

機関番号：12103
 研究種目：若手研究(B)
 研究期間：2007～2010
 課題番号：19700627
 研究課題名(和文) 聴覚障害者のためのリアルタイム字幕呈示システムにおける最適情報呈示に関する研究
 研究課題名(英文) An optimum method of displaying speakers' face images and captions for a real-time speech-to-caption system for the deaf
 研究代表者
 黒木 速人(KUROKI HAYATO)
 筑波技術大学・産業技術学部・准教授
 研究者番号：00345159

研究成果の概要(和文)：復唱音声認識方式を用いたリアルタイム字幕システムにおいても精度100%の字幕の提供は困難である。そこで、字幕呈示過程において、誤認識を含む不完全字幕と話者の発話情報を最適な形で呈示することで内容理解を向上させる試みを行った。結果、不完全字幕に対し話者の顔を付加することで内容理解が向上する結果を得た。さらに、不完全字幕を話者の顔の呈示に対して先行呈示させることにおいても内容理解が向上する結果を得た。

研究成果の概要(英文)：Our throughgoing real-time speech-to-caption system that using speech recognition technology with a "repeat-speaking" method showed that the accuracy of the captions is about 97% in Japanese-Japanese conversion. We found the idea to display information of captions and speaker's face movement images with a suitable way to achieve a higher comprehension. The results showed that the mouth movements contribute to improve the comprehension of the captions, and the sequence "to display the caption before the speaker's face image" improves the comprehension of the captions also.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	900,000	0	900,000
2008年度	800,000	240,000	1,040,000
2009年度	800,000	240,000	1,040,000
2010年度	700,000	210,000	910,000
年度			
総計	3,200,000	690,000	3,890,000

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学・教育工学

キーワード：ヒューマン・インタフェース、障害者支援、音声認識、字幕、ノンバーバル情報、教育工学、認知科学、情報工学

1. 研究開始当初の背景

音声認識を用いたリアルタイム字幕呈示システムにおいて、現状の技術レベルでは精度100%の字幕の生成は困難であり、字幕は誤認識を含む不完全な文となる。しかしながら重要なのは、最終的な伝達内容に対する理解度を100%に近づけることであ

る。そのためには、字幕の不完全部分を補うための情報を内容理解の阻害なく、最適な状態で呈示することが重要となる。

2. 研究の目的

本課題では、現在までのシステム開発で行ってきた字幕生成過程における精度向上のみならず、字幕呈示過程においてもヒト

が情報を阻害なく補完・統合するための最適情報呈示方法に関する研究を行うことを目的とする。具体的には、音声認識処理後の誤認識を含む不完全文と、話者の発話時のノンバーバル情報の2つの情報を呈示させる際に、顔の部位や情報間の呈示タイミングなどと言った、バーバル情報とノンバーバル情報の2つの情報間における最適呈示に関する方法を探ることを目的とする。本課題を通して得られた結果は、研究代表者が行っているリアルタイム字幕システムのみに適応される訳ではない。リアルタイムに字幕を呈示させる用途において、リアルタイム性と字幕精度はトレードオフの関係にあるため、一般に呈示させる字幕には誤認識が付きまとう。このような不完全字幕を呈示する用途においても適用可能な結果になり得る。

3. 研究の方法

呈示させる不完全文の作成は、実際のシステム使用場面を考慮し、実験者が音声認識装置に対して定型文を読み上げることで作成した。読み上げの際の音声認識結果は発声音声の明瞭さと顔・口元情報の読み取りやすさの両方を反映するであろうと言う仮定に基づき、読み方を一定に保つ指標としても用いた。作成する不完全文の文完全率が、正解文に対しておよそ8割となるように尤度閾値を調整した。

不完全文と一緒に呈示させる発話時の話者映像は上記の読み上げ時に、話者の顔映像を撮影したものを用いた。不完全文と話者顔映像を実験の主旨に合わせて映像編集ソフトにて事前に編集し、実験試料映像集としてテープに記録して用いた。

呈示させる話者の顔映像に関しては、呈示する領域的な検討と時間的な検討に分類して検討する必要がある。そこで、検討すべき内容を以下の小項目4つに分類して検討を行った。

(1) 不完全な文に対し話者の顔などのノンバーバル情報の活用による効果とその最適表示の方法

① ノンバーバル情報として顔のどの領域を呈示するのが良いか

② ノンバーバル情報と字幕との呈示タイミングをどう制御するのが良いか

③ ノンバーバル情報と字幕を同期させて呈示する際、どの程度の精度が要求されるか (ヒトの時間ずれの検知限の把握)

④ ノンバーバル情報と字幕をある時間間隔で呈示する際、2つの方法をどこまで統合した情報として保持できるか (ヒトの情報に対する時間的統合限界)

4. 研究成果

本課題を通じての実験被験者は、聴覚障害者が5名、健聴者が6名の計12名であった。本課題では、先の実験方法において分類し

た項目のなかの、概ね3つの小項目に関して成果を得たと考える。

①では、不完全文からなる字幕のみの呈示条件(字幕のみ)を基準として、不完全文に対し話者の発話時の顔全体映像を付加した条件(字幕+顔)、同様に口元近接映像を付加した条件(字幕+口元)の3種で比較した。実験では、完全文(正解文)に対する回答文の正答率を文理解向上の指標として用いた。健聴者と聴覚障害者に対し実験を行った結果、正答率はいずれの被験者群においても、「字幕のみ」<「字幕+顔」<「字幕+口元」となる傾向を示した。「字幕のみ」はコントロール条件であるため、「字幕のみ」の条件に対して、条件差を設けた呈示方法である「字幕+顔」および「字幕+口元」において、統計的有意差のある結果として得られた。このことから、不完全文に対してノンバーバル情報を「付加」することが文理解に貢献すると言う結果が得られたと言える。

②では、ノンバーバル情報として話者の発話時の顔全体映像を用い、ノンバーバル情報に対して不完全文の呈示に時差(-5秒~+5秒)を設けた試料に対し、いずれの時差が文理解の向上に貢献するかを比較した。

リアルタイム字幕システムにおいて、本課題のように誤認識字幕とノンバーバル情報を統合させて呈示しようと試みた場合、ノンバーバル情報はリアルタイムで呈示が可能だが、字幕においては音声認識処理を施す必要があるためどうしても処理時間分リアルタイムから遅れてしまう。ノンバーバル情報よりも字幕が遅れて呈示される条件では、情報取得者にとってどのような影響を及ぼすのかを確認する必要がある。さらに、2つの情報を呈示する時間差において、実験で得られる知見をより拡張させる意味合いも含めて、ノンバーバル情報よりも字幕が先行呈示させる場合も条件に含めて実験を行った。実験では、話者の顔情報に対して誤認識字幕が先行呈示される場合を負(-)の呈示時差とした。同様に、誤認識字幕に対して字幕が先行表示される場合を正(+)の呈示時差とした。

①の実験と同様に、実験では、完全文(正解文)に対する回答文の正答率を文理解向上の指標として用いた。正答率をそれぞれの被験者群で求めた結果、正答率は時差なし(±0秒)ではあまり向上が認められなかった。時差なし(±0秒)は話者の顔情報と字幕が同時に呈示される条件であるため、日常において最も自然な条件と考えられる。しかしながら不完全文に対する回答文の正答率の向上は数%であった。この条件は、小項目①の「字幕+顔」と同じ実験条件であり、実際、小項目①の「字幕+顔」の条件における実験結果と同じ値を示した。さらに興味深い結果として、字幕先行側に

1秒ずらした場合(-1秒(字幕先行1秒))が、全ての呈示時差を通して最も高くなる傾向を示した。この理由として、字幕は時間的に更新される情報ではあるが、ある一定時間は「字幕」として画面に停留される。停留することにより情報取得者は読み返しが可能になる。読み返したあとで話者顔映像が呈示されることで情報理解を促進させたのではないかと推察される。これらの呈示時差が-1秒の場合が、今回の呈示条件では最も適していたと言える。

また、字幕先行側に外れるほど(呈示時差が負(-)の値になればなるほど)正答率が単調増加する傾向を示した。これらの傾向は、何れの被験者群においても統計的有意差として確認された。顔先行側(呈示時差が正(+))の値)においては、顕著な傾向は示さなかった。

いずれの時差においても大局的な傾向は被験者を通じて似た傾向を示したが、細かくは個人差依存が大きいことが伺えた。

③では、話者の顔情報と不完全字幕の呈示時差(-5秒~+5秒)を意図的に設け、その呈示時差を被験者が検知できるかどうかを計測することを試みた。実験では、時差があることが「わかる」「わからない」「気になる」「気にならない」を指標にして、被験者に数直線上の値に記入することで評価した。その結果、被験者により大きく異なる傾向を示した。ただ、概ね呈示時差0~+2秒の間に許容限があると推察される結果となった。しかしながら、この定性実験の結果、検知限には個人差が大きいことがわかった。今後は定量実験の実施方法と解析方法を再考する必要があると思われる。

デジタル補聴器が開発された1980年代において、生の顔情報に対して、補聴器を通して得られる音声情報がデジタル信号処理時間分遅れて取得されることを懸念した、視聴覚情報統合に関する実験がいくつか行われた。本課題で行った実験のように、字幕・発話情報と言った、視覚におけるバーバル・ノンバーバル情報の情報統合に関する検討を行った実験は過去見当たらない。その意味で本課題の結果は、リアルタイム字幕システムのパフォーマンス向上と言った実用面のみならず、視覚的なバーバル・ノンバーバル情報の情報統合と言う脳科学の分野においても非常に興味深い結果が得られたと言えよう。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計6件)

- ① H.Kuroki, S.Ino, S.Nakano, K.Hori, T.Ifukube, A Method for Determining the Timing of Displaying the Speaker's Face and Captions for a Real-Time Speech-to-Caption System, SICE JCMSI, Vol.3, No.6, pp.402-408, 2010, 査読有
- ② 黒木速人, 三好茂樹, 白澤麻弓, 河野純大, 萩原彩子, 磯田恭子, 田中美希, 金澤貴之, 大倉孝昭, アメリカにおける聴覚障害者のための文字による情報保障技術に関する報告, 筑波技術大学テクノレポート, Vol.15, pp.181-185, 2008, 査読無
- ③ 中野聡子, 金澤貴之, 牧原功, 黒木速人, 上田一貴, 井野秀一, 伊福部達, 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)-改行効果に焦点を当てて-, ヒューマンインタフェース学会論文誌, Vol.10, pp.435-444, 2008, 査読有
- ④ 中野聡子, 金澤貴之, 牧原功, 黒木速人, 上田一貴, 井野秀一, 伊福部達, 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-, メディア教育研究, Vol.5, pp.63-72, 2008, 査読有
- ⑤ 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, 音声認識技術を用いた聴覚障害者向け字幕呈示システムの課題-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, Vol.J90-D, No.3, pp.808-814, 2007, 査読有
- ⑥ 黒木速人, 井野秀一, 中野聡子, 堀耕太郎, 伊福部達, 音声同時字幕システムにおける内容理解の向上を目的とした話者の顔情報の呈示方法, ヒューマンインタフェース学会論文誌, Vol.9, pp.205-212, 2007, 査読有

[学会発表] (計1件)

- ① H. Kuroki, et al., Effective Timing of Displaying Speakers' Face and Captions to Achieve Higher Comprehension, 23rd Annual International Technology and Persons with Disabilities Conference (CSUN2008), 2008.3.14, Los Angeles, USA.

[その他]

ホームページ

筑波技術大学・機関リポジトリ

<http://www.a.tsukuba-tech.ac.jp/rep/dspace/>

6. 研究組織

(1) 研究代表者

黒木 速人 (KUROKI HAYATO)
筑波技術大学・産業技術学部・准教授
研究者番号：00345159

(2) 研究分担者

なし

(3) 連携研究者

伊福部 達 (IFUKUBE TOHRU)
東京大学・先端科学技術研究センター・教授

研究者番号：70002102

中野 聡子 (NAKANO SATOKO)
東京大学・先端科学技術研究センター・特任助教

研究者番号：20359665

堀 耕太郎 (HORI KOTARO)
株式会社ビー・ユー・ジー