

日本語字幕ルビ適応に向けた聴取音声の影響調査

阿部 晋矢^{1,a)} 藤井 翔子¹ 小松 佑人¹ 美野 秀弥¹ 後藤 淳¹ 藤津 智¹ 松村 欣司¹ 藤沢 寛¹

概要: 本来聴覚に不安のあるユーザー向けに用意されるクローズドキャプションなどの字幕が、聴覚に不安のないユーザーにも現在多く利用されている。本稿では、聴覚に不安のないユーザーを対象とした字幕の適応、特に漢字のルビのユーザー適応に注目する。聴覚に不安のない成人を対象とした実験から、字幕と同内容の音声を聴取しているかが漢字に付与するルビの要不要に影響するかをウェルチのt検定により明らかにした。また、実験結果をベースとした字幕漢字にルビを付与すべきかを予測する適応モデルを提案し評価した。

キーワード: 字幕, クローズドキャプション, ユーザー適応

Listening Audio Impact Survey for Adaptive Japanese Subtitles and Captions Ruby

SHINYA ABE^{1,a)} SHOKO FUJII¹ YUTO KOMATSU¹ HIDEYA MINO¹ JUN GOTO¹ SATOSHI FUJITSU¹
KINJI MATSUMURA¹ HIROSHI FUJISAWA¹

Abstract: Subtitles and closed captions, which are prepared for users with hearing concerns, are now widely used by users who do not have hearing concerns. In this paper, we focus on the adaptation of subtitles and captions for users who do not have hearing concerns, especially the adaptation of kanji ruby. From our experiments on adults who do not have hearing concerns, Welch's t-test was used to clarify whether listening audio with the same content affects the necessity of kanji ruby. In addition, we proposed and evaluated an adaptive model to predict whether ruby should be added to captions kanji based on experimental results.

Keywords: Subtitles, Closed captions, User adaptation

1. はじめに

現在、多くのコンテンツに字幕やクローズドキャプションが付与されている。主に音声は聞こえるが外国語であるなどユーザーが音声を理解できない場合にコンテンツに付与する文字情報を字幕 (Subtitles) と呼び、聴覚に不安のあるユーザーなどに向けて音声情報を補助する文字情報をキャプション (Captions) と呼ぶ。特に、ユーザーが選択して表示/非表示を切り替えるキャプションをクローズドキャプション (Closed captions) と呼ぶ。これらのコンテンツに付与された字幕やクローズドキャプションは、聴覚

に不安のないユーザーにも現在多く利用されている [1]。以降は、特に字幕とキャプションを区別しない場合、本稿では共通して字幕と記載する。

幅広いユーザーが字幕を利用している現在、ユーザーに合わせた字幕の提供が検討されている。Adaptive Subtitling [2] は英国での視聴者を対象としたユーザー適応字幕である。出演者のうちユーザーが任意に選択する特定の人物の発言のみを字幕として表示する。Adaptive Subtitling は聴覚に不安のないユーザーも対象とした数少ない字幕に関するシステムであるが、英語圏における地域での発音差などを考慮した手法であり日本語ならではの問題を考えてはいない。また、日本語を対象としたユーザーに合わせて字幕の提供を行う例として、学年別ルビ付加機能を有する字幕ソフトウェア [3] が提案された。大学講義での教授の

¹ 日本放送協会 放送技術研究所
NHK Science & Technology Research Laboratories

a) abe.s-jg@nhk.or.jp

発言の書き起こしを提供するソフトウェアで、聴覚に不安のある大学生を対象に、すべての漢字にルビを付与するよりも学年に応じてルビを表示の方が受け入れやすいことを検証した。学年別ルビ付加機能を有する字幕ソフトウェアは漢字へ付与するルビという日本語ならではの問題を扱うが、聴覚に不安のあるユーザーでのみ検証しており、またルビ適応のパラメータは学年のみである。

本稿では、聴覚に不安のないユーザーを対象とした字幕の適応、特に漢字のルビのユーザー適応に注目する。聴覚に不安のない成人を対象とした実験から、字幕と同内容の音声を聴取しているかが漢字に付与するルビの要不要に影響するかをウェルチのt検定により明らかにした。また、実験結果をベースとした字幕漢字にルビを付与すべきかを予測する適応モデルを提案し評価した。

2. 関連研究

現在、多くのコンテンツに字幕やクロズドキャプションが付与されている。特にクロズドキャプションは、総務省による放送分野における情報アクセシビリティに関する指針 [4] も示され、放送番組への付与が進められている。海外でも同様の情報アクセシビリティに関するガイドラインがあり、たとえば欧州の HbbTV[5] や W3C[6] でも、ガイドライン内で字幕について触れられている。字幕についても、作成のガイドラインが提案されており BBC は BBC Subtitle Guidelines[7] を公開している。日本でも、公開される字幕のガイドラインはないが、内容理解度に関する評価 [8] など、利用しやすい字幕に向けた取組みが報告されている。これらを提供する従来の字幕フォーマットとして、日本の放送で利用される字幕フォーマット [9], [10], TTML Profiles for Internet Media Subtitles and Captions (IMSC)[11], WebVTT[12] がある。

字幕、特にキャプションは聴覚に不安のあるユーザーの補助として、用意されるものの字幕ユーザーは必ずしも聴覚に不安のあるユーザーだけでないことが調査されている [1], [13]。また、字幕を使う理由もさまざまであることが調査されている [1], [2]。たとえばウェルズにおける字幕の利用調査 [14] では、英語が第 1 言語でないこと、出演者の発言がつぶやくように小さくて聞き取りづらいこと、出演者の音声と比較し背景音が大きすぎること、歌番組の歌詞を確認したいなど、様々な字幕の利用意向があげられている。その他、字幕の利用者の好みも様々であり、たとえば好みの文字サイズが異なることが調査されている [15]。このように、字幕の利用意向や利用者の好みも様々であることから、文字サイズや表示位置のパーソナライズ [16]、出演者ごとの表示/非表示の選択 [2] という字幕のパーソナライズ手法が提案された。また、日本語を対象としたユーザーに合わせて字幕の提供を行う例として、学年別ルビ付加機能を有する字幕ソフトウェア [3] が提案された。大学

講義での教授の発言の書き起こしを提供するソフトウェアで、聴覚に不安のある大学生を対象に、すべての漢字にルビを付与するよりも学年に応じてルビを表示の方が受け入れやすいことを検証した。

関連研究では、海外を中心に聴覚に不安のないユーザーも含めて、個人に合わせた字幕の適応が検討されている。一方で、日本語ならではの字幕の個人適応は、これまで聴覚に不安のあるユーザーのみに注目されてきた。これまで対象とされなかった聴覚に不安のないユーザーの場合、コンテンツに含まれる音声と同時に字幕を視聴することになる。この際、たとえば日本語音声に日本語字幕を付与する字幕と同内容の音声を聴取する場合と、英語音声に日本語字幕を付与する字幕と同内容の音声を聴取しない場合で、適切なルビの振り方など最適な字幕適応結果が異なる可能性がある。後述する字幕利用に関する調査も受け、我々は聴覚に不安のないユーザーも対象とする日本語の字幕の個人適応に注目する。

3. 字幕利用に関するオンライン調査

関連研究の通り、英国ではテレビ視聴者の 10% が毎日字幕を使用し [16]、また字幕ユーザーの 80% は聴覚障害者ではない [13]。これらの調査結果をもとに、英国にてユーザー適応字幕の需要があるとして Adaptive Subtitling[2] が提案された。

同様に、日本における需要を確認したい。そこで我々は、日本語を対象としたユーザーに合わせた字幕の提供、特に聴覚に不安のないユーザーも含めて、日本人の字幕利用状況についてオンライン調査を実施した [1]。

オンライン調査は 2022 年 5 月 17 日から 20 日で実施された。20 歳以上を対象に性別と年代について日本の人口構成比に合った回収を目指し、1053 名から回答を得られた (表 1)。

表 1 オンライン調査の回答者数

Table 1 Number of online questionnaire samples.

| 年齢 | 男性 | 女性 |
|-----|-----|-----|
| 20代 | 68 | 67 |
| 30代 | 78 | 75 |
| 40代 | 101 | 96 |
| 50代 | 103 | 105 |
| 60代 | 89 | 87 |
| 70代 | 84 | 100 |

調査では以下の 5 項目について質問した。

- Q1. あなたが普段テレビ番組やインターネット動画、DVD ブルーレイディスクなどのコンテンツを視聴している際に利用したことがあるものを選択してください
- Q2. それぞれの字幕機能について、利用する理由を選択し

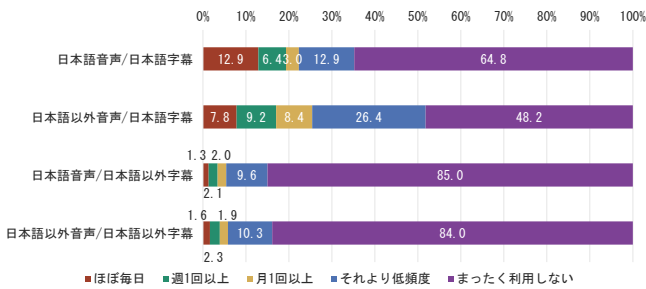


図 1 Q1 による各字幕の利用割合

Fig. 1 Percentage of each subtitle/caption used.

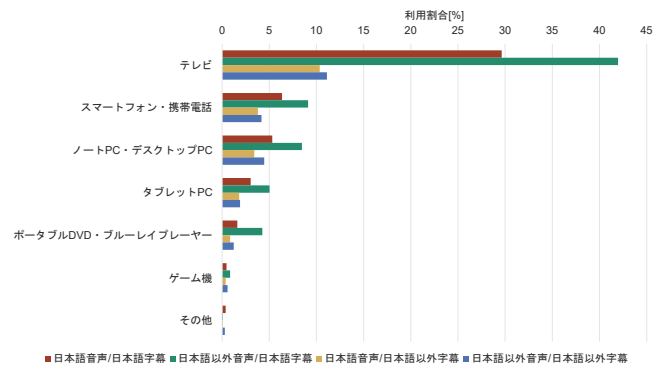


図 3 Q3 による各字幕の利用デバイス

Fig. 3 Devices used for each subtitles/captions.

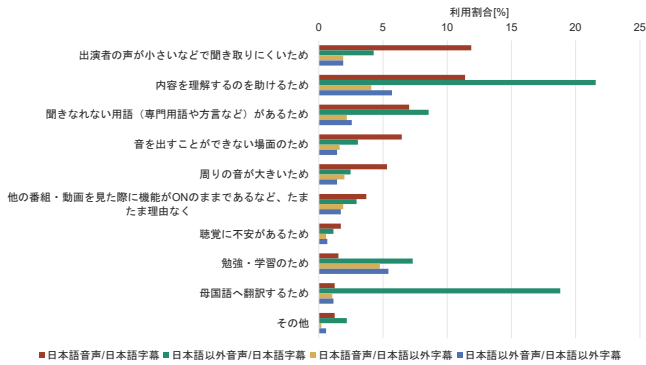


図 2 Q2 による各字幕の利用理由

Fig. 2 Reasons for each subtitle/caption used.

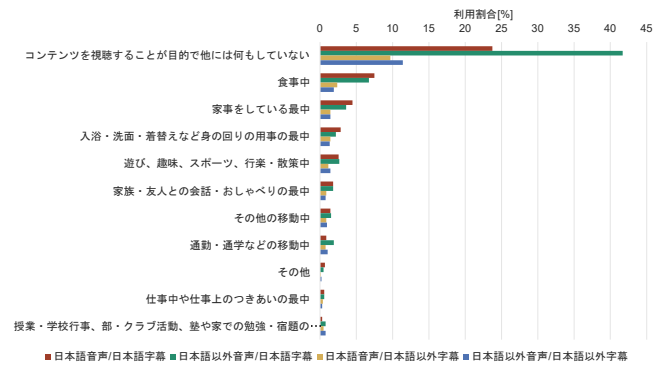


図 4 Q4 による各字幕の利用状況

Fig. 4 Using context for each subtitles/captions.

てください

- Q3. それぞれの字幕機能を利用する際に使用するデバイスを選択してください
- Q4. それぞれの字幕機能を利用してコンテンツを視聴する際の状況を選択してください
- Q5. それぞれの字幕機能を利用して視聴したコンテンツのジャンルを選択してください。コンテンツは複数のジャンルにまたがるものもありますので、当てはまると思われるジャンルをすべてお選びください

各質問では以下の4通りの字幕について回答を得た。

- (1) 日本語音声の番組・動画を再生中に日本語字幕を利用 (例: テレビ番組視聴中に字幕機能を利用)
- (2) 日本語以外の音声の番組・動画を再生中に日本語字幕を利用 (例: 英語の映画を視聴中に日本語字幕を利用)
- (3) 日本語音声の番組・動画を再生中に日本語以外の字幕を利用 (例: 日本語吹替映画を視聴中に英語字幕を利用)
- (4) 日本語以外の音声の番組・動画を再生中に日本語以外の字幕を利用 (例: 英語の映画を視聴中に英語字幕を利用)

図 1 より、日本語音声のコンテンツに日本語の字幕を利用するユーザーが約 35%おり、12.9%は毎日利用していることが分かる。その割合は、図 2 の聴覚に不安があるため字幕を利用する割合より多く、聴覚に不安のないユーザー

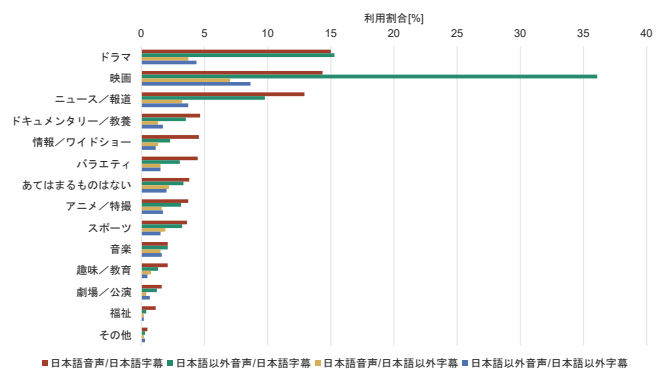


図 5 Q5 による各字幕の利用ジャンル

Fig. 5 Content genre used for each subtitles/captions.

も字幕を利用していることが分かる。図 3 より様々なデバイスが利用されていることが分かるが、最も多いデバイスはテレビである。図 4 からコンテンツのみを利用するケースが多いこと、図 5 からドラマ、映画、ニュース/報道で比較的多く字幕が利用されていることが分かる。

4. Web ベース日本語適応字幕システムによるルビ適応

オンライン調査の結果、聴覚に不安のないユーザーも字幕を利用していることが分かる。また、約 35%のユーザーが字幕を利用しており、母数も決して少なくない。多くのユーザーがいる場合、すでに聴覚に不安のあるユーザーを

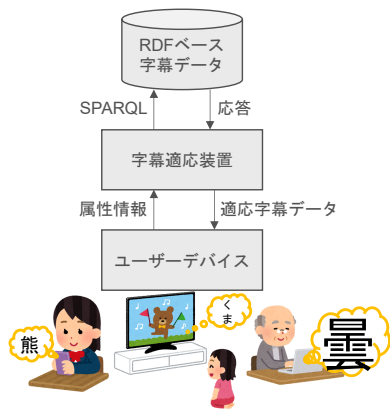


図 6 Web ベース日本語適応字幕システム

Fig. 6 Web-based Adaptive Japanese Subtitles and Captions.

対象に必要性が検証されたルビの最適化など、ユーザーの属性や状況に合わせた字幕が求められる可能性がある。

そこで、我々は様々なユーザーを対象とした日本語適応字幕を目指し、Web ベース日本語適応字幕システムを提案する。提案システムを図 6 に示す。従来の字幕フォーマットとして、日本の放送で利用される字幕フォーマット [9], [10], IMSC[11], WebVTT[12] がある。これらのフォーマットはテレビやスマートフォンなど、放送、インターネット動画などで利用され普及しているが、個人に合わせて字幕を提供することへの考慮は不十分である。そこで、提案システムでは元の字幕データを Resource Description Framework (RDF)[17] で用意し、ユーザーの属性や状況に応じて必要な情報を抽出する。抽出した情報を既存の字幕フォーマットに合わせて各種プレイヤーへ提供することで、既存のデバイスを活用しながらユーザーに合わせた字幕を提示できる。

RDF はセマンティック Web[18] に向けた W3C の技術仕様で、Web ページのつながりを継続的に表すため拡張性高く、機械可読なメタデータである。RDF ではデータを主語、述語、目的語のトリプルで表し、その構造は有向グラフで表される。Web ページだけでなく、近年 RDF はデバイスの情報 [19], コンテンツのメタデータ [20], [21] も表す語彙が提案されている。これらの取組みと同様に、我々は個人適応可能な字幕として必要な情報を含める RDF データを提案する。

提案する Web ベース日本語適応字幕システムのはじめのユースケースとして、ユーザーに合わせてルビを振る字幕システムを対象とする。ユーザーに合わせてルビの最適化は、既存研究で聴覚に不安のあるユーザーを対象に必要性が検証されている。一方、聴覚に不安のないユーザーを対象とした場合、たとえば聴取音声はルビの最適化の変数となるかなど、検証が十分でない。以降は、聴覚に不安のないユーザーを対象に、提案システムにおける最適化の変数について調査する実験と、実験結果をもとに生成したル

ビ適応モデルについて紹介する。

5. 聴取音声を与える影響調査と実験結果に基づくルビ適応モデル

本章では、提案する Web ベース日本語適応字幕システムのはじめのユースケースとして、ユーザーに合わせてルビを振ることを目指した調査実験とその結果に基づくルビ適応モデルについて紹介する。

5.1 聴取音声を与える影響調査

オンライン調査の結果を受け、提案システムでは聴覚に不安のないユーザーも対象とする。過去の研究では、聴覚に不安のあるユーザーを対象に、年齢や本人の漢字の識字能力に合わせたルビを提示することの効果を検証された [3]。一方で、聴覚に不安のないユーザーの場合、コンテンツに含まれる音声と同時に字幕を視聴することになる。この際、たとえば日本語音声に日本語字幕を付与する字幕と同内容の音声を聴取する場合と、英語音声に日本語字幕を付与する字幕と同内容の音声を聴取しない場合で、適切なルビの振り方が異なる可能性がある。

そこで我々は以下の仮説を検証する実験を行った。

- H_0 : 字幕と同内容の音声を聴取する場合と、字幕と同内容の音声を聴取しない場合とで、適切なルビの振り方に差はない。
- H_1 : 字幕と同内容の音声を聴取する場合と、字幕と同内容の音声を聴取しない場合とで、適切なルビの振り方に差がある。

実験に向け表 2 の通り、全漢字にルビなし/あり、同内容の音声あり/なしの 2 条件を組合せ 4 種類の字幕付き動画を用意した。

表 2 実験向け動画の設定
Table 2 Video settings for experiments.

| | 全漢字にルビなし | 全漢字にルビあり |
|----------|----------|----------|
| 同内容の音声あり | 動画 1 | 動画 2 |
| 同内容の音声なし | 動画 3 | 動画 4 |

その他の動画の条件は以下の通り。

- 映像: Big Buck Bunny[22](開始 28 秒から 68 秒の 40 秒間)
- 解像度: 1920px×1080px
- 音声:
 - 背景音: Big Buck Bunny の背景音を利用 (同内容の音声あり/なし共通)
 - 発話部分: VOICEVOX:ずんだもん, VOICEVOX:四国めたん [23](同内容の音声あり設定のみ)
- 字幕:

```
@base <http://example.org/> .
@prefix schema: <https://schema.org/> .
...
<#Caption1-1>
  schema:startTime "00:00:04" ;
  schema:endTime "00:00:10" ;
  schema:position 1;
  schema:hasPart [
    schema:position 1 ;
    <#term> <#むかしむかし> ] .
<#Caption1-2>
  schema:startTime "00:00:04" ;
  schema:endTime "00:00:10" ;
  schema:position 2;
  schema:hasPart [
    schema:position 1 ;
    <#term> <#この> ] ;
  schema:hasPart [
    schema:position 2 ;
    <#term> <#密林> ] ;
  schema:hasPart [
    schema:position 3 ;
    <#term> <#には> ] ;
  schema:hasPart [
    schema:position 4 ;
    <#term> <#智慧> ] ;
  schema:hasPart [
    schema:position 5 ;
    <#term> <#の> ] ;
  schema:hasPart [
    schema:position 6 ;
    <#term> <#ある> ] ;
  schema:hasPart [
    schema:position 7 ;
    <#term> <#モノケ> ] ;
  schema:hasPart [
    schema:position 8 ;
    <#term> <#が> ] ;
  schema:hasPart [
    schema:position 9 ;
    <#term> <#いた> ] .
...
```

図 7 Turtle による字幕 RDF データの擬似コード
Fig. 7 Turtle pseudo code of caption RDF data.

- フォント: IPAexGothic
- 本文サイズ: 72px
- ルビサイズ: 36px
- 文字色: 白
- 背景色: 黒

実験で利用した字幕の RDF データの擬似コードを図 7, またシステムの出力例となるルビなしの IMSC ファイルを図 8 に示す. 実験を円滑に進めるため, 実際の再生には図 8 の IMSC ファイルを再生した場合と同等の表示となるようあらかじめ映像に字幕内容を埋め込んだ. ルビありの字幕を表示した例を図 9 に示す.

実験は 2022 年 6 月 30 日から 7 月 19 日に実施した. 参加者は NHK 技研の職員, スタッフから募り 41 名を得た. 年齢は 25 歳から 53 歳 (平均 34.3 歳, 標準偏差 6.87) だった. 実験に関する報酬は無かった.

参加者を以下の 2 つのグループに分けた.

グループ 1: 同内容の音声あり (動画 1 および動画 2) を視聴, 20 名, 26 歳から 53 歳 (平均 34.8 歳, 標準偏差

6.90)

グループ 2: 同内容の音声なし (動画 3 および動画 4) を視聴, 21 名, 25 歳から 47 歳 (平均 33.8 歳, 標準偏差 6.97)

実験は以下の手順で行われた.

1. 説明者より実験の概要が説明され, 実験への参加同意書を記入する.
2. 割り当てられたルビなしの動画を視聴する.
 - グループ 1 は動画 1 を視聴する.
 - グループ 2 は動画 3 を視聴する.
3. 割り当てられたルビありの動画を視聴する.
 - グループ 1 は動画 2 を視聴する.
 - グループ 2 は動画 4 を視聴する.
4. 動画に出現する 20 の漢字について, ルビの要/不要を回答する. その際, 手順 2, 3 で視聴した動画を何回見返してもよい.
5. 20 の漢字にすべて回答し, 実験終了.

ルビが必要と答えた漢字の数を集計した結果を図 10 に示す. 平均値の差をウェルチの t 検定により検定した. $p = 0.0228$ であり, 有意水準 0.05 で H_0 が棄却され, H_1 の字幕と同内容の音声を聴取する場合と, 字幕と同内容の音声を聴取しない場合とで, 適切なルビの振り方に差があることが確認された.

5.2 ルビ適応モデル

聴取音声を与える影響調査の結果を用いて, ルビの適応モデルを作成した. ロジスティック回帰を選択し, 各変数の影響度についても評価した. 変数は以下とした.

- 目的変数: 漢字にルビが必要かの確率
- 説明変数:
 - 視聴者の年齢
 - 同内容の音声の有無
 - 漢字の難易度
 - 漢字の画数
 - 漢字が当て字かどうか

実験で使用した漢字に関する情報を表 3 に示す. 難易度は, 小学生については学習指導要領 [24], 中学生と高校生については日本漢字能力検定 [25] を基準とした.

聴取音声を与える影響調査の結果のうち, 8 割のデータを訓練用, 2 割のデータを検証用に分けロジスティック回帰を適用した. ステップワイズ法により変数を選択し, 漢字の画数は説明変数から取り除かれた. ロジスティック回帰の結果を表 4 に示す. 検証用データにおける正解率は 88.4%, 適合率は 91.5%, 再現率は 74.1%であった.

ロジスティック回帰により, 今回の実験におけるルビ適


```
<?xml version="1.0" encoding="UTF-8"?>
<tt xmlns="http://www.w3.org/ns/ttml" xmlns:tts="http://www.w3.org/ns/ttml#styling" xmlns:ttp="http://www.w3.org/ns/ttml#parameter" xml:lang="en" ttp:timeBase="media">
  <head>
    <styling>
      <style xml:id="defaultStyle" tts:fontFamily="IPAexゴシック" tts:fontSize="100%" tts:textAlign="center" />
      <style xml:id="spanDefaultStyle" tts:backgroundColor="#000000" tts:color="#FFFFFF" />
    </styling>
    <layout>
      <region xml:id="rBottom" tts:origin="10% 10%" tts:extent="80% 80%" tts:displayAlign="after" />
    </layout>
  </head>
  <body style="defaultStyle">
    <div begin="00:00:04" end="00:00:10">
      <p region="rBottom">
        <span style="spanDefaultStyle">むかしむかし</span>
      </p>
      <p region="rBottom">
        <span style="spanDefaultStyle">この密林には智慧のあるモノケがいた</span>
      </p>
    </div>
    <div begin="00:00:15" end="00:00:21">
      <p region="rBottom">
        <span style="spanDefaultStyle">僕は兎の尻</span>
      </p>
      <p region="rBottom">
        <span style="spanDefaultStyle">月をみていて遅くなったから眠いなあ</span>
      </p>
    </div>
    <div begin="00:00:21" end="00:00:24">
      <p region="rBottom">
        <span style="spanDefaultStyle">起きるぞ、喝！</span>
      </p>
    </div>
    <div begin="00:00:25" end="00:00:29">
      <p region="rBottom">
        <span style="spanDefaultStyle">外は眩しいなあ、泣ける</span>
      </p>
    </div>
    <div begin="00:00:29" end="00:00:33">
      <p region="rBottom">
        <span style="spanDefaultStyle">腹が減って黒洞みたいだな</span>
      </p>
    </div>
    <div begin="00:00:34" end="00:00:39">
      <p region="rBottom">
        <span style="spanDefaultStyle">あ、蚊に刺されてる</span>
      </p>
      <p region="rBottom">
        <span style="spanDefaultStyle">痣になっちゃうなあ</span>
      </p>
    </div>
  </body>
</tt>
```

図 8 すべての漢字にルビ無しの場合の IMSC ファイル
 Fig. 8 Output IMSC caption file without ruby.

表 3 漢字に関する情報
 Table 3 Parameters of kanji.

| 漢字 | 読み | 難易度 | 難易度基準 | 画数 | 当て字か |
|----|---------|-----|----------------------------|----|-------|
| 密林 | ジャングル | 11 | 当て字 | 19 | True |
| 智 | ち | 9 | JIS 第 1 水準 (日本漢字能力検定準 1 級) | 12 | False |
| 慧 | え | 9 | JIS 第 1 水準 (日本漢字能力検定準 1 級) | 15 | False |
| 僕 | ぼく | 8 | 高校生 | 14 | False |
| 兎 | うさぎ | 9 | JIS 第 1 水準 (日本漢字能力検定準 1 級) | 7 | False |
| 尻 | こがらし | 10 | JIS 第 2 水準 (日本漢字能力検定 1 級) | 6 | False |
| 月 | つき | 1 | 小学 1 年生 | 4 | False |
| 遅 | おそ(く) | 7 | 中学生 | 12 | False |
| 眠 | ねむ(い) | 7 | 中学生 | 10 | False |
| 起 | お(きる) | 3 | 小学 3 年生 | 10 | False |
| 喝 | かつ | 8 | 高校生 | 11 | False |
| 外 | そと | 2 | 小学 2 年生 | 5 | False |
| 眩 | まぶ(しい) | 10 | JIS 第 2 水準 (日本漢字能力検定 1 級) | 10 | False |
| 泣 | な(ける) | 4 | 小学 4 年生 | 8 | False |
| 腹 | はら | 6 | 小学 6 年生 | 13 | False |
| 減 | へ(る) | 5 | 小学 5 年生 | 12 | False |
| 黒洞 | ブラックホール | 11 | 当て字 | 20 | True |
| 蚊 | か | 8 | 高校生 | 10 | False |
| 刺 | さ(され) | 7 | 中学生 | 8 | False |
| 痣 | あざ | 10 | JIS 第 2 水準 (日本漢字能力検定 1 級) | 12 | False |

応モデルを得られた。結果、漢字の難易度、視聴者の年齢、同内容の音声の有無、漢字が当て字かの説明変数がルビ適応に向けて重要であることが分かった。

一方で、今回のモデルは実験条件のための制限がある。今回、漢字の画数はステップワイズ法により説明変数の候補から取り除かれた。しかし、今回の実験は 25 歳から 53

歳に実施しており、より低年齢や高齢の参加者が得られた場合、文字の表示サイズと画数により文字を視認できないなど、漢字の画数もルビの要/不要に影響する可能性がある。また、視聴者の年齢については、今回のルビ適応モデルでは年齢が上がるほどルビが必要となるという結果だった。しかし、学校において漢字を習得する学生の年齢にお



図 9 字幕の表示画面例

Fig. 9 An display example of captions.

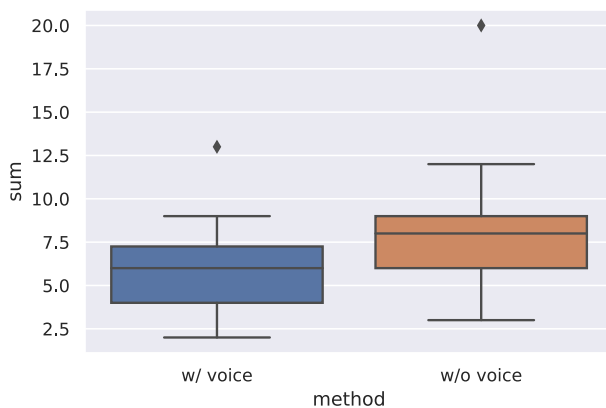


図 10 ルビが必要である漢字数の集計

Fig. 10 Boxplot of the number of ruby required kanji.

表 4 ロジスティック回帰の結果

Table 4 The results of logistic regression.

| 説明変数 | 係数 | p 値 | [0.025 | 0.975] |
|-------|--------|------------------------|--------|--------|
| 定数 | -10.5 | 3.33×10^{-22} | -12.6 | -8.37 |
| 年齢 | 0.0732 | 1.01×10^{-5} | 0.041 | 0.106 |
| 同内容音声 | -0.877 | 1.14×10^{-4} | -1.32 | -0.431 |
| 難易度 | 0.934 | 7.02×10^{-23} | 0.748 | 1.12 |
| 当て字 | 1.32 | 0.0414 | 0.052 | 2.59 |

いては、年齢が下がるほどルビが必要となることが容易に想像できる。より広い年代、より多くの漢字での検証が今後の課題である。

また、今回は重要な説明変数の特定のため、ロジスティック回帰による適用モデルを提案し検証した。さらなる予測精度の向上へ向け、より精度の高いモデルの適用、提案も今後の課題である。

6. おわりに

本稿では、聴覚に不安のないユーザーを対象とした字幕の適応、特に漢字のルビのユーザー適応に注目し、聴覚に不安のない成人を対象とした実験から、字幕と同内容の音声を取っているかが漢字に付与するルビの要不要に影響するかをウェルチの t 検定により明らかにした。また、実験結果をベースとした字幕漢字にルビを付与すべきかを予

測する適応モデルを提案し評価した。

今後の課題として、より広い年代、より多くの漢字での検証、より予測精度の高いモデルの提案があげられる。

参考文献

- [1] 阿部晋矢他: 個人向け字幕提示に向けた字幕利用調査, 映像情報メディア学会冬季大会 (2022).
- [2] Gorman, B. M. et al.: Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption, *Proc. of CHI 2021*, pp. 1–11 (2021).
- [3] 小林正幸他: 学年別ルビ付加機能を有するソフトウェアを利用した発話内容提示システムの構築と評価, 映像情報メディア学会誌, Vol. 62, No. 4, pp. 595–605 (2008).
- [4] 総務省: 放送分野における情報アクセシビリティに関する指針 (2018).
- [5] EBU: *Guidelines for Delivering Accessibility Services Using HbbTV*, Tech Report 065 (2021).
- [6] W3C: Making Audio and Video Media Accessible - Captions/Subtitles, <https://www.w3.org/WAI/media/av/captions/>.
- [7] BBC: BBC Subtitle Guidelines, <https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/>.
- [8] 小峯一見他: 内容理解度によるリアルタイムニュース字幕の評価, 映像情報メディア学会誌, Vol. 52, No. 4, pp. 504–511 (1998).
- [9] ARIB: デジタル放送におけるデータ放送符号化方式と伝送方式, ARIB STD-B24 6.5 版 (2022).
- [10] ARIB: デジタル放送におけるマルチメディア符号化方式 (第 2 世代), ARIB STD-B62 2.2 版 (2019).
- [11] W3C: TTML Profiles for Internet Media Subtitles and Captions 1.2, <https://www.w3.org/TR/ttml-imsc1.2/>.
- [12] W3C: WebVTT: The Web Video Text Tracks Format, <https://www.w3.org/TR/webvtt1/>.
- [13] Ofcom: Television Access Services, <https://www.ofcom.org.uk/consultations-and-statements/category-1/accessservs> (2006).
- [14] S4C: *Research into the Demand for Welsh Language Subtitling in Wales* (2001).
- [15] Hughes, C. J. et al.: Responsive Design for Personalised Subtitles, *Proc. of Web4All 2015*, pp. 1–4 (2015).
- [16] Armstrong, M. et al.: Understanding the Diverse Needs of Subtitle Users in a Rapidly Evolving Media Landscape, *SMPTE Motion Imaging Journal*, Vol. 125, No. 9, pp. 33–41 (2016).
- [17] W3C: RDF 1.1 Primer, <https://www.w3.org/TR/rdf11-primer/>.
- [18] Berners-Lee, T. et al.: The Semantic Web, *Scientific American*, Vol. 284, No. 5, pp. 34–43 (2001).
- [19] W3C: W3C Web of Things, <https://www.w3.org/WoT/>.
- [20] Urakawa, M. et al.: Rikamap-An Educational Application Using RDF-Formatted Learning Paths., *Proc. of ISWC 2017* (2017).
- [21] Fujii, S. et al.: Extracting Basic-and-Advanced Relationships Based on the Structure of Teaching Material Documents, *Proc. of GCCE 2022* (2022).
- [22] Blender Foundation: Big Buck Bunny, www.bigbuckbunny.org.
- [23] VOICEVOX, <https://voicevox.hiroshiba.jp/>.
- [24] 文部科学省: 小学校学習指導要領 (平成 29 年告示) (2017).
- [25] 日本漢字能力検定協会: <https://www.kanken.or.jp/kanken/>.