テレビ放送への聴覚障害者向け字幕付与の自動化

Automatic Closed Captioning on TV Broadcasting for Hearing Impaired People

江原暉将*、沢村英治**、福島孝博**、丸山一郎**、門馬隆雄**、白井克彦***
* 通信・放送機構 渋谷上原リサーチセンター（ＴＡＯ）／ＮＨＫ放送技術研究所
eharate@shibuya.tao.go.jp; http://www.shibuya.tao.go.jp
** ＴＡＯ
*** ＴＡＯ／早稲田大学
Terumasa Ehara*, Eiji Sawamura**, Takahiro Fukushima**,
Ichiro Maruyama**, Takao Monma**, Katsuhiko Shirai***
* Telecommunications Advancement Organization of Japan (TAO) /
NHK Science and Technical Research Laboratories
** TAO
*** TAO / Waseda University

１．はじめに

通信・放送機構渋谷上原リサーチセンターでは、字幕付きテレビ番組など視聴覚障害者向け放送ソフトの制作技術に関する研究開発を平成８年度から５ヵ年のプロジェクトですすめている。具体的には、聴覚障害者向けの字幕（クローズドキャプション）付きテレビ番組を制作する作業を支援する技術の中で、自動要約、自動同期、システム技術の３つのテーマについて研究を行っている[1] [2]。本文では、本プロジェクトの研究内容およびこれまでに得られている研究成果、今後の課題について報告する。

２．字幕作成の手順

本論に入る前に、字幕付きテレビ番組の制作過程を考察する。これによって、技術支援がどのように可能になるかを考察することができる。字幕のないテレビ番組に字幕を付与して字幕付きテレビ番組を制作する過程を観察すると、以下のステップからなることがわかる。ただ、実際の字幕制作現場では、これらのステップを順次進めるわけではなく、全ての作業をまとめて行う方法もある。また、実際には、これらの作業の後で、試写と修正が行われ、字幕付きテレビ番組が完成することにも注意する必要がある。

（１）音声を文字化する

番組の音声を文字化する。この場合、動物の鳴き声などの効果音や背景音についても必要に応じて文字化や記号化が行われる。テーマ音楽の部分にも音楽であることを示す記号が挿入される。文字化にあたっては、適切な表記を選ぶ必要がある。漢字の使用や外国語の表記については、親映像の字幕（オープンキャプション）と同じように表記規準が定められている。

（２） 要約を行う

番組によっては、発話速度が早いため、文字化した音声を全て字幕とすると読み切れない場合がある。その場合は、発話内容を要約して適切な字幕表示量とする。たとえば、ニュースのアナウンサーは、毎分４００文字以上の速度で話す場合があり、これは、字幕の最高表示速度の目安である毎分３００文字を越える。適切な要約が必要なゆえんである。

（３） 字幕画面を制作する

文字化された音声を適切に区切って、字幕画面を制作する。読みやすい字幕画面とするためには、１画面に１文が入ることが望ましいし、改行位置も文節末であることが望ましい。また、ドラマや対談番組などでは、字幕の内容が、どの発話者の発言であるかを分かりやすく表示する必要がある。そのために、字幕の表示位置を変えたり、字幕に色をつけるなどの工夫が行われる。さらに、親映像にすでにオープンキャプションがあって、クローズドキャプションと重なる場合は、オープンキャプションを避けた位置にクローズドキャプションを配置することも必要である。

（４） 音声と字幕の同期

以上のステップで制作された字幕画面は、最後に音声にあわせてタイミング良く親映像に重畳される。これを、本文では、音声と字幕の同期と呼んでいる。同期では、音声と字幕の提示タイミングが合っていることが基本であるが、画面の切り替えなどがある場合は、提示タイミングをわざとずらす場合もある。また、字幕の提示時間長も、文字数や隣接の字幕画面との関係を考慮して適切に決められる。

３．研究項目の選択

前節で述べたように、字幕付きテレビ番組の制作過程は、さまざまな要素から構成されており、人間の高度な能力に負っている部分が多い。従って、機械で支援するにあたっては、細心の注意が必要である。本節では、機械支援が可能な部分について、特に本プロジェクトで採用している研究項目を中心に述べる。

（１）音声の自動文字化

字幕制作の中でも、手間のかかる番組音声の文字化の部分を、音声認識を利用して自動的に行うことが考えられる。しかし、本プロジェクトの開始時点では、成果の見通しが十分たたなかったため、研究テーマとしては採用しなかった。従って、本プロジェクトでは、何らかの手段で番組音声が文字化された電子化原稿が事前に存在する番組を対象にして研究を行っている。

（２）自動要約

自然言語処理技術を利用して、要約を自動的に行う自動要約が考えられる。字幕制作のための要約と一口にいっても、番組の種類によって、その手法が異なる。発話が機知に富んでいるドラマやバラエティ番組の自動要約は、困難性が高い。一方、ニュースは、文体が比較的に標準化されているため、自動要約の実現の可能性が高い。本プロジェクトでは、ニュースを対象にした自動要約を研究テーマとしている。

（３）自動同期

音声認識技術を中心とする音声処理技術を使って、番組音声と電子化原稿を対比することで「今どこの部分を発話しているのか」という自動同期のための基本的な情報を得ることができる。音声認識と同様に自動同期は、発話の明瞭度や背景音の有無によって技術的困難性が変わる。本プロジェクトでは、比較的明瞭な発声をしているニュース番組やドキュメンタリー番組を対象に自動同期を研究している。

（４）自動字幕画面制作（システム技術）

自然言語処理技術および音声処理技術を用いて、適切な字幕画面を自動的に作成する手法を研究テーマとしている。改行、改ページの挿入場所を自動的に設定したり、字幕提示時間長の制御を行うことがテーマである。

ニュースを対象にした場合、親映像の字幕との重なりが問題になる場面が多くある。この問題に対処するために、親映像の画面外に字幕表示のための専用領域を設ける表示方法が考えられる。この表示法を含めて、さまざまな字幕表示方法で実験番組を制作し、評価実験を行うことによって適切な字幕提示法を求めることもシステム技術の研究テーマの一つとした。

以上の研究項目を達成した暁には、図１に示すような自動字幕放送制作システムが完成する。左から電子化原稿が入力され、自動要約、自動字幕画面制作（自動字幕化）を経て字幕文が作成される。次いで、自動同期によって音声と字幕が同期され、最終的に字幕付きテレビ番組が完成する。自動同期は字幕の多重を行うタイミングを決定する認識フェーズと、認識フェーズで得られたタイムコードに従って実際に字幕を送出する同期フェーズの２つのパスで構成される。

４．研究の状況

４．１　自動要約の研究状況

従来の自動要約の研究は、大きく３種類に分類できる。第１の方法は、言語理解とそれに続く言語生成による要約であり、第２の方法は、文間の関係や段落構造などを利用して、要約するものである。これらの方法は、テレビニュースに適用するのは困難である。第３の方法は、重要語が集中する部分を重要部分と判断し、その部分を切り出すことで要約を行うものであり、広範な文章にロバストに適用できる。そのため、少なくとも現時点では、第３の方法を中心に研究を行っている[3]。この方法で、切り出す単位を文としたものは重要文抽出法と呼ばれ、広く一般に用いられている。本研究でも、重要文抽出法を用いている。第３の方法で、抽出単位を語（形態素）や連語にしたものを形態素単位文字数圧縮法あるいは単に、文字数圧縮法と呼んでこれを用いている。ただ、この場合は、重要な部分を抽出するというより不要な部分を削除するといった方が適切である。現在までに、文字数圧縮法によって約９０%、重要文抽出法によって約８０%の圧縮を行い、全体で７０%の圧縮を可能にしている。

４．２　自動同期の研究状況

次に、自動同期について述べる。自動同期は、番組音声とその原稿をもとに、原稿を話しているタイミングを検出し、ひいては字幕を送出するタイミングを検出する技術である。これに対応できる検出手法として、ワード列(単語列)のペアを用いたワードスポッティング法であるワード列ペアモデルを考案した[4]。ワードスポッティングの手法として、フォワード・バックワードアルゴリズムにより単語の事後確率を求め、その単語尤度のローカルピークを検出する方法が提案されている。ワード列ペアモデルは、これを応用して同期点の前後の複数の単語（ワード）を連結し、そのワード列の中点で尤度を観測して、そのローカルピークを検出する方法である。ワード列は、音素 HMM の連結により構成され、同期点のワード列以外を吸収するガーベジ部分は全音素 HMM の並列な枝として構成されている。また、アナウンサーが原稿を読む場合、内容が理解しやすいように息継ぎの位置を任意に定めることから、ワード列間にポーズを挿入している。実験から、SN 比が 20dB 以上あるような背景音が小さい場合には、沸きだしがほとんどなく検出率は９５%を得ることができた。これは、実用上、十分な精度で

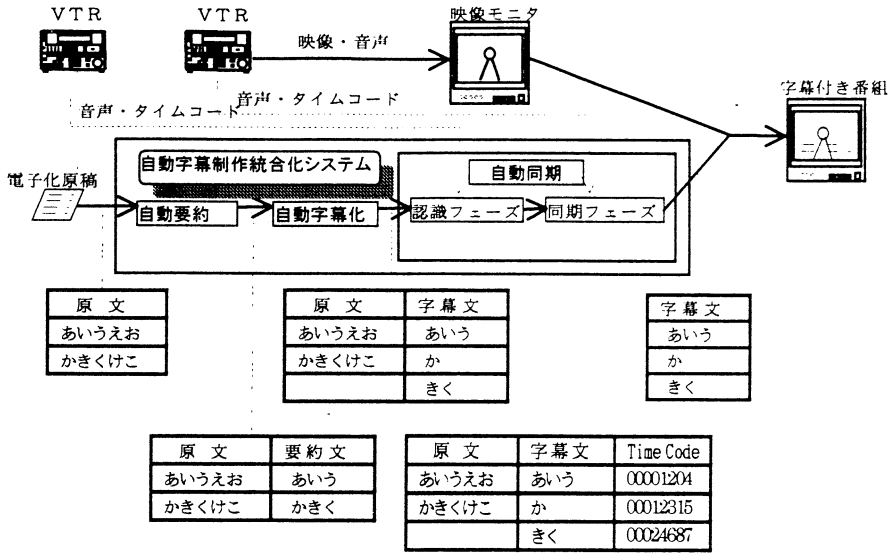

図１　自動字幕番組制作システム概念図

自動同期が可能であることを意味している。

４．３　システム技術

システム技術の研究では、自動要約、自動同期、自動字幕画面制作の各要素技術を統合化し、自動字幕番組制作システムを構築するとともに、様々な字幕表示方法の中でどれが好ましいかについて聴覚障害者などの協力を得て、主観評価実験を行った[5]。

自動字幕番組制作システムとして、２つのシステムを構築した。シミュレーションシステムとプロトタイプシステムである。前者は、自動字幕番組制作のための要素技術である自動要約、自動同期、自動字幕画面制作の３つの技術が完成したと仮定して、字幕制作を自動的に行うものであり、プロジェクトの１年目と２年目に作成した。次に、プロトタイプシステムは、上記３要素技術のプロトタイプの完成を受けて、平成１０年度に作成したもので、自動要約と自動同期はワークステーション上で実行するものである。

プロトタイプシステムは、今後の新しい字幕制作システムの構築に向け、字幕の提示方法や字幕制作システムの総合機能・性能の検証とその評価を行うことができ、シミュレーションシステムの主要ハードウエアと実験用字幕制作技術を中核とし、高機能化、高画質化を図った。要素技術である自動要約システムと自動同期システムとを有機的に統合化し、目標としていた字幕制作の全プロセスをほぼ自動的に行い得るものである。平成１１年度は、上記プロトタイプをベースにして、ヒューマンインターフェースの改善や可搬性の向上を行い、実用に近いシステムとして実証モデルを作成している。

最適な字幕提示方法を求めるために、聴覚障害者を含む多数の評価者の協力を得て、種々のパラメータを有する字幕提示方法による、実験用字幕番組の評価を行い、字幕提示方法に関する多数の貴重な主観評価データを得た。

評価実験は、主要な字幕パラメータを組合わせた３８種類の字幕形式、実際に放送された５種類のニュース番組を素材として、計１９０本の実験用字幕番組を制作し、健聴者３６名と聴覚障害者３５名（聾者と中途失聴者の割合は約半分）の協力を得て行った。評価項目は、「字幕の見やすさ」、「映像の見やすさ」、「全体の良さ」の３項目で、それぞれについて５段階で評価してもらった。図２に評価画面の例を示す。評価の結果、字幕専用領域を設けた提示方法の優位性が示されるなどの結果が得られた。

５．今後の課題

５．１　自動要約

重要文抽出と形態素単位文字数圧縮によって、ニュース記事に対して一応７０％の自動要約が可能になっている。しかし、重要文抽出は文単位の要約であるため、字幕作成の目的には粗すぎる場合がある。一方、形態素単位の文字数圧縮では、今後の改良を加えたとしても圧縮率が８５％程度が限界であると予想される。そこで、現在、文節単位の文字数圧縮法を検討している。これは、文と形態素の中間の単位である文節に着目して、重要度の低い文節を削除する手法である。しかし、もし削除する文節に係っている文節が削除されずに残ると係り先のない文節が存在してしまい

字幕専用領域なし

字幕専用領域あり

図２　実験用字幕制作例

文の意味が不明となる。つまり、文節単位文字数圧縮のためには、形態素解析に加えて係り受け解析を利用する必要があり、高精度な係り受け解析がキー技術の一つとなる。

さらに、ニュース以外の番組に自動要約を適用することも今後の重要な課題であり、そのためには、各種番組の音声と字幕をデータベース化し、どのような要約が行われているかの基礎調査から行う必要がある。

５．２　自動同期

ワード列ペアモデルを用いてSN比が20dB以下である低レベルの背景音であれば、十分な精度で自動同期が可能であることが分かった。今後の課題として、背景音が大きい場合に対応させることがあげられる。そのために、ワード列ペアモデルとビタビ照合を組み合わせたハイブリッドモデルやワード列ペアモデルを直列に接続した時間軸モデルを検討している。自動同期においてもニュース以外の番組に適用することが課題であり、一部ドキュメンタリー番組については研究を開始している。さらに、現在の自動同期システムは処理速度がリアルタイムの７～８倍を要しており、高速化も課題である。

自動同期のベースとなる音声処理技術は音声認識と共通する部分が多く、音声認識を利用した番組音声の自動文字化の研究も自動同期の発展形として考えられる。

５．３　システム技術

自動字幕番組制作システムは、シミュレーションシステム、プロトタイプシステムを経て、現在実証モデルを作成している。実証モデルはパソコンをベースにして可搬性の良いものとなっているが、実用的なものとするには、まだ課題がある。その一つとして現在の手作業での字幕制作・試写システムと統合化があげられる。本プロジェクトの残りの期間で、手作業でのシステムと粗結合による統合化を進める予定だが、実用機とするには、両者を一体化することが望まれる。

字幕表示方法の主観評価実験に関する今後の課題としては、ハイビジョンでの望ましい字幕表示方法の検討があげられる。ハイビジョンやデジタル放送での字幕システムのあり方は、放送システム全体に関わることであり本プロジェクトのみでなく広範囲での検討が必要となる。

６．まとめ

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研究状況と課題について報告した。研究用に放送番組を使用することを許可いただいた日本放送協会および日本テレビ放送網（株）に感謝する。

参考文献

[1] 江原暉将、沢村英治、若尾孝博、阿部芳春、白井克彦：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、Mar., 1997。

[2] 若尾孝博、江原暉将、沢村英治、丸山一郎、白井克彦：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、言語処理学会第４回年次大会、Mar., 1998。

[3] 福島孝博、江原暉将、白井克彦：文単純化のための文字数圧縮規則、言語処理学会第５回年次大会、Mar., 1999。

[4] 丸山一郎、阿部芳春、江原暉将、白井克彦：ワード列ペアモデルによる字幕送出タイミング検出の検討、電子情報通信学会音声研究会、June, 1998。

[5] 門馬隆雄、沢村英治、江原暉将、白井克彦：多様な提示方法の実験字幕番組の自動制作手法と字幕評価実験概要、映像情報メディア学会マルチメディア情報処理研究会、May, 1999。