

字幕アーカイブ構想を考慮した字幕・音声ガイドの記述様式の検討

内野 每美[†] 松井 祐太郎[‡] 坂根 裕[‡] 穉川 友宏[‡]

[†] 静岡大学情報学部 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: [†] ia05012@s.inf.shizuoka.ac.jp [‡] hal@inf.shizuoka.ac.jp

あらまし 筆者らは字幕アーカイブ構想を考慮して、汎用性の高い字幕・音声ガイドの記述様式の検討を行っている。これまでに、字幕と音声ガイドの統一したオーサリングや、プラネタリウムなどの異種メディアへの字幕の転用などに対応すべく、そのつど記述様式を拡張し、ツールを試作してしのいできた。本稿では、その中間報告をするとともに、今後の構想についても整理する。

キーワード 字幕アーカイブ、アクセシビリティ、字幕、音声ガイド

A study of versatile describing format of subtitles and audio description for multimedia programs

Maimi UCHINO[†] Yutaro MATSUI[‡] Yutaka SAKANE[‡] and Tomohiro HARAICAWA[‡]

[†] Faculty of Information, Shizuoka University 3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka, 432-8011 Japan

E-mail: [†] ia05012@s.inf.shizuoka.ac.jp [‡] hal@inf.shizuoka.ac.jp

Abstract This intermediate report is about versatile describing format of subtitles and audio description for the subtitle archives project. The authors have tried to design unified describing format and tools for co-authoring of subtitles and audio description, and have also tried to modify the format and the tools to less-commonly media such as planetarium movie.

Keyword Subtitle Archives, Accessibility, Subtitles, Audio description

1. はじめに

障害のある人に、見えていない、聞こえていない情報を伝えることを情報保障というが、自治体等のユニバーサルデザイン (UD: Universal Design) 運動の一環として、情報保障に配慮した映画上映の企画が増えてきている。UD コンテンツを増やしていくためには、情報保障素材制作の負担を軽減し、さらに、メディアをまたいだ再利用を促進する必要がある。後者は字幕アーカイブ構想としてキュー・テックの川野氏が推進している。

筆者らの所属する静岡大学情報学部では、自治体や福祉系 NPO との地域連携活動として、聴覚障害者が映画を楽しむための日本語字幕の制作と、字幕つき上映の現場支援を行ってきた。一般に、音は聞こえているという前提のもとで言語の橋渡しのみを期待される洋画の日本語字幕とは異なり、聴覚障害者のための日本語字幕は、情景音の文字化をはじめ、話者の区別しやすさ、常用しない漢字へのルビふりなど、いくつもの配慮が求められる。そのため、一般に流通している字幕制作ツールや記述様式では機能不足が生じ、筆者らはその都度補助ツールを開発し、記述様式を拡張して凌いできた。

その活動を続ける中で、視覚障害者が映画を楽し

むための音声ガイドを制作していた「シーンボイスはままつ」という団体と交流をもつようになり、字幕と音声ガイドは、制作や上映支援に多くの共通点を持つことがわかってきた。肉声に近い合成音声エンジンも追い風となり、字幕と音声ガイドを同時にオーサリングすることで、UD 上映用の素材を効率的に制作できる可能性が見えてきた。

さらに、記述様式の工夫次第では、映画以外の、たとえばテレビ放送でも、HDD レコーダなどのコンシューマー機器の助けを借り、字幕や音声ガイド付き番組を数多く増やせると筆者らは考えるようになった。

筆者らはそのための字幕・音声ガイドの汎用的な記述を定義しようとしており、さまざまな UD 上映会に参画して情報を実践的に収集している。本稿ではいくつかの事例と展望を紹介する。

2. 字幕・音声ガイドつき上映の実際

映画のフィルムは、秒間 24 コマのスピードで映写されるが、デジタル素材と異なり、機械式映写機ではどうしても±0.2%程度の誤差が生じる。これは、90分の映画だと 10 秒強もの差として現れるため、フィルムの速度誤差にあわせて上映する必要がある。以前は字幕送出用のツールを使っていたが、現在では音声ガイドを含めた他団体での再利用の容易さから、可変

表 1: 情報保障に応じた上映形態

情報保障の形態	配信方式	映像	字幕	音声
通常の上映	一斉	本編映像	OFF	主 (ステレオ)
聴覚情報保障のみ		黒画面※	解説字幕	主 (ステレオ)
視覚情報保障のみ		黒画面※	OFF	主副混合 (ステレオ)
	個別	黒画面※	OFF	主副分離 (モノラル)

※映写機への重畳投影でない場合は、本編映像とする

映像選択 本編映像・黒画面
 字幕選択 OFF・字幕・解説字幕
 音声選択 本編音声・
 解説音声 (主副混合)・
 解説音声 (主副分離)

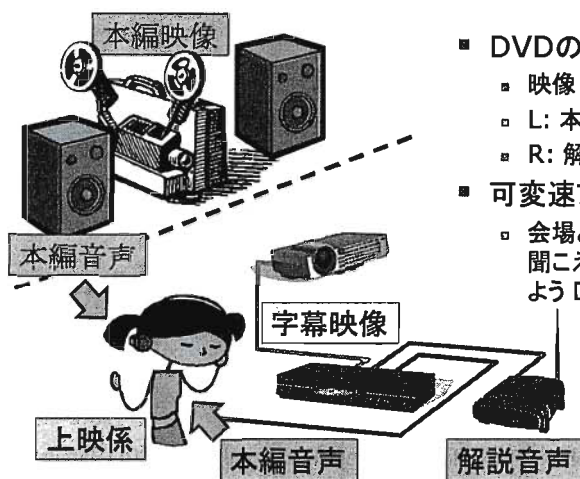


図 1: 同期再生のしくみ

- DVDの仕様
 - 映像: 黒画面に字幕
 - L: 本編 (→ ヘッドホン)
 - R: 解説音声 (→ 無線機)
- 可変速プレーヤーで重畳
 - 会場とヘッドホン両方から聞こえる本編音声が入合うよう DVD の速度を調整

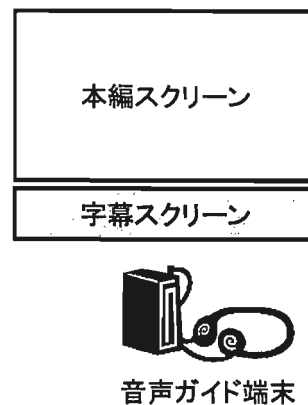


図 2: 上映形態

速 DVD プレーヤーを用いて、同期上映用にオーサリングした DVD を映写機とは別のプロジェクタから重畳投影する方法を採用している。

2.1. 同期投影用 DVD の構成

字幕と音声ガイドを別の団体が制作した場合、同期上映用の素材は別々のメディアに格納されているので、映写機にあわせた速度調整も個別に行うことになる。隣の音声ガイドのチームとは、お互いにまったく同じ作業をしているおかしさから、現場での工数をどちらかにまとめられないかという運びになった。

さまざまな情報保障の形態を考えると、DVD に求められる上映形態は表 1 のようになり、映像は「本編映像」、「黒画面」の 2 種類、字幕は「OFF」、「字幕」、「解説字幕」の 3 種類、音声は「本編音声」、「解説音声つき (主副混合)」、「解説音声つき (主副分離)」の 3 種類から選択可能であればいいことがわかる。上映形態により、これらを選択し、必要な情報のみを記した形式を用いるのである。

映像の本編映像と黒画面との切り換え可能にするためには、DVD のマルチアングル機能を利用する。第 1 の映像に本編を入れておき、第 2 の映像に黒画面を入れておけば、いつでも本編映像を消すことが可能になる。

2.2. 同期再生

字幕や解説音声を本編に同期させる研究はいくつかあるが^{[1],[2]}、多くの現場では、上映係は図 1 のよう

な原理で同期再生を行っている。上映係のヘッドホンに繋いだ左音声チャンネルからは本編音声、無線送信器に繋いだ右音声チャンネルからは解説音声が出力される。上映係は可変速プレーヤーを利用し、会場から流れてくる本編音声と、ヘッドホンから聞こえてくる本編音声はずれないように DVD の速度を微調整する。そうすることで、字幕と音声ガイドを本編に同期させることが可能となる。可変速プレーヤーとしては Numark の VJ-01 を使用している。

2.3. 字幕素材の制作工程

ボランティア団体によっては、字幕制作工程や映写速度誤差の吸収はまだ敷居が高く、PowerPoint のスライドショーのように字幕を映画の進行に合わせてめくるような方式もとられている。一方で、筆者らは再利用性を高めるため、可変速プレーヤーで再生する DVD を、以下のプロセスで制作している。

2.3.1. セリフの入力

字幕の制作工程は、Excel へのセリフ入力から始まる。秒あたりの文字数など、各種の目安があるが、いずれかの字幕仕様にあうように要約することとなる。最初の段階では、要約は考えずに、セリフを書き起こし、要約の仕上げは、複数人で通読するかたちをとる。登場人物のセリフ以外に、ナレーション等の朗読箇所がある場合は、このときにセリフとナレーションの表記を分別しておく。

2.3.2. 情景音の追加

セリフを書き起こした後、情景音を文字化して追加する作業を行う。情景音の表記にはさまざまなものがあるが、筆者らは（ ）で囲むという規則のもと入力を進める。

2.3.3. 時刻情報の追加

セリフや情景音を文字化した後、時刻情報を追加していく。一般には、Subtitle Workshop というツールがよく使われているようである。筆者らは映画のデジタル素材を扱うこともあるため、DF (ドロップフレーム) と NDF (ノンドロップフレーム) のどちらのタイムコードでも区別なく作業できるように、VisualSubSync というツールを使い、波形を見ながら音声を基準にタイミングデータを確定していく。タイミングデータは波形の始末端にぴったりと合わせて切る。字幕のフェードイン/アウトのような前後の余裕は、あとから計算で求める。VisualSubSync の SRT 字幕形式での書き出しにはバグがあるため、スクリプト言語によるプログラムを作成することにより、これを回避した。

表示時間に対する文字数が多いなどの調整点があるため、作業終了後、変換プログラムを用意して、字幕情報を再び Excel に戻し、文字送りの調整などを行った後、時刻を確定する。情景音に（ ）をつけているため、台詞と情景音は自動で区別することが可能である。

字幕情報を確定した後、用意した変換ツールを用いて、STL (Spruce Technology 社の字幕形式) に変換し、Mac 上の DVD Studio Pro に転送する。映画に重畳する場合と DVD 単体のいずれでも字幕付き上映が可能になるよう、本編映像と黒画面をマルチアングルで焼きつけておく。マルチアングルのオーサリングが可能で、かつマルチフォントの字幕レイヤが扱えるものは、DVD Studio Pro のみのようである。

2.4. 音声ガイド素材の制作工程

音声ガイドは、もともとはまったく独立した工程で制作されており、その具体的な手順は以下のとおりである。

2.4.1. セリフの入力

音声ガイドを制作する際、おもにセリフの合間に音声ガイドを挿入するため、セリフをあらかじめ入力しておく必要がある。入力形式は字幕作成時のセリフ入力と似ているが、配役名の情報が入る点が音声ガイドの特徴である。

2.4.2. 情景の追加

セリフ入力が終了したら、情景を文字化して追加していく。情景情報から必要なものを、熟練者が取捨選択し、朗読原稿を作成していく。

2.4.3. 朗読原稿の録音

朗読原稿を録音し、読み誤りなどは編集して 1 本の音声ファイルにする。これを速度調整しながら再生する。

2.5. 音声ガイド素材の制作工程の改善

ここまでの工程を表 2 にまとめる。字幕と音声ガイドは、同期投影のための作業のみならず、制作段階でも共通する作業が多い。このため、両者が共同で作業をする環境を整えれば、UD 上映用の素材はきわめて効率的に制作することができる。

表 2: 字幕および音声ガイド制作工程

字幕制作工程	音声ガイド制作工程
・ Excel でセリフを入力	・ Excel でセリフを入力
・ 情景音を文字化し追加	・ 情景を文字化し追加
・ 時刻情報を追加	・ 朗読原稿を録音
・ 字幕情報を Excel に戻し読速度を調整	・ テープ編集・音声保存
・ STL 形式に変換	・ タイミング・音量調整
・ DVD Studio Pro に転送	・ DVD Studio Pro に転送

音声ガイドのグループと同時に作業をするようになり、制作過程は以下のように変化した。

2.5.1. セリフの入力

字幕のために入力した原稿を流用する。作業状況によっては、先に音声ガイドのグループにセリフを入力し、その原稿を字幕素材として流用することもある。

2.5.2. 情景の追加

この作業は従前どおりに行う。ただし、字幕化された情景音を参考にしてガイド原稿を考えることもできる。

2.5.3. 朗読原稿の録音

字幕のグループはもともと映像編集を伴うので、朗読原稿もデジタルビデオを用いて録音する。プラズマテレビの画面を左右に 2 分割し、右画面に本編映像、左画面に原稿を表示できるプロンプタを即席で作成し、原稿が進行にあわせてスクロールするようにした。ナレータはヘッドホンで本編音声を聞きながら、デジタルビデオに接続したマイクで声の収録を行う。これを一般の映像編集ソフトで編集する。筆者らは、Final Cut Pro という Mac 上の映像編集ソフトを使い、編集を行った。

2.5.4. ビデオ編集・音声保存

先に行ったナレーションの録音において、一度の失敗もなく音声を収録できるのであれば、デジタルビデオを用いる利点はない。しかし、通常は読み間違いによる録直しを行う。その際に、音声を収録しているビデオを一時停止し、本編映像を巻き戻し、停止前の少し前から収録し直す。デジタルビデオでは、テープ

をまるごと素材として取り込んだあと、一時停止をかけたポイントで素材を切ってくれる機能があるため、この作業によって、映像の切れ目を手がかりに、失敗した部分を結合・修正することが可能となる。デジタルビデオには本編映像も記録されているため、朗読と本編との同期をとることは録音物に比べて容易である。



図 3: 音声と同時に記録されたプロンプタ映像

また、朗読部分以外を無音にすることにより、ノイズを取り除く作業を行う。

2.5.5. タイミング・音量調整

先に編集した解説音声で、主音声の音を遮らないようタイミング調整をする。タイミングを取り終えた後、音声を1本の素材として書き出し、音量を調整する。解説音声は、本編の背景音の大きさにあわせて調整する。こうして編集したものを、DVD Studio に転送する。

2.6. 字幕・音声ガイド作成ツール



図 4: 時刻情報付与および音声ガイド合成ツール

筆者らは、以上のように字幕と音声ガイドの並行制作の工程を踏まえたうえで、図4のような、字幕と音声ガイドを統一的にオーサリングできるツールを試作した。本ツールでは、両者を同じタイムライン上に

並べて編集することができ、音声合成を用いて音声ガイドのシミュレーションや、あるいは VoiceText (Pentax 社) のような高品質音声合成ソフトを用いてガイドを作ることも可能である。形式は、ASS (Advanced Substation Alpha) フォーマットをもとにした拡張 ASS フォーマットを使用しており、ふりがなや行間の調整、マルチフォント時のベースライン補正などが可能となる。

3. メディア間の素材の再利用

先に述べた手順を用い、一般的な平面画面での映画に字幕を付与できることが分かった。筆者らはさらに、プラネタリウム映画「銀河鉄道の夜」の字幕制作を通じて、より汎用的な字幕・音声ガイドの記述形式を検討した。黒背景に明るい色で書いた38分の字幕映像を、別途用意するプロジェクタを用い、本編に同期して再生することで字幕付き上映を実現する。これは、一般の35mmもしくは16mmフィルムに対する字幕つき上映と同等の原理である。

3.1. 位置データの作成

プラネタリウム映画が一般の映画と最も違うのは、全天周スクリーンが広過ぎ、画面の下方に字幕を出しても、いままさに解説している事物とともに参照することが困難であることである。そのため、プラネタリウムに字幕を出す場合は、ジョイスティックでプロジェクタを移動させ、話題の中心を追いかけながら投影するのが一般的である^[3]。

当初は、プラネタリウムの字幕と同様に、プラネタリウム映画の字幕も、科学館設置の X-Y 投影機（電動プロジェクタ）を用い、動きはジョイスティックを用いて手でつける計画でいた。しかし、実際に試してみると手ぶれや位置ずれにより映画の雰囲気は損なってしまうことがわかった。そこで、字幕は、プラネタリウムに設置されている電動プロジェクタを使い、話題の中心を追いかけながら、かつ明るいところを避けながら投影することとした。

このため、最適な字幕位置を割り出し、X-Y 投影機をプログラムしてしまうことにした。たとえ全天周映像であったとしても、際限なく試写を繰り返して位置を調整することが可能であれば、必ず適切な位置に字幕をつけることができるが、それでは、限られた試写の時間をすぐに使い切ってしまう。

そこで、これまで使ってきた字幕記述様式に位置の情報を含めることとし、タイプセッタ側でシミュレーション映像を生成させることで、最適な字幕位置が割り出せるようにした。最初の試写では、方眼を描いた画像を用意し、字幕が必要そうなエリアに、特製のレーザーマーカによってこの方眼を貼り合わせながら魚眼レンズで撮影した。さらに、版元の KAGAYA ス

タジオから善管注意義務のもとでの再撮の許可を得たうえで、試写の様子を同じレンズで持ち帰って合成し、最適な字幕位置の割り出しを行った。

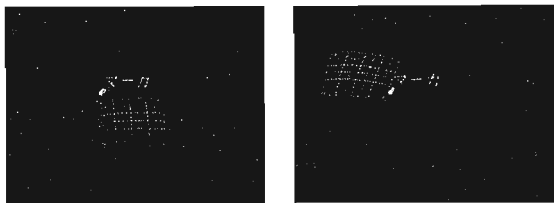


図5: 全天に方眼を描く
(著作権の関係から本編映像は不掲載)

3.2. シミュレーションツール

拡張 ASS フォーマットを ASS フォーマットに変換するプリフォーマットを試作し、Gabest の textsub タイプセッタを用いて字幕重畳のシミュレーションを行った。割り出された最適な字幕位置をもとにプラネタリウム制御用の SPICE プログラムを作成し、映画に同期して全天を適切に動く字幕が実現でき、無事に上映会に供することができた。

4. まとめ

汎用性の高い字幕・音声ガイドの記述様式について、実践結果をまとめた。上映後のアンケート調査では「すべての漢字にふりがなをつけてほしい」など多数の意見をいただき、ルビ機能や話者に対応した字幕の移動など、都度、記述様式の拡張とプリフォーマットの改良を続けている。

平成 20 年度以降の視聴覚障害者向け放送の普及拡大に向け、総務省は「視聴覚障害者向け放送普及行政の指針」を策定している^[4]。同資料の「策定の経緯」によれば、平成 18 年度実績では、字幕付与可能な番組への字幕付与は NHK で 100%、民放キー局は平均で 77.8% 達成されている。ただし、この数値は午前 7 時から午後 12 時の時間帯のみが対象で、かつ表 1 のものを除くとした統計であり、全放送時間からすれば NHK で 43.1%、民放キー局では 32.9% にすぎない。

同資料の別紙 2 にも現れるように、ローカル局の状況は、字幕付与の委託先に乏しいなど、特に厳しい。発言内容をリアルタイムに字幕化する、いわゆる PC 要約筆記者は地方にもおり、ボランティアに名乗りを上げているところもあるが、字幕が誤った場合の責任問題に対して局が慎重であり、結果として採用は進んでいない。直後に放送波で訂正がかかるか、局の公式サイトに正誤表を掲載すれば免責とするといった社会的な合意形成や、正誤表をもとに録画済番組の字幕に対して自動で修正をかけられる家庭用 HDD レコーダに期待するところは大きい。

一方で、同指針では、視覚障害者のための解説放送についても数値目標が定められたが、NHK 総合と民法で 10%、NHK 教育で放送の 15% にとどまり、音声解説の付与がより困難であることを裏付けている。

字幕や音声解説をさらに普及させていくためには、専門性の高いリアルタイム処理を極力減らして情報保障の担い手を増やし、公式な字幕や音声解説とボランティアなそれらが相補的に運用できるような仕組みを考えなければならない。たとえば、放送時には本編のみを先に HDD レコーダに録画しておき、事後のボタン操作により、HDD レコーダがその番組に対応する字幕・音声解説を局のサイトから取得し合成するなどはそのひとつの手法である。「昼間録画しておいたワイドショーは、帰宅時に音声解説付きで見られればよい」ということであれば、字幕や音声ガイドを後付けするためのフォーマットやプロトコルを適切に標準化することによってリアルタイム処理を避けて通る道が開かれ、手の施しようがあまりない生番組の音声解説なども数多くサポートできるようになる。同様に、ローカル局の字幕制作もボランティアに任せ、局の目を通してからリリースできるようになる。

このようなさまざまな可能性を考えながら、より汎用的な字幕フォーマットを練り上げ、今後の字幕・音声ガイドつき上映に生かしていきたい。

文 献

- [1] 浅野大介, 和田守美穂, 飯塚育生, “音響分析による映写機の再生ピッチ抽出法,” 電子情報通信学会技術研究報告, vol.104, no.497(20041203) pp. 55-58
- [2] 和田守美穂, “映画の副音声同期化システムの開発,” “映像情報メディア学会 2003 年冬季大会”, no5-14, pp.1, March.2004.
- [3] 野田学, 服部完治, 毛利勝廣, 鈴木雅夫, 小林修二, 北原政子, “字幕付きで見るプラネタリウム,” 名古屋市科学館紀要, no.32, pp.13-15, 2006.
- [4] 総務省報道資料 http://www.soumu.go.jp/s-news/2007/070629_9.html

(遅筆により不鮮明な資料のまま期限を迎えてしまいました。発表にて補足させていただきたく思います)