

## 聴覚障害者のための遠隔講義支援に関する実践的研究

|       |       |                       |
|-------|-------|-----------------------|
| 代表研究者 | 竹内 義則 | 名古屋大学情報連携統括本部情報戦略室准教授 |
| 共同研究者 | 皆川 洋喜 | 筑波技術大学産業技術学部教授        |
| 共同研究者 | 若月 大輔 | 筑波技術大学産業技術学部助教        |

### 1 背景と目的

大学などの高等教育機関に進学する聴覚障害者の増加に伴い、大学の講義や学会の公演発表など専門性の高い場面での情報保障のニーズが高まっている。聴覚障害者が健聴者の中に混ざって講義などを受講する場合の支援方法の1つとして、手話や読話(読唇)、筆談、リアルタイム字幕、要約筆記といった情報保障が行われている。最近では、ネットワーク技術の発展に伴い、遠隔地から支援を行う情報保障の1つとして、遠隔パソコン要約筆記が普及してきている。通常、パソコン要約筆記は支援者が講義室で行う。しかし、地域によって支援体制に大きな差があり、支援体制が整っていない地域へ支援者を派遣するには大きなコストがかかる。そこで、インターネットを利用して、講師の音声と映像を設備と人員の整った拠点へ送り、遠隔地でパソコン要約筆記を行い、要約文を講義室内の聴覚障害者に提示することで、コストを削減する試みが行われている。

遠隔で講義を視聴するため、映像の解像度や視野角によって制約が生じている。例えば、複数の数式の中の1つを指して「この式にありますように」というように、講師が指示語を伴って指示動作を行ったとする。通常、講師が発話してから、要約筆記文が聴覚障害者に提示されるまでに数秒の遅れが生じる。この遅れにより、「この」という指示語が要約文に現れるとその要約文を見た聴覚障害者は指示語の意味する内容が分からないということが起こる。そのため、要約筆記者は「この」という指示語をそのまま入力せずに、「○番目の数式」と指示語を置き換えて入力するようにしている。しかし、現状の遠隔パソコン要約筆記では、限られたネットワークの通信速度内で音声と映像を送っているために、要約筆記者が見る映像は解像度やフレームレートを落とさなければならない。そのため、音声を聞き、文字入力に集中している要約筆記者が、講師の指示動作を見逃してしまい、指示語を置き換えられないという問題が起こってしまう。

この問題を解決するために、講師音声から指示語を、講義映像から指示動作と指示対象を抽出し、指示語と指示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。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、映像処理では講義映像から指示語発話時の指示対象を抽出する。さらに、講師音声からの指示語抽出結果と講義映像からの指示動作抽出結果を統合し、指示語発話を伴って指示動作が行われている場合を抽出する。そして、指示語発話を伴う指示動作が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、講師が指示するために指示棒を使用し、スライドをスクリーンに映して行う講義を対象とした。そして、そのような講義において、主に講義映像から指示対象を抽出することを目的とした。さらに、指示語に対応する指示対象を要約筆記者に提示する支援方法が有効であるか評価することも目的とする。

聴覚障害者のための情報保障に関する研究は、数多く行われている。加藤ら[1]は、要約筆記者を支援するために、重要なキーワード提示についての検討を行った。三好ら[2]は、聴覚障害者のための遠隔地から支援が可能なリアルタイム字幕提示システムの開発を行っている。このシステムを、特殊なキーボードを用いて文字入力を行う速記タイピストが利用することで、遠隔地からリアルタイムに講師の発話内容を字幕化して聴覚障害学生がいる教室へ送り、講義の支援が実現できる。他にも、音声認識を用いたリアルタイム字幕の研究も行われている。伊藤ら[3]のように市販の音声認識ソフトを利用するケースや、M.Waldら[4]のように、研究中の音声認識エンジンを利用しているケースもある。

指示対象を抽出する研究では、先山ら[5]が、指示棒を利用する講義の様子を観察することで、指示棒の動きと指示動作との関連性を調べ、その特徴を利用したポインタ位置の補正と重み付き投票法に基づく指示領域の検出手法を提案している。また、丸谷ら[6]は、指示棒先端と軌跡・指示領域の種類・講師の体の位置・講師の体の向きと指示動作との関係を整理し、それに基づきセンサ情報から自動で指示動作の種類と指示領域を抽出する手法を提案している。

本研究では、要約筆記者を支援するために、講師の発話した指示語に対応する指示対象を抽出する。指示

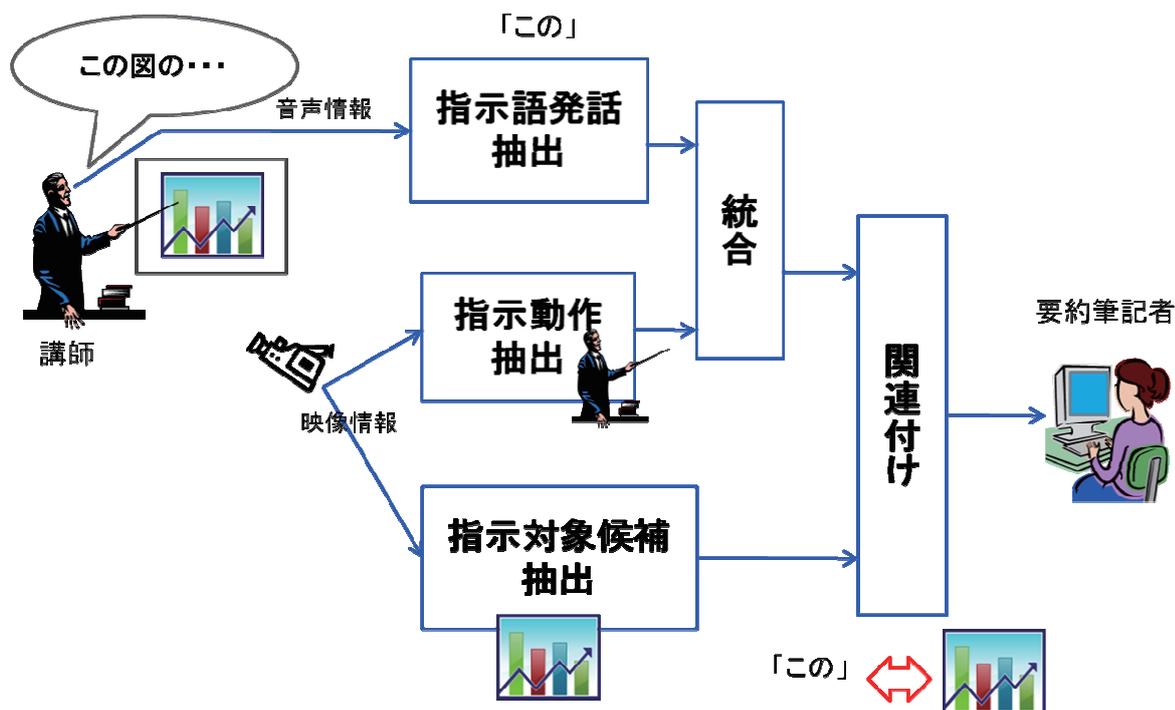


図1 システムの概要

対象を抽出する研究は行われてきたが、それを発話された指示語と対応付ける試みはこれまで行われていない。また、丸谷ら[6]は、講義の両肩及び指示棒の手元部分2箇所、計4箇所にビーコンをつけ、それらの位置情報を講義室天井に設置した超音波位置センサにより取得している。しかし、遠隔パソコン要約筆記は要約筆記者の派遣コストを削減するために行われており、将来的には一般の大学の講義室での講義に対して行われることも期待されている。よって、本システムは可能な限り低コストで実現する必要があり、本研究ではセンサなどは用いず、画像処理によるアプローチを試みる。

## 2 指示語に対応する指示対象抽出システム

### 2.1 システムの概要

指示語に対応する指示対象抽出システムの概要を図1に示す。本システムは、講師はパソコンでスライドをスクリーンに映し、指示棒を用いて指示をする講義において、指示語に対応する指示対象を抽出するシステムである。システムはまず講師音声から指示語を、講義資料と講義映像から指示動作を抽出する。その後、指示語と指示動作の統合を行い、指示語を伴う指示動作が行われている場合の指示対象を抽出する。そして、講義映像と本システムで抽出した指示対象を要約筆記者に提示することで、要約筆記の支援を行う。本研究では、主に講義資料と講義映像から指示動作及び指示対象の抽出を行う。

### 2.2 指示語の抽出

指示語の抽出処理は伊藤ら[8]の方法を使用した。以下で、その概要を示す。指示語は音声認識により抽出を行う。本システムでは、音声認識エンジンJuliusであり、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オンモデルを使用した。日本語話し言葉コーパスは、日本語の自発音声を大量に集めて、形態論情報や文節音・イントネーションラベルなど多くの研究用情報を付加した話し言葉研究用のデータベースである。

そして、指示語の認識精度を上げるために、以下の2点の改良を行った。

- 言語モデルの改良

言語モデルは、日常的によく見かける単語列には高い確率を割り当てるようになっている。逆に、文法的に誤っている、または滅多に見かけないような単語列には低い確率を割り当てる。そこで、まず「これ」「この」「ここ」「こう」「こちら」といった指示語の確率を上げて認識しやすくした。さらに、指示語発話時には、500ms以上のポーズがあり、その次に「まあ」「え」などのフィラーが発話されるこ

とがある。その特徴を利用し、「ポーズ+指示語」や「ポーズ+フィラー+指示語」といった形の単語列

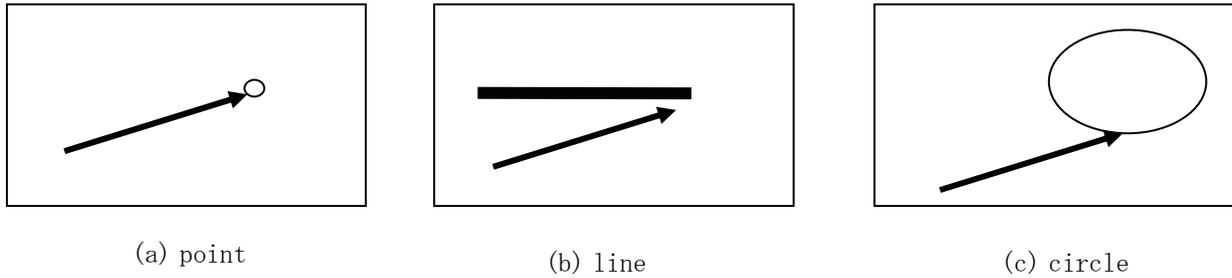


図2 指示棒先端の軌跡

の確率も上げた。

● 認識パラメータの選定

Juliusの実行パラメータである以下のパラメータ値を変更し、より多くの指示語が認識されるようにした。変更したパラメータは、ビーム幅と言語重み、単語挿入ペナルティである。

▶ ビーム幅

ビーム幅は、小さく設定すると、計算する仮説の範囲が狭くなり高速化される。逆に、大きく設定すると、広い範囲を探索するため処理時間が長くなる。指示語を抽出する際は、単語の探索範囲が広がるようにビーム幅を大きい値に設定した。

▶ 言語重み

言語モデルと音響モデルの影響度を表すもので、指示語の抽出のために言語モデルの評価を重くするようにした。

▶ 単語挿入ペナルティ

単語挿入ペナルティを変更すると、長い単語あるいは短い単語を優先して認識するように設定することができる。これを小さい値にすることによって、指示語のような短い単語が認識されやすくなるようにした。

## 2.3 指示動作の抽出

### (1) 講義で行われる指示動作

講義の各時刻において講義室前面にパソコンで提示されるスライドは、テキストで表現された個々の項目あるいは文を単位とする文領域や、個々の図を単位とする図領域の組み合わせにより構成されている。これらを総称して図字領域と呼ぶ。これに対して指示対象とは、図字領域の内容と講師による指示動作に基づいて伝達されるスライド中の領域であり、スライド中の特定の場所に講師が指示棒を向ける、指示棒を動かすなどの行為をして表現される。

Lilianらの研究では、講義における指示動作を人により客観的に分類・解釈している[7]。それを計算機で利用することを念頭にまとめると、指示動作時の指示棒の使い方には、講師により個人差が存在するが、スライド中の図字領域上で止める・なぞる・囲むという方法がある。これにより、図2のような点・直線・楕円の3種類の特徴的な軌跡が現れる。本研究では、このように指示点の軌跡が点・直線・楕円となるように指示棒を使う行為を指示動作とする。

### (2) 指示動作の分類

各フレームで得られた指示棒先端の座標を保持することで、指示点の軌跡を得る。この軌跡が点・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 $N$ 点を用いて、 $N$ 点の軌跡の一時的な状態を判別する。一時状態を以下のように分類した。

POINT  $N$ フレームのうち指示棒先端が半数以上抽出され、すべての点が小正方形領域内にある。

LINE  $N$ フレームのうち指示棒先端が半数以上抽出され、すべての点が細長い方形領域内にある。

OTHER  $N$ フレームのうち指示棒先端が半数以上抽出され、POINT、LINEの条件を満たさない。

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。

一時状態は指示棒先端の座標とともに毎フレーム保持される。次に、保持している軌跡全ての点についての状態を判別する。これを最終状態と呼ぶ。最終状態は、以下のように分類される。

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

POINTの条件を満たす。

LINE 保持している全てのフレームにおいて、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LINEの条件を満たす。

CIRCLE 保持している全てのフレームにおいて、全ての一時状態がLINEまたはOTHERで、抽出された点群のうち、前後のフレームを除き距離が近いものが存在する、または、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

OTHER 保持している全てのフレームにおいて、全ての一時状態がLINEまたはOTHERで、CIRCLEの条件を満たさない。

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。その際、(ある最終状態に初めてなった時刻)-N+1 を指示動作開始時刻とし、(その最終状態を満たさなくなった時刻)-N+1 を指示動作終了時刻とする。

## 2.4 指示語と指示動作の統合

上記の方法で抽出した指示動作を指示語発話と統合する。伊藤ら[8]は指示語発話と指示動作の時間関係を分析し、指示語発話は、指示動作開始時刻の1.5s前から指示動作終了時刻までの間に開始される確率が99.1%であることを報告している。

本研究では、この手法に基づいて、抽出した指示動作の開始時刻-1.5s から終了時刻の間に、指示語発話開始時刻があった場合に統合を行った。ただし、指示語が開始されてから1.5s経つ間に指示動作が終了した場合、それまでの指示動作とこれからの指示動作のどちらを指示語と統合するのか2つの選択肢がある。この場合は、指示動作の最終状態の優先度が高い方を採用する。最終状態の優先度は、その状態になる条件の厳しさと考え、以下のようにした。

*CIRCLE > LINE > POINT > OTHER > NOTHING*

また、指示語発話が始まってから1.5s経つ間に再び指示動作が終了した場合も、上記の優先度を考え、指示語と対応する指示動作を選ぶ。以上により、1つの指示語に対して複数の指示動作を抽出することを防いだ。

## 2.5 指示対象の抽出

指示対象とは1つの図字領域と一致するとは限らず、図字領域の一部や複数の図字領域であることもある。しかし、点で指示を行った場合、指示語発話と指示点の軌跡を用いて、指示した図字領域を指示しているのか、その図字領域の一部を指示しているのかを判別することは困難である。そこで、指示点の軌跡により指示された図字領域上に、指示点の軌跡を描いたものを要約筆者に提示することを考えた。これにより、要約筆者は抽出された領域を見て、その領域内のどこを指示していたかが分かると考えられる。よって、本研究において抽出する指示対象を指示点の軌跡と図字領域を内包する矩形領域と定義する。

そこで、まずは軌跡により、指示された図字領域を抽出する。図字領域は、画像内の座標として事前に得られることを前提とした。遠隔パソコン要約筆記では、講義前にスライドなどの講義資料が要約筆者に渡される。また、スライドの画像と講義映像を照合することで、スライドの画像における2次元座標系から講義映像内のスライドにおける2次元座標系への射影が可能であると考えられる。よって、講義前に図字領域の位置が画像内の座標として得られることは前提とする。

図字領域と指示点を内包する矩形領域を抽出し、その領域に指示点を重ね合わせたものを要約筆者に提示する。その例を図3に示す。図3の下半分が抽出された指示対象である。左側の指示対象は「こちらがx(n)で離散時間信号です」とCIRCLEで指示されたもので、右側の指示対象は、その約3秒後に「こちらがそのx(n)のz変換です」とPOINTで指示されたものである。図3は右側の指示対象が抽出された時刻の状況である。

## 2.6 要約筆者への指示対象の提示

上述したように、指示語に対応する指示対象として指示点の軌跡と図字領域を内包する矩形領域として抽出し、それを講義映像とともに要約筆者に提示する支援法を提案する。提示する指示対象を表示する時刻は、指示対象が決定したときを開始とし、一定時間表示し続ける。本研究では、表示時間を5秒とした。

提案する支援法を用いれば、要約筆者は講師の指示動作を見逃してしまい、指示語に対応する指示対象が分からない場合でも、指示対象を確認することができる。このことは、要約文を入力する際の指示語の言い換えがしやすくなり、指示語に対応する指示対象が何であったかという不安を解消することにも繋がる。

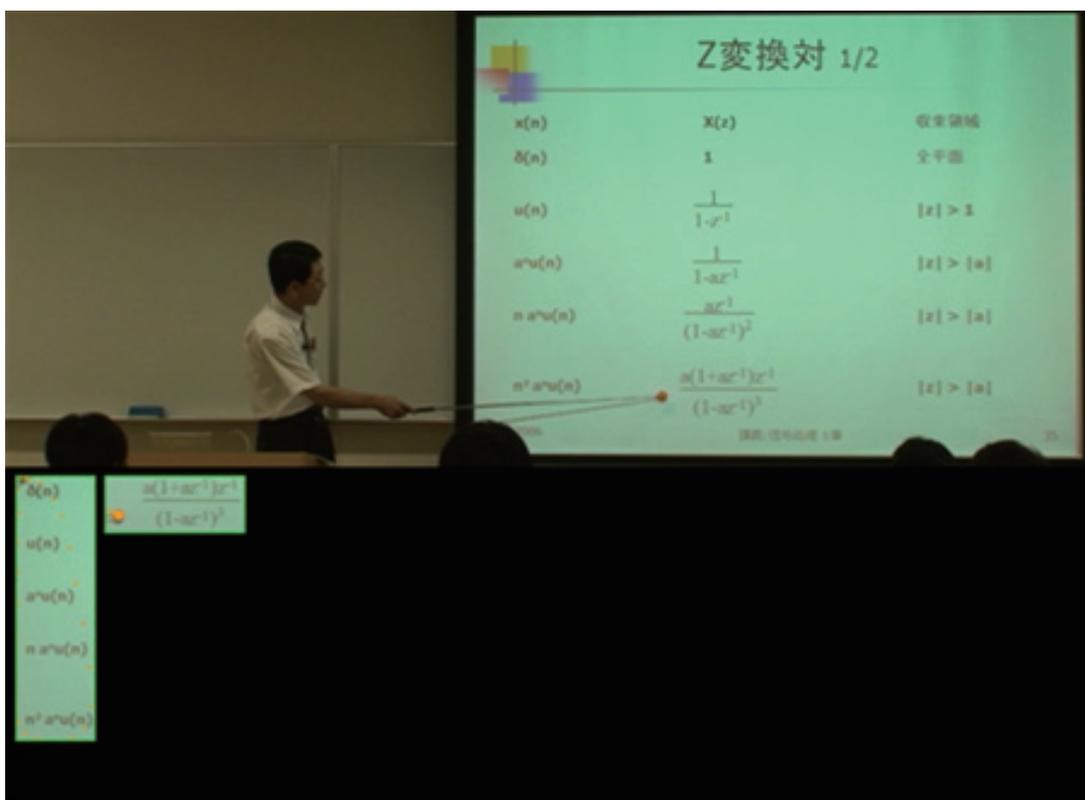


図3 システムの出力例：上段は入力，下段は出力

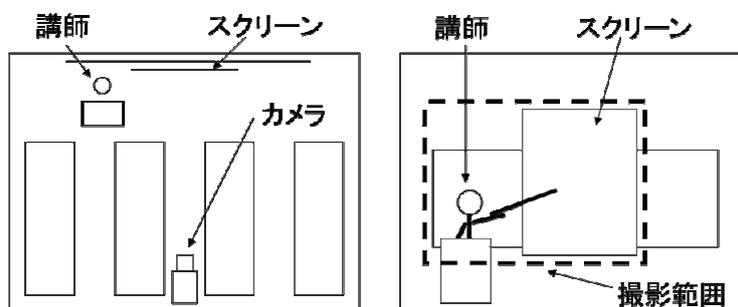


図4 撮影環境

### 3 指示対象の抽出実験

#### 3.1 講義データの収録

講義のデータベースを構築するために、図4のような前面にスクリーンのある講義室で、HDVカメラ(SONY HDR-SR8)を講義室後方に固定し、スライドを利用した講義の撮影を行った。音声は、マイクロフォン(SONY ECM-CZ10)を講師胸元に設置し、録音を行った。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。また、指示棒の先端を画像処理により抽出しやすくするために、指示棒の先端にオレンジ色のピンポン球をつけて録画を行った。収録した講義は信号処理という講義である。この講義では、数式の説明が多く、指示語を使って説明する回数が多い。また、本研究では、スライドに対しての指示のみを対象としており、板書に対しての指示は対象外としている。

#### 3.2 方法

講義映像に対して、今までに述べた手法により、指示対象を抽出した。実験に用いた講義映像は約65分であり、そのうち板書に対して指示を行っている部分は実験対象から除外した。なお、指示語の抽出は手動で

表1 指示語に対応する指示対象の抽出結果

|                    |                |
|--------------------|----------------|
| 講師が指示語を伴う指示動作を行った数 | 189個           |
| 本システムが抽出した指示対象の数   | 191個           |
| 正解数                | 162個           |
| 再現率                | 85.7%(162/189) |
| 適合率                | 84.8%(162/191) |

表2 抽出したが、指示対象が間違っていた原因と個数

|                       |    |
|-----------------------|----|
| 複数個の指示対象をPOINTで順に指示した | 9個 |
| 指示語と指示動作の統合を誤った       | 9個 |
| 真値が統合条件を満たしていなかった     | 6個 |
| 指示対象を指示しなかった          | 3個 |

行い、指示語発話開始時刻を0.1秒単位で記録した。また、スライド切り替えのタイミングや図字領域の画像内における座標も講義映像から手動で取得したものを用了。

### 3.3 結果

本システムの抽出結果を評価するために、指示語発話を伴う指示動作及びその指示動作により指示された対象を手動で取得し、これを真値とした。そして、真値と本システムで抽出した指示対象が合致していれば正解とした。本システムによる指示語に対応する指示対象の抽出結果を表1に示す。

### 3.4 考察

指示語に対応する指示対象の抽出結果は、再現率が85.7%、適合率が84.8%となった。指示点の軌跡が点・直線・楕円となる場合を指示動作とすることで、指示語に対応する指示対象を高い精度で抽出できた。

抽出したが、指示対象が間違っていた27個を表2にまとめる。

「この対応」や「これら」という発話に対して、複数個の指示対象をPOINTで順に指示するという場合が9個あった。本手法では、1つ目をPOINTとして指示した時点で1つの指示動作が終了したと判断し、抽出をしてしまった。このようなケースに対応するには、「対応」といったキーワードに対して2つの指示対象を抽出するといった処理が必要となる。また、「これら」といった複数を意味する指示語は、単数を意味する指示語とは区別して考えることが必要である。本手法では、音声処理の結果として、指示語発話の開始時刻のみを使用しているため、今後、音声処理の改良が必要だと考えられる。

指示語と指示動作の統合を誤った9個のケースでは、指示語発話開始時刻に行っていた指示動作を統合したが、本当はその後の指示動作が正解であった場合である。指示語発話開始時刻に指示動作を行っており、指示語発話開始時刻から1.5秒以内に別の指示動作に移る場合に、このケースは起こっていた。指示語発話開始までに行っていた指示動作と指示語発話開始後に開始された指示動作のどちらを選択するか判断するときに、指示動作の種類だけではなく、指示動作時間などの指示動作の重要度をを用いることが考えられる。

指示語と指示動作の統合に失敗した6個のうち、1個は指示動作を終えてから指示語を発話する場合であった。残りの5個は指示語が発話されてから1.5秒経過後に指示動作が開始される場合であった。指示語開始時刻は指示動作開始時刻の1.5秒前以降、または指示動作終了時刻までにあるという伊藤ら[8]の分析結果を用いているが、もちろんこれに全てが該当するわけではない。この統合条件を緩めて、より多くの指示動作を候補に上げ、それから統合すべき指示動作を選択するという方法が考えられる。

指示棒先端が指示対象まで届いておらず、別の図字領域を抽出してしまったケースが3個あった。遠隔パソコン要約筆記を行う講義では、講師が指示対象をしっかりと指示するように注意することが対策として考えられる。

本来、抽出すべきでないのに抽出した2個の場合は、指示語発話時に指示動作をしていないのにスライド上に指示棒先端があり、そのときの指示棒先端の軌跡から指示動作を抽出してしまったときであった。指示動作はせずに、「このように」とそれまでに説明したことを意味するケースであった。これに関しても、講師が指示する意思がないときは指示棒をスライド上から外すように気をつけることが対策として考えられる。

指示対象の抽出は、講師が気をつけるケースを除いた全てのケースをできることが目標である。よって、「複数個の指示対象をPOINTで順に指示した」、「指示語と指示動作の統合を誤った」、「真値が統合条件を満たしていなかった」、のケースへの対策を講じる必要がある。

## 4 要約筆記者の支援法の評価

### 4.1 評価実験

#### (1) 方法

指示語に対応する指示対象を要約筆記者に提示する支援法の評価を行うために、講義映像のみの場合と、講義映像と指示語に対する指示対象が表示される場合で、被験者が要約筆記を行った。そして、要約文に含まれるべき項目のうち要約文に適切に記述された項目数及び指示語が適切に言い換えられた項目数を調査した。

被験者は、情報系の学生 12 名とした。被験者は全員が講義に出てくる用語を理解しており、人並み以上のタイピング能力を持つ。要約文に含まれるべき項目とは、講師の発話した内容の要点を箇条書きにまとめたものである。要約文に記述された内容が要約文に記述されるべき項目と合っていて、かつ指示語がある場合はその正しい内容に置き換えられている場合にその項目が要約文に記述されたと判断した。実験を行うにあたって、表 3 のような 2 つの動画を作成した。動画 A と B は、講義内容は異なるが、要約筆記の難しさが同程度となるようにシーンを選んだ。被験者は図 3 の画面を見ながら実験を行った。指示対象を表示する時間は 5 秒とした。指示対象を出力するタイミングはシステムが出力したタイミングに基づいている。

最初に練習として長さが約 1 分である講義映像のみの練習用動画で練習を行った。その後、表 4 のように 4 つのグループに分け、要約筆記を行った。

#### (2) 結果

要約文に記述されるべき項目に対する記述された項目の割合を表 5 に示す。また、要約文に記述された項目のうち、指示語をその内容に言い換えることが必要な項目を表 6 に示す。要約文に記述された項目数は被験者ごとにばらつきが見られたものの、指示対象を表示したほうが記述された項目数の割合も指示語を適切に言い換えた項目数も大きいケースが多く見られた。

#### (3) 考察

指示対象を提示するかしないかで要約文に記述された項目数の割合や指示語が適切に言い換えられた項目数の割合に差があるかどうか検定を行った。検定にはウィルコクソンの符号付順位和検定を用い、有意水準を 1% とした。その結果、指示対象を提示した場合としない場合で有意な差があると確認された。また、指示語の言い換えができた項目数の割合も同様に指示対象を提示した場合としない場合で有意差が確認された。よって、指示対象を提示したほうがしない場合より要約文に項目が多く記述されると言える。

### 4.2 要約筆記者へのアンケート

実際に講義の要約筆記経験者に対して、本手法で抽出した指示対象が付加情報として表示される図 3 のような講義映像を約 9 分間見てもらい、アンケート調査を行った。

#### (1) アンケート内容

Q1. パソコン要約筆記の経験年数を教えてください。

表3 実験に用いた動画

|                | 動画A   | 動画B  |
|----------------|-------|------|
| 時間             | 3m50s | 4m20 |
| 平均発話速度(モーラ/分)  | 245   | 249  |
| 要約文に記述されるべき項目数 | 15個   | 17個  |
| 指示語を言い換えるべき項目数 | 5個    | 5個   |

表4 被験者のグループ分け

| 被験者 | 実施内容 |      |
|-----|------|------|
|     | 1回目  | 2回目  |
| 1   | 動画A  | 動画B  |
| 2   | 指示対象 | 指示対象 |
| 3   | 提示あり | 提示なし |
| 4   | 動画B  | 動画A  |
| 5   | 指示対象 | 指示対象 |
| 6   | 提示なし | 提示あり |
| 7   | 動画A  | 動画B  |
| 8   | 指示対象 | 指示対象 |
| 9   | 提示なし | 提示あり |
| 10  | 動画B  | 動画A  |
| 11  | 指示対象 | 指示対象 |
| 12  | 提示あり | 提示なし |

表5 要約文に記述された項目の割合

| 被験者 | 指示対象なし     | 指示対象あり     |
|-----|------------|------------|
| 1   | 47%(8/17)  | 60%(9/15)  |
| 2   | 59%(10/17) | 60%(9/15)  |
| 3   | 47%(8/17)  | 47%(7/15)  |
| 4   | 59%(10/17) | 73%(11/15) |
| 5   | 53%(9/17)  | 60%(9/15)  |
| 6   | 41%(7/17)  | 53%(8/15)  |
| 7   | 33%(5/15)  | 65%(11/17) |
| 8   | 33%(5/15)  | 53%(9/17)  |
| 9   | 40%(6/15)  | 59%(10/17) |
| 10  | 60%(9/15)  | 88%(15/17) |
| 11  | 47%(7/15)  | 47%(8/17)  |
| 12  | 33%(5/15)  | 82%(14/17) |

表6 指示語が適切に言い換えられた項目の割合

| 被験者 | 指示対象なし   | 指示対象あり   |
|-----|----------|----------|
| 1   | 20%(1/5) | 40%(2/5) |
| 2   | 0%(0/5)  | 60%(3/5) |
| 3   | 40%(2/5) | 40%(2/5) |
| 4   | 20%(1/5) | 80%(4/5) |
| 5   | 20%(1/5) | 60%(3/5) |
| 6   | 40%(2/5) | 60%(3/5) |
| 7   | 40%(2/5) | 60%(3/5) |
| 8   | 20%(1/5) | 80%(4/5) |
| 9   | 0%(0/5)  | 40%(2/5) |
| 10  | 60%(3/5) | 80%(4/5) |
| 11  | 40%(2/5) | 60%(3/5) |
| 12  | 20%(1/5) | 80%(4/5) |

Q2. 指示された対象が表示されるという機能は必要ですか？

はい ・ いいえ

その理由 ( )

Q3. 要約筆記をする際に、指示語に対応する指示対象が表示される方法として良いと思う順に順位をつけて下さい。最も良いものが1 となります。「その他」が無ければ、その順位は空欄で構いません。

- ( ) スライド全体が表示され、指示領域が色づけされる。
- ( ) スライド全体が表示され、指示棒先端等の軌跡が色づけされる。
- ( ) 指示領域が切り取られて表示される。
- ( ) 指示領域が切り取られて表示され、指示棒先端等の軌跡が色づけされる。
- ( ) その他 ( )

Q4. 将来的には容易な操作で抽出したものをそのまま要約文に埋め込む機能を実現したいと考えています。そのような機能は必要ですか？

はい ・ いいえ

(2) 結果

Q1 に関して、3名の要約筆記者について表7に示す。

Q2 に関しては、3名とも「はい」という答えだった。その理由として、「指示されているのがどの部分か・どの範囲か分かれば、入力時に指示語の言い換えがしやすい」というものがあつた。

Q3 に関して、集計結果を表8に示す。その他に何か別の方法を書いた要約筆記者はいなかった。「抽出された指示対象がスライドのどこにあるのか分かったほうがよい」という意見を得た。また、「スライドのどこが抽出されたのか分かるように色づけ等されて表示されれば、方法3の指示領域を切り取る方法でもよい」という意見を得た。

Q4 に関しては、3名とも「はい」という答えだった。「記号や式を表したいときがあると便利」や「数式以外への入力漏れがなくなる」という意見を得た。

表7 パソコン要約筆記者の経験歴

| 要約筆記者          | A    | B  | C   |
|----------------|------|----|-----|
| 経験年数           | 1.5年 | 7年 | 3ヶ月 |
| 数式のある講義の要約筆記経験 | あり   | あり | なし  |

表8 アンケートQ3の結果

| 要約筆記者 | A  | B | C |
|-------|----|---|---|
| 方法1   | 1  | 1 | 3 |
| 方法2   | 2  | 3 | 4 |
| 方法3   | 空欄 | 2 | 2 |
| 方法4   | 空欄 | 4 | 1 |

### 4.3 考察

アンケートの結果から、3名の要約筆記者は指示語に対応する指示対象抽出の提示を必要としている。指示語に対応する指示対象が表示されることで、講師の指示動作を見逃した場合もそのときの指示対象が分かるので、指示語の言い換えが容易になると考えられる。このことはほとんどの要約筆記者に該当することであるとされる。

指示対象の表示方法については、概して、スライド全体のうちでどこが指示されたかが分かることが良いと分かる。しかし、指示対象が分かっても数式などの入力にかかる手間は変わらない。本支援法の効果をさらに高めるためには、入力に手間がかかるものを容易に要約文に埋め込むことのできる仕組みが必要である。

## 5 結論

本研究では、指示語発話開始時刻と指示点の軌跡から指示語に対応する指示対象を抽出するシステムを提案し、さらに、抽出した指示対象を遠隔パソコン要約筆記者に提示するという要約筆記者の支援法を提案した。

指示語に対応する指示対象抽出システムは、講師の音声から音声認識を用いて指示語発話の抽出を行う。また、講義映像から指示棒の先端を抽出し、指示点の軌跡から指示動作を抽出する。そして、指示語と指示動作を統合することで指示語と指示動作により指示されたスライド上の図字領域を対応付ける。その図字領域上に指示点の軌跡を描いたものを、要約筆記者に提示する。

指示対象の抽出実験では、再現率85.7%、適合率84.8%という精度が得られた。また、要約筆記者の支援方法の評価では、講義映像のみの場合と、講義映像に加えて指示対象を提示した場合で、要約文に記述された項目数を比較した。その結果、指示対象を提示したほうが記述されるべき項目に対する記述された項目の割合が多かった。そして、要約筆記者へのアンケートでは指示語に対応する指示対象が提示される機能や、それを基に指示対象を要約文に埋め込む機能が期待されていることが分かった。

指示対象の抽出実験の結果と要約筆記者のアンケートから、指示対象が複数の場合、セットにして要約筆記者に提示する必要がある。そのためには、複数を意味する指示語やキーワードを認識し、複数の指示対象を順に指示した場合でもセットで抽出する必要がある。また、統合手法についても見直す必要がある。指示語に対して指示動作に複数の候補がある場合は、指示動作の重要度により選択することが必要である。

今回は、スライドで行う講義で指示棒を用いて指示することを想定したが、板書の場合、レーザーポインタや手で指示する場合にも対応させる。また、スライドの切り替えタイミングや図字領域の位置については映像内のスライド部分を認識し、自動で取得できるようにする。

今後は、抽出した指示対象を要約文に埋め込む機能も実現し、システム全体がリアルタイムで動作するよう実装を行う。そして、実際の講義で提案する要約筆記者の支援法を適用し、その効果を評価する。最終的には、聴覚障害者に提示するヒューマンインタフェースについても検討していく必要がある。

### 【参考文献】

- [1] 加藤伸子, 河野純大, 三好茂樹, 西岡知之, 村上裕史, 皆川洋喜, 若月大輔, 白澤麻弓, 石原保志, 内藤一郎: “聴覚障害者の情報保障におけるパソコン要約筆記入力者に対するキーワード提示” ヒューマンインタフェース学会論文誌, Vol.19, No.2, pp.195-203, 2007.
- [2] 三好茂樹, 河野純大, 西岡知之, 加藤伸子, 白澤麻弓, 村上裕史, 皆川洋喜, 石原保志, 内藤一郎, 若月大輔, 黒木速人, 小林正幸: “遠隔講義保障におけるリアルタイム字幕作成者を支援するための映像情報提示手法に関する基礎的研究” 電子情報通信学会論文誌, Vol.J91-D, No.9, pp.2236-2246, 2008.
- [3] 伊藤英一, 石原剛志, 旭洋一郎, 根岸則子, 宮本晃太郎, 荒川健一: “高等教育における障害学生のための授業支援” 第20回リハビリテーション工学協会, No.93C06, pp.342-343, 2005.
- [4] M. Wald and K. Bain: “Universal access to communication and learning: role of automatic speech recognition” Universal Access in the Information Society, Vol.6, No.4, pp.435-447, 2007.
- [5] Takuro Sakiyama, Masayuki Mukunoki, Katsuo Ikeda: “Detection of the Indicated Area with an Indication Stick”, 3rd. Int. Conf. on Multimodal Interfaces, pp.480-487, 2000.
- [6] Takafumi Marutani, Satoshi Nishiguchi, Koh Kakusho, Michihiko Minoh: “Making a lecture content with

deictic information about indicated objects in lecture materials ”

The 3rd AEARU Workshop on Network Education, pp.70-75, 2005.

[7] L.L. Pozzer-Ardenghi and w.M. Roth : “ Gestures:Helping students to un-derstand photographs in lectures ”  
Connections'03, pp.1-30, 2003.

[8] 伊藤綾鹿, 竹内義則, 大西昇, 飯塚重善, 中嶋信弥 : “ 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”

電子情報通信学会技術研究報告. WIT, Vol.108, No.435, pp.5-10, 2009.

[9] IPtalk, <http://iptalk.hp.infoseek.co.jp/>

### 〈発 表 資 料〉

| 題 名  | 掲載誌・学会名等   | 発表年月              |
|--|--|-------------------|
| 遠隔パソコン要約筆記における指示語に対応する指示対象抽出   | 2009年映像情報メディア学会年次大会  | 2009年8月           |
| 指示語に対応する指示対象抽出による遠隔パソコン要約筆記支援の提案   | 電子情報通信学会福祉情報工学研究会  | 2010年3月           |
| Extraction of Displayed Objects Corresponding to Demonstrative Words for use in Remote Transcription | 12th International Conference on Computers Helping People with Special Needs | 2010年7月<br>(発表予定) |
|  |  |                   |