平成22年度　メディア科学専攻修士論文要旨

大西 研究室	氏　　名	川 口　弘 哲
修士論文題目	リアルタイム字幕作成支援のための数式抽出	

背景

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つにリアルタイム字幕提示システムがある．話者発話内容の全てを文字に変換し，即座に字幕として提示するためのシステムで，高度なキーボード入力技能を有する字幕作成担当者が行う．このシステムの課題の１つとして数式が挙げられる．発話される数式には複数の解釈がある場合がある．例えば,「x マイナス２分の１」と講師が発話したとする．このとき，音声情報のみでその数式を解釈しようとした場合，$x-\frac{1}{2}$ と $\frac{1}{x-2}$ というような２通りの解釈が可能である．そのため，聴覚障害者がどちらの解釈か判断に困る場合がある．また,「1 と 8」や「l と m と n」といった紛らわしい発話数式が，字幕作成者にとって字幕入力が困難な場合もある．

目的

その問題を解決するため，リアルタイム字幕提示システムを使用する際に，数式入力の困難さを解決する手法を提案する．本研究では発話された数式に対応するスライド中の数式画像を自動で抽出する手法を用いている．この手法は数式発話時に講師が発話数式に対応するスライドの数式を指示する，という特徴を利用している．

抽出手法

音声認識エンジン Julius，日本語話し言葉コーパスモデルから得られた音響モデル・言語モデルを用い，講義資料に含まれる数式要素を辞書登録することで，数式発話抽出を行う．また，映像から取得した指示棒先端の軌跡から，講師の行っている指示動作を抽出する．そして，数式発話抽出結果と指示動作抽出結果を統合する．統合方法は，指示動作開始時刻の２秒前から，指示動作終了時刻までに数式発話開始時刻がある，という条件で統合を行う．統合を行った後，指示動作に応じて指示対象とそのときの軌跡を抽出することで，発話数式に対応する数式画像を抽出する．抽出した数式画像の例を図１に示す．

実験・結果

音声認識における数式要素の抽出率を調査した．その結果，約 71%の再現率，約 90%の適合率を得た．また，実際に収録したデータに対して，音声認識によって抽出した数式要素を用いて数式画像抽出処理を行った．その結果，約 70%の再現率，約 91%の適合率を得た．この結果を分析したところ，音声認識の数式要素の誤認識によって，数式画像の抽出率を下げている，と考えられる．

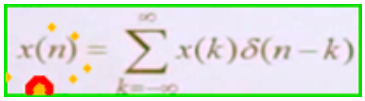

図１　抽出した数式画像

表１　数式画像抽出における実験結果

講義データ	データ１	データ２
発話された数式の個数	205 個	68 個
音声認識が抽出した数式の個数	173 個	50 個
数式画像を抽出した数式の個数	170 個	48 個
対応する数式を抽出した個数	150 個	45 個
非対応・未抽出の個数	55 個	23 個
再現率	73.1%	66.2%
適合率	88.2%	93.8%