

ストリーミング配信動画字幕表示システムの改良

Revised Closed Caption System for Streaming Video

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について述べた[2][3][4]。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(<http://www.tuins.ac.jp/~takao/cc/ccr.cgi>)、さまざまな知見を得ることができた。これまでの字幕表示クライアントシステムでは、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)は1種類に固定であった。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)によってフォントの色が固定されていた。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ていた。なお、字幕放送の場合のフォント色については、主役は黄色、準主役はシアン、その他の脇役は白色とすることが慣例となっており、これらは番組制作側で決定しているので視聴者側で変更することはできない。

本研究では前述のアンケート調査の結果をふまえ、字幕表示クライアントシステム側でロール(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試みた。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[5]を開発した。

2. 「字幕放送に関する調査」の結果

「字幕放送に関する調査」の全26個の質問のうち、「字幕放送に望むこと」の回答結果を図1に示す(有効回答数64)。ニュース番組、スポーツ番組などの生放送番組やCM、番組予告など、字幕を付ける番組の充実を望む回答が多かっただけでなく、字幕とテロップが重ならないようにするとともに、字幕フォントなど字幕を見やすくする配慮に関する期待・要望も少なからずあった。

自由回答においても、「登場人物の音声ごとに字幕の色が違うとわかりやすい」、「ユーザが自分で字幕表示位置、文字フォントスタイルなどを調整できるとよい」、「発言する人の字幕を区別できるように」、「字幕のレベルを選べるように」などの意見がいくつか見られた。これら結果を受け、

字幕表示クライアントの機能として、ドラマなどの場合に役者やロール(役、登場人物)に応じて字幕フォントスタイルをユーザ側で別々に独立して設定することができるように改良した。

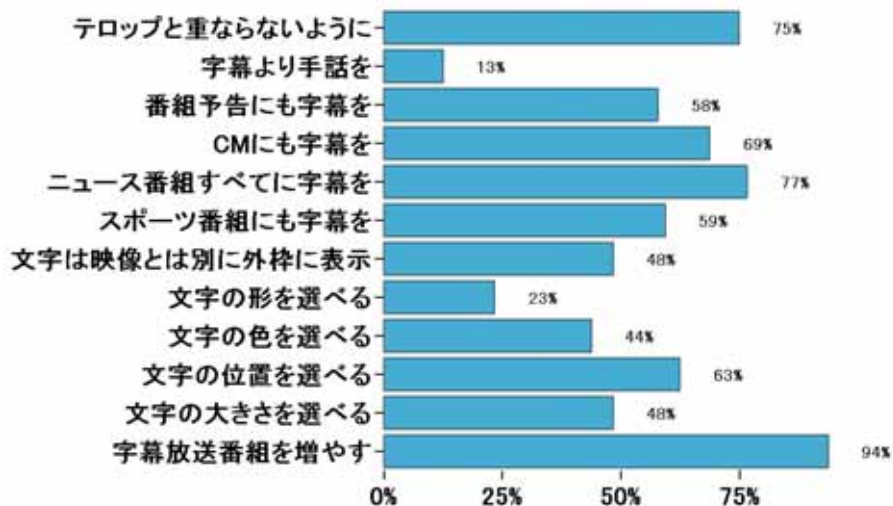


図1 字幕放送に望むこと

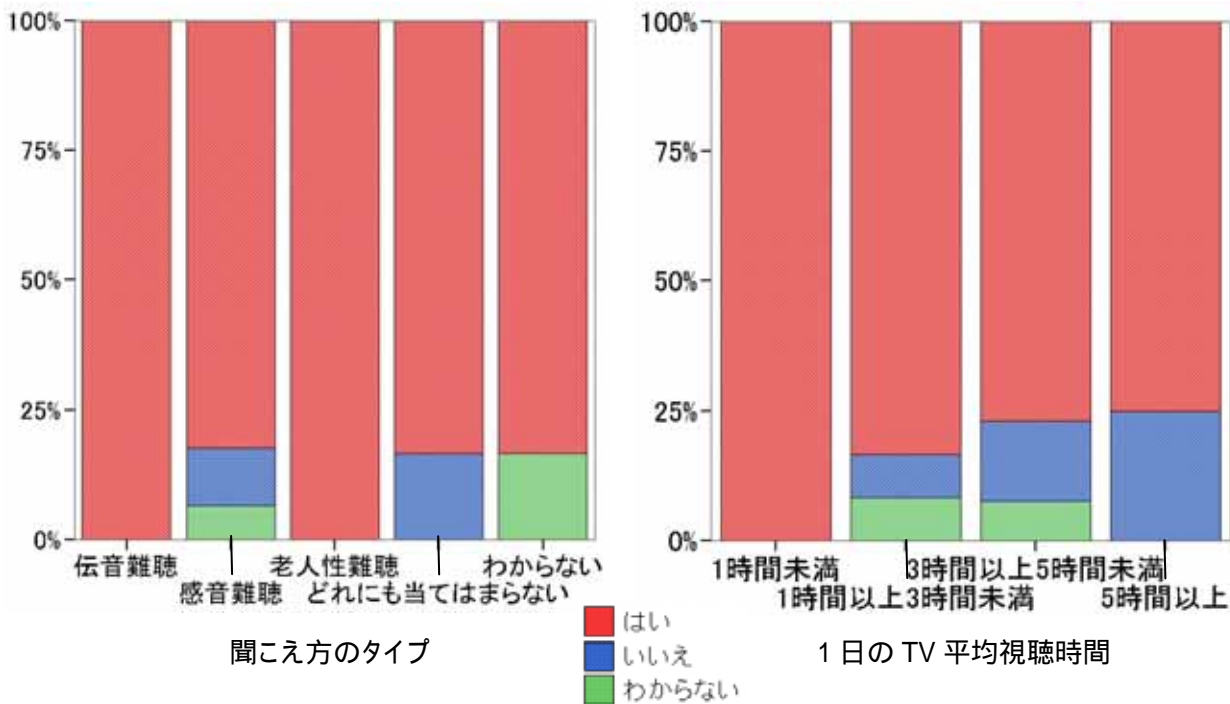


図2 「映像」と「ナレーションの字幕」のズレが気になるかどうか。

また、ニュース番組などのリアルタイム字幕について、「映像」と「ナレーション」のずれについて調査した結果の一部を図2に示す。聴覚障害者の聞こえの障害のタイプ別(伝音難聴、感音難聴、老人性難聴、その他)では、字幕による情報保障を必要とするしないに関わらず気になるとの回答が圧倒的に多い。その中で、聴覚障害の程度にもよるが、ナレーションの音声から情報を得ることが多いほど気になる割合が増える傾向にある。また、高度難聴者や聾者の場合(感音難聴あるいはどれにも当てはまらないのいずれかを選択することになる)は音声の認識自体が難しいた

め、ズレの有無を判断できるまでには至らないことがある。そのような場合でも映像とナレーション字幕の対応がとれていないと気になる割合が高くなる。

1日のTV視聴時間が1時間未満の場合は全員が映像とナレーション字幕のズレが気になるとの回答であったが、視聴時間が長くなるほどしだいにズレが気にならなくなるといった興味深い結果となった。これは短時間に集中して視聴するか、ながら的な長時間視聴など、バックグラウンド的な視聴をするかなど、視聴のしかたに関係していると推定できる。

字幕の表示方式[6]については、2行総入れ替え方式と2行上方向スクロール方式を採用したが、ドラマやドキュメンタリーなどでは2行総入れ替え方式、ニュースなどでは2行上方向スクロール方式が好まれる結果となった。これは、ドラマなどのセリフの場合は2行で一人分の発言として表示できることが多く、視聴する側としてはひとまとまりで把握することが容易なためと考えられる。ニュースなどのリアルタイム字幕では同時要約が難しく、アナウンサーのナレーションをほぼそのまま字幕化することが多く、その結果として字幕文字列数が多くなり、視聴者が読みこなすのに時間がかかるためと考えられる。



字幕表示時刻と
字幕テキストの抽出

```

...
12:12:29 7~9月のGDP成長率は、
12:12:35 0.5%のプラスとなりました。
12:12:41 年率に換算しますと、
12:12:48 2.0%のプラスになります。
...
    
```

```

<sami>
<head>
<samiparam>
  media {news}
  metrics {time: ms; duration: 900000;}
</samiparam>
...
<sync start=749000>
<p class=jpcc>7~9月のGDP成長率は、<br>
<sync start=755000>
<p class=jpcc>0.5%のプラスとなりました。<br>
<sync start=761000>
<p class=jpcc>年率に換算しますと、<br>
<sync start=768000>
<p class=jpcc>2.0%のプラスになります。
<br>
...
</sami>
    
```

SMIL ファイル自動生成

図3 字幕表示用ファイルの自動生成

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試作した(図3)。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[6][7]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た。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。

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[7]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[8]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。

4. 字幕表示クライアントシステムの改良

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4に示す。映像領域の下側に横書き字幕表示領域、左側に縦

The screenshot shows a video player interface. At the top, there is a video frame showing two people in a room. Below the video, there is a subtitle area with the text "実はなかなかつらい。 さようか。". Below the subtitle area, there is a control panel with various settings. The control panel includes buttons for "再生/一時停止" (Play/Pause), "停止" (Stop), and "コントロールパネル なし" (No Control Panel). It also has dropdown menus for "動画サイズ" (Video Size), "字幕域サイズ" (Subtitle Area Size), "背景" (Background), "字幕表示方法" (Subtitle Display Method), and "言語" (Language). There is a table for subtitle font settings with columns for "字幕フォント" (Subtitle Font), "色" (Color), "種類" (Type), "スタイル" (Style), and "サイズ(ポイント)" (Size (Points)). The table has three rows: "基本" (Basic), "主役" (Main Role), and "準主役" (Sub-role). Below the table, there are buttons for "現在の設定を保存" (Save Current Settings), "保存した設定を反映" (Apply Saved Settings), "設定消去" (Clear Settings), and "字幕テキスト表示" (Show Subtitle Text). At the bottom, there is a section titled "動画選択" (Video Selection) with a list of video titles and their durations.

字幕フォント	色	種類	スタイル	サイズ(ポイント)
基本	白	ゴシック	標準	20
主役	黄	丸文字	イタリック	22
準主役	シアン	UI Gothic	イタリック	28

動画選択

- 左端の○をクリックして、再生ボタンをクリックすると上記画面で再生します。
- タイトルを選択するとWindows Media Playerで再生します。字幕表示は横書きのみです。
- 1. 『利家とまつ』(字幕上方向スクロール版)(約3分)
- 2. 『NHKニュース21(2005.06.28)』(字幕上方向スクロール版)(約3分)
- 3. 『シリーズ世界遺産100 知床(2005.07.24)』(字幕上方向スクロール版)(約3分)
- 4. 『NHK手話ニュース845(2006.10.12): 万葉仮名』(縦書き横書き混在字幕)(字幕上方向スクロール版)(1分40秒)

図4 本システムによるロール別字幕フォント指定表示例

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ダーや音量スライダ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。

今回のユーザによる評価を受けて行なった改良はロール(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点である。デフォルトでは、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、フォントの種類、スタイル(ノーマル、イタリック、ボールドなど)、サイズは基本字幕に設定した属性を継承している。図4ではロールによって別々に設定して表示させた例である。ロールによって、フォント色、種類、スタイル、サイズがそれぞれ異なっていることがわかる。

ロールの指定は、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おいて、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。クラス指定は複数の要素に対して同じ属性を適用させることができる[9]。その記述例を図5に示す。ここでは、主役のクラス名はR1、準主役のクラス名はR2としている。とには含まれたテキストがそのロールの字幕となる。なお、基本字幕およびその他の脇役の字幕についてはクラス指定は不要である。これにより、理論上はクラス名をいくつでも指定できるので、全てのロールについてそれぞれスタイルを設定することも可能である。

動画再生中にクラス名R1、R2などが設定された要素がブラウザのドキュメント内に出現すると、あらかじめクラスごとに設定されたスタイルシートが適用され、その結果として望む結果が表示される。

各ロールについて字幕フォントの属性値(色、種類、スタイル、サイズなど)が変更されると、たとえ動画再生中であっても即座にそのロールのクラス名に対応

```

...
[00:08:500] 承りました。
[00:10:800] <span class=R1>それでよい。</span>
[00:12:800] お体は？
[00:15:700] <span class=R1>実は なかなか つらい。</span>
[00:17:700] さようか。
...
[00:43:684] <span class=R2>がっかりした様子で </span>
[00:45:738] <span class=R2>食も 細くなりまして...。</span>
[00:49:000]
[00:51:700] (はる)家康殿は いつ 来られるのですか？
[00:55:300]
...

```

図5 字幕テキストファイルにおけるロール指定

するスタイルシートがJavaScriptのcreateStyleSheet関数によって再定義され、この時点以降

に表示される字幕テキストのスタイルに自動的に反映される。

5. まとめ

主に聴覚障害者を対象とし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および閲覧できる環境を実現するシステムについて、ユーザの評価結果を受けてロール(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改良した結果について述べた。この改良された字幕表示クライアントシステムの評価については、現在「字幕放送に関する調査(その2)」として主に高齢者および聴覚障害者を対象としてアンケートを実施中である(<http://www.tuins.ac.jp/~takao/cc/ccr2.cgi>)。

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。このような時、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。

技術的に字幕付加が困難なリアルタイム字幕については、今後検討を行なっていく予定である。事前字幕が作成困難なニュース番組やスポーツ中継放送などにおけるリアルタイム字幕については、現在は音声認識を利用したリスピーク方式(アナウンサーやナレーターの発声を字幕キャストが要約して言い直したものを音声認識し、さらに認識誤り修正などの後編集を行なう)とスピードワープロを利用した高速オペレーター方式によって行なわれている。リアルタイム字幕と映像・音声の対応については、映像・音声のタイムシフト(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)を利用する方法が検討されているが、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)など課題も残されており、今後の研究が期待される。

なお、本研究の一部は(財)放送文化基金およびユニバーサ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) Dick C. A. Bulterman, Lloyd Rutledge, SMIL 2.0, Springer-Verlag (2004)
- (7) Shirai, et al.: Program and Proceedings of TAO WORKSHOP on TV Closed Captions for the hearing impaired people, pp.9-29, (1999)
- (8) <http://www.w3.org/AudioVideo/>
- (9) 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)

