

ストリーミング配信動画向きの字幕表示システム

Closed Caption System for Streaming Video

高尾 哲康

Takao Tetsuyasu

1. はじめに

近年のブロードバンド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]。リアルタイム通信としてはこれまでも IP 電話やインターネット放送・中継、ネットワークビデオ会議、VoD(Video on Demand)などが利用されてきた。しかし、これらのシステムではネットワークが輻輳しているなど通信帯域幅制限時には映像配信より音声配信が優先されるため、聴覚障害者や高齢者の一部には欲しい情報がなかなか得られにくいので不評であった。最近の現象としてはマスコミやポータル系以外でも個人で動画を発信するサイトが増加していることである。放送局系のサイトでは過去の TV 放送番組の配信を行なっているところもあるが、字幕放送番組であってもインターネット配信動画やストリーミング配信動画には字幕が付加されていないことがほとんどである。ここで、インターネット配信とは主に HTTP プロトコルを利用してダウンロードしながらプレイヤーで再生するやり方(ファストスタートともいう)である。またストリーミング配信とは MMS(Multimedia Messaging Service)プロトコルを利用してストリーミングサーバから配信することによりダウンロード完了か否かにかかわらず動画の任意位置から再生可能なやり方であり、長時間動画の再生に適している。なお、ここで述べる字幕(クロズドキャプション)とはアナログ放送用の字幕デコーダ装置や字幕表示機能が内蔵されているデジタル放送用チューナなどのアダプタを介することで表示するものであり、あらかじめ映像に埋め込まれて常時表示する字幕(オープンキャプション)とは別ものである。一部の外国映画では日本語字幕がオープンキャプションとしてつけられているものもあるが、この場合は日本語以外の音声となっている。一方、音声吹き替え版では日本語字幕が表示されないものも多い。このため、聴覚障害者はもとより、一部の高齢者にとっては視聴の楽しみがそこなわれる結果となっている。2006 年 4 月から開始予定の携帯電話向け地上デジタル放送の「ワンセグ」では、XML ベースの技術を用いて動画以外にもテキストなどのデータを放送できるので字幕表示についても期待されており、放送局によっては字幕放送の延長として行なわれる予定である。本研究ではこれらのニーズを背景に、インターネット接続環境と一般的な PC を利用して個人でも容易に字幕付き動画を配信・表示するシステム[2][3]を開発した。

2. 本システムの特徴

本システムの特徴を簡単にまとめると以下のようになる。

- 大多数が所有している標準的な PC 機器・ソフトウェア(Windows2000/XP または MacOSX、ブラウザは Internet Explorer または Firefox、必要に応じて動画再生ソフト Windows Media Player)のみで実現する。
- クローズドキャプション字幕の表示部分(字幕表示方式、フォントの種類、フォントサイズ、フォント色、背景色、字幕言語)は視聴者側で設定・選択でき、変更した設定を保存・反映できる。字幕表示方式については、聴覚障害者による評価が高い、2 行総入れ替え表示方式と 2 行上方向スクロール表示方式を採用した[4]。スクロール方式は字幕が新しく変わっているかどうか容易にわかるという利点がある。
- 基本的にウェブブラウザ上で字幕表示し、必要に応じて全画面表示にしたり、Windows Media Player などの外部プレイヤーでも字幕表示できる。
- チャンネルを切り替えるように、再生する動画を簡単に選択できる。
- 制作側は、動画と字幕テキスト(各行が字幕表示時刻(動画先頭からの経過時間、time-tag)と字幕文字列で構成されるテキストファイル)さえ用意すれば、W3C 推奨の SMIL 仕様(Synchronized Multimedia Integration Language)[5][6]字幕表示用ファイルをプログラムによって自動生成する。なお、字幕を消すタイミングは空文字列を表示することで行なう。
- 字幕テキスト内に HTML タグを含めることができる。これにより、ルビを付けたり、話者識別や発言の強弱、背景音などを表わすためのフォントの一時的な変更を行なえる(フォント色、フォントサイズ、スクロール表示、字幕表示位置など)。また、画像やリンクなどのタグも含めることもできるので、双方向機能として視聴者からの応答を受け付けたり、スポット広告などにも利用できる。
- 検索のポータルサイトである Google Video Search(<http://video.google.com/>)や Yahoo Video Search(<http://www.yahoo.com/>)など、クローズドキャプションを利用した動画検索用インデックス作成への利用が期待される。字幕ファイルには動画の経過時間情報が含まれているので、動画内の一部のシーンのみを対象として検索・再生することも可能である。

3. 本システム導入の背景

字幕放送の現状

現在の TV による放送では、公共放送と民間放送のいずれにおいてもデジタル放送、アナログ放送を問わず、字幕放送の割合が増えてきている。それにより字幕(クローズドキャプション)表示を考慮しないオープンキャプションが増加してきている。ニュース番組やスポーツ中継などのリアルタイム字幕を付ける場合は図 1 のように、オープンキャプションとクローズドキャプションが重なることが特に多くなっている。このような場合、視聴者は頻繁に字幕表示のオン・オフを繰り返す操作を強いられることになる。字幕表示方式・位置・色などは送信側で決めるために視

聴者側でコントロールできないことも字幕を利用しにくくしたり視聴にストレス感が残ったりする。なお、一部の事前収録番組では字幕とオープンキャプションが重ならないようにクローズドキャプションを表示する位置を変えたりするものもある。このような番組では字幕表示位置が頻繁に変わったり、出演者の顔に近い部分に字幕が表示されたりとかえって見苦しくなる場合もある。また、字幕言語は国内では日本語のみである。デジタル放送時代に入り、字幕に第2言語を選択できるようになったが、その他の多くの問題はそのまま残されている。本システムでは図1のように映像の外側の下側に字幕を表示させている。携帯電話向け地上デジタル放送の「ワンセグ」では、映像・音声とともに BML(Broadcasting Markup Language)で記述されたデータを放送でき、この中に字幕も含めることができるようになってきているので、映像領域とは別に字幕表示域を持つことが可能になっている。なお、この字幕表示機能については、携帯電話の機種や機能に依存する部分もある。最近の機種でも字幕表示機能は字幕放送の延長として表示できるようにしただけであり、字幕のフォントや色、スタイルなどは自由に変更できないようになっている。今後、字幕の重要性が認識されてくれば、ユーザ側で柔軟な設定ができたり、番組の録画の字幕テキストから見たい部分を検索できたり、字幕テキストをインデックスにして頭出しができるようになっていくものと予想される。これは、いわばウェブブラウザを通して放送を視聴していることになり、本システムでの字幕表示方式と非常に似たものとなっている。



改正著作権法（平成 12 年 12 月 8 日公布 平成 13 年 1 月 1 日施行）

第三十七条の二（聴覚障害者のための自動公衆送信）

聴覚障害者の福祉の増進を目的とする事業を行う者で政令で定めるものは、放送され、又は有線放送される著作物について、専ら聴覚障害者の用に供するために、当該著作物に係る音声を文字にしてする自動公衆送信（送信可能化のうち、公衆の用に供されている電気通信回線に接続している自動公衆送信装置に情報を入力することによるものを含む。）を行うことができる。

著作権法が改正され、視覚障害者向けの点字に関する第 37 条に聴覚障害者のための自動公衆送信に関する条項が追加された。これにより、聴覚障害者向けに音声を文字に変換して自動公衆送信することが制度面でも容易に行なえるようになった。なお、法令に定める事業者としては、聴覚障害者情報提供施設、全日本難聴者・中途失聴者団体連合会、全日本聾啞連盟、日本障害者リ

図1 オープンキャプションとクローズドキャプション

ハビリテーション協会、全国盲ろう者協会などがある。

聴覚障害者向きの字幕表示

字幕は主に聴覚障害者にとって重要な情報であるが、字幕を作成したり、表示のタイミングを決めたりするのはほとんどが健聴者である。音声やナレーションをベースにして決めているためである。そのため、どのような字幕表示が聴覚障害者にとってベストなのかはほとんど考慮されていないのが現状である。そこで、通常の日本語能力をもつ聴覚障害者に対してフィールド調査を行なった。その結果を下記に示す。

- 外国映画などの一般的な字幕(音声の主)

音声の開始より 0.1~0.2 秒早めに字幕を表示し、音声の終了とともに字幕を消すと流れを妨げない(文字情報は音声に比べ、出現してから脳で認識するまでに少し時間がかかる)。

「読んだ気がしない字幕」(字幕の存在を意識させない)がベスト。

- 聴覚障害者向きの字幕(字幕の主)

字幕と音声とのずれがあまり気にならないので、流れを妨げない程度に長く表示されたほうがよい(視点移動に余裕が欲しい)。出演者の顔を見て読唇しやすいように早めの字幕表示が好まれる。

リアルタイム字幕では読唇は困難になるが、映像と字幕内容がずれなければよい。

字幕表示内容の切替わりのタイミングが分かりやすい方がよい。従って、字幕入れ替え表示方式よりスクロール方式が好まれる。入れ替え時にブリンクしたり、色や明度を変えたりすることも良いかも知れない。また、横書き字幕の左右スクロール方式では慣れないと酔いやすい。

横書きを受け付けない高齢聴覚障害者などがいる。

4. 本システムの開発

字幕表示のユーザインタフェースの部分の作成には HTML、CSS(スタイルシート)、JavaScript、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 2 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(ブラウザのクッキーとして保存)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図3のようなポップアップウィンドウが表示される。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。

字幕表示時刻(動画先頭からの経過時間、time-tag)付き字幕テキスト作成には Sub Station Alpha、Subbなどの字幕作成ソフトや PC 用 MP3 プレイヤーの WinAmp 用プラグイン(もともとは歌詞の表示・入力用)などのフリーソフトを利用した。

ストリーミング動画配信用字幕ファイル作成には 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)から SMIL(Synchronized Multimedia Integration Language、スマイル)仕様ファイルや他形式字幕ファイル(SAA, SRT, LRC)に変換する

ことにした。SMIL は W3C にて標準化(<http://www.w3.org/AudioVideo/>)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。字幕表示方式(入れ替え方式、スクロール方式、縦書き、横書き)ごとに字幕ファイルが作成され、配信時に設定情報に応じて自動選択される。

ストリーミング配信用動画ファイルについては、動画ファイルと字幕ファイルを 1 つに結合(MUX: マルチプレクサ)する場合と動画ファイルと字幕ファイルを別々にする場合がある。1 つの

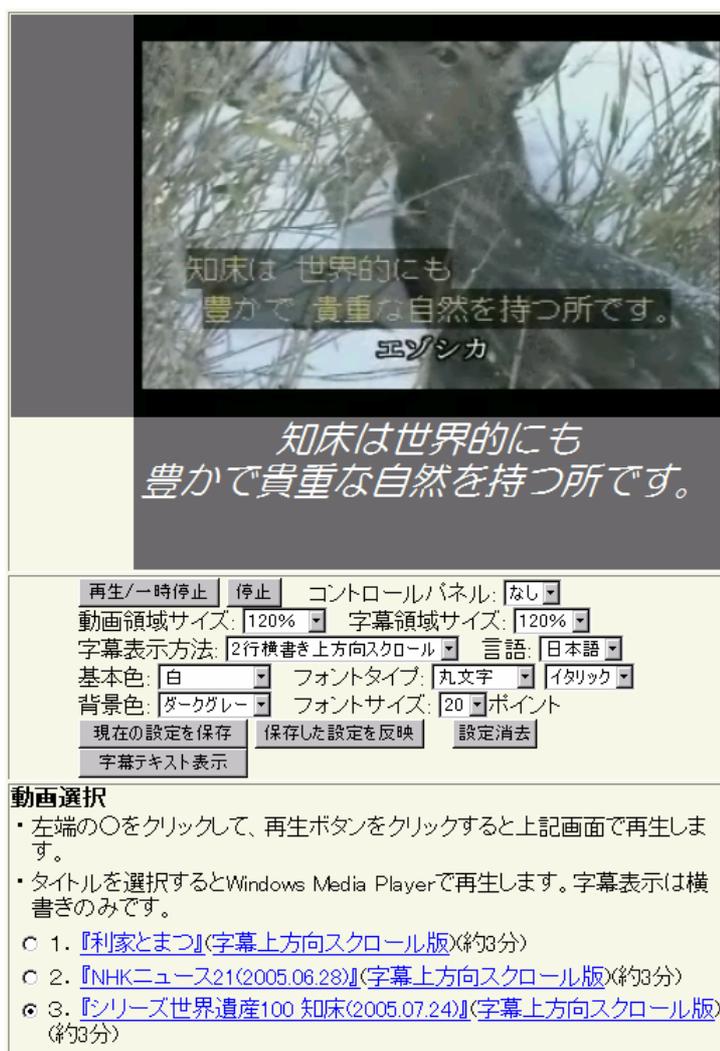


図2 本システムによる字幕のブラウザ内表示例

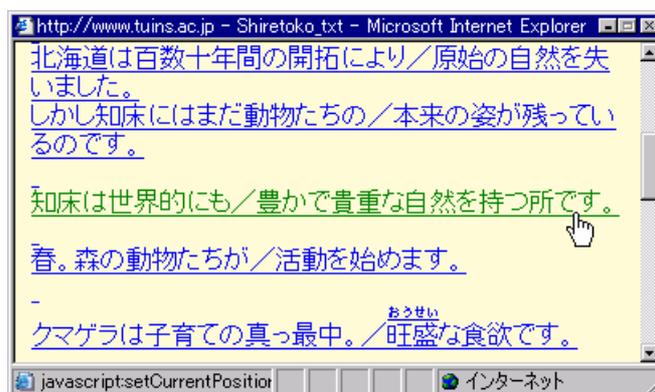


図3 字幕テキストウィンドウの表示例

ファイルにする場合は VobSub などのフリーソフトを利用するが、ネット配信が容易になるという利点がある。動画ファイルと字幕ファイルを別々にする場合は Media Player の ActiveX コントロールを利用するかまたは、動画ファイルの URL の引数として字幕ファイルを指定することになる。

本システムを利用する場合のいくつかの注意点については次の通りである。

動作環境としては Windows2000/XP では、ブラウザ内の埋め込み Windows Media Player コントロール(Windows 版 Internet Explorer や GreenBrowser(<http://www.morequick.com/indexen.htm>), Firefox で利用可能)を利用して字幕表示が可能である。Firefox の場合は、拡張機能 IE Tab をインストールし、ページを右クリックメニューの IE Tab で開く。IE の機能を使いたくない場合は拡張機能 ActiveX plugin for Firefox をインストールする。ページに EMBED タグを使用している場合、複数の動画が含まれている場合もあるので、WMP の再生メニューから前へ(Ctrl-B)や次へ(Ctrl-F)を選ぶと動画を切り替えることができる。動画の順番は、～.asx ファイル内に記述されたエントリタグの順番となる。

また、DivX 形式ビデオの再生には、DivX コーデック(codec)が必要となる。DivX Video for Windows のサイト(<http://www.divx.com/divx/windows/?lang=ja>)からダウンロードできる。

WindowsPC にどのコーデックがインストールされているかは、コントロールパネル システム ハードウェア デバイスマネージャ サウンド、ビデオ、およびゲームコントローラ ビデオ CODEC プロパティにて確認できる。

Windows Media Player にて字幕を表示するには、再生メニュー キャプションと字幕 利用できる場合はオンにチェックを入れる必要がある。また、必要ならば Windows Media Player において、ツール オプション セキュリティタブにて、「ローカルキャプションがある場合は表示する」にチェックを入れる。あるいは、上記のリンクをコピーし、Windows Media Player を起動してファイル URL を開くを選択し、コピーしたリンクを貼り付ける。その場合、Windows Media Player において asx ファイル読み込み時に動画ファイルが見つからないとなる場合は、ツール オプション ファイルの種類にて Windows Media ファイル(asf)にしっかりチェックが入っていることを確認する。背景がグレーのチェックの時はチェックを入れ直す。

MacOSX 対応については次のようにすれば字幕付きで再生できる。MacOS9 以前については非対応である。まず、Microsoft 社から MacOSX 版 Windows Media Player をダウンロードしてインストールする。次に動画選択にて見たい動画をクリックすると Windows Media Player が自動的に起動する(Internet Explorer5.2 や Safari、Firefox など、いずれのブラウザでも可)。起動しない場合は動画選択で右クリック(1 ボタンマウスでは長押しクリック)し、メニューからリンク URL をコピーを選択する。次に Mac 版 Windows Media Player を起動する。表示メニューから、「キャプションを表示」にチェックを入れる。ファイルメニューから、「URL を開く」を選択し、コピーした URL を貼り付ける。すると字幕付きで動画の再生が始まる。字幕フォントのスタイルやサイズなどの変更は再生メニューの「言語」から選択することでできるが、ルビが表示できないなどいろいろ制限がある。なお、ブラウザから Windows Media Player を自動起動した場合は、再生メニューの言語の中の項目が表示されないことがあるが原因は不明である。

図4に本システムを利用した字幕表示のデモ機の外観である。図では本字幕表示システムの全画面表示を行なっている。このデモ機は無線LAN機能内蔵のペン感圧型タッチパネル型PCであり、基本的な操作はペンまたは指を利用して行なう。OSはWindows2000/XPの双方のシステムが利用できる。必要に応じて本体をクレードルから取り外して携帯でき、画面を90°



図4 全画面表示例とシステム全体図

や180°回転させて表示することも可能である。キーボード・マウスは必要に応じて接続できる(PS/2とUSBいずれも可)。内蔵の外部記憶装置としては、2.5インチ型ハードディスク、コンパクトフラッシュ、マイクロドライブなどが選択できる。

開発のための技術資料

本システムの開発にあたっては下記の技術資料を参考にした。

- Windows Media Player SDK(日本語)
http://msdn.microsoft.com/library/ja/default.asp?url=/library/ja/wmplay/mmp_sdk/default.asp
- Windows Media SDK コンポーネント
<http://www.microsoft.com/japan/windows/windowsmedia/mp10/sdk.aspx>
- SAMI 1.0 の理解
http://www.microsoft.com/japan/msdn/accessibility/general/atg_samiarticle.asp
- デジタル メディアへのクローズド キャプションの追加
http://www.microsoft.com/japan/msdn/windowsmedia/production/wmp7_sami.asp
- Internet Explorer ActiveX Components Gallery
<http://activex.microsoft.com/>
- SMIL(W3C)
<http://www.w3.org/AudioVideo/>
- ActiveX plugin for Firefox
<http://www.iol.ie/~locka/mozilla/mozilla.htm>

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述べた。今後、システムの評価・改良を行なう予定である。評価については、現在「字幕放送に関する調査」として主に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/ccr.cgi>)。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ター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。NHKなどが実施)とスピードワープロを利用した高速オペレーター方式(日本テレビなどが実施)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) Shirai, et al.: Program and Proceedings of TAO WORKSHOP on TV Closed Captions for the hearing impaired people, pp.9-29, (1999)
- (5) <http://www.w3.org/AudioVideo/>
- (6) Dick C. A. Bulterman, Lloyd Rutledge, SMIL 2.0, Springer-Verlag (2004)