

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 27 日現在

機関番号：53901

研究種目：若手研究(B)

研究期間：2016～2017

課題番号：16K16119

研究課題名(和文)字幕の理解しやすさとの関係を考慮した音声データの漸進的な構造化に関する研究

研究課題名(英文) Incremental Structuring of Speech Data Considering Relationships with Readability of Transcription

研究代表者

村田 匡輝(MURATA, Masaki)

豊田工業高等専門学校・情報工学科・助教

研究者番号：30707807

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：研究に使用するデータの収録，構造化データの一部構築を進めた．既存の講演データに対し，解析結果例の作成，タグ付けの仕様に関する検討を行った．また，授業形式のデータについては，5講義分の音声・映像データの収録を行い，音声データの書き起こしを進めた．字幕を提示するシステムのプロトタイプシステムの開発を行った．パワーポイントVBAを使用し，音声認識の結果をパワーポイント上に表示するシステムを開発した．被験者評価の結果，字幕の表示時間，読みやすさについて，一定の評価が得られた．

研究成果の概要(英文)：I have promoted the recording of research data and the construction of a part of structured data. I have constructed examples of parsing results and considered annotation guidelines for existing discourse data. Also, I have recorded speech and image data of five classes. A part of speech data was transcribed into texts. I have developed a prototype of a captioning system. The system displays the result of speech recognition on PowerPoint by use of a VSTO add-in for PowerPoint. A subjective evaluation shows a certain evaluation for display time and readability of caption.

研究分野：情報学

キーワード：音声言語 字幕生成 構造化 漸進的解析

1. 研究開始当初の背景

ニュース番組でのアナウンサーや講演の場での講演者の発話音声を、文字化して字幕として提示する試みが行われており、字幕を自動で生成するシステムの開発も行われている[中野 2007]。これらの字幕は、高齢者や聴覚障がい者の方による内容理解を支援する目的で作成されており、音声を一字一句正確に文字化し、過不足なく提示することが重視されている。しかし一方で、読みやすい字幕を提示するためには、その提示方法も重要となる。例えば、改行の挿入、読点の挿入といった提示方法に関する整形処理を施すことで、音声の文字化テキストの内容を変更することなく、字幕を読みやすくすることが可能である。

話の内容理解を支援するという観点で見れば、音声の全てを忠実に文字化することは必ずしも必要ではなく、話の構造を反映し整形した字幕が有効であると考えられる。例えば、論文やメールなどの書き言葉の文章においては、見出しの付与、箇条書き、重要語句の強調等の構造を明示することが行われている。また、講演で提示されるスライドにおいても同様のことが行われている。このように、字幕の理解しやすさと文章の構造とは密接な関係があると考えられる。スライドの自動作成の試み[安村 2003]、談話構造解析技術[Lin 2010]、並列構造の解析技術[吉本 2015]等、個々の技術の開発は行われているものの、字幕の理解しやすさに着目して構造化を体系的に行うものではない。

2. 研究の目的

本研究では、理解しやすい字幕提示を行うための要素技術として、音声の構造化手法の開発を行う。本手法では、音声言語文を読みやすさという観点から分割した処理単位に対して、

- 単位内の構造（主語述語の関係、重要語句・キーワードの抽出、主題の抽出、言い換え・並列関係の検出等）の解析
- 単位間の構造（見出し・結論の抽出、並列構造の抽出、埋め込み・補足文の抽出、抽象化・具体化文の抽出等）の解析

を順に行うことにより、音声の構造化を実現する。構造の解析を音声の入力に対して漸進的に行うことにより、リアルタイム性が求められる字幕提示システム等において利用できるようにする。

3. 研究の方法

(1) コーパスの整備

構造化の対象となる音声データを収集し、コーパスとして整備する。本研究では、構造化の対象として、授業及び講演における音声を用いる。講演音声としては、名古屋大学で構築された「同時通訳データベース」を用いる。授業コンテンツとしては、既存のデータ

で利用可能なものが存在しないため、新たに授業音声を録音、書き起こしの付与を行う。書き起こしテキストに対して、韻律情報の付与、および、テキストの読みやすさに関係する構造情報の注釈付けを行い、構造化コーパスとして整備する。

コーパスの観察に基づき、字幕の理解しやすさと音声内容の構造との関係について、体系化を実現する。構造解析の処理単位への分割技術の開発を行う。

(2) 統計的な構造解析手法の開発

整備したコーパスを用いて、統計的な構造解析手法の開発に関する研究を推進する。構造解析の処理単位への分割技術、単位内の構造解析技術、単位間の構造解析技術を順に開発する。さらに、音声言語文の構造解析実験を実施し、手法の評価・実験結果の考察を行う。

①構造解析の処理単位への分割技術の開発

読みやすい字幕提示を行うための改行挿入手法を拡張することにより、音声言語文を単位内・単位間の構造解析に適した処理単位へと分割する手法の開発を行う。

②単位内の構造解析技術の開発

読みやすさを考慮して定める単位に対して、単位内の構造を解析する。具体的には、(2)の体系化に従い、単位中の重要個所、キーワード、主題、補足情報、不要情報などを解析する。

③単位間の構造解析技術の開発

単位内の構造を解析した処理単位の情報も使用し、単位そのものの情報、及び、単位間の構造を解析する。具体的には、並列構造、主題-述部の関係、結論などを解析する。

(3) 実験・評価

これまでに開発した技術を統合し、テキストの構造解析技術を実現する。コーパスからテスト用データセットを設け、評価実験を実施し、実験結果の分析を行う。また、解析結果を用いて理解しやすい字幕提示を行うプロトタイプシステムを開発し、提示される字幕について、被験者評価を実施する。これにより、構造化技術の応用可能性を評価する。

4. 研究成果

本研究では、「字幕の理解しやすさ」という観点から、音声の構造化を行う手法の開発を目的とし、研究を推進した。

(1) データの収録

研究に使用するデータの収録、構造化データの一部構築を進めた。本研究は、講演形式、授業形式の音声言語を対象に構造化を行うことを目標としていたため、講演形式のデータについては、既存の独話による講演データである名古屋大学「同時通訳データベース」を使用し、解析結果例の作成、タグ付けの仕様に関する検討を行った。具体的には、音声の

書き起こしテキストに対して、並列箇所や主題-述部の関係、重要箇所、補足箇所といった、解析すべき内容の整理を行い、書き起こしテキストに追記する形で記述した。例を図1に示す。授業形式のデータについては、豊田工業高等専門学校情報工学科の講義5講義分の音声・映像データの収録を行い、音声データの書き起こしを進めた。また、初年度から最終年度前半にかけて、外注によるデータのアーノテーションを実施した。収録・作成したデータは、構造解析手法開発のための重要なデータになると考えられる。

1[<正確に分けること 主題>は<難しい 述部>でございます]
 2[私は<1945年から1972年 見出し>]
 3[<1972年 キーワード>というの]
 4[沖縄が日本に帰ってきた年であり]
 5[また日中の中国との国交正常化ができた年でございます 3,4,5:埋め込みと補足 4,5:並列]
 6[その辺までを<第一期戦後処理 結論> 1,6:見出しと結論]
 7[<復興或は国交回復の時代 言い換え>といっているのではないかと 6,7:具体化]
 8[という気がいたします]

図1. 解析結果例

(2) 字幕提示のためのプロトタイプシステムの開発

字幕を提示するシステムのプロトタイプシステムの開発を行った。本システムは、講演・講義の場で利用することを前提に、プレゼンテーションファイル、具体的には、Microsoft社のPowerPointファイルと同一画面に字幕を表示することが可能である。具体的には、PowerPoint VSTOを用いて字幕提示用のPowerPoint VSTOアドインを作成する。音声を音声認識ソフトウェア Juliusによって文字化し、自動的に追加されるテキストボックスに認識結果文字列を追加することで字幕として提示する。字幕は、1文字あたり0.25秒、1行最大30文字提示する。

本システムによる実際の講義音声をを用いた字幕提示を行い、アンケートによる5段階評価を行った。結果を表1に示す。結果より、字幕の表示時間や見やすさについては、一定の評価が得られている。しかし、内容理解についての評価は低い。これは、音声認識誤りに起因する評価結果であると考えられる。

表1. 視認性に関するアンケート結果

評価項目 (人)	1	2	3	4	5
字幕の表示時間	0	0	2	15	3
字幕の文字の大きさ	0	8	4	6	2
字幕の文字数	0	5	9	3	3
字幕の内容の理解度	13	6	1	0	0
字幕の見やすさ	0	3	12	5	0

本システムを用いることによって、システム使用者は、特別に意識することなく、通常通りパワーポイントを使用した講演・授業を行うだけで、字幕提示が可能となる。システム開発により、構造解析技術が開発されれば、すぐに字幕提示に利用できる環境が整ったといえる。

<引用文献>

- [中野 2007] 音声認識技術を用いた聴覚障害者向け字幕提示システムの課題—話し言葉の性質が字幕の読みに与える影響—, 電子情報通信学会論文誌, J90-D(3), 808-814, 2007.
 [安村 2003] 論文からのプレゼンテーション資料の作成支援, 人工知能学会論文誌, 18(4), 212-220, 2003.
 [Lin 2010] A PDTB-Styled End-to-End Discourse Parser, Arxiv preprint arXiv:1011.0835, 2010.
 [吉本 2015] 並列構造解析に向けた依存構造解析アルゴリズムの拡張, 情報処理学会研究報告, 音声言語情報処理, 2015-SLP-106(5), 2015.

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0件)

[学会発表] (計0件)

[図書] (計0件)

[産業財産権]

○出願状況 (計0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0件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 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

村田 匡輝 (MURATA, Masaki)
豊田工業高等専門学校・情報工学科・助教
研究者番号：30707807

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

重久 直輝 (SHIGEHISA, Naoki)