ユーザプロファイルに応じた字幕表示システム

Closed Caption System using User Profiles

高 尾 哲 康
Takao Tetsuyasu

1.　はじめに

ブロードバンドインターネットの普及により動画や音楽などストリーミング配信するコンテンツが増加しつつある[1]。筆者は「ストリーミング配信動画向きの字幕表示システム」において、ストリーミング配信動画をユーザ好みのスタイルの字幕付きで配信・閲覧できる環境を実現するシステムを構築した[2][3]。その後、字幕表示クライアントシステムの評価および改良を行なうために、「字幕放送に関する調査」として主に聴覚障害者を対象とするアンケート調査を実施した[4]。そこで得られたさまざまな知見をもとに、字幕として表示するテキストの基本スタイル(フォントの種類、サイズ、色、スタイルなど)を変更したり、また、ドラマなどの場合に、主役、準主役、その他の脇役など、ロール(役、登場人物)によってそれぞれ別々に字幕フォントスタイルの指定ができるようにシステムの改良を行なった[4]。これらは、ストリーミング配信用字幕ファイル内で XML 形式のタグを事前に付けることにより実現した。さらに、字幕放送や既存の字幕付き動画配信システムからストリーミング配信用字幕ファイルを容易に作成できるように、ストリーミング配信動画用字幕作成支援システムを開発した[5][6]。

アナログ放送時代においては、字幕放送番組が提供するクローズドキャプションとしての字幕を表示するには専用のアダプタを導入する必要があった。デジタル放送時代を迎えた現在では字幕表示機能が受像器の標準機能として提供されるようになった。しかし、アナログ放送時代の字幕表示方式をそのまま移行したため、依然として字幕放送の字幕の表示位置やフォントの種類・色などはユーザ側で変更することができない(アナログ放送とデジタル放送の受像器のハード的な違いにより若干の違いはある)。また、ニュース番組やトーク番組など、生放送で行なわれるリアルタイム字幕(ライブ字幕)では映像と字幕のずれが生じているだけでなく、話し言葉をそのまま字幕にしたものが多く、しかも要約がほとんどなされていないために表示される字幕テキストが多すぎて読むのが困難となり、視聴に集中できないのが現状である。字幕テキストとしての表層面だけでなく、重要語や聞き間違いやすい語が含まれている部分をわかりやすく表示したり、字幕の「ななめ読み」をしやすくするなど、音声か字幕かの二者択一にとどまらず、視聴支援としての字幕の活用にも焦点を当てることにした。

本研究では現在の字幕放送の字幕表示の特徴を調査し、より見やすく読みやすくしたり、聞きづらさなどの多様なユーザの特性に合わせた柔軟な字幕表示方式を検討し、実際にシステムに組

開始時刻	終了時刻	時間間隔	ミリ秒	文字数	行数	表示文字列
59:50.2	59:54.3	00:04.1	4146	24	2	》本場大阪のお好み焼きはやっぱり一味違いました。
59:54.3	00:04.1	00:09.8	9814	16	1	藤》中に入れないでかけるんですね
00:04.1	00:07.3	00:03.2	3195	27	2	堤》ほとんど大阪の基本のお好み焼きは混ぜ込むんですが、
00:07.3	00:10.3	00:02.9	2944	24	2	強調したい具は挟んだり上にかけたりするんですね。
00:10.3	00:23.1	00:12.8	12809	17	2	早速丸いお皿のジャガイモ入りから。
00:23.1	00:26.1	00:03.0	3004	26	2	藤》ジャガイモをバターで炒めた段階で既においしそう。
00:26.1	00:28.5	00:02.4	2403	13	1	堤》ジャガバターにした後で
00:28.5	00:31.5	00:03.0	3005	8	1	挟み込みました。
00:31.5	00:36.8	00:05.3	5337	19	2	堤》周りがフンワリとしてますでしょう？
00:36.8	00:39.4	00:02.5	2534	16	1	藤》ちゃんとおイモの食感が残って
00:39.4	00:41.8	00:02.4	2404	14	1	オイモが、おいしいおいしい！
00:41.8	00:44.3	00:02.5	2533	19	2	堤》これ大阪で大人気なんだそうですよ。
00:44.3	00:47.1	00:02.8	2804	18	2	岡》スペイン風オムレツみたいですね。
00:47.1	00:51.9	00:04.8	4807	12	1	堤》ちょっと近いですね。
00:51.9	00:55.2	00:03.3	3335	30	2	お家にジャガイモは常備されていると思いますのでボリュームアッ
00:55.2	00:57.0	00:01.7	1732	3	1	プに。
00:57.0	00:59.6	00:02.6	2604	17	2	藤》子どもたち喜びそうです、これ。
00:59.6	01:02.3	00:02.7	2734	22	2	堤》食べ盛りのお子さんにはいいと思いますよ。
01:02.3	01:04.7	00:02.3	2343	11	1	そして、シソ入りです。

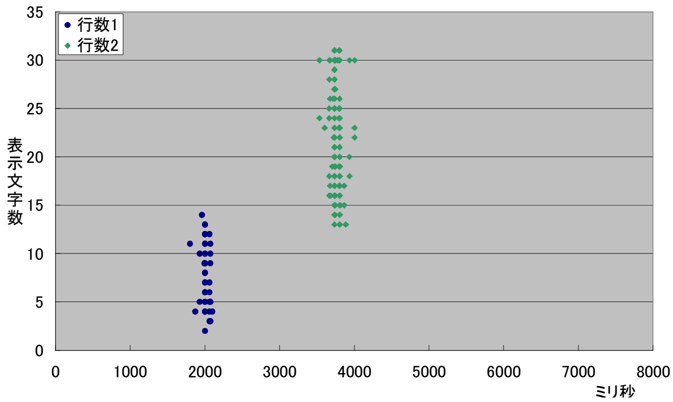
図１．字幕テキストの記録例

み込むことにより利用者が評価できる環境を実現した。

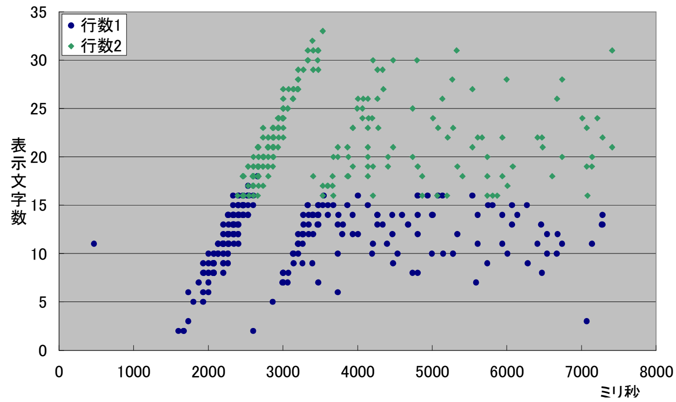
2.　現在の字幕放送における字幕表示の特徴

現在の字幕放送において聴覚障害者向けに付加されているクローズドキャプションとしての字幕がどのようなタイミングでどのように表示されているかを調査した。字幕が表示・消去された時刻をミリ秒単位で字幕テキストとともに記録するシステムを実現した。実際に記録された例を図１に示す。図において、表示文字列は複数行にわたって表示された字幕を１画面分としてまとめたものである(データとしては 1 行目、2 行目とそれぞれ別々に記録している)。なお、字幕の１行分は全角文字 15 文字相当であるが、事前収録字幕では英数字や記号、カタカナ文字には画面表示上は半角相当文字として表示されるものもある。その一方でリアルタイム字幕では入力作業の時間的制約があるために半角文字が使われることはほとんどない。オープンキャプションとしての字幕では多くは項目としての表示であるが、その文体が文章であっても句点が使われることはなく、また読点もほどんど使われないが、クローズドキャプションとしての字幕では普通に使われている。また、表示行数は 1 行または 2 行表示がほとんどであるが、場合によっては 3 行表示となることもある。ドラマなどの事前収録字幕では文字色として、主役は黄色、準主役はシアン、司会者や解説者などは緑色表示となる。リアルタイム字幕では文字色は白のみであり、背景色は黒が多いが時には字幕テキストの背景の映像が見えるようにするためにグレーの半透明となることもある。

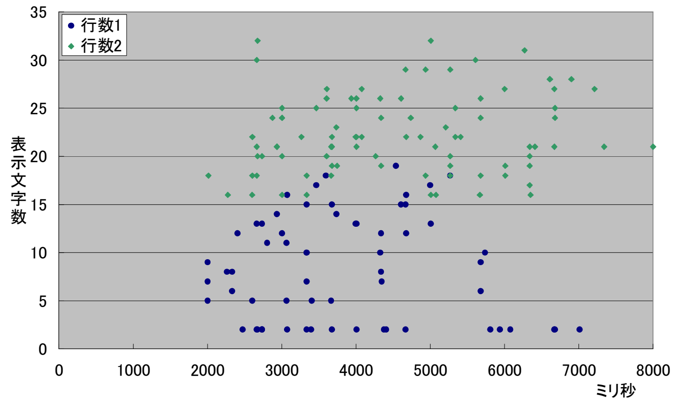
１画面分として表示される字幕の表示継続時間を測定した結果を図２に示す。(a)は NHK の「正午のニュース」のリアルタイム字幕、(b)は TBS 系列のトーク番組「はなまるマーケット」のリアルタイム字幕である。NHK のリアルタイム字幕では、1 画面内の字幕が 1 行の場合は 2 秒間、2 行にわたる場合は倍の 4 秒間と、1 行あたりの文字数の多少に関係なく決め打ちとなっている。そのため、視聴していて 2 文字程度の短い字幕がなかなか消えない時があったかと思えば長い字幕がすぐに消えてしまうような印象を受け、視点移動がやりづらくなり、その結果読みづらいものになってしまっている。一方、TBS 系列のリアルタイム字幕では、字幕テキスト長に応じて柔軟



(a) リアルタイム字幕(NHK 正午のニュース)



(b) リアルタイム字幕(TBS はなまるマーケット)



(c) 事前収録字幕(NHK 朝ドラ)

図２．字幕の表示継続時間

な表示時間となっている。これが視聴者にとって比較的読みやすくなっている理由のひとつである。後述の字幕表示クライアントで字幕を表示する際には、字幕文字数・行数によって表示継続時間を調整するようにしている。また、(b)で右上･左下方向にプロットがない谷間が見られるが、これはトーク番組において、話者の切り替えのタイミングに関係し、一人の話者が話し終えて次の話者に切り替わる間(ま)が現われていることがわかった。視聴者側への配慮として、あいさつ文や感嘆詞などを除いて複数人が同時に話すようなことがないようにしているためと考えられる。(c)はドラマなどの事前収録字幕である NHK 朝ドラの場合の字幕表示継続時間を測定したものであり、ドラマのセリフらしく字幕の文字列数とは全く関係がないことがわかる。また、事前収録字幕にかかわらず、字幕文字数が 2～3 文字で表示継続時間が数秒以上のものがあるが、これは BGM を表わす「♪～」や電話が鳴っていることを表わす電話マークのように、主に非言語テキストの場合である。

3.　字幕配信用ファイルの自動生成

字幕配信用ファイルを容易に作成するために、字幕放送の字幕テキストを字幕配信用ファイルに自動変換するシステムを実現した(図３)[6]。字幕放送やストリーミング配信動画の他形式のフォーマットの字幕テキストを字幕表示時刻情報(日本時間)と字幕テキストのペアを自動的に取り込み、本システムで利用可能な SMIL フォーマット字幕ファイル[9]に自動変換する。これにより、既存資産の再利用が容易に可能になっている。なお、SMIL フォーマットファイルの字幕表示時刻は、映像の先頭からの経過時刻となる。さらに、リアルタイム字幕の場合は映像と字幕表示とに

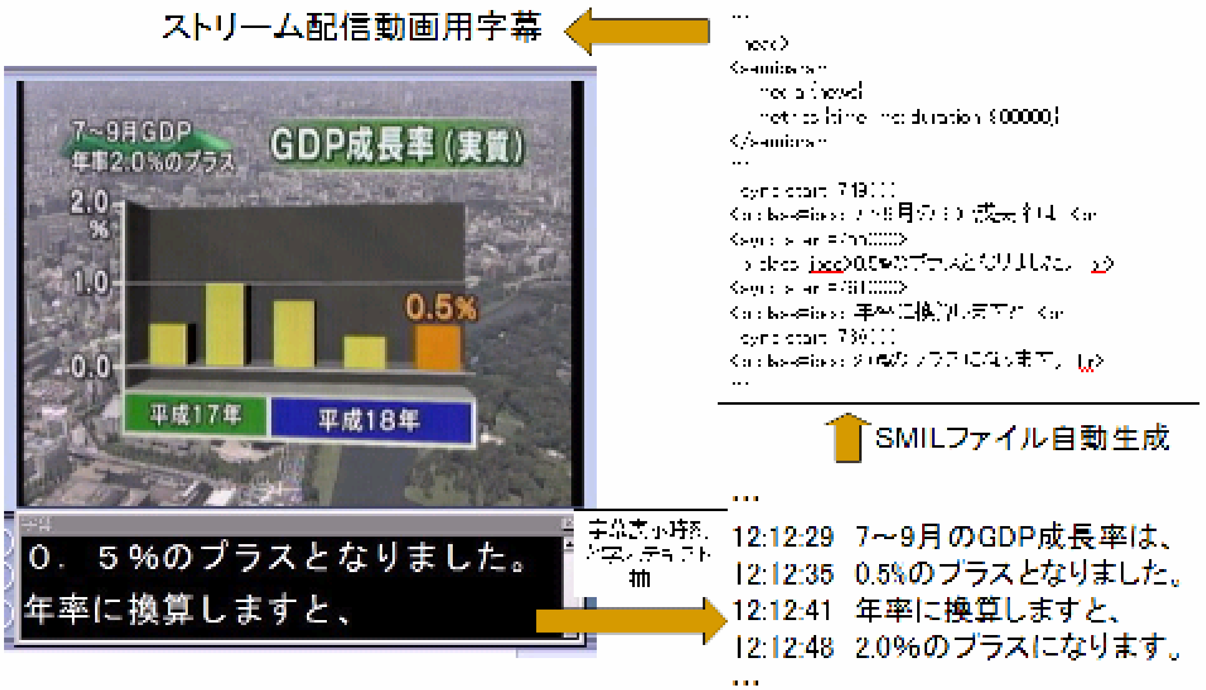


図３．ストリーム配信動画用字幕の自動生成

ずれが生じるので、これをパラメータで補正できるようにした。音声認識を利用した音声と字幕テキストとのアライメント機能は実現していないので、現在のところは番組ごとの経験則により時間ずれを固定パラメータで補正している。さらに字幕表示クライアント側でも微調整を可能にしている。

ストリーミング動画配信用 SMIL フォーマット字幕ファイルの自動生成には Perl 言語を利用して time-tag 付き字幕テキストファイル(日本語、英語などの言語ごとに)から SMIL(Synchronized Multimedia Integration Language、スマイル)仕様ファイル[9]や他形式字幕ファイル(SAA, SRT, LRC)に変換する。SMIL は W3C にて標準化[10]している映像・音声・画像・テキストなどのさまざまなメディアのレイアウトやハイパーリンク、再生タイミングなどの設定を行なうためのマークアップ言語である。特徴として、枠組みは XML をベースにしており、映像、音声、画像、テキストなど複数のメディアを Web ページのようにレイアウトしたり、各メディアの表示タイミングのコントロールやハイパーリンクを各メディアに結びつけることが可能になっている。

本システムでは、特に字幕テキストに言語処理を施すことで得られるさまざまな情報を XML のタグとして付加することにより字幕表示クライアント側でより柔軟なオプションが利用できるなど、さらなる発展性が得られた。

4.　ユーザプロファイルと言語処理を利用した字幕表示システムの開発

システムの開発においては閲覧にブラウザを利用するため、ユーザインタフェース部分には Ajax(HTML、CSS、JavaScript)、ActiveX コントロールを利用した。柔軟な字幕表示を実現するために字幕テキストに対して形態素解析、構文解析などの言語処理機能を積極的に導入することにより、読み、発音、品詞、係り受け関係などの情報を XML 形式のタグ付けを行なった。柔軟な表

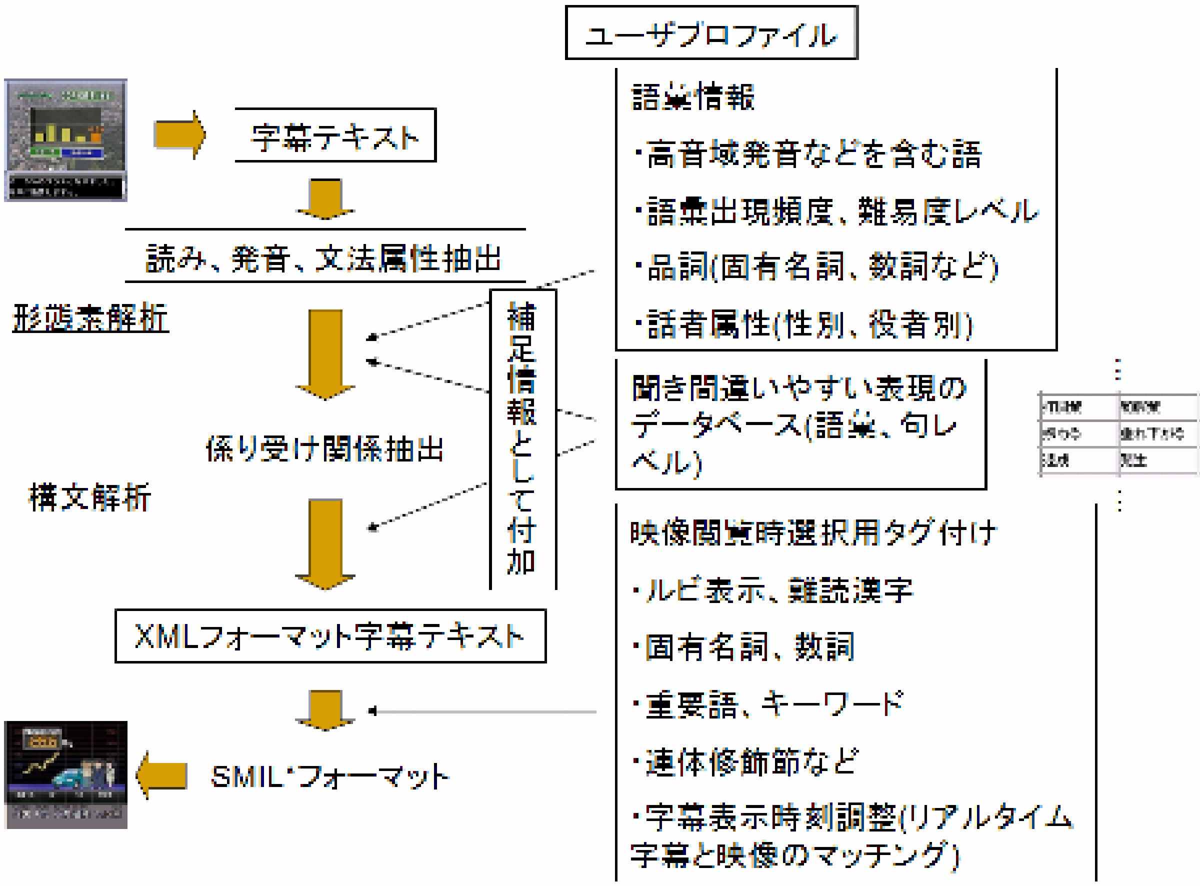


図４．ユーザプロファイルと言語処理を利用した字幕表示

示機能オプションはユーザプロファイルとして、字幕を表示する話者の選択や固有名詞や数詞などの品詞選択、聞き間違いやすい語彙や句をデータベースとして利用、また視聴者の聞こえ方の特性に合わせるなど、さまざまなオプションを設定・保存できるようにしている。字幕テキスト全体としてはXMLフォーマットをベースとする配信用SMIL(スマイル)仕様字幕テキストをリアルタイムに作成できるようにした。開発したシステムの概要図を図４に示す。

5.　字幕表示クライアントシステム

字幕表示のユーザインタフェースの部分の作成にはHTML、XML、CSS、JavaScriptとActiveXコントロールを利用して開発した。ブラウザ内に表示した例を図５に示す。映像領域の下側に横書き字幕表示領域、左側に縦書き字幕表示領域がある。その下には字幕表示に関するさまざまな設定や設定の保存、設定の反映が行なえるポップアップメニューやボタンがある。コントロールパネルを「あり」に選択すれば、動画の再生位置スライドバーや音量スライドバーなどのコントロールパネルを表示する。PCの全画面に表示したいときは、動画再生中に動画領域サイズから全画面表示を選択する。再生終了時には全画面表示が解除されもとのブラウザ画面に戻る。途中で中断したい場合はESCキーを押すか、全画面終了ボタンをクリックする。

字幕テキスト表示をクリックするとポップアップウィンドウに字幕テキストが表示される。ここで、字幕テキストをクリックすることで再生中の動画の頭出しができる。また、字幕テキスト

図５．字幕表示クライアント

の全体が閲覧できるので、文字列検索により見たいシーンを検索することも可能である。

ロール(役、登場人物)によって字幕フォントスタイルが自由に設定できる機能もすでに実現している。デフォールトでは、現在の字幕放送で慣例となっている、主役：黄色、準主役：シアン、その他の脇役および基本字幕：白色になっており[11]、フォントの種類、スタイル(ノーマル、イタリック、ボールド、縁取り色など)、サイズは基本字幕に設定した属性を継承している。

さらに、字幕テキストに対して言語処理を行なうことで可能になった機能として、ルビ表示、固有名詞や数詞部分の表示、聞き間違いやすい語の表示、意味的な最小のひとまとまりである連体修飾句の表示などがある。図５ではルビ付き、数詞表示によって重要な情報を優先的に表示している例である。音声だけでなく字幕を補助的に利用することで理解性を高めることができる。字幕テキストは外国人の日本語学習にもよく利用されるが、ルビが付かないリアルタイム字幕であっても漢字にルビを付けることで学習効率をさらに高めることができる。字幕テキストを言語処理することでルビを容易に付けることができるが、「黄金」(おうごん、こがね)など、複数の読み方を持つ語彙があることや文として読み方が変わる場合もあるため、品質向上のためには音声認識を利用したり、構文解析・意味解析を利用する方式を検討する必要がある。

各種オプションの指定は、XML フォーマットのテキストから経過時刻情報と字幕テキストのペアを行単位でもつ字幕テキストファイルに変換する際に、ブラウザ内に表示するドキュメントオブジェクトのインライン要素の一部に CSS のクラス指定をすることで容易に設定できるようにしている[12]。

その他のオプションとしてはリアルタイム字幕用として字幕表示時刻の微調整が可能になっている。これも本来ならば、音声と字幕テキストとのアラインメントを行なう必要があるが、計算機の処理能力などの制約から現在はタイムシフトのみとなっている。

また、各種オプションについて字幕フォントの属性値(色、種類、スタイル、縁取り色、サイズ、表示・非表示など)やルビ付きなどさまざまな属性がユーザによって変更されると、動画再生中であっても即座にその反映されるようになっている。

6.　まとめ

ストリーミング配信動画をユーザ好みのスタイルなど多くのプロファイルごとに字幕付きで閲覧できる環境を実現した。ロール(役、登場人物)ごとに字幕フォントスタイルを任意に指定できるようにし、性別、役者等の特定話者のみの字幕を表示できる機能、聞きづらい語やフレーズ、重要語、固有名詞、数詞などを優先的に選択表示する機能、意味的なひとまとまりの句として連体修飾関係にある部分を表示する機能などが実現できた。さらに、既存のリアルタイム字幕放送から得られた知見から、字幕文字列長や表示行数を考慮し、タイムシフト(映像・音声データをハードディスクなどにバッファリングすることにより再生位置をずらす)を利用した映像とリアルタイム字幕の簡易マッチング再生も実現した。

要約筆記においては「ことばを追うな。意味を追え」と言われるように、話しことばを単に文字化するのではなく、利用者が読みやすいように意味主導型の要約筆記へと発展することが望まれている。しかし、現在の計算機処理では意味処理を十分に行なえるまでには至っていないが、ある程度の言語処理機能を導入することでもユーザごとの特性を生かしたよりコンパクトな字幕を表示することで読みやすさが格段に向上することが可能になりつつあると考える。

この字幕表示クライアントシステムの実現には一般的なブラウザが提供する JavaScript、スタイルシートなどを利用しているので、適用範囲は広い。ワンセグを含めてデジタル放送時代には動画だけでなくさまざまなデータ放送の内容を表示させたり、双方向機能を利用したりする機会が増えてくると考える。そのような時には TV も一種のブラウザを通して閲覧するようになる。本研究で試作した機能なども容易に適用できるので、今後の発展性は高いものになると期待される。

リアルタイム字幕と映像・音声の対応については、現在は計算機処理速度の面から、映像・音声のタイムシフトを利用し、クライアント側で微調整できるようにする方法で実現したが、今後は音声認識を利用するなど、不完全なデータ(リアルタイム字幕には入力誤りや中断、一部省略、極端な要約などが入る割合が高い)に基づくアラインメント方法(対応のマッチング)など、より高品質化に向けた課題も残されており、今後の研究が期待される。

本研究の一部は(財)放送文化基金およびユニベール財団の助成によって行なった。また、動画コンテンツの研究・調査への利用については、NHK マルチメディア局・著作権センターなどから許諾をいただいた。

参考文献

(1) 櫻井智明、平 明弘、実践！ブロードバンドストリーミング、オーム社 (2002)
(2) 高尾哲康、ネット配信動画向きの字幕表示システム、平成 17 年度電気関係学会北陸支部連合大会, E-35, (2005)
(3) 高尾哲康、ストリーミング配信動画向きの字幕表示システム、第 4 回とやま産学官交流会ポスターセッション、PB-15, (2005)
(4) 高尾哲康、ストリーミング配信動画向きの字幕表示システム、富山国際大学地域学部紀要、

pp.127-134, Vol.6, (2006)
(5) 高尾哲康、ストリーミング配信動画用字幕作成支援システム、第 5 回とやま産学官交流会ポスターセッション、PB-09, (2006)
(6) 高尾哲康、ストリーミング配信動画字幕表示システムの改良、富山国際大学地域学部紀要、pp.87-93, Vol.7, (2007)
(7) 香取淳子、高尾哲康、高齢者・聴覚障害者からみた字幕表示のあり方に関する研究、ユニベール財団研究報告, 2007
(8) 高尾哲康、ユーザプロファイルに応じた字幕表示システム、第 6 回とやま産学官交流会ポスターセッション、46, (2007)
(9) Dick C. A. Bulterman, Lloyd Rutledge, SMIL 2.0, Springer-Verlag (2004)
(10)http://www.w3.org/AudioVideo/
(11)Shirai, et al.: Program and Proceedings of TAO WORKSHOP on TV Closed Captions for the hearing impaired people, pp.9-29, (1999)
(12)水津弘幸、石井　歩、HTML+CSS HANDBOOK、C&R 研究所、ソフトバンクパブリッシング、(2003)