自然言語処理　144－17
(2001. 7. 17)

聴覚障害者向け字幕番組の制作技術

江原 暉将*　沢村 英治**　福島 孝博***
丸山 一郎†　和田 裕二**　門馬 隆雄**　白井 克彦††

* 通信放送機構／ＮＨＫ
** 通信放送機構
*** 通信放送機構／追手門学院大学
†通信放送機構（現　三菱電機（株））
††通信放送機構／早稲田大学

あらまし　通信・放送機構で平成８年度から１２年度まで実施した「視聴覚障害者向け放送ソフト制作技術の研究開発プロジェクト」の研究成果と残された課題について報告する。本プロジェクトの目的は、聴覚障害者のための字幕付きテレビ放送番組を効率的に制作するための技術基盤を確立することである。具体的な研究項目として、自動要約、自動同期、統合化システム技術がある。自動要約については、ニュース記事を対象に文字数にして 70%にすることを目標にして研究を進め、「重要文抽出法」と「形態素単位文字数圧縮法」を併用して目標を達成した。自動同期については、ニュースおよびナレーション主体のドキュメンタリー番組を対象に研究し、ナレーションと背景音の比が 20dB 以上の番組に対しては自動同期が可能であることを示した。統合化システム技術では、適切な点で字幕の改行・改ページを加える自動字幕画面制作技術を研究し、自動要約、自動同期とあわせて自動字幕制作システム実証モデルを構築した。本実証モデルを用いて評価実験を行い、性能評価を行うと共に実用化のための課題を明らかにした。
キーワード　字幕放送、自動要約、自動同期、自動字幕画面制作

Program production technologies for TV closed captions for hearing impaired people

Terumasa Ehara*　Eiji Sawamura**　Takahiro Fukushima***
Ichiro Maruyama†　Yuji Wada**　Takao Monma**　Katsuhiko Shirai††

* TAO / Japan Broadcasting Corporation (NHK)
** Telecommunication Advancement Organization of Japan (TAO)
*** TAO / Otemon Gakuin University
† TAO (Now Mitsubishi Electric Corporation)
†† TAO / Waseda University

Abstract Telecommunication Advancement Organization of Japan proceeds "Research Project for TV Production for the Seeing and Hearing Impaired" from 1995 to 2001. The purpose of the project is to establish the technologies of producing closed captions for hearing impaired people on TV programs efficiently. We have three research issues in the project: automatic text summarization, automatic synchronization with speech and captions and system engineering. Automatic text summarization summarizes Japanese news text to 70% volume. Important sentence extraction, morphem-based text shortening and bunsetsu-based text shortening are used. Automatic synchronization uses HMM-based word spotter and DP-based synchronizing point search. The method can be applicable to news and narration programs in which signal strength ratio between speech and background sound is more than 20dB. System engineering research results automatic changing method of new page and new line at a point easy to read. We integrate these elementary technologies to the automatic captioning system and evaluate it by caption creators and end users. From this evaluation experiments, we can know the system performance and future research issues.
Key words Closed-Caption, Automatic Summarization, Automatic Synchronization

1 はじめに

通信・放送機構では、「視聴覚障害者向け放送ソフト制作技術の研究開発」を行っている。このプロジェクトでは、聴覚障害者のための字幕付きテレビ放送番組を効率的に制作するための技術基盤を確立することを目的として、自動要約技術、自動同期技術、統合化システム技術の３つの項目を研究している。プロジェクト期間は平成８年度から１２年度までであり、本年３月に終了したが、第２フェーズとして更に３年間の予定で研究している。ここでは第１フェーズの研究成果と残された課題を中心に報告する。

2 研究開発対象

本プロジェクトの具体的な研究開発対象としては、（１）字幕原稿の自動要約技術、（２）字幕と映像の自動同期技術、（３）字幕番組制作の統合化システム技術の３つの項目を対象とする。本プロジェクトの研究成果を利用することで、図 1に示すような自動字幕番組制作システムを作成できることが期待される。なお、本プロジェクトでは、あらかじめ原稿のある放送番組を対象としており、スポーツの生中継など、原稿のない番組は対象としていない。また、オフラインの字幕制作が対象であり、リアルタイム制作は範囲外である。

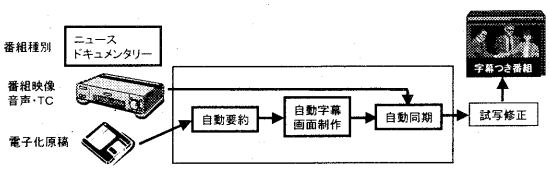

図 1 自動字幕番組制作システム

2.1 自動要約技術

ニュースなど情報番組のアナウンス速度は最高、毎分４００文字以上と、非常に高速であり、これをすべて字幕にすると読み切れない可能性がある。実際、字幕の最高表示速度の目安は毎分３００文字とされている。そこで、自然言語処理技術を用いて、原稿を自動的に要約する技術を研究開発する。現在、人手に頼っている要約を含む字幕原稿の作成を自動化あるいは半自動化することが本テーマの目的である。具体的にはニュース記事を対象にして文字数にして７０％にすることを目標にする。

2.2 自動同期技術

字幕データは映像に同期してタイミング良く付加されなければならない。現在は、この同期作業を人手で行っており、この部分を音声認識技術を中心とする音声処理技術を用いて自動化するのが本テーマの目的である。プロジェクト開始当初は、比較的静かな環境での発声が多く、しかも事前制作番組としてニュースのＶＴＲインサート部分を対象番組としたが、研究途中で対象番組にドキュメンタリー番組を追加した。ドキュメンタリー番組は、ナレーターの発話が主体であるが、背景音がニュースより大きく、さらに、長い非音声区間があるなど当初目標より難易度が高い。

2.3 統合化システム技術

自動要約技術と自動同期技術を中核部分とするシステムインテグレーションを行い、字幕放送番組の自動制作システムを構築することを主目標とする。本目標を達成するために、電子化テキストに適切な改行・改ページ点を挿入し、字幕画面を自動制作する技術の研究が必要であるため、これも目標に加えた。さらに、最終年度において、字幕制作の専門家や聴覚障害者などの協力を得て、自動字幕制作システム実証モデルの機能・操作性の評価および同モデルを用いて制作した字幕付き番組の評価を研究細目に加えた。

3 研究開発の成果

3.1 自動字幕要約技術

ニュース記事を対象に文字数にして７０％にすることを目標に自動要約を研究した。具体的成果は以下のとおりである。

3.1.1 重要文抽出法

記事中から重要語を抽出し、重要語を多く含む文を重要文として抽出することで要約を行うものである。本システムを利用して記事を構成する各文について重要度の順位付けを行い、人手による順位付けと比較した。ニュース記事では、第１文がリード文として最重要である場合が多い。そこで、システムが最重要と判定した文が第１文であるかどうか調査した。その結果を図 2に示す。

また、システムと人手での順位づけを比較したところ、上位２文についての一致度は７２％であり、下位２文の一致度は６０％であった。

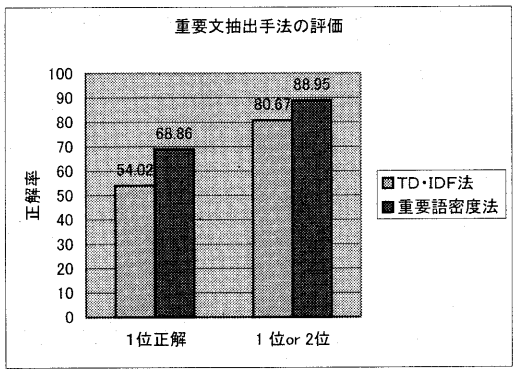

図 2 重要文抽出法の評価結果

3.1.2 自動短文分割

重要文抽出法は、文単位に要約するものであるため、記事中のある部分が大きく抜けることになる。これは、字幕のための要約としては望ましいものではない。そこで、長文を事前に自動短文分割してから要約を行う手法を提案した。短文分割により、1文あたりの文字数を平均96文字から66文字にすることが出来、比較的均等な要約が可能となった。短文分割後に重要文抽出を行った結果、下位2文の一致度が3%向上し、要約の精度面からも有効であった。

3.1.3 形態素単位文字数圧縮法

文末の丁寧表現を簡潔な表現にするなど、形態素単位で文字数を圧縮する自動要約手法である。例えば、「行きます」を「行く」とすることで文字数を2文字少なくすることができる。圧縮のための規則を207種類用意して実験を行ったところ、約92%に文字数を少なく出来た。重要文抽出法と合わせて、目標の70%の要約率を達成した（表 1）。

表 1 文字数圧縮の評価結果

	圧縮率
文字数圧縮のみ	0.947
短文分割＋文字数圧縮	0.919

3.1.4 文節単位文字数圧縮法

より均等な要約を行うために、文節単位文字数圧縮法を研究した。これは、修飾文節など、比較的重要でない文節を削除することで文字数を少なくするものである。当初研究細目にはなかったテーマであるが、ニュース記事の要約を一層一様にする目的で追加した。本手法のシステム化は行ったが、圧縮規則の充実や要約性能の評価は未着手であった。

3.2 音声認識・自動同期技術

ニュース番組を対象に、原稿と音声の同期を自動的に行うことを目標に研究した。研究途中から対象範囲をドキュメンタリー番組にも拡張した。字幕送出タイミング検出手法の流れを図 3に示す。

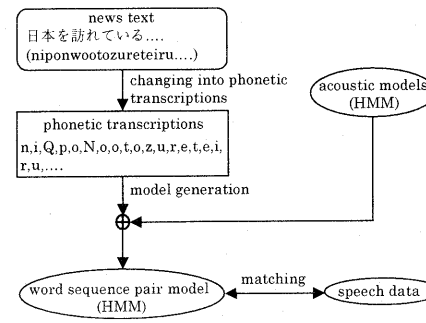

図 3 字幕送出タイミング検出の流れ

具体的成果は以下のとおりである。

3.2.1 ワード列ペアモデル

ニュース番組では、原稿とアナウンスが異なる場合がある。そのような場合にも対応できる自動同期手法として、キーワードスポッティグ手法を拡張したワード列ペアモデルによる手法を提案し、システム化した（図 4）。

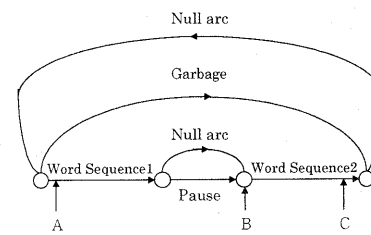

図 4 ワード列ペアモデル

本手法を背景音の少ないニュース音声に適用したところ、許容誤差を0．1秒とした場合で同期点検出率100%、沸き出し誤りが1～10

false-alarm/keyword/hour という性能が得られ有効性が確認できた。

3.2.3 ＤＰマッチングとの併用

ドキュメンタリー番組はニュースと異なり、長い非音声区間が存在したり、大きな背景音が存在したりするため、自動同期の困難性が高い。そこで、ワード列ベアモデルで同期点検出に失敗した場合、ビタビ照合を行うハイブリッド型同期システムを研究開発した。

また、長い非音声区間がある場合に、音声／非音声の区分けを行って、番組を複数の音声区間に分割し、さらに原稿の文字列がどの音声区間で発話されているかを推定する音声区間推定の手法を考案した。また、文の発話順序や音声らしさを考慮するスコアとワードスポッティングのスコアを組み合わせてＤＰマッチングにより最適同期点を求める手法を考案し、システムに組み込んだ。その結果、ドキュメンタリー番組のクリーン音声に対して許容誤差を１秒とした場合で９９％以上の同期点検出率が得られた。この実験では、同期点を１カ所だけ出力しているため、沸き出し誤りの評価は必要ない。一方、ドキュメンタリー番組の放送音声に対しては、許容誤差を１秒とした場合で９６．９％の同期点検出率が得られた。実験結果を図 5に示す。

3.2.4 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して各種の音声データベースを構築した。これは、収録した音声を厳密に書き起こし、さらに発話時刻の情報などを付与したものである。構築した音声データベース表 2に示す。

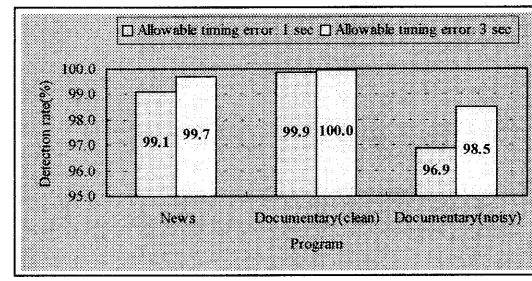

図 5 自動同期実験結果

表 2　音声データベース

年度	番組種類	内容	時間数[時]
平成8	シミュレーションニュース	スタジオ内におけるアナウンサーによるニュース読み上げ(男 6,女 14)	7.5
平成9	シミュレーションニュース	スタジオ内におけるアナウンサーによるニュース読み上げ(男 2,女 2)	3
	ニュース番組	NTV テレビニュース、NHK ラジオニュース	20
平成10	情報/ドキュメンタリー番組	NHK「生きもの地球紀行」	11.25
平成11	ハイビジョンニュース番組	NHK「ハイビジョンニュース」	2.3
平成12	バラエティ番組	ビデオ「昭和名人芸大全」「ひとり芝居」他	9.9
	情報/ドキュメンタリー番組	NHK「生きもの地球紀行」「四国八十八ヶ所」「シルクロード」、「街道をゆく」	27.6
	教育番組	NHK 教育「ふしぎ研究所」、「くらし発見」、「マセマティカ」他 2 番組	21

3.3　字幕スーパープロトタイプシステム技術

自動要約、自動同期、自動字幕画面制作の３種類の要素技術を統合し、自動字幕制作システムを構築することを目標にした。さらに、自動字幕画面制作技術も本テーマの中で研究した。

3.3.1 自動字幕制作システムプロトタイプ

自動要約、自動同期、自動字幕画面制作の３つのサブシステムを統合化し、さらに、字幕制作のヒューマンインターフェースも加えて字幕番組制作統合化システムのプロトタイプを作成した。自動要約と自動同期はワークステーション上で

動作し、自動字幕画面制作と全体制御はパソコン上で動作する。両者はネットワークを介して結合した。

3.3.2 自動字幕制作システム実証モデル

自動同期部分を除く残りの部分全体をパソコン上に実装した実証モデルおよび自動同期も含めてシステム全体をパソコン上に実装した実証モデル２を構築した。これらの実証モデルは実用化を視野に入れたものであり、既存の電子化原稿の作成システムや試写・修正システムとのインターフェースも考慮して開発した。本実証モデルを用いて3.3.4に示すシステム評価実験を行った。

3.3.3 自動字幕画面制作技術

電子化原稿は音声に沿ったテキストであるため、字幕画面とするためには、適切な個所で字幕の改ページを行う必要がある。また、１ページの字幕には、１行に最大１５文字のテキストが複数行提示できるので、適切な個所で改行を行う必要がある。これらの改行・改ページを自動的に行う自動字幕画面制作技術を研究した。日本語形態素・文節解析を利用して、文節末などできるだけ読みやすい個所で改行・改ページを行うようにした。

3.3.4 自動字幕制作システム実証モデルの評価

実証モデルの評価としてａ）実証モデルを用いて制作した字幕番組の品質評価、ｂ）実証モデルを用いて字幕制作を行う作業過程において必要な機能性能の評価　の２種類の評価を行った。その結果、当初目標は達成しているものの、実用化のためには、映像との適切な関係を保つなどの課題があることが分かった。また、電子化原稿中には長いポーズ情報や話者種別の情報など付加的な情報が必要であることが分かった。本システム評価は、システム評価ワーキンググループを設置して行った。ワーキンググループには、字幕制作会社、放送局、聴覚障害者、学識経験者などに加わっていただいた。

4　今後の課題

自動字幕制作システム実証モデルと実証モデル２の構築、およびそれらを用いたシステム評価の結果、以下のような今後の課題が明らかになった。

4.1　自動要約技術

（１）　文節単位文字数圧縮法のシステム化

現在プロトタイプレベルであるので、これを完成させシステム化する。これによってニュース記事に対する自動要約は一応の完成とみることができよう。

（２）　ニュース記事以外の自動要約

ドキュメンタリーやバラエティ、ドラマなどニュース以外の番組に対しても自動要約が必要である。これらの番組の要約は、ニュースの場合とかなり異なる。予備的な検討の結果、ナレーション部分よりも会話部分で要約の必要性が高いことが分かり、ニュースにはない表現の要約手法が必要となる。

（３）　分かりやすい表現への自動変換

音声をそのまま字幕とすると分かりにくい場合がある。特に会話で、そのような場合が考えられる。そこで、より分かりやすい表現に自動変換することが考えられる。これは、一種の言い換え処理であり、これまで取り組んでいる文字数圧縮法と比較して、より高度な処理が必要となる。言い換え処理の中には、難しい漢字をかな表記にしたり、ルビをふる処理も含まれる。漢字の使用は子供向け番組の場合と大人向け番組の場合で当然異なるものである。これらの言い換え処理によって、理解しやすい字幕にすることができる。

4.2　自動同期技術

（１）　会話調発話対応

現在の同期システムはニュースやナレーションなどの読み上げ音声を主対象にしているが、多くの番組を対象にするには会話調発話に対応させる必要がある。

（２）　背景音が大きい場合の自動同期

現在、音声と背景音の比率が20dB以上ある場合には自動同期が可能であるが、番組によっては、背景音が一層大きいものもあり、これらの高背景音に自動同期技術を対応させることが課題である。

（３）　複数話者対応

番組によっては、複数の話者が同時発話を行う場合がある。現在のシステムは逐次発話が基本であり、相づちなど字幕としては不要な一部の現象を除いて同時発話には対応していない。ドラマなどを対象とした自動同期の実現のためには、同時発話を含む複数話者対応が必要である。

（４）　高速化の実現

現在、番組時間の３倍程度の処理時間を要する自動同期処理を高速化することで利便性を一層向上できる。番組時間の１倍の処理を目標に高速化を行うことが課題である。

4.3　自動字幕画面制作技術

（1）　改行・改ページ点決定手法の改良

現在の改行・改ページ点の決定手法には、固有名詞の途中で改行が行われるなど問題点がある。これらの点を改良する新しい改行・改ページ手法を求めることが課題である。

（2）　イン点・アウト点決定手法の改良

現在のイン点・アウト点（字幕画面の提示会誌と終了のタイミング）決定は、自動同期の結果をほぼそのまま利用するのが基本である。しかし、故意に同期結果をずらす必要がある場合がある。例えばナレーターの発話の場合、イン点は早めにアウト点は遅めに設定する方が適切な場合がある。このようにシステム的な調整を行うきめ細かなイン点・アウト点決定手法の研究が課題である。

（3）　親映像との関係調整

親映像にオープンキャプションがある場合や表など文字情報がある場合には、字幕によって、それらの映像が隠されてしまうので、親映像と字幕画面との関係を適切に制御する必要がある。さらに、画面内に特に重要な映像がある場合も、字幕で親映像を隠すのは望ましくない。この場合も字幕画面の調整が必要となる。

（4）　背景音情報などの字幕化

現在のシステムでは、音声のみを字幕化しているが、現実の字幕では音楽マークや背景音情報（犬の鳴き声など）なども字幕で表す必要がある。さらに話者マークや字幕色による話者区別なども行われる。これらを含んで総合的な字幕画面制作を行う必要があり、本テーマの課題である。現時点では、この課題に対して自動で行うことが困難であると考えられ、電子化原稿の中に何らかの付加情報を事前に記入しておくことなどが一方法であろう。

4.4　統合化システム技術

（1）　実用システムの提案

現在の手作業による字幕制作装置との整合性を取り、実用システムを提案する。その中には、字幕色の選択、半角文字の利用、改行幅の設定など実証モデル２では実現できていない機能を実現する必要がある。

（2）　電子化原稿作成支援

本システムは電子化原稿の存在を仮定したものである。放送局から電子台本が提供されることは、今後進むと予想されるが、提供された台本だけでは、情報が不足する可能性がある。例えば、字幕化が必要な背景音の情報などが必要である。そこで、字幕制作に十分な電子化原稿の作成を支援するシステムの開発が課題である。

さらに、番組音声から自動的に電子化原稿を作成する技術もニーズが高い。しかし、背景音を含む音声の自動認識は精度が不十分であるため、印刷台本を併用することで電子化原稿を自動または半自動で作成する方法を考えている。

以上、今後の課題について示したが、現在行われている人手による字幕制作を観察すると、高度な人間の能力に依存していることが分かる。そこで自動字幕制作技術を実用化するにあたって、人間の能力と機械の能力を上手にタイアップさせることが重要であると考える。そのためには、どの部分を人間が担当し、どの部分を機械で行うかを十分に検討してから課題に取り組む必要があろう。

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」について報告した。本プロジェクトは、当初目標を達成できたものの実用化するには課題が残った。

本研究を進めるにあたって、ＮＨＫおよびＮＴＶからは番組の研究使用を許諾していただいた。連絡会、システム評価ＷＧの構成員の皆様および研究フェローの皆様には、ご指導、ご協力をいただいた。これら関係各位に深く感謝いたします。

総務省（旧　郵政省）の指針によると２００７年までに字幕付与可能な放送番組に対して１００％字幕を付与することが要請されている。本プロジェクトの成果をベースにして、システム評価の結果明らかとなった課題を解決し、効率的な字幕制作を技術的に支援する自動字幕制作システムの実用機を提案して行くことで、字幕放送の一層の充実に寄与したい。

参考文献

本報告の詳細は各年度の「研究開発報告書」と「最終報告書」を参照されたい。その内容は下記のWWWサイトからでも入手できる。

http://www.shibuya.tao.go.jp/report.htm