

平成20年度 メディア科学専攻修士論文要旨

| | | |
|--------|--------------------------------------|---------|
| 大西 研究室 | 氏 名 | 伊 藤 綾 鹿 |
| 修士論文題目 | 遠隔パソコン要約筆記のための 指示語発話の抽出及び指示動作との統合 | |

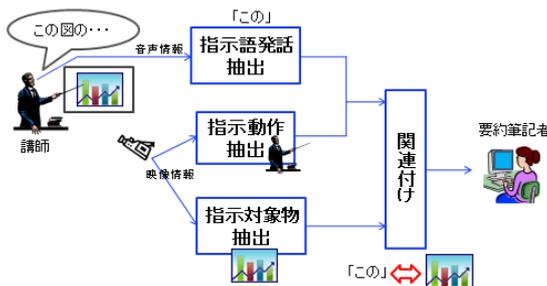
大学での講義において、聴覚障害者のための情報保障の1つに遠隔パソコン要約筆記がある。これは、講義の音声と映像を遠隔地にいる要約筆記者に送り、その情報を基に作成した要約筆記文を講義室内の受講生(聴覚障害者)に提示するものである。しかし、例えば、複数の数式の中の一つを指して「この数式にありますように」というように、講師が指示動作を行いながら指示語を発話した時に、要約筆記者がその指示動作を見ていないと、指示語に対応する指示対象物が分からなくなってしまうという問題が起こりうる。

この問題を解決するために、講師音声から指示語を、講師映像からは指示動作及びその指示対象物を抽出し、指示語と指示対象物を関連付けて要約筆記者に提示するシステムの研究開発を行っている。そのシステムの概要を図1に示す。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、予想される指示対象物の候補を抽出する。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し、指示語を発話しながら指示動作が行われている場合を抽出する。そして、指示語抽出結果と指示動作抽出結果の統合結果から、指示語と指示動作が同時に行われている場合の、指示語と指示対象物を対応付ける。その結果を、要約筆記者に提示し、要約筆記文にその内容を反映しようと考えている。

本研究では、まず、講師音声からの指示語発話抽出を行う。この指示語発話抽出手法として、指示語発話の前にある程度の長さのポーズがあることに着目し、韻律情報を特徴量として判別分析によって指示語を抽出する手法と、音声認識を利用する手法の2種類の手法を提案している。講義データを用いて、2種類の手法でそれぞれ講師音声からの指示語発話抽出を行ったところ、韻律情報を利用する手法より、音声認識を利用する手法の方が、高い精度を得られることがわかった。

そして、音声認識を利用した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。指示語発話と指示動作の時間の関連性を利用した統合手法を提案し、指示語抽出結果と指示動作抽出結果に対して統合処理を行ったところ、再現率91%、適合率71%を得ることができた。その結果を表1に示す。

表1 統合結果



| | 音声処理 | 映像処理 | 統合処理 |
|-----------|-------|-------|-------|
| 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% |

図1 システムの概要