

音声認識技術を利用した字幕呈示システムの活用に関する研究 —聴覚障害者のニーズに即した呈示方法—

中野 聡子¹⁾・金澤 貴之²⁾・牧原 功²⁾・黒木 速人³⁾・
上田 一貴¹⁾・井野 秀一⁴⁾・伊福部 達¹⁾

本稿では、聴覚障害者のニーズに即した音声同時字幕システムの呈示方法について述べた。聴覚障害者にとって、音声認識字幕の文章が読みにくい理由を説明し、これまでの運用経験及び実験から、1) 字幕精度は100%に近いものでなければならない、2) 短い文字数での改行が読みやすさにつながるについて述べた。また、今後の課題として、1) 対象者の日本語力や場面に応じた復唱の方法、2) 英語で行われる授業での活用、3) ポータブル機器への呈示について提案した。

キーワード

聴覚障害者、音声同時字幕システム、字幕精度、改行パターン、復唱

1. 音声認識技術を利用した字幕呈示システムに求められる字幕精度

1.1 音声認識技術を使用した字幕呈示システムの概要

情報保障は聴覚障害者の社会参加にとって必要不可欠なものである。近年、音声認識ソフトウェアの認識率が向上し、また、手頃な価格で入手することが可能になったため、聴覚障害者向けに、これを利用して話者の音声を字幕化する情報保障手段が注目されている。なぜなら、音声認識技術を利用した字幕呈示は、パソコン要約筆記と比較した時、よりリアルタイムに、発話速度に追従して多くの音声情報を呈示できる方法だからである。従って、話者の音声を正確に字幕化することができれば、講義を最も臨場感あふれる形で伝えることができる。

音声認識技術を利用した字幕呈示方法の共通点は、音声認識ソフトウェアを利用することにある。現在国内で利用可能な主なソフトウェアには、Via Voice 10.5(IBM製)、Dragon Naturally Speaking 2005 Professional/Select (株アセンディア エー・エス・アイ・カンパニー製)、Ami Voice 2008 Es (株アドバンスト・、メディア製)などが挙げられる。こうした音声認識ソフトウェアに、話者の音声を直接認識させるか、復唱によって特定話者の音声に変換するか、また、認識結果の修正をするかしないかで、字幕精度や字幕呈示までのタイムラグ、そしてコストにも差が出てくる(図1参照)。

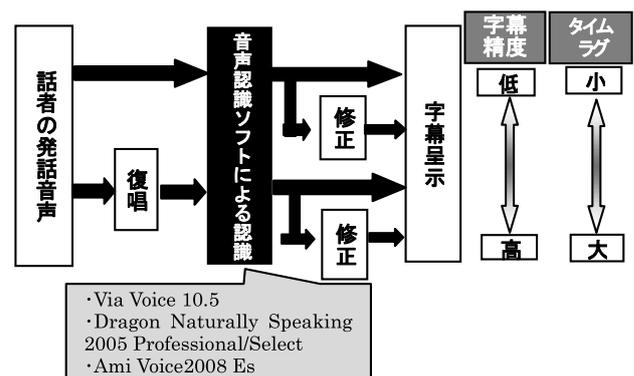


図1 音声認識技術を利用した字幕呈示システムの概要

1.2 音声同時字幕システムとは

東京大学先端科学技術研究センターの伊福部達の研究グループと株式会社ビー・ユー・ジーが共同開発した「音声同時字幕システム」は、話者の音声を訓練された復唱者が復唱し、若干の同時修正を加えることで精度の高い字幕を呈示することができる。なお、「音声同時字幕システム」は、音声認識エンジンにVia Voice (IBM製)を利用した独自開発のソフトウェアを用いている。訓練を受けた復唱者を入れる理由は、1) 不特定話者環境を特定話者環境に変換する、2) 適切な発音と一定の発声強度を保つことにより、音声認識率を高めるためである。

現在の技術レベルでは、通常の会議や講義での発話スタイルのまま不特定話者が音声認識ソフトウェアを利用した場合、その認識精度は60~70%台、また未経験者が明瞭に発話するように意識した場合は80%台、ある程度経験を積んで初めて95%前後という高い認識率

¹⁾ 東京大学先端科学技術研究センター
²⁾ 群馬大学
³⁾ 筑波技術大学障害者高等教育研究支援センター
⁴⁾ 産業技術総合研究所

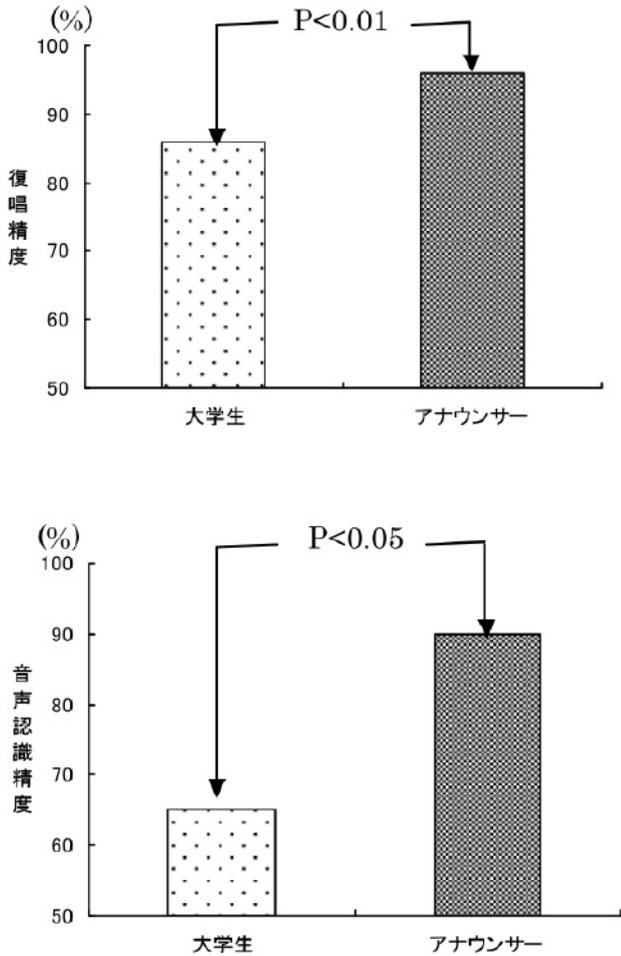


図2 被験者別復唱精度および音声認識精度 (井野ら, 2003)

を実現できるようになるとされている (三好, 2008)。

また復唱については、話者の音声を聞き取ると同時に復唱を脱落なく、かつ音声認識ソフトウェアにとって正しく認識しやすい発音・発声強度で行うことはたやすいことではない。

井野・黒木・加藤・渡邊・堀・伊福部 (2003) は、発声・発話トレーニングを受けた民放アナウンサーと一般の大学生を対象に、復唱課題を行ったところ、復唱精度、音声認識精度共に、両者には大きな開きがあったと報告している (図2参照)。

同システムでは、アナウンサーなど、発声・発話トレーニングを受けた者に復唱させることによって、復唱精度は95%以上、音声認識精度は90%程度での文字化を実現させている。これに修正クライアントを利用して同時修正を行うことにより、97.2%の精度で字幕呈示を行うことが可能となっている (黒木・井野・中野・加藤・渡邊・堀・伊福部, 2003)。

1.3 音声認識技術を利用した字幕文章の誤変換に見られる特徴

聴覚障害者向けの情報保障としての音声認識技術を利用した字幕呈示システム運用に際して留意しなければならないのは、字幕の読み手である聴覚障害者は聴者とは異なり、話者の音声フィードバックがないということである。誤変換のパターンが聴者にとって容易に正変換を予測できるものでも、聴覚障害者には困難なものも多く含まれていることが挙げられる。

菊池 (2006) は、母音変化を起こした誤変換を対象として、聴覚障害者の誤変換の推測について検討した。母音変化を起こした誤変換を4つのタイプに分け (表1参照)、それぞれのタイプの誤変換を含む文章を、聴覚障害者と聴者の両者に呈示して正しい変換を推測する課題を行った (図3参照)。

その結果、全体的な傾向として、聴覚障害者は回答に時間がかかり、また正しい変換の回答率が聴者よりも低かったとしている。各タイプ別の正答率については、図3に示す。特に、語のつながりの変化や音素の変化が起きているときに、より推測が困難になることが明らかになった。

このことは、誤変換を含んだ字幕の文章を読むとき、聴覚障害者には聴者以上に負荷がかかっていることを示唆している。さらには、後節で詳しく述べるが、話者の発話そのものに含まれる文法的エラーがそのまま字幕化されることもあって、聴覚障害者は聴者以上に、音声認識ソフトウェアによって字幕化される文章に読みにくさ

表1 母音変化を起こした誤変換のタイプ (菊池, 2006)

		母音的な音素変化	
		なし<1>	あり<2>
語のつながりの変化	なし<Ⅰ>	I -<1> 例：助成 (女性)	I -<2> 例：相談五等 (相談事)
	あり<Ⅱ>	Ⅱ -<1> 例：共和 (今日は)	Ⅱ -<2> 例：脳症割って (直し終わって)

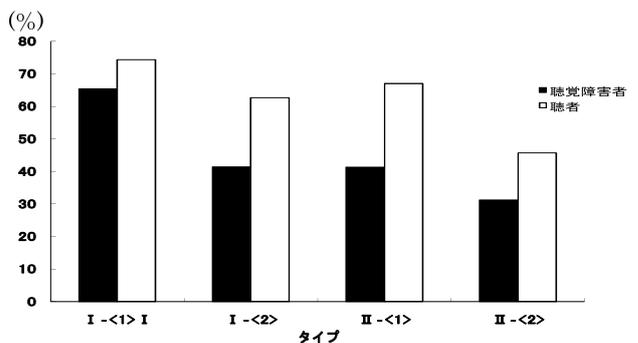


図3 誤変換タイプ別正答率 (菊池, 2006をもとに作成)

を感じていると言えよう。

そのため、我々はこれまでの音声同時字幕システムの運用経験から、聴覚障害者向けの情報保障を目的として使用する場合、その字幕精度はほぼ100%に近いものでなくてはならないと考えている。修正が字幕精度の向上に寄与するのは、修正者の人数にもよるが全体の10%以下に過ぎない。黒木ら(2003)のシステム運用では4名の修正者で修正前後に7%の差をもたらしている。こうしたことから、現在の技術レベルでは、ある程度訓練を受けた者による復唱、そして修正作業をはさむことが、聴覚障害者にも理解できる字幕精度の達成には欠かせないと考えられる。すなわち、タイムラグよりも字幕精度の向上を優先させる呈示手順をふむことが大切であると言える。

2. なぜ誤変換をなくしても字幕の文章は読みづらいのか

しかしながら、「音声同時字幕システム」において、95%以上の精度で字幕が産出されているときでさえ、「読みにくい」という聴覚障害者の声もある。

音声認識ソフトウェアを利用して呈示する字幕文章が読みにくくなる原因の1つとして、話者の音声そのまま字幕化されることが考えられる(中野・牧原・金澤・菊池・黒木・井野・伊福部・福島, 2006; 中野・牧原・金澤・中野・新井・黒木・井野・伊福部, 2007)。書き言葉は周到に練り上げられた高度に圧縮された表現であるのに対して、話し言葉はとぎれとぎれに冗長な表現で文が産出され、表現の順序を時間的にさかのぼって修正できないため、通常の構文や語順とは異なる、反復や倒置などの不整表現が現れやすくなると述べている(なお、本字幕システムでは、意図的でない反復については復唱段階で省略しているため、字幕に反復が出ることはほとんどない)(半澤, 1990; 定延, 2005)。

中野・牧原・金澤・菊池・黒木・井野・伊福部・福島(2005)は、音声認識ソフトウェアを利用して呈示される字幕文章のわかりにくさを構成すると考えられる要因について、以下のように整理している。なお、字幕文章の検討には、群馬大学において、音声同時字幕システムが運用された講義「障害者情報保障論」(2004年4~8月, 全15回)、「障害児教育福祉論」(2004年10月~2005年3月, 全14回)のうち、修正作業を含め、安定してシステム運用が行えた1015分間の字幕ログデータを用いた。

1) 従属度の高い従属節の中に従属度の低い従属節が含まれるような文構造

文章自体も長くなる傾向にあり、非常に意味が理解しづらい。ただし、話者の音声記録を再生してみると、さほどわかりにくさを感じない。音声言語では、このような構造になっていても、ポーズが、かかっていく先が遠

くにあることを示すサインとなっている可能性がある。

2) 呼応関係の消失

呼応関係が消失していることで、最後まで文を読み終わっても意味が理解できない。ただし、すべての場合において、意味が理解できなくなるとは限らない。

3) 主語の消失

話し言葉、書き言葉ともに一文の中で主語が省略されることはある。しかし、話し言葉を文字化すると、主語がないことで内容理解に混乱が生じる例があった。

4) 句読点の問題

句読点が打たれていないことで、複数の意味の解釈が生じてしまう場合がある。例：私は資料を読みながら通訳を受ける学生の問題について考えてみたい。

5) 省略

省略された文章では、文法におかしく、意味理解困難な文章になることがある。例：本当にだけれども→本当に必要だけれども(下線部を省略)

6) 言い間違い・言い直し

話者自身も訂正せずに話している言い間違いで、受動と能動の不統一、主語・述語の係り受けのねじれ、といった文のねじれは、文字化された時に不自然で理解しにくい文となる。

言い直しで、意味不明となってしまっていた例としては、文の中止するところまで発話が続いた後で、「あっ」「え〜」「じゃなくて」といったことばをはさみず、言い直しの発話が続いたものが挙げられる。このケースでは、おそらく、言い直しを示すサインは、音声以外の手段、すなわち、首振り、手振りなどの身振りによって、前文の打ち消しを伝えたためではないかと思われる。

7) 引用表現部分のわかりにくさ・抽出のしにくさ

引用の中に会話のやりとりがあると、1人の会話がどこで終わるのかわからず、意味理解に混乱が生じることがあった。

実際の文章記録全体としては、1つの文章の中に上述した要因が混在することで、ますますわかりにくくなっていた。

以上に述べてきたような字幕文章のわかりにくさは、句読点や記号の挿入によって解決できるものもあるが、文法的エラーは、話しことばの中にも含まれるものである。音声言語の場合にはこうした不整表現を含みつつ、多くの節が続く長大な文となっていくうちに、話者自身もエラーに気づかず話していることが多い。しかし、文字には、はっきりとエラーが現れるので、非常に読みづらい字幕となってしまう。

これまで、聴覚障害者からの「音声認識による字幕はわかりにくい」といった声に対して、それは聴覚障害者自身の日本語力の問題ではないかと考えられてきた。しかし、これらの例からもわかるように、字幕の読みにくさの原因を、読み手である聴覚障害者の言語能力の問題

にのみ帰することができないのは明らかである。それはこうした字幕を聴者に見せた場合にも、同じ感想が聞かれることからわかる。

しかし、注目に値すべきことは、聴者が、話者の実際の音声聞きながら字幕を見ているときには、こうした文法エラーは全く気にならないということである。このことは、音声言語の認知過程と文字言語の認知過程が必ずしも同一ではないことにより、音声言語の認知過程では問題とならないエラーが文字化された際には大きな障害となるという可能性を示している。

とはいえ、文全体の構造に影響するような文法的エラーは、修正者が容易に修正できないものである。とすると、文法的エラーを含む文をいかに理解しやすい形態で呈示できるか、が課題となってくる。

3. 改行挿入実験から明らかになった聴覚障害者の読みのストラテジー

前節では、音声認識ソフトを利用して文字化を行った字幕文章には、話者の文法的エラーが含まれ、また一文が長く多数の節が連なる冗長な文章として産出される傾向があることについて述べた。

我々は、字幕の文章がこうした特徴を抱えていることから、通常書き言葉の文章とは異なったスタイルでの改行が読みやすさの向上につながるのではないかと考えた。そこで、聴覚障害者と聴者を対象に、どのような規則の下に行われる改行がもっとも字幕の読みやすさを向上させているか、1) シェッフエの対比較法による平均嗜好度、2) 字幕を読んでいる時の眼球運動の特徴を分析する以下のような実験を行った(中野・牧原・金澤・黒木・上田・井野・伊福部, 2008)。なお、実験1, 2ともに、2007年10月に実験を実施した。

3.1 実験1：対比較法による改行の嗜好度分析

3.1.1 被験者

先天性聴覚障害者8名。年齢は、25-37歳(平均31.4, SD4.7)。聴力は良耳で78-130dB(平均105.3, SD14.6)。本研究では、字幕の読みやすさについて分析するため、高い日本語能力を持ち、短大卒以上の者を被験者とした。

統制群として、聴者8名。年齢は、21-23歳(平均20.8, SD1.3)。東京都内にある大学及び大学院に在籍中の学生。

3.1.2 呈示文

群馬大学において、音声同時字幕システムが運用された講義「障害者福祉論」(2004年10月-2005年3月, 全15回)のうち、修正作業を含め、安定してシステム運用が行えた字幕ログの中から、一部分を取り出しても、内容が十分にわかる1つの文章を選んだ(文字数: 370字)。誤認識はすべて修正した上で、以下に示す5つの

改行パターンによって文を作成し、呈示文とした。

[A]: 改行なし

[B]: 句点で改行

[C]: 句読点で改行

[D]: 句読点及び/ten/または/te/で終わり/n/が付加されて/ten/に転化しやすいもので改行

[E]: 句点及び25文字前後の読点で改行

呈示文は、パソコン要約筆記用ソフト「IPTalk」(Ver. iptalk9i109)を用いて、実験条件に合わせた改行をした上で、実際の字幕呈示と同じような文字再生となるようにした。

3.1.3 実験手続

実験は個別に行った。被験者は机においた2台のノート型パーソナルコンピュータ(Apple製iBookG4)の前に座り、2つの呈示文がランダムにディスプレイに呈示されるのを読んで、各被験者の判断により読みやすさについて5段階のスコアによる視覚的主観評価を用意された回答用紙に記入した。

5つの呈示文は明らかに改行の仕方に違いがあることから、呈示文の位置を考慮しないこととし、被験者1名につき、10通りの組み合わせについて比較してもらった。

3.1.4 結果及び考察

改行条件を変えた5つの呈示文について、シェッフエの対比較法(中屋変法)によって得られた全体群における平均嗜好度は主効果が有意であり($F(F(4, 96) = 39.79, p < .01)$)、5つの改行条件の読みやすさには有意差があった。全体群での平均嗜好度は、[E], [C], [B], [A], [D]の順であり、[E]が最も読みやすく、[D]が最も読みにくいと感じていることが明らかになった。このことから、改行は必ずある方が良いが、文章の途中で突然、語が切れる改行には違和感がある、ということが明らかになった。

聴覚障害者群と聴者群に分けて平均嗜好度を産出した結果を図4に示す。聴覚障害者群と聴者群に共通していたのは、[D]条件のように、文章の途中で語が切れる改行は最も読みにくく、句読点及び25文字前後で改行する[E]条件のように、あまり長くなりすぎないところで改行が入るものが読みやすいという評価であった。

2つの群を比較すると、句点改行の[B]条件と句読点改行の[C]条件に関して、嗜好度の順番が入れ替わっており、また聴覚障害者群は2つの条件の嗜好度の差が聴者群よりも大きかった。

嗜好度の個人差の結果から、句読点改行の[C]条件の嗜好度の中央値-0.0706より高い評価をした被験者は聴覚障害者が4名(D1: 0.3294, D4: 0.7294, D7: 0.7294, D8: 0.5294)、聴者が3名(H3: 0.1294, H5: 0.1294, H6: 0.5294)であり、聴覚障害者の評点が聴者に比べて

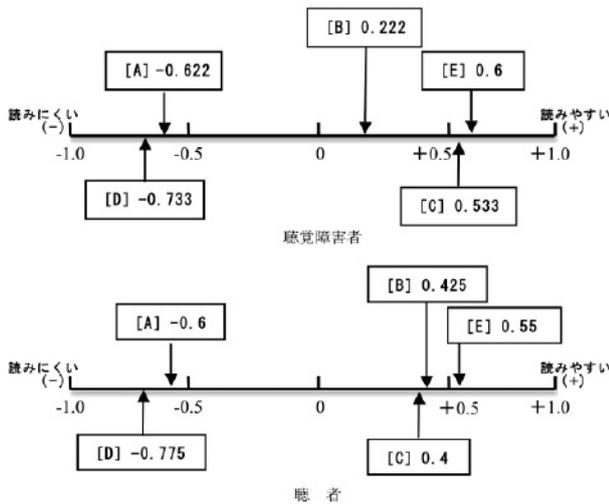


図4 シェッフェの一対比較法による字幕の読みやすさについての平均嗜好度 (中野ら, 2008をもとに作成)

中央値との差が大きい者が多いことがわかった。ただし、分散分析を行ってみたところ、統計的な差は認められなかった。今後人数を増やして検討したい。

また、嗜好度の個人差の結果を見ると、それぞれ改行条件に対する評価にはばらつきがあることから、各改行条件での読みやすさの評価は、実際の字幕での読み方とも深く関連しているのではないかと考えられた。そこで、実験2では、各改行条件での呈示文を読んでいるときの眼球運動を調べ、実験1の読みやすさの評価と合わせて考察することとした。

3.2 実験2：眼球運動測定による字幕の読み方分析

3.2.1 被験者

実験1と同じ。聴覚障害者、聴者ともに、裸眼またはソフトコンタクトレンズによる矯正視力で0.8以上であり、眼球運動測定装置の使用に問題はなかった。

3.2.2 呈示文

実験1と同じく、群馬大学での試運用における字幕ログの中から一部分を取り出しても、内容が十分にわかる文章を5箇所抽出し、誤認識はすべて修正した上で、以下に示す5つの改行パターンによって文を作成し、呈示文とした。なお、実験1で用いたのは別の箇所を使用した。

1つの呈示文は、213-415文字から成り、16-33行で構成されていた。呈示時間は52.13-83.66secであった。1行あたりの文字数が少ない呈示文では、呈示文の行数が多くなることから、同じ文字数の場合、改行の少ない呈示文に比べて呈示時間が長くなる。そのため、文字数よりもむしろ呈示時間に大きな差が出ないように注意し、すべての呈示文の呈示時間は、平均呈示時間±2SDの範囲内におさまるようにした。

また実験1と同じく、呈示文はパソコン要約筆記用ソ

フト「IPTalk」(Ver. iptalk9i109)を用いて、実験条件に合わせた改行をした上で、実際の字幕呈示と同じような文字再生となるようにした。

3.2.3 実験手続

実験は個別に行った。被験者に、キャリブレーション終了後、ディスプレイに流れる字幕を黙読してもらい、字幕を読んでいる時の眼球運動を測定した (EMR-AT VOXER, ナックイメージテクノロジー製を使用)。呈示順序は乱数を用いて順序効果が出ないようにし、各文の字幕呈示は1回のみとした。また、呈示文の意味を理解しながら読むための動機づけを目的として、各文の呈示後内容理解に関する質問用紙に回答をもらった。

3.2.4 結果及び考察

文字呈示画面を図5に示したように(I)-(V)の領域に分けて、各被験者のそれぞれの領域における停留点位置及び回数に着目して分析を行った。なお、文字は画面の一番下の行、左から順番に呈示されてゆき、次の行にわたる際にスクロールアップされて行が少しずつ上にあがってゆく。常に新しい文字が呈示されるディスプレイ画面の下部に停留点が多い読み方をするタイプとそうでないタイプに分けた。すなわち、(Ⅲ)(Ⅳ)(Ⅴ)領域における停留回数の呈示文全体に占める割合が被験者全体の平均 (13.61%, SD8.06) 未満であった前者のタイプを追従型、平均以上であった後者のタイプを非追従型とした。

なお、追従型の呈示文全体に対する(Ⅲ)(Ⅳ)(Ⅴ)領域の停留回数割合の平均は5.10% (SD1.79)、非追従型は、20.22% (SD2.64)であった。

聴覚障害者と聴者の読み方タイプを分析し、フィッシャーの直接法によって検定を行ったところ、聴覚障害者は、非追従型の読み方をするタイプが多い傾向がみられた ($p < .10$) (表2参照)。

読み方タイプと実験1で測定した個々の嗜好度の関連について、少ない文字数で改行をする[C]条件と[D]条件の2つの呈示文の嗜好度に着目し、各読み方タイプ別

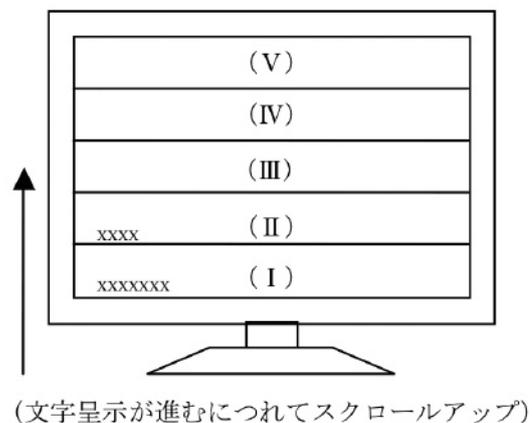


図5 停留位置の領域 (中野ら, 2008)

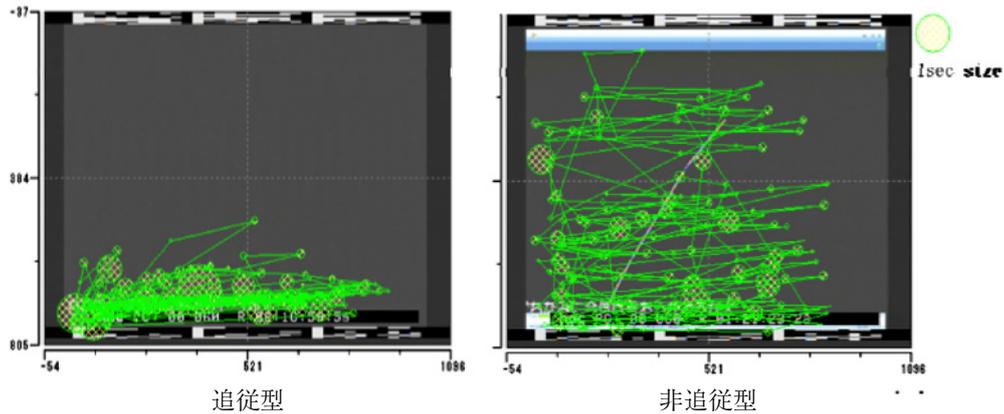


図6 停留点軌跡の例

表2 被験者群別読み方タイプの分布 (中野ら, 2008) (n=16)

	聴覚障害者	聴者
文字追従型	1	6
文字非追従型	7	2

表3 読み方タイプと呈示文の嗜好性 (中野ら, 2008) (n=16)

呈示文の嗜好性	読み方のタイプ	
	文字追従型	文字非追従型
[A][B][E] > [C][D]	6	3
[A][B][E] < [C][D]	1	6

に[C]条件と[D]条件の嗜好度を加算したものと、改行なしの[A]条件、句点で改行する[B]条件、句点及び25文字前後の読点で改行する[E]条件の嗜好度を加算したものと比較してみた(表3参照)。

フィッシャーの直接法による両側検定の結果、有意傾向がみられ ($p < .10$)、追従型の読み方をするタイプの被験者は、少ない文字数での改行をあまり好まず、非追従型の読み方をするタイプの被験者は、逆に少ない文字数での改行を好む傾向が認められた。

以上をふまえると、聴覚障害者では、非追従型の読み方をする者が多いことから、少ない文字数で改行する呈示文を好む傾向にあるのではないかと推測された。

行をまたぐ読み返し頻度は、被験者群や読み方タイプ、呈示文と関連があることが明らかになった。

群(聴覚障害者・聴者)×呈示条件([A][C][E])の分散分析を行ったところ、各要因の単純主効果は、群($F(1, 42) = 0.003, p < .05$)、呈示文条件($F(2, 42) = 0.032, p < .05$)で有意であった。また、交互作用も有意であった($F(5, 42) = 0.016, p < .05$)。多重比較の結果、聴覚障害者群では、句読点で改行する[C]条件での行をまたぐ読み返しが、改行なしの[A]条件や句点及び25文字前後で改行する[E]条件よりも有意に高かった。聴者群で

は3つの呈示文間に有意な差はみられなかった。

また、読み方タイプ(追従型・非追従型)×呈示条件で分散分析を行ったところ、単純主効果は各条件で有意となり(読み方タイプ($F(1, 42) = .001, p < .05$), 呈示文条件($F(2, 42) = .049, p < .05$)), 交互作用も有意であった($F(2, 42) = .001, p < .05$)。多重比較の結果、非追従型群において[C]条件での行をまたぐ読み返しの割合が[A][E]条件よりも有意に高かった。追従型群では、呈示文間に有意な差はみられなかった。

これらの結果から、非追従型の読み方をする多くの聴覚障害者群では、句読点という少ない文字数での改行をする[C]条件で、行をまたぐ読み返しが多いことが明らかになった(図7参照)。

以上の結果について留意すべきなのは、非追従型群の被験者では、[C]条件の嗜好度が高く、読み返しの多さが読みにくさにつながっていないと考えられることである。そこで、非追従型タイプの読み方をし、かつ[A],[E]

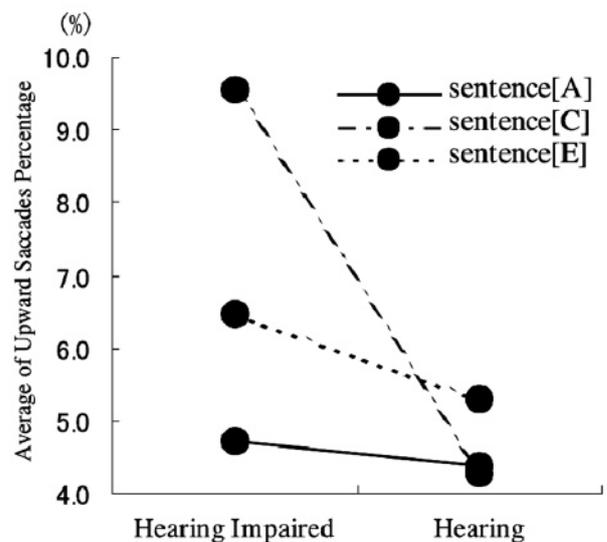


図7 呈示文[A][C][E]の読みにおける上方サッカードの割合の平均(中野ら, 2008)

条件よりも[C]条件を読みやすいと評価した聴覚障害者、D1, D4, D5, D7, D8の5名について、呈示文[A][C][E]の読み返し時を中心とした停留時間を解析してみた。文字呈示画面を図4に示したように(I)-(V)の領域に分けて、主に読み返し時に停留が生じる(III)(IV)(V)の領域について、1つの領域における総停留時間が5sec以上であった度数を求めたところ、呈示文[A](改行なし)：5名、呈示文[C](句読点で改行)：1名、呈示文[E](句点及び25文字前後で改行)：2名であった。句読点で改行する[C]条件では、行をまたぐ読み返し頻度は高くても読み返しにかかる時間は[A][E]条件に比較して短いことがうかがえる。個々のそれぞれの呈示条件における読み返し時の停留をみると、句読点で改行する[C]条件に比較して1行あたりの文字数が多い[A],[E]条件では、1回あたりの停留時間が長いことが多かった。また同一行内での読み返し頻度について、[C]条件での同一行内読み返し頻度が2SDを超えたD8を除いた4名で分散分析を行ったところ、主効果は有意であり($F(2,9) = .020, p < .05$)、多重比較の結果、[A],[E]条件は[C]条件に比較して同一行内での読み返し頻度が高い傾向が認められた($p < .10$)。このことから、句読点で改行する[C]条件は読み返しがしやすく、そのことが[C]条件への嗜好度を高める結果になったのではないかと考えられる。

文字呈示は2-15文字単位ずつ呈示されるので、追従型では文字呈示にそって視線を動かしてゆけば、1行あたりの文字数が多いことはさほど気にならないとも考えられる。また追従型では語尾の位置で停留して次の文字が出てくるのを待つ読み方をする者が多く、あまり読み返しをしない。このことも、読み直したい箇所を探す必要がないので、1行あたりの文字数がある程度多くてもよいと考えられる。逆に、句読点で改行する[C]条件のような呈示文では、視線の行変え運動をする頻度が高くなり、聴者に多い追従型ではあまり好まれなかった可能性がある。

これに対し、聴覚障害者においては、非追従型の読み方をするタイプが多く、比較的少ない文字数で改行する呈示文をより好む可能性があることが示唆された。また読み返し頻度における分析から、聴覚障害者は少ない文字数での改行をする字幕は読み返しがしやすく、読みやすいと感じているのではないかと考えられた。

今回実験の対象とした先天性聴覚障害者と聴者における改行の嗜好や読み方タイプに見られた相違の理由の1つとして、内的音声化の影響が考えられる。今回の実験で、呈示文を読むときはすべて黙読を条件としていたが、黙読時においても文字情報から音韻情報への変換過程が存在し、それが読みの過程に影響していることを示唆する報告がある(近藤・神長・馬塚・林, 2007)。先天性聴覚障害者は、通常の音読においても聴覚フィードバック

が困難なため、黙読時において内的音声化を行うことは少ないのではないかと考えられる。内的な音声変換を行わない場合、視覚情報(文字形態情報)から直接的に語彙アクセスや意味理解を行っていると考えられるので、それが非追従型の読みにつながり、かつ少ない文字数で改行する[C]条件の読みやすさ及び読み返しのしやすさに結びついた可能性がある。

4. 利用者のニーズと場面に応じた利用—音声認識技術を利用した字幕呈示システムの今後

前節では、短い文字数での改行が聴覚障害者にとっては、音声認識ソフトウェアを利用して字幕化される文章の読みやすさにつながる可能性について述べた。

音声認識ソフトウェアに関しては、各個人の音声特徴を登録することで高い認識率を維持しているわけだが、現在不特定話者への対応や口語的な表現への対応等を視野に入れた開発も進められている。(三好, 2008)。

今後、ソフトウェアの認識率向上、復唱者用及び修正者用ソフトウェアのインタフェース改良、機器接続の簡略化による利用しやすさの向上によって、聴覚障害者の情報保障手段として活用できる場合は益々広がってゆくと考えられる。こうしたことを見据えながら、大学等高等教育機関における、音声認識技術を利用した字幕呈示システムの今後の活用の可能性について以下に3点あげる。

4.1 復唱者の養成

1.2で述べたように、話者の音声を聞き取ると同時に復唱を脱落なく、かつ音声認識ソフトウェアにとって正しく認識しやすい発音・発声強度で行うことはたやすいことではない。しかし、音声認識ソフトウェアによる誤変換を、修正段階で字幕精度をあげられる幅はあまり大きくないため、訓練を積んだ復唱者の養成がこのシステム運用には欠かせない。

また、話者の音声をそのまますべて復唱するか、要約して復唱するか、という問題がある。話者の音声をすべて復唱して字幕化した文章は、イントネーションやプロソディなどのバラ言語情報が脱落しているとはいえ、その話者の発話に含まれる微妙な言い回しや語尾表現はそのまま呈示される。ただし、発話に含まれる文法的エラーや冗長な表現もそのまま呈示されるため、書きことばに比べて読みにくい文章ではある。一方要約して復唱する場合には、発話に含まれる特徴的な表現は要約によって削られてしまうが、文章としてはまとまっており字幕は読みやすくなる。

聴覚障害者には、日本語が得意でない人も多い。近年ではAO入試や推薦入試、社会人入学等によって大学で学ぶ聴覚障害者も多種多様となってきている。このような聴覚障害者向けへの字幕システム活用を考えると、

文章を理解しやすいうようにまとめて要約して復唱すべきではないかという考えもある。また、使用場面については、基礎知識を学ぶ概論のような講義では、語尾の詳細な言い回しや繰り返しは省略して要約し、演習やゼミなど発言者の見解や考え方を捉えなければならない講義ではすべて復唱する、といった使い分けが必要だろうと思われる。小畑・西川・高橋(1985)は、言語能力の高い聴覚障害者の場合、テレビ番組において台詞に忠実な字幕を用いた方が、台詞を要約した字幕よりも内容の理解が優れていたと報告している。音声認識対象となる聴覚障害者の日本語能力、そして使用場面に応じて復唱の仕方を変えていくことも検討する必要がある。

要約して復唱する場合については、要約の仕方についてのガイドラインを設けなければならないであろう。また、復唱の際に要約するかしないかでは、その情報処理の相違ゆえ、復唱の訓練の方法が異なってくると考えられる。復唱の適性を見極めるスクリーニングテスト作成と共に復唱の訓練方法についてもマニュアル等を設けていく必要がある。

4.2 英語の授業やゼミでの使用

現在、大学で行われる英語の授業は、実践的な英語コミュニケーション能力 (practical English communication ability) を身につけることを目的として、日本語ではなく英語で行われることが多くなっている。

しかしながら、こうした授業に対する聴覚障害学生への情報保障はノートテイクで行われていることが多い。PC要約筆記ではなくノートテイクである理由は、英語でのタイピング速度の問題であると思われるが、手書きであること、日本語話者である支援者の聞きとり及び要約能力の制限などから、伝えられる情報量は日本語の授業におけるノートテイクと比べてもさらに少なくなってしまうことは否めない。

音声認識ソフトウェアを利用して英語で字幕を呈示するシステムの活用には、ネイティブの英語話者である復唱者の確保が鍵になってくる。

Hoogenboom, Uehara, Kanazawa, Nakano, Kuroki, Ino, and Ifukube (2008) は、群馬大学で2007年度に行われた講義「Applied Linguistic Research I」(応用言語学研究 I) に、音声同時字幕システムを2回運用して英語の字幕呈示を行った結果について報告している。日本語話者の教員が英語で行う講義であったが、ネイティブの英語話者である教員が復唱をし、英語科の学生が修正を行う体制で試運用を行った。なお、音声認識ソフトウェアには、英語版 Via Voice (Via Voice for Windows Pro USB Edition Release 10, IBM 製) を使用した。

その結果、復唱については以下のような課題が報告されている。

1) 話し言葉の典型的な特徴である話者の発話に含まれ

る文法的エラーや冗長的表現をそのまま復唱することに対する難しさ。

- 2) 「,」「.」「?」などを復唱時に加えなければならないことへの慣れ。
- 3) 英語話者であっても地域によってアクセントの違いがあり、音声認識ソフトウェアに正しく認識されやすい発音で復唱しなければならないこと。
- 4) 話者の発話にある日本語の単語や固有名詞などが英語版の音声認識ソフトウェアでは正しく認識されないこと (例: gengo-gaku → gang again go back to)。
- 5) 別室で復唱するため、教室の様子が変わらず受講生らの反応をふまえて話す話者の発話のスピードがつかみにくいこと。

1)~3) については、英語での字幕運用に際しても英語のネイティブスピーカーなら誰でもよいのではなく、日本語での運用と同様、復唱における訓練が必要であることを示唆している。4), 5) についてはハード面の問題として今後改善が必要である。

しかしながら、字幕システム運用で伝えられる情報量はノートテイクやPC要約筆記に比べて圧倒的に多く、英語で授業を受ける聴覚障害学生にとっては非常に有効な情報保障手段であることが確認されたとしている。

4.3 ポータブル機器への呈示

大学における授業活動は、学生の専攻する分野によっては、教室で座って講義を受けるだけでなく、実験、実習、フィールドワーク等、字幕呈示機器を持ち運びできる方が使いやすい授業形態も多い。

高校までいかなる情報保障も受けた経験がない聴覚障害学生の中には、教室内で自分にだけ支援者や支援機器がおかれていることに抵抗を持つ場合もある。また、例えば教育学部における小中学校での教育実習などでは、子どもたちが聴覚障害者とのコミュニケーションにすんなりとして入っていけるよう、支援機器はあまり目立たなくすることが必要であることも多い。こうした場合、ポータブル機器への字幕呈示ができれば、音声認識ソフトウェアを用いた字幕呈示システムは、より利用しやすいと思われる。

IP Talk を用いたPC要約筆記では、プレイステーション・ポータブル (SONY 製) へ字幕を呈示する方法が試みられている。このシステムでは、プレイステーション・ポータブルの使用者自身が、自分で字幕をスクロールさせて読み返すことも可能である。また、速記タイプを利用した字幕呈示システムにおいては、携帯電話への呈示も可能になっている (小林正幸・西川 俊・三好茂樹・石原保志, 2007; 小林正幸・西川 俊・三好茂樹・石原保志・宮岡真也・岡田圭司・福岡寛之, 2008)。こうした事例と同様、音声認識ソフトウェアを利用した字幕呈示システムでも、ポータブル機器へ字幕呈示を行うため

のソフトウェア作成が待たれる。

謝 辞

東京大学障害教員支援者として、本研究を支えて下さった手話通訳士の三宅三枝子氏に深く感謝する。

なお、本研究は、日本学術振興会科学研究費補助金（科研基盤研究(S)17100009）の補助を受けた。また、字幕システムの運用に関して、一部同補助金（基盤研究(C)(2)18530748）（基盤研究(C)20530879）の補助を受けた。

引用文献

- 半澤幹一（1990）. 文章と談話のあいだ 寺村秀夫・佐久間まゆみ・杉戸清樹・半澤幹一編 ケーススタディ 日本語の文章・談話 おうふう.
- Hoogenboom, R.B., K. Uehara, T. Kanazawa, S. Nakano, H. Kuroki, S. Ino, and T. Ifukube (2008). An application of real-time captioning system using automatic speech recognition technology to college EFL education for deaf and hard-of-hearing students. *Gunma University Annual Research Reports, Cultural Science Series*, **57**, 95-113.
- 井野秀一・黒木速人・加藤士雄・渡邊括行・堀耕太郎・伊福部達（2003）. 聴覚障害者の会議参加支援を目的としたリアルタイム音声字幕化システムの設計, 計測自動制御学会第18回生体生理工学シンポジウム論文集, 221-224.
- 井野秀一・黒木速人・加藤士雄・渡邊括行・堀耕太郎・伊福部達（2003）. 聴覚障害者の会議参加支援を目的としたリアルタイム音声字幕化システムの設計, 計測自動制御学会第18回生体生理工学シンポジウム論文集, 221-224.
- 菊池真里（2006）. 音声認識を活用した聴覚障害学生の情報保障のあり方に関する研究－誤認識の推測に注目して－群馬大学大学院教育学研究科修士論文.
- 近藤公久・神長伸幸・馬塚れい子・林安紀子（2007）. 読みの過程における音韻長および韻律の影響, 信学技報, **TL2007-8 (2007-04)**, 41-46.
- 小林正幸・西川俊・三好茂樹・石原保志（2007）. 聴覚障害者のための携帯電話を用いたリアルタイム字幕提示システム FIT2007（第6回情報科学技術フォーラム）講演論文集, 601-604.
- 小林正幸・西川俊・三好茂樹・石原保志・宮岡真也・岡田圭司・福岡寛之（2008）. 聴覚障害者のための携帯電話を用いたリアルタイム文字提示システムによる講義支援 FIT2008（第7回情報科学技術フォーラム）講演論文集,（第3分冊）551-553.
- 黒木速人・井野秀一・中野聡子・加藤士雄・渡邊括行・堀耕太郎・伊福部達（2003）. 聴覚障害者の国際会議参加支援のための遠隔型音声字幕化システム－札幌－横浜間におけるシステム運用とその評価－, ヒューマンインタフェース・シンポジウム2003論文集, 729-732.
- 三好茂樹（2008）. 音声認識技術を用いた情報保障(20) PEPNet-Japan Tip Sheet.
- 中野聡子・牧原功・金澤貴之・菊池真里・黒木速人・井野秀一・伊福部達・福島智（2005）. 音声認識技術を活用した高等教育機関における聴覚障害者の情報保障(2)－音声言語と文字言語の性質の違いを中心とした検討－, 日本特殊教育学会第43回大会論文集, p556.
- 中野聡子・牧原功・金澤貴之・菊池真里・黒木速人・井野秀一・伊福部達・福島智（2006）. 音声認識技術を利用した

字幕提示システムの現状と課題－音声言語と文字言語の性質の違いに焦点をあてて, 群馬大学教育実践研究, **23**, 251-259.

- 中野聡子・牧原功・金澤貴之・中野泰志・新井哲也・黒木速人・井野秀一・伊福部達（2007）. 音声認識技術を用いた聴覚障害者向け字幕提示システムの課題－話し言葉の性質が字幕の読みに与える影響－, 電子情報通信学会論文誌D, **J90-D (3)**, 808-814.
- 中野聡子・牧原功・金澤貴之・黒木速人・上田一貴・井野秀一・伊福部達（2008）. 聴覚障害者向け音声同時字幕システムの読みやすさに関する研究(1)－改行効果に焦点をあてて－ ヒューマンインタフェース学会誌（印刷中）.
- 小畑修一・西川俊・高橋秀和（1985）. 聴覚障害者のための字幕挿入に関する研究－台詞に忠実な字幕挿入の可能性と効果－, 特殊教育学研究, **23**, 1-11.
- 定延利之（2005）. 「雑音」の意義, 月刊言語, **34 (1)**, 30-37.



なかの さとし
中野 聡子

2001年筑大大学院博士課程心身障害学研究科修了。博士（心身障害学）。2002年10月から2年間東大先端研バリアフリープロジェクトにて研究に従事(特任助手)。以来、聴覚障害者にとって読みやすいリアルタイム字幕提示の研究に取り組んでいる。現在東大先端研人間情報工学分野特任教員(助教), 群馬大学, 共愛学園前橋国際大学, 東京福祉大学非常勤講師。日本特殊教育学会等会員。日本手話学会理事。



かなざわ たかゆき
金澤 貴之

1998年筑大大学院博士課程心身障害学研究科中退。群馬大教育学部助教授。聴覚障害者のコミュニケーションや情報保障に関する社会学的な研究に従事。日本特殊教育学会等会員。日本発達障害学会編集委員, 障害学会編集委員。



まきはら つとむ
牧原 功

1997年筑大大学院博士課程文芸言語研究科中退。国立台湾大学文学部講師, 東海大学文学部講師を経て2000年より群馬大学留学生センター助教授。日本語の統語論, 語用論, 及び日本語教育の研究に従事。日本語学会, 日本語用論学会等会員。



くろき はやと
黒木 速人

1996年北海道大学大学院工学研究科生体工学専攻修士課程修了。同年セコム株式会社IS研究所研究員。2002年東京大学先端科学技術研究センター科学技術振興特任教員(特任助手)。2006年筑波技術大学障害者高等教育研究支援センター特任助手。2007年同大特任助教。人間情報工学に基づくアシスティブテクノロジーの研究, 特にリアルタイム字幕システムの研究に従事。日本生体医工学会等会員。



うえだ かずなが
上田 一貴

2004年広島大学大学院生物圏科学研究科博士課程後期修了。2002年日本学術振興会特別研究員。2005年広島大学大学院教育学研究科附属心理臨床教育研究センター助手。2007年東大先端研特任助教。博士(学術)。専門は精神生理学, 認知心理学。認知, 感情に関する脳機能イメージング研究に従事。日本心理学会, 日本臨床神経生理学会, 日本認知心理学会, Society for Neuroscienceなど各会員。



いの しゅういち
井野 秀一

1988年北大・工・電子卒。1993年同大大学院工学研究科博士後期課程修了。1991年日本学術振興会特別研究員。1993年北大電子科学研究所助手。1995年同大講師。2003年東大先端研助教授。2008年産総研主任研究員。工博。生体情報システム論に基づくアシステイブテクノロジーやバーチャルリアリティに関する研究に従事。日本生体医工学会、日本機械学会、ヒューマンインタフェース学会、日本VR学会、日本ロボット学会、バイオメカニズム学会、IEEE等各会員。



いふくべ とおる
伊福部 達

東大先端研教授。工学博士。専門は生体工学、福祉工学、音響工学。電子情報通信学会フェロー。約30年間にわたり、ヒトの感覚生理学・音響心理物理学に関する基礎研究に基づき、各種バリアフリー機器の開発を行う。著書：音声タイプライタの設計（CQ出版，1983）、音の福祉工学（コロナ社，1997）、人工現実感の評価（培風館，2001，編著）、福祉工学の挑戦（中公新書，2004）他。

Issues on the Use of Real-Time Captioning System Using Speech Recognition Technology: Suitability to the Needs of the Hearing Impaired

Satoko Nakano¹⁾ · Takayuki Kanazawa²⁾ · Tsutomu Makihara²⁾ ·
Hayato Kuroki³⁾ · Kazutaka Ueda¹⁾ · Shuichi Ino⁴⁾ · Tooru Ifukube¹⁾

This paper discusses a method of presenting captions that suits the needs of the hearing impaired. Considering possible reasons for difficulty with processing captions presented through Real-Time Captioning System Using Speech Recognition Technology, based on our empirical data we conclude as follows: (a) Captions' accuracy level must be as close as possible to 100%; and (b) captions are more comprehensible when limited to small numbers of characters presented between line breaks. Issues for future research using Real-Time Captioning System Using Speech Recognition Technology are: (a) Methods of re-speaking suitable to the situation and to the Japanese-language ability of the hearing impaired; (b) application of the system to English-speaking classes and/or conferences; and (c) methods of presenting captions with portable equipment.

Keywords

the hearing impaired, real-time speech-to-caption system, accuracy of captions, line-break pattern, re-speaking

¹⁾ Research Center for Advanced Science and Technology, The University of Tokyo

²⁾ Gunma University

³⁾ Research and Support Center on Higher Education, National Universities Corporation Tsukuba University of Technology

⁴⁾ National Institute of Advanced Industrial Science and Technology