隣に在籍する聴覚障害者の方々に協力いただきました。ここに記して深く感謝の意を表します。

参考文献

- 1) 福田友美子、坂本幸、黒木総一郎：読話に対する残聴の併用効果、日本音響学会誌、32巻4号、1975
- 2) 福田友美子：口型による音声情報の伝達 -各種の読話の補足手段-、日本音響学会誌、43巻5号、1987
- 3) 福田友美子：人工内耳埋め込み患者の音声知覚の計算機利用による検査・自習訓練システム、電子情報通信学会、ET-90-122、1991
- 4) 福田友美子、四日市章：聴覚障害者の視覚と聴覚による音声知覚の評価、音声言語医学、Vol.33No2、1992
- 5) (社会法人)全日本難聴者・中途失聴者団体連合会：豊かなコミュニケーションに向けて -読話のためのビデオテキスト-、「家族編」p64～p90、「生活編」p63～p73、1998

Improvement of accuracy in Lip Reading by showing a top letter of a spoken word

Tsuguyuki Oda, Hidehisa Oku*

Graduate School of Engineering,

**Department of Information and Computer Engineering,*

Faculty of Engineering,

Okayama University of Science,

Ridai-cho 1-1, Okayama 700-0005, Japan

(Received November 4, 1999)

It has been reported by previous studies that additional audio information by using hearing aids significantly improves accuracy in lip reading by hearing impaired persons. However, this method have been not effective for a certain group of hearing impaired persons who don't have hearing capability enough to receive audio information by using hearing aids.

In these circumstances, the authors have suggested a new method to improve accuracy in lip reading by using not audio information but linguistic information. The linguistic information is given by showing a top letter of each spoken word, with movement of lip. Effectiveness of the suggested method has been evaluated by measuring accuracy of lip reading in both the suggested method and the ordinary method that don't have additional information. Subjects in the experiment were both hearing impaired persons (7 males and 2 females) and able persons (16 males). A hundred of moving pictures were prepared for the experiment. The content of each moving picture was brief scene, from 2 to 3 seconds, that a person was speaking a word. The fifty of the moving picture were without additional information, and each of the rest had a top letter of a spoken word. In the experiment, each subject was asked to answer what word was spoken in each moving pictures.

The result indicates that the suggested method significantly improves accuracy of lip reading than the ordinary method does ($p=5\%$).