

平成21年 4月30日現在

研究種目：若手研究(B)
 研究期間：2007～2008
 課題番号：19700470
 研究課題名（和文） 聴覚障害者の講義における遠隔地パソコン筆記でのカメラ制御に関する研究
 研究課題名（英文） A Study on Camera Control for Remote PC Captioning in the lecture for Hearing Impaired
 研究代表者
 竹内 義則 (TAKEUCHI Yoshinori)
 名古屋大学・情報連携統括本部情報戦略室・准教授
 研究者番号：60324464

研究成果の概要：

本研究では、講師音声からの指示語発話抽出手法として、音声認識を利用した手法を提案し、さらに、指示語抽出結果と指示動作抽出結果の統合手法も提案する。この統合手法は、指示語発話と指示動作の時間の関連性を利用している。そして、実際に収録した講義データに対して、指示語・指示動作抽出処理を行い、さらに、統合処理を行った結果、再現率約 91%、適合率約 71% を得ることができた。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,100,000	0	1,100,000
2008年度	1,000,000	300,000	1,300,000
総計	2,100,000	300,000	2,400,000

研究分野：福祉情報工学

科研費の分科・細目：人間医工学 リハビリテーション工学・福祉工学

キーワード：遠隔パソコン要約筆記，指示発話抽出，指示動作抽出，情報保障

1. 研究開始当初の背景

本研究では、聴覚障害者が大学での高等教育を受ける際の情報保障について取り扱う。一般に聴覚障害者集団を対象にした教育は、手話で行われているが、手話で講義が出来る教育者は限られている。専門知識を持った外部の非常勤講師に講義を依頼する場合、その講師は、手話ができないことが多い。その場合、最近になって、パソコン要約筆記によって情

報保障が行われてきた。これは、健聴者のボランティアをその講義室に配置し、その場で講師の声をパソコンでタイプし、前のスクリーンにタイプした文字を投影する手法である。専門用語は手話で表現しにくいですが、この手法では文字として容易に表現できるので、大学での講義の情報保障として適している。また、パソコン要約筆記は、文字を早くタイプでき

ればよいと、手話通訳に比べ習熟が早く、ボランティアの確保も容易であるという利点がある。

また、マイクに入力された講師の声と、カメラで撮影した、黒板、スライド、ホワイトボードなどの映像を、自宅などの遠隔地に送り、遠隔地側で文字を入力して送り返す「遠隔地パソコン筆記」が行われている。これにより、ボランティアが現場に出向いたり、その場でパソコンを準備したりするコストを削減でき、より支援しやすい環境が整いつつある。しかし、遠隔地のボランティアは、パソコンへの文字入力だけでなく、カメラの制御を行う必要があり、これが新たな負担となっている。講師の声を入力しただけでは、難解な専門用語を誤って入力するかもしれないし、指示語に対応することができなくなる。したがって、黒板等を見ながら入力することが必要である。

このような背景の元、これまでに聴覚障害者集団の講義を対象とした研究を行ってきた。この研究では、聴覚障害者の学生が発言するときの問題点を解決するために、画像処理技術を応用し、質問した学生を発見、撮影するシステムを開発してきた。本研究では、画像処理の技術を、黒板、スライド、ホワイトボードなどの撮影に応用するものである。これまでの研究により、映像情報は、聴覚障害者の講義での情報保障として有用であることも分かっている。

2. 研究の目的

本研究では、遠隔地パソコン筆記においてカメラ制御を自動化し、講義の場面に応じた撮影をすることを目的とする。これにより遠隔地の入力者は、文字入力に集中でき、必要なときに黒板等を参照できるようになる。

本研究は、筑波技術大学産業技術学部産業

情報学科の皆川洋喜助教授の協力の下で、実際に大学で行われている聴覚障害者集団での講義を対象とする。筑波技術大学では、一部の講義を外部の非常勤講師に依頼しているが、手話が出来ないためにパソコン筆記による情報保証が行われている。この講義では、黒板、パワーポイントなど講師が作成したファイルを投影するスクリーン、講師と学生が筆談を行うためのホワイトボード、入力された文字を表示するための大型ディスプレイが使われている。

このような状況のもとで、講義室の後ろから講師をカメラで撮影し、その映像を解析し認識することによって、自動的にカメラの撮影方向やズームを変えることを主要な目的とする。黒板に文字を書いているときはその文字を、スライドで説明しているときはスライドを映すようにカメラを制御する。特に、講義では黒板に書いた文字や図を指して説明することがよくある。例えば「これとこれを合わせると、…」のように、指示語をそのまま入力した文字だけでは全く理解できないので、指示語の内容を補って入力する必要がある。このとき、講師が指し示している黒板の箇所が遠隔地から読めなければならない。したがって、講師の指差し動作を認識して、説明している部分を撮影することが重要である。

本研究では、音声認識を利用して講師音声から指示語を抽出する手法を開発する。また、映像情報から、講師領域を抽出し、その中から手や指示棒の領域を抽出することによって、指示動作を抽出する手法を開発する。さらに、講師音声からの指示語抽出結果と、講師映像からの指示動作抽出結果を統合するために、指示語と指示動作の時間的な関連性を利用して統合する手法を開発する。

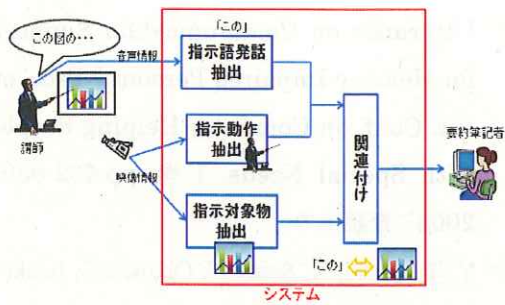


図1. システムの概要

3. 研究の方法

講師発話中の指示語とその指示対象物の対応付けを行うために、講師音声と講師映像の両方を利用する手法を検討した(図1)。

まず、音声処理では講師音声から指示語を抽出する。並行して、映像処理では講師映像から指示語発話時の指示動作と指示対象物を抽出する。そして、指示語と指示対象物を対応付ける。この結果を、要約筆記者に提示し、要約筆記者にその内容を反映しようと考えている。

講師音声からの指示語抽出については、音声認識を利用して、「これ」「こう」「この」「ここ」「こちら」といった指示語の抽出を行う。一方、講師映像からの指示動作抽出では、背景差分を用いて講師領域を求め、さらに、その領域から手・指示棒領域を求める。そして、講師領域に対する手・指示棒領域の割合から、講師が指示棒を使用しているかどうか判断し、指示棒を使用していれば指示動作を行っているとした。

そして、指示語と指示動作の時間的関係を用いて、抽出された指示語と指示動作の発生時刻を基に、指示動作を伴う指示語かどうかを判断する。また、指示動作を伴う指示語であれば、抽出された指示語と指示対象物の対応付けを行う。

4. 研究成果

講師音声からの指示語抽出結果と、講師映

表2 統合実験結果

	音声	映像	統合
指示語+指示動作	218 個	229 個	213 個
指示語+非指示動作	32 個	23 個	23 個
非指示語	90 個	963 個	66 個
再現率	92.8%	97.4%	90.6%
適合率	64.1%	20.0%	70.5%

像からの指示動作抽出結果の統合処理の評価実験を行った。1 講義分(60 分)の講義データに対して、音声認識を用いた指示語抽出結果と、指示動作抽出結果を統合し、再現率と適合率を計算した。この講義で、指示動作と共に行われた指示語発話(指示語+指示動作)の数は、235 個である。ここで、講師がスライドではなく板書に対して指示を行っている部分は、実験対象から外している。

結果を表 2 に示す。表中の「指示語+指示動作」は、この実験で正解となる、指示動作を伴う指示語の認識数を示す。また、「指示語+非指示動作」は、指示動作を行わずに指示語発話した時を、「非指示語」は、指示語以外の単語を発話した時を、間違っ指示動作を伴う指示語だと認識した数を示す。

再現率は、統合によって、音声処理のみから約 2 ポイント、映像処理のみからは約 7 ポイント下がっていることがわかる。また、適合率は、統合によって、音声処理のみから約 6 ポイント、映像処理のみからは約 51 ポイント上がっていることもわかる。

ここで、適合率を上げることができた理由は、音声・映像処理それぞれの結果において、一方の処理が、間違えて「指示動作を伴う指示語である」と認識しているところを、もう一方の処理が、正確に「指示動作を伴う指示語ではない」と認識できているため、間違っ

た認識結果を減らすことができているからである。

逆に、再現率を下げた理由は、一方の処理が、正確に「指示動作を伴う指示語である」と認識しているところを、もう一方の処理が、間違っ「指示動作を伴う指示語ではない」と認識してしまうところがあり、正解の認識数を減らしてしまっているからである。音声処理のみで正解である「指示語+指示動作」の認識数は、統合処理によって5個減っている。この5個のデータを見ると、映像処理が認識できていなかった場合と、指示語を発話してから指示動作の開始まで、時間がかかっているために「指示語+指示動作」であると認識できなかった場合の2種類あった。一方、映像処理のみで正解である「指示語+指示動作」の認識数は、統合処理によって16個減っている。この16個のデータには、音声処理で、早口でしゃべっているために指示語を認識できず、映像処理と統合することができなかった場合が多く見られた。

今後、さらによい精度を得るためには、指示語・指示動作抽出処理の精度向上や、統合手法の検討が必要である。

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 3件)

- ① Y. Takeuchi, K. Saito, A. Ito, N. Ohnishi, S. Iizuka, S. Nakajima, "Extracting Pointing Object with Demonstrative Speech Phrase for Remote Transcription in Lecture," Proc. of Int. Conf. on Computer Helping People with Special Needs, 1巻, pp.624-631, 2008. 査読あり
- ② A. Ito, K. Saito, Y. Takeuchi, N. Ohnishi, S. Iizuka, S. Nakajima, "A Study on Demonstrative Words Extraction in Instructor

Utterance on Communication Support for Hearing Impaired Persons," Proc. of Int. Conf. on Computer Helping People with Special Needs, 1巻, pp.632-639, 2008. 査読あり

- ③ Y. Takeuchi, K. Saito, N. Ohnishi, S. Iizuka, S. Nakajima, "Camera Control for Remote Transcription System," Proc. of the 10th IASTED Int. Conf. Computer Graphics and Imaging, 1巻, pp.88-93, 2008. 査読あり

[学会発表] (計 5件)

- ① 伊藤綾鹿, "遠隔パソコン要約筆記のための指示語と指示動作との統合手法の検討", 電子情報通信学会福祉情報工学研究会, 2009年2月20日, 松山
- ② 伊藤綾鹿, "聴覚障害者向け情報保障のための講師発話内指示語および指示動作抽出手法の検討", ヒューマンインタフェース学会, 2008年12月3日, 東京
- ③ 伊藤綾鹿, "遠隔パソコン要約筆記のための指示動作の検出", 画像電子学会年次大会, 2008年6月20日, 東京
- ④ 伊藤綾鹿, "聴覚障がい者向け情報保障のための講師発話からの指示語抽出手法の検討", 情報処理学会全国大会, 2008年3月15日, つくば
- ⑤ 斉藤健, "遠隔PC要約筆記のための指示発話と動作の関連性分析", 電子情報通信学会福祉情報工学研究会, 2008年1月25日, 松江

6. 研究組織

(1) 研究代表者

竹内 義則 (TAKEUCHI Yoshinori)
名古屋大学・情報連携統括本部情報戦略室・准教授
研究者番号: 60324464