

聴覚障害者のための字幕つきテレビ放送作成 プロジェクト

若尾 孝博
通信・放送機構 (TAO)
渋谷上原リサーチセンター
wakao@shibuya.tao.or.jp

江原 暉将
NHK放送放送技術研究所
/ TAO

丸山 一郎
通信・放送機構
渋谷上原リサーチセンター

沢村 英治
通信・放送機構
渋谷上原リサーチセンター

白井 克彦
早稲田大学 / TAO

1. はじめに

聴覚障害者がテレビ放送を楽しめるように字幕つきテレビ放送(字幕放送)が行われている。字幕放送は、テレビ番組中の音声元にして字幕を作成して、テレビ番組に多重して放送するものである。

この字幕放送は、日本ではあまり普及していない。米国では、4大ネットワークの70%以上に字幕が付与されているし、英国では公共、民放を問わずテレビ番組全体の3分の1以上に字幕が付けられている。日本では、年々少しずつ増えてきているが、10%程度にすぎない。

この状況を打開するため通信・放送機構 (Telecommunications Advancement Organization of Japan 略称TAO) は「視聴覚障害者向け放送ソフト制作技術研究開発プロジェクト」を開始した。このプロジェクトの概要については、平成9年の言語処理学会第3回年次大会など国内外の学会で発表している([1] [2] [3])。

本プロジェクトの研究目的は、自然言語処理技術や音声認識の技術を用いて、字幕作成を効率的に行う技術を得ることにある。

本文では、プロジェクトの研究項目、研究計画を概観し、前回の発表後に行った新たな研究成果を、自動要約技術と自動同期技術の研究を中心に報告する。

2. 研究項目

本プロジェクトの研究項目は以下の通りである。

- 自動要約技術の研究
- 自動同期技術の研究
- 字幕作成システム技術の研究

目標としているシステムは、図1に示すような字幕放送作成システムである。

本システムでの処理の流れを説明する。入力は、ニュース番組の電子化原稿と映像(音声付き)である。電子化原稿は、自動要約を行うサブシステムで自動的に要約され字幕の原稿(図1字幕データ)となる。

一方、音声部分は自動同期を行うサブシステムで、元の電子化原稿との対応が測られてタイミング情報(タイムコード)が加えられる(図1認識フェーズ)。

次に、元原稿と字幕原稿、それに、タイミングの情報から、字幕と音声(映像)との同期がとられ(同期フェーズ)、最終的に字幕がテレビ番組に付与される。

尚、テレビ放送番組全般を研究対象としているが、現状において聴覚障害者の要望が強いニュース番組を、第一の目標としている。本文では、専らテレビニュース番組を対象にして記述する。

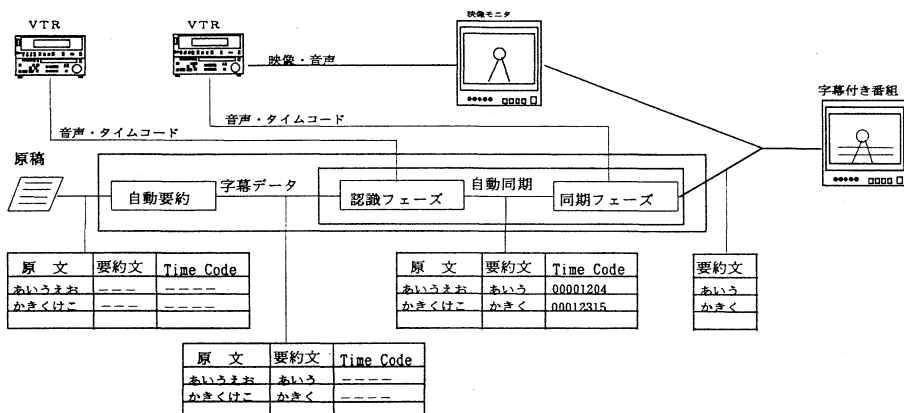


図 1 自動字幕放送作成システム処理フロー

3. 研究計画

本プロジェクトは平成8年度よりの5ヶ年計画で進められており、前期（3年）と後期（2年）の2段階がある。前期は、下記に述べる研究項目の要素技術を研究し、また、それらを統合してプロトタイプシステムを作成する。後期では、プロトタイプシステムを改良するとともに、字幕の表示の方法について研究・評価を行い実用システムを提案する予定である。

平成8年度、9年度の研究項目は以下の通りである。

- 自動要約技術
 - ・短文分割システム
 - ・重要部切り出しシステム
 - ・つなぎ部分調整システム
 - ・パターン方式文字数圧縮システム
- 自動同期技術
 - ・読みつけ、音声モデル合成システム
 - ・最尤照合システム
 - ・音声データベース構築
- 字幕作成システム技術
 - ・字幕作成統合化シュミレーションシステム
 - ・字幕の画面上での提示方法

4. 研究成果

次に、各研究項目毎に最近の研究成果を述べる。これらは、昨年の言語処理学会での発表後に、新たに行われた研究である。

4.1. 自動要約技術

前回は、テキスト中の重要語・重要文の自動抽出について研究を報告したが、今回は、重要だと判断された文の文字数を削減する方法についての研究成果を述べる。

まず、現在放送されている数少ない字幕付きニュース番組として、NHK教育テレビの「手話ニュース845」に注目した。この番組では、その日の主なニュースに字幕と手話が付与されている。字幕は、音声として読まれる原稿と比べると文字数に制限があり、結果として音声原稿の要約となっている。この音声原稿と字幕に注目をして、そこで使われている要約の方法を調べた〔4〕。

この調査により判明したことは、要約文は、原文と比べてみると、

- ・複雑な構文をしていない、基本的に単文
- ・文頭、文末が省略／簡潔化されている
- ・文中の重要な名詞句は字幕中に残る
- ・固有名詞はなるべく簡潔に表現

などの特徴があった。

次に、この調査に基づき要約のための規則、つまり、文字数を削減する異なる規則の収

集を試みた。収集した規則は表層の文字列からの情報を基礎にして適用出来るものを中心にした。これらの規則は大きく分けて5つのグループに分割出来る。

- 1) 文末の削除、言い換え
 - サ変動詞はサ変名詞に変え、その後来る語句は削除
「… 疑いで逮捕したものです。」
→ 「… 疑いで逮捕」
 - 丁寧助詞「ます、まし」は削除する
「… 見合わせました。」→ 「… 見合わせ」
 - 報告文の文末に来る報告動詞の部分は削除
「… 売ろうとしていたと言うことです。」
→ 「… 売ろうとしていた」
- 2) 文の部分を残す
 - 文の一部を残し、あとは削除する
「特捜部では、利益提供は株主総会を乗り切るため行われたものとみています。」
→ 「利益提供は株主総会を乗り切るため」
- 3) 意味を変えずに別の語句で言い換え
 - 短縮形を使う
「駐留軍用地特別措置法」→ 「特措法」など
- 4) 文頭語句の削除
 - 接続詞や文頭のつながりの語句は削除
「しかし」、「一方」、「その一方で」など
- 5) 比較日時は削除
 - 「今日」「先月」などの比較日時を表す語句は削除

次に、別途用意した「手話ニュース」からの音声原稿文と、それに対応する字幕文を192ペア用意し、そこで、上記のどの要約の規則がどれだけの頻度で使われているかを、また各要約の規則を適用することにより、どれだけの文字数が削減されたのかを調査した。その結果が表1である。

使用の頻度から見るとグループ1が最も頻繁に用いられていた。しかし、削減される文字数から見るとグループ2や3がより有効であることが分かった。

現在は、これら規則の幾つかを自動的にテキストに適用し、字幕文を自動生成するシステムの研究を進めている。

表 1 要約規則の使用頻度と削減文字数

規則グループ	使用頻度	削除文字数 (1規則当り)
1	105	4.30
2	80	8.33
3	34	5.62
4	25	3.32
5	9	2.22

4.2. 自動同期技術

自動同期技術の研究では、ニュース音声と字幕を同期させる手法について研究を進めている。ニュース音声を発音記号列に変換し、実際のニュース音声（「実音声」）と比較することで同期のタイミングを検出する方法を用いている。

次節に説明するニュース音声データベースの一部を学習データとして、音声モデルを作成した。これを用いて各話者毎に異なる3ループ8混合分布の音韻HMMを学習した。この音韻HMMを基に、ワード列ベ

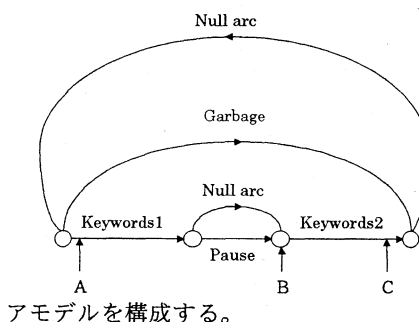


図 2 ワード列ペアモデル

ワード列ペアモデルは、図1のように同期点の前後で単語列1(keywords1)と単語列2(keywords2)とを連結し、中点Bで尤度を観測して、そのピークを検出するものである。前年の報告では、単語列として1文節相当のものをしていたが、今回の実験では2文節相当のものに拡大した。また、文節間にポーズが挿入される場合を考慮してモデルを改良した。このモデルでは、実音声を通した時、同期点以外の音声はガーベジ部分を通り、同期音声だけが単語列ペアの部分を通るようになっている。このため、同

期音声部分が通る場合にのみ尤度が相対的に高くなり、尤度のローカルピークから同期点を検出することが可能となる。

次に、このモデルの評価を行った。単語列のペアとして34種類を選択し、話者4名で、合計136箇所の同期点の検出を行い、検出の精度（検出率）とキーワードあたりの沸きだし数を測った（[5]）。

検出閾値	検出率 (%)	沸き出し率 (FA/KW/Hour)
-10	34.56	0
-20	44.12	0
-30	54.41	0
-40	60.29	0
-50	64.71	0.06
-60	69.12	0.06
-70	69.85	0.06
-80	71.32	0.12
-90	78.68	0.18
-100	82.35	0.18
-150	91.18	0.54
-200	94.85	1.21
-250	95.59	1.81
-300	99.26	2.41

表 2 同期点検出結果

4.3. 音声データベース

自動同期技術研究のための基礎データとなるニュース音声データベースを収集している。平成8年度は、シュミレーションによって音声データを収集した。男女合計20名、トータルで約7.5時間のニュース音声を収録することが出来た。平成9年度はシュミレーションによるものに加えて実際のテレビ・ラジオ番組から其々10時間のリアル音声の収集を行った。

4.4. 字幕作成システム技術

自動要約、自動同期技術の研究と同時に、字幕作成の技術についての研究も進めている。ここでは、聴覚障害者にとって、より良い、より見やすい字幕とは、どういうものかを調べ、字幕作成統合化シュミレーションシステムに反映させて行く予定である。

平成9年度は、数種類のテレビ番組に手で字幕を付与し、健聴者と聴覚障害者の方々に、その字幕付き番組を見てもらい、どの

字幕の提示の方法が良いかを主観評価してもらった（[6]）。

調査の項目としては、以下のようなパラメーターを設定した。

- ・ 字幕の文字：書体、サイズ、間隔、色
- ・ 単語・行：行数、文字数／行、行末処理
- ・ 字幕全体：表示向き、位置、タイミング
- ・ 親映像：字幕を重ねる、重ねない

この調査の詳細については、機会を改めて、発表する予定である。

5. おわりに

視聴覚障害者向け放送ソフト制作技術研究開発プロジェクトの概要について述べた。今後は、字幕作成用システムを構築することを進め、そのシステムを使って字幕作成のための技術を確立していきたい。

参考文献

- [1] 江原、沢村、若尾、阿部、白井「聴覚障害者のための字幕つきテレビ放送制作への自然言語処理の応用」言語処理学会第3回年次大会pp 489-492.
- [2] Wakao, T., Ehara, E., Sawamura, E., Abe, Y., Shirai, K. (1997) *Application of NLP technology to production of closed-caption TV programs in Japanese for the hearing impaired* ACL 97 workshop, Natural Language Processing for Communication Aids, pp 55-58.
- [3] Ehara, T., Wakao, T., Sawamura, E., Maruyama I., Abe Y., Shirai K. (1997) *Application of natural language processing and speech processing technology to production of closed-caption TV programs for the hearing impaired* NLPRS 1997.
- [4] 若尾、江原、白井「テレビニュース番組の字幕に見られる要約の手法」情報処理学会、自然言語処理研究会、NL-122-13.
- [5] 丸山、阿部、江原、白井「字幕・音声同期におけるワード列ペアモデルのスポットティング精度改良」日本音響学会、平成10年度春期研究発表会
- [6] 棚橋「聴覚障害者向けTVニュース用字幕の表示に関する研究」早稲田大学理工学研究科 情報科学（修士論文）1998.