

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

伊藤 綾鹿[†] 大西 昇[†] 竹内 義則[†] 斉藤 健[†] 飯塚 重善^{††} 中嶋 信弥^{†††}

名古屋大学大学院情報科学研究科[†] NTTサイバーソリューション研究所^{††}

国土館大学理工学部^{†††}

1. はじめに

講義における、聴覚障がい者のための情報保障の 1 つに遠隔地での要約筆記がある。聴覚障がい者は、遠隔地にいる要約筆記者が作成した要約筆記文を読むことで、講師の発話内容を理解する。要約筆記者を支援する 1 つの方法として、キーワード提示がある^[1]。しかし、聴覚障がい者にとっては、講師が発話してから、要約筆記文が提示されるまでに若干時間を要するため、要約筆記文中の「これ」「この」といった指示語と、その時指示していた対象物(以下、指示対象物)との対応付けがしにくく、内容の理解が困難になるという問題がある。

この問題を解決するため、筆者らは、講義の音声と映像を用いて、指示語とその指示対象物を抽出し、それらを関連付けて提示することで、聴覚障がい者が講義内容を理解しやすくする手法の検討を行っている。

本稿では、音声からの指示語抽出処理として、指示語発話に関する特徴、及び判別分析による指示語抽出処理とその結果を示す。

2. アプローチ

指示語とその指示対象物の関連付けを行うために、筆者らは、音声と映像の両方を利用する手法を検討している(図 1)。

まず、音声処理では指示語のみを抽出する。並行して、映像処理では指示動作と指示対象物を抽出する。そして、抽出した指示語と指示動作の関連付けを行い、指示語と指示対象物を対応付ける。この結果を、聴覚障がい者に提示しようと考えている。

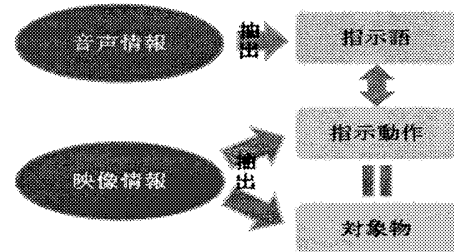


図 1 本研究のアプローチ

次章からは、判別分析を用いた音声からの指示語抽出処理について示す。

3. 指示語発話時の特徴分析

まず、講師発話データベースを構築するために、ある一人の講師の講義を、約 4.5 時間収録した。そして、指示語発話時の特徴を分析するため、この収録データに対して、音声編集ソフトを用いて、指示語発話部分のラベリングを手動で行い、約 1200 個のラベルデータを得た。

ここで、収録を通して、“指示語を発話する前に、ある程度のポーズ(無声区間)がある”という仮説を得ていたため、その検証を行った。その結果、指示語発話全体のうち、300ms 以上のポーズ後の 500ms 以内に指示語の発話が始まるケースが、約 75% を占めていることが確認できた。なお、この約 75% の中には、ポーズと指示語の間に、「ま」などのフィラーが発話されたケースも含んでいる(図 2)。

このことから、抽出対象を「300ms 以上のポーズ後の 500ms 以内に発話された指示語」に限定して抽出処理方式の検討を進めることにした。

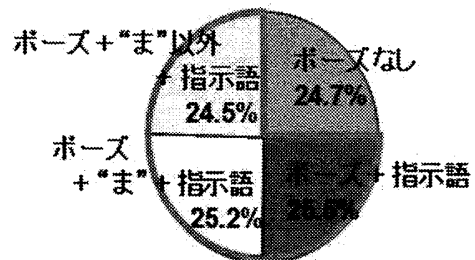


図 2 指示発話時の特徴分析

A study on extraction of deictic words in teacher's utterance on informational support for hearing impaired persons

[†] Graduate School of Information Science, Nagoya University

^{††} NTT Cyber Solutions Laboratories

^{†††} School of Science and Engineering, Kokushikan University

4. 処理方式

指示語を抽出する処理として、基本的な手法である判別分析を用いて、指示語と指示語でないもの(以下、非指示語)を判別することにした。

判別分析を行うためには、非指示語のラベルファイル群も必要となるため、この作成を行った。これは、ラベリングに利用した音声編集ソフトから出力されるピッチファイルを用いて、300ms以上のポーズ後の500ms以内に指示語ラベルの始点がない場合、その500ms区間に非指示語のラベリングを自動で行った。

続いて、判別分析に用いる特徴量の取得を行った。特徴量を取得する分析区間は、3章で示した指示語発話時の特徴を利用して、300ms以上のポーズ後の500msとした。特徴量としては、ピッチファイル中に、10ms単位で出力されている「ピッチ(基本周波数)F0」、「振幅A」、及びそれらの組み合わせを用いることにした。さらに、分析区間を50msごとに分割し、ピッチファイルから得られるF0とAの値を用いて、計16種類の特徴量を取得した。

判別分析では、判別データ(判別対象)と、指示語群と非指示語群それぞれとのマハラノビス距離を算出し、距離の小さい方を判別結果とした。

5. 判別分析の試行

前章で述べた方法について、まず、学習データと判別データに、同じ講義データを使った判別分析(クローズテスト)を実施した。その講義データは、1講義(約1時間)から得た指示語(「これ」「これは」「これを」「これが」のみ)と、非指示語から構成されている。ここでは、特徴量の選定のために、16種類の特徴量の全ての組み合わせ($2^{16}=65536$ 通り)を用いて、判別分析を行い、適合率と再現率を求めた。適合率が最も良かった特徴量の組み合わせ(F0(10), F0max(1), Δ F0min(1), Δ F0max- Δ F0min(1), A(10), Amax(1), Amin(1), Amax-Amin(1), Δ Amin(1) (数字はその特徴量の次元数, maxとminは最大値と最小値))での判別結果を表1中の(1)に示す。

クローズテストでは、比較的良好な結果が得られたので、次に、判別データに、学習データとは異なる講義データを使った判別分析(オープンテスト)を実施した。学習データは、2講義(約2.5時間)から得た指示語(「これ」「これは」「これを」「これが」のみ)と、非指示語から構成されている。判別データは、クローズテストに用いた講義データと同じものを使用した。表1中の(2)に、判別分析結果を示す。この表からわ

かるように、オープンテストでは、クローズテストに比べて、再現率と適合率が共に大幅に下がってしまった。

つまり、この試行で用いた特徴量と学習データの組み合わせでは、学習データ自身の特徴はうまく捉えられていたが、一方で、汎用性に欠けていたと考えられる。

表1 判別分析の試行結果

		(1)	(2)
学習データ	指示語	51個	116個
	非指示語	864個	3214個
判別データ	指示語	51個	51個
	非指示語	864個	864個
再現率 (正解数/指示語総数)		56.9% (29/51)	11.8% (6/51)
適合率 (正解数/抽出数)		87.9% (29/33)	31.6% (6/19)

6. おわりに

聴覚障がい者が、講義の内容をより分かりやすくするために、講師発話内容の要約筆記文中の指示語と、その指示対象物を関連付けて提示する方法の検討を行っている。本稿では、音声からの指示語抽出処理において、「指示語発話前にある程度のポーズがある」点に着目し、音声特徴量を用いて行う判別分析と、その試行結果について示した。その結果、学習データと判別データを同じにしたクローズテストの場合は、適合率は約90%あり、ある程度の精度が得られたが、学習データを違うものにしたオープンテストの場合は、適合率が約30%と低くなってしまった。これは、今回用いた学習データに、汎用性が欠けているという問題があると考えられる。

今後は、汎用的な学習データの構築や判別分析に用いる特徴量の再考、判別方法の検討等により、指示語抽出処理の改善を行う必要がある。さらに、映像処理と結合し、システム全体の構築を進め、本処理方式の有効性を確認する方向である。また、リアルタイムでのピッチファイル生成方法の確立も行っていく予定である。

参考文献

[1] 加藤伸子, 他: 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.9, No.2, pp.125-133 (2007).