

**科学研究費助成事業 研究成果報告書**

平成 27 年 5 月 21 日現在

機関番号：33907

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500642

研究課題名(和文) パソコン要約筆記における要約筆記文の改善に関する研究

研究課題名(英文) Research on Improvement of Sumarized Transcript for PC Transcription

研究代表者

竹内 義則 (Takeuchi, Yoshinori)

大同大学・情報学部・准教授

研究者番号：60324464

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：聴覚に障害のある学生が高等教育機関に通い、健聴者と一緒に講義を受ける機会が増えている。その際の支援の一つに要約筆記がある。現在の要約筆記の問題点として、指示語と数式の問題がある。そのため、まず本研究では、講義映像中から指示対象物をリアルタイムで抽出するシステムを作成する。そして、指示対象物を要約筆記者に提示し、要約筆記文中に直接埋め込むことができるシステムを提案する。要約筆記文作成システムを用い、要約筆記実験及び、その評価を行った。その結果、システムの画像挿入の機能を使用して行った要約筆記の方が、使用しない場合に比べ、要約筆記文の作成環境が改善されている傾向が見られた。

研究成果の概要(英文)：The occasions are increasing for hearing impaired students to attend the lecture with hearing students. Summarized transcript is a one of a support for hearing impaired. One problem in current transcription is that transcribers are not able to assign demonstrative words with the pointed object. Another problem is that inputting complex mathematical formulas is hard for transcribers. In this study, we make a system which extracts pointed or directed objects from lecture videos in real time. Then, we propose a system that can show directed objects to transcribers and embed them directly in the transcription. We conducted an experiment to extract the directed objects corresponding to the instructor's demonstrative word in real time. Then, we conducted experiments to transcribe the lecture with the proposed system. As a result, there was a tendency that the environment for creating transcription is improved by using the function of inserting images.

研究分野：福祉情報工学

キーワード：福祉情報工学 要約筆記 指示対象抽出 指示動作認識

### 1. 研究開始当初の背景

近年、聴覚に障害のある学生が、大学などの高等教育機関に通い、健聴者と一緒に講義を受ける機会が増えている。その際、学生が講義の内容を理解し、健聴者と同じように講義を受けられるような情報保障が求められる。手話で講義が出来る教育者は限られており、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。この場合、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。

また、講師の声をマイクロホンから入力し、黒板、スライドなどの映像をカメラで撮影し、それら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「遠隔パソコン要約筆記」が行われている。遠隔パソコン要約筆記では、講師の声だけでなく、スライドや黒板などの映像も重要である。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。例えば、「この式について」といったように、講師が指示語を伴った指示動作をした時を考える。音声を聞いて文字入力に集中している要約記者が講師の指示動作を見逃してしまい、指示語と指示対象との関連付けが困難になる。大学などの高等教育機関で使用される数式は複雑なものが多く、その入力には時間がかかり、要約記者にとって大きな負担となる。

このような背景の下、これまでに遠隔パソコン要約筆記で問題となる指示語を検出し、指示対象物を抽出する研究を行ってきた。要約筆記では、要約文中の指示語の指示対象物が要約文だけではわからないため、入力者が指示語を指示対象物に置き換えて入力している。遠隔パソコン要約筆記では、カメラの画角、解像度の問題で、講師が指示語を発生したときに指示対象物を見落とすことがある。そこで、指示対象物を抽出し、要約記者へ提示することにより、補うことができる。

### 2. 研究の目的

本研究では、パソコン要約筆記において、数式や表といった視覚コンテンツを要約文の中に挿入するシステムを開発することを目的とする。まず、講義を撮影したカメラから得られる講師の動作を解析し、要約筆記に必要な情報を抽出する。それを入力者に提示し、入力者が簡単な操作によって要約文に挿入することを目的とする。これにより入力者は、入力の手間が軽減され、さらに作成した要約文もわかりやすくなることが期待される。

### 3. 研究の方法

講師の発話した指示語に対応する指示

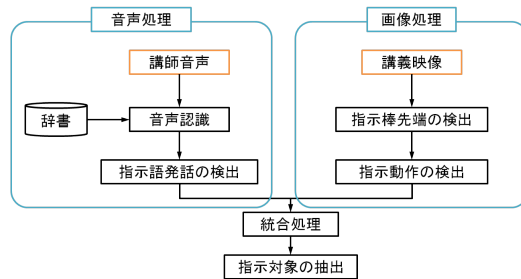


図1 指示対象抽出システムの概要

対象物を抽出するシステムを開発する。その概要を図1に示す。入力には講師の音声と、講師とスライドを撮影した映像である。講師の音声から指示語発話を検出し、講義映像から指示動作を検出する。その後、指示語発話と指示動作との統合を行い、指示語を伴った指示動作が行われた時に、指示対象となった図字領域を抽出する。

講師が発話した音声からの指示語発話の検出は、音声認識によって行う。本システムでは、音声認識システム Julius(バージョン 4.1.5)を使用する。また、音響モデルと言語モデルには、日本語話し言葉コーパス(CSJ: Corpus of Spontaneous Japanese)の性別非依存音韻トライフォンモデルを使用する。

リアルタイムで講師の発話した音声を認識するために、入力として、講師の持つマイクからの入力音声を使用する。今回、検出の対象とした指示語発話は、「これ」、「この」、「ここ」、「こう」、「こちら」の5つとした。これらの指示語が含まれていると検出された場合、講師が指示語発話を行ったと判断する。

大学で一般的に行われているスライド使用した講義を対象とする。この形式の講義では講義室の前面にスクリーンがあり、講師はPCを持ち込み接続して使用する。スクリーンに投影された各スライドは、テキストで表現された個々の項目や、文章が記述された文領域、個々の図を単位とする図領域から構成されている。これらを総称し、図字領域と呼ぶ。そして、指示対象とは、講師の指示動作によって指示されたスライド中の図字領域を指す。

講師が指示棒を使用して指示動作をする際、指示棒の使い方には、個人差はあるが、大きく分けて3つの方法がある。それは、指示棒をスライド中の図字領域で止める、なぞる、囲むという方法である。その結果、点、直線、楕円の3種類の特徴的な軌跡が現れる。指示点がこのような軌跡を描くように講師が指示棒を使った時を、指示動作とする。

指示動作の検出のために、まず、指示棒先端を検出する。検出した指示棒先端の座標を保持することで、指示棒先端の軌跡が得られる。その指示棒先端の軌跡が点・線・楕円を描いたときに、指示動作がなされたと判断する。まず、現在フレームを含めた過去  $N (= 5)$  フレームの指示棒先端の座標から  $N$  点の軌

跡を得る．その N 点から，軌跡の一時的な状態を判別する．これを一時状態と呼ぶ．一時状態の条件は，以下のとおりである．

POINT N フレームのうち指示棒先端が半数以上抽出され，抽出された点群が点条件を満たす．

LINE N フレームのうち指示棒先端が半数以上抽出され，抽出された点群が線条件を満たす．

OTHER N フレームのうち指示棒先端が半数以上抽出され，抽出された点群が点条件及び線条件を満たさない．

NOTHING N フレームのうち指示棒先端が半数未満しか抽出されない．

ここで，点条件は，「点群が，1 辺の長さが 1 である正方形内にある」，線条件は，「点群が，縦の長さが 1 である帯内にある」と定義した．実験では， $1 = 15[\text{pixels}]$  としている．スライドの大きさは，約  $700 \times 500[\text{pixels}]$  である．一時状態は，各フレームにおいて保持される．

次に，各フレームにおいて最終状態を判別する．最終状態の条件は，以下のとおりである．

POINT 保持しているフレームにおいて，全ての一時状態が POINT であり，抽出された点群全てについて点条件を満たす．

LINE 保持しているフレームにおいて，全ての一時状態が LINE であり，抽出された点群全てについて線条件を満たす．

OTHER 最終状態 POINT，LINE の条件は満たさず，保持しているフレームにおいて，全ての一時状態が LINE，または OTHER である．

NOTHING 保持している全てのフレームにおいて，一時状態が NOTHING である．ある最終状態の条件を満たして，その最終状態の条件を満たさなくなるまでを 1 つの指示動作と判断する．その最終状態の条件を満たさなくなった場合は，保持しているフレームを破棄し，新たにフレームを保持し直す．その結果，各フレームにおいて，必ずいずれかの最終状態が決定される．

指示点の軌跡が楕円になるような指示動作を抽出するために，最終状態 OTHER である間，以下の包囲条件の少なくとも一方を満たすときに，囲む指示動作があったと判断し，その動作を最終状態が CIRCLE である指示動作として抽出する．指示点の軌跡が必ずしも厳密な楕円とならないことがあるため，包囲したかどうかを判定することで楕円のように囲む指示動作の分類を行った．

[包囲条件 1] 保持している点群のうち，

現在から 1 つ過去の点を除いて，現在フレームの指示点からの距離が 1 以内である点が存在する．

[包囲条件 2] 保持している点群のうち，任意の連続する 2 点を結ぶ線分が他の連続する 2 点を結ぶ線分と重なりを持つ．

指示棒の軌跡により，指示された図字領域を抽出する．講義で使用するスライドは，あらかじめ入手可能であるため，そのスライドを分析することにより，スライドのどの部分に図字領域が含まれるかがわかる．これは，準備として事前に手動で行う必要がある．図字領域は，それを囲う長方形の左上，右下の 2 点をデータとして抽出する．入手したスライドファイルから自動的に図字領域を抽出する手法の開発は，今後の課題である．また，スライドの画像と講義を撮影したカメラ画像を照合することで，どのスライドがスクリーンに投影されているかを知ることができ，スライドの座標系から講義画像の座標系への変換も得ることができる．現在のシステムでは，この照合は字幕入力者が手動で行う必要がある．この問題は，講義中のスライドの領域が，事前に入手したスライドのどの画像に近いかを判断することである．これは，類似画像検索の一種であるので，その技術を利用してこの照合を自動化することは，今後の課題である．

図字領域の抽出は，最終状態により次のように行う．

POINT 軌跡の点群の重心から最も距離の近い図字領域を抽出する．

LINE 軌跡の点群を内包する矩形領域を 1 だけ上方に並行移動した領域と重なる図字領域全てを抽出する．

CIRCLE 図字領域の中心が，軌跡の点群からなる多角形内にある図字領域全てを抽出する．図字領域が見つからない場合は，軌跡の点群の重心から最も距離の近い図字領域を抽出する．

OTHER 軌跡の点群を結んでできる折れ線が通過する図字領域全てを抽出する．

ここで，重心と図字領域の距離は，重心から図字領域の長方形の各辺までの最短距離とする．重心が図字領域の内部にある場合は，距離は 0 とする．軌跡が通過する図字領域は，その軌跡が図字領域の長方形に含まれるかどうかを判断する．抽出された数式は，字幕作成者に送られ，字幕作成時の補助情報として用いられる．

講義映像中から抽出した指示対象物を要約筆者に提示し，要約筆記文作成の際に，その指示対象物を要約筆記文中に直接埋め込むエディタを提案する．このシステムを使用した際の画面は図 2 のようになる．

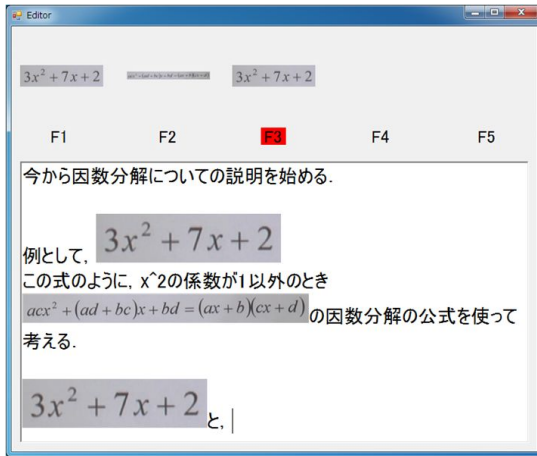


図 3 要約筆記作成エディタ

このエディタは大きく分けて2つの部分から構成される。1つが上部にある「画像提示部」で、もう1つが下部の「テキスト入力部」である。まず、画像提示部に講義映像から抽出した画像を表示する。画像提示部には最大5枚の画像を、F1からF5までのラベルの上に順番に表示できる。F5まで画像が表示された後は、再びF1、F2と順番に画像が表示される。また、文字ラベルの色を変えることで、その画像が最新のものであることがわかるようにする。

テキスト入力部には、要約筆記者が講義の様子を見ながら要約筆記文を入力する。その際、画像提示部に表示されている画像は、それぞれの対応するキーを入力することで、テキスト入力部に画像として直接挿入することができる。

つまり、この要約筆記作成システムを使用することにより、多くの労力を必要とした数式等の入力を簡単なキーボード操作により、画像として要約筆記文中に入力することが可能になる。また、文字ラベルの色を変えることで、要約筆記者が講師の指示動作を見逃してしまった場合でも、指示語と指示対象とを関連付けることが可能となる。

#### 4. 研究成果

前面にスクリーンのある講義室で、講義の音声・映像を記録した。HDVカメラ(SONY HDR-SR8)を講義室後方に固定し、映像を撮影した。音声は、マイクロフォン(SONY ECM-CZ10)を講師胸元に設置し、録音した。また、この音声は量子化16bits、48kHzのサンプリングでデジタル化した。スライド全体と講師を含む範囲を撮影し、映像のサイズは1440×1080、フレームレートは10fpsとした。

収録した講義は「信号処理」であり、数式の説明が多く、指示棒を用いて講義を行なっている。また、スライドを使用して説明している場面を対象としており、ホワイトボードを使用した説明や小テストなどのスライドを使用していない場面は対象外としている。

講師が発話した数式に対応する数式画像

表 1 記述された項目の割合

被験者	グループ	画像挿入機能を使用しない	画像挿入機能を使用する
1	A	25% (15/60)	58% (35/60)
2	A	45% (27/60)	63% (38/60)
3	A	27% (16/60)	75% (45/60)
4	A	27% (16/60)	70% (42/60)
5	A	27% (16/60)	48% (29/60)
6	B	50% (30/60)	65% (39/60)
7	B	37% (22/60)	55% (33/60)
8	B	20% (12/60)	25% (15/60)
9	B	30% (18/60)	37% (22/60)
10	B	40% (24/60)	43% (26/60)
11	B	30% (18/60)	40% (24/60)
12	C	67% (40/60)	70% (42/60)
13	C	47% (28/60)	60% (36/60)
14	C	30% (18/60)	30% (18/60)
15	C	40% (24/60)	25% (15/60)
16	C	22% (13/60)	27% (16/60)
17	D	65% (39/60)	88% (53/60)
18	D	10% (6/60)	38% (23/60)
19	D	27% (16/60)	72% (43/60)
20	D	22% (13/60)	33% (20/60)
21	D	38% (23/60)	57% (34/60)
22	D	25% (15/60)	47% (28/60)

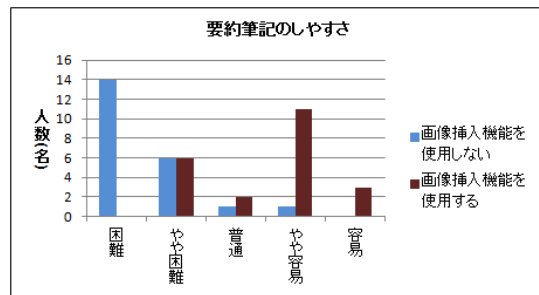


図 2 要約筆記のしやすさのアンケート結果

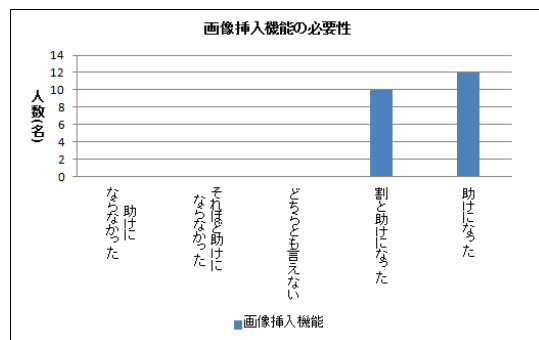


図 4 画像挿入機能の必要性に関するアンケート結果

を抽出した結果、71.1%の再現率、86.2%の適合率を得た。

要約筆記作成エディタの評価のため要約筆記実験を行った。エディタの画像提示部に表示される画像は、事前に手動で抽出したものを使用した。

実験には2つのディスプレイを使用した。一方には要約筆記作成エディタを表示し、もう一方には模擬講義映像を表示した。実験の

被験者は情報系学生 22 名で、全員が講義映像内で使用される用語を理解している。各被験者は 2 回、別の評価用映像を見せ、要約筆記を行った。また、それぞれの評価用映像を用いて要約筆記を行う前に、練習用映像を用いて要約筆記を行った。実験終了後には、被験者に対して、行った要約筆記についてのアンケートを採った。そして、作成された要約筆記文について、要約筆記文中に記述されるべき項目のうち、適切に記述された項目の数を調べた。この調査は 4 名の学生が行った。

各被験者が作成した要約筆記文で、要約筆記文に記述されるべき項目に対する、適切に記述された項目の割合を表 1 に示す。割合の高かった方を色付きで表した。

そして、アンケートに関して、「エディタの画像挿入の機能を使用せず / 使用して行った要約筆記のしやすさ」についての回答結果を図 3 に示す。また、「画像挿入の機能は、要約筆記をする上で助けになったか」についての回答結果を図 4 に示す。

表 1 より、被験者 22 名中 20 名について、要約筆記の際に、システムの画像挿入機能を使用した場合の方が、要約筆記文中に記述された項目の割合が高かった。この要因として、画像挿入機能により、数式等の入力時間が削減され、より多くの文章が入力できたこと、講師の指示に対応した指示対象物を適切に要約筆記文に入力できたことが挙げられる。このことから、本システムにより、要約筆記文の作成環境が改善されると考えられる。また、画像挿入機能を使用した場合でも、記述されていた項目の割合が低かった被験者が 2 名いた。この被験者はどちらもグループ C に属しており、評価用映像 A では画像挿入機能を使用し、評価用映像 B では画像挿入機能を使用せず要約筆記を行った。評価用映像はどちらも同程度の難易度となるように作成していたが、要約筆記のしやすさという観点からみると、評価用映像 B の方が要約筆記を行いやすかった可能性が考えられる。

そして図 3 のアンケート結果より、画像挿入機能を使用しない場合は、ほぼ全ての被験者が要約筆記のしやすさを「困難」、もしくは「やや困難」と回答した。しかし、画像挿入機能を使用することで、14 名の被験者が「やや容易」、もしくは「容易」と回答した。この結果からも、画像挿入機能により、要約筆記文を作成する環境が改善されたと考えられる。また図 4 より、全ての被験者が、画像挿入の機能は要約筆記の際に「助けになった」、または「助けになった」と回答した。よって、要約筆記文の作成の際に画像挿入を用いることは、要約筆記文作成環境の改善のために有効であり、非常に役に立つと言えると考えられる。

##### 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)

Yoshinori Takeuchi, Yuji Sato, Kazuki Horiike, Daisuke Wakatsuki, Hiroki Minagawa, Noboru Ohnishi, Captioning System with Function of Inserting Mathematical Formula Images, Proceedings of the 14th international conference on Computers Helping People with Special Needs, 査読有, 1, 2014, pp.33-40

川口弘哲, 竹内義則, 松本哲也, 工藤博章, 大西昇, リアルタイム字幕作成におけるスライド中の発話された数式抽出, 電子情報通信学会論文誌 D, 査読有, J97-D, 2014, 1035-1043

Yoshinori Takeuchi, Hironori Kawaguchi, Noboru Ohnishi, Daisuke Wakatsuki, Hiroki Minagawa, A System for Matching Mathematical Formulas Spoken during a Lecture with Those Displayed on the Screen for Use in Remote Transcription, Proceedings of the 13th international conference on Computers Helping People with Special Needs, 査読有, 1, 2012, pp.142-149

##### [学会発表](計 3 件)

堀池一希, 竹内義則, 工藤博章, 松本哲也, 大西昇, 指示対象挿入機能を用いた要約筆記文作成システムの開発, 電子情報通信学会福祉情報工学研究会, 2015 年 3 月 13 日~2015 年 3 月 14 日, 「筑波技術大学春日キャンパス(茨城県・つくば市)」

堀池一希, 竹内義則, 大西昇, 工藤博章, 松本哲也, 数式挿入機能を備えた要約筆記文作成エディタの開発, 電子情報通信学会福祉情報工学研究会, 2014 年 3 月 7 日~2014 年 3 月 8 日, 「筑波技術大学春日キャンパス(茨城県・つくば市)」

堀池一希, 竹内義則, 大西昇, 工藤博章, 松本哲也, 要約筆記文改善のための講義中の指示対象物挿入法に対する評価, 電気関係学会東海支部連合大会, 2013 年 9 月 24 日~2013 年 9 月 24 日, 「静岡大学浜松キャンパス(静岡県・浜松市)」

##### [図書](計 件)

##### [産業財産権]

出願状況(計 件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:

国内外の別：

取得状況（計 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

#### 6．研究組織

##### (1)研究代表者

竹内 義則 (TAKEUCHI, Yoshinori)  
大同大学・情報学部・准教授  
研究者番号：60324464

##### (2)研究分担者

( )

研究者番号：

##### (3)連携研究者

皆川 弘樹 (MINAGAWA, Hiroki)  
筑波技術大学・産業技術学部・教授  
研究者番号： 00273285

若月 大輔 (WAKATSUKI, Daisuke)  
筑波技術大学・産業技術学部・准教授  
研究者番号： 50361887