

字幕映像と吹き替え映像における 内容理解および視覚情報の認識に関する検証

小川莉莉香[†] 中島 悠貴[†] 八藤後美羽[†] 三上 弾[†]

[†] 工学院大学 情報学部 コンピュータ科学科
東京都新宿区西新宿 1-24-2

E-mail: [†]{jx20058,jx20193,jx20291}@g.kogakuin.jp, ^{††}mikami.dan@cc.kogakuin.ac.jp

あらまし 外国語の映像コンテンツにおける情報提供の方法として字幕と吹き替えが一般的である。マルチメディアコンテンツにおいては、シーン中から視覚的に取得できる情報（シーン情報）と、登場人物の発話などから得られる意味的な情報（内容情報）とがある。我々の興味は、字幕提示中にシーン情報を十分に得ることができるのか、という問題である。そこで、実験では、複数の、同一内容の字幕版、吹き出し版コンテンツを作成し、視聴後に、シーン情報、内容情報を正しく取得できているかを確認した。加えて、その時の視線の動きをアイトラッキングにより計測し、字幕、吹き替えによる視聴傾向の違いについて、詳細に分析を行った。その結果、エキスパートの付与した適切な字幕と、ボランティアによる字幕の場合、コンテンツジャンルの違いなど様々な要因で視聴スタイルが変化し、それにより、シーン情報、内容情報の把握に影響があることが示唆された。

キーワード 字幕、吹き替え、アイトラッキング、シーン情報

Can people understand semantic and visual information simultaneously in a subtitled video? –Verification by comparing dubbed video–

Ririka OGAWA[†], Yuki NAKAJIMA[†], Miu YATOGO[†], and Dan MIKAMI[†]

[†] 1-24-2 Nishi-shinjuku, Shinjuku-ku, Tokyo

E-mail: [†]{jx20058,jx20193,jx20291}@g.kogakuin.jp, ^{††}mikami.dan@cc.kogakuin.ac.jp

Abstract Subtitling and dubbing are standard methods of providing information in foreign-language video content. In multi-media content, there are two types of information: visual information obtained from scenes (scene information) and semantic information obtained from characters' speech (content information). We are interested in the question of whether it is possible to obtain sufficient scene information on subtitling content. In our experiment, we created multiple subtitled and dubbing versions of the same content and confirmed whether scene and content information could be obtained correctly after viewing the content. In addition, we used a glasses-type eye tracker to measure eye movements during viewing. We detailedly analyzed the differences in viewing tendencies between subtitled and dubbed content. The results suggest that viewing styles change depending on various factors, such as differences in content genres between appropriate subtitles provided by experts and those provided by volunteers. These factors affect the understanding of scene and content information.

Key words Subtitling, Dubbing, Eye tracking, Scene information

1. はじめに

映像コンテンツにおいて、登場人物の発話の理解が困難な視聴者に対する情報補償が必要なケースがしばしばある。例えば総務省聴覚障がい有する視聴者を含む視聴覚障害者等の情報アクセス機会の確保に向けた指針を定めている。また、コンテンツの多様化により、発話が母国語ではないなどの理由で理解

できないケースも多くある。この場合の対応策として大きく2通りが一般的である。第1は吹き替えであり、第2は字幕である。映画を吹き替えで見るか、字幕で見るかはしばしば議論がおこるほどどちらにも一長一短ある表現方法となっている。

さて、映画をはじめとした映像作品は、マルチメディアのコンテンツであり、主に視覚情報と聴覚情報とから構成されている。以降、人物の表情や風景など視覚的にとらえる情報をシー

ン情報、内容など文字的情報（マルチメディアコンテンツにおいては主に音声で提供されることが多い）を内容情報と呼ぶこととする。ここで、情報と情報源、情報の受容器官の関係で考える。内容情報の情報源は主に音声であり、自然な状態では耳で内容情報を受け取る。その際、目を用いてシーン情報の入手が可能である。吹き替えで視聴した場合、内容情報は変わらず音声で届けられ、耳から情報を受け取る。一方、字幕にした場合、この関係に変化が生じる。内容情報の情報源が音声から、テキストに切り替わり、視覚的に内容情報を取得することになる。そのため、本来シーン情報を獲得するために利用することが期待されていた視覚が、内容情報のために利用されてしまうのである。

このような問題意識の下、字幕提示では様々な工夫がなされてきている。江草らは字幕の提示位置を下部固定とするのではなく、発話者の顔の付近に吹き出しとして提示した。シーン情報の中でも重要な要素のひとつである表情と、内容情報とが隣接して提示されることになるため、情報取得に対する負荷が軽減するものと考えられる。江草らは、視線の移動量を評価尺度として評価している [3]。さらに、この文脈から、視線の移動量を軽減させるための研究としていくつかの手法が提案されている。例えば江草らは吹き出し型字幕による情報保証と、TV、映画、演劇などで一般的に用いられる提示位置が固定された字幕による情報保証の特性的な違いを定量的に表すことで、吹き出し型字幕の有効性を検証した。吹き出し型字幕は、上部字幕による情報保障を行う映像と比べ、参加者の映像鑑賞中における視線移動量の軽減に寄与することが示された。さらに、赤堀らは、視線追跡データを用いて映像の中で視線が集まりやすい領域を算出し、その領域の下部に字幕を配置することで、視線移動の少ない字幕付き映像の鑑賞方法を提案した。その結果、実験で用いた多くの映像において、字幕を画面端に固定して配置をするより提案した手法を使うことによって、映像と字幕との双方を快適に鑑賞できることが明らかになった。

しかしこれらの研究は、主に聴覚障害者に対する情報補償を念頭においており、シーン情報と内容情報の双方の理解度には注目するのではなく、視線移動量を尺度として、鑑賞における負荷が評価されている。これに対して本研究では前提として字幕映像を鑑賞した場合と、吹き替え映像を鑑賞した場合で、シーン情報および内容情報の理解度にどれほどの差が生じるか検証することを目指す。そこで、本研究では、同一内容の、字幕と吹き替えにおけるシーン情報および内容情報の理解度の差を複数の映像とアンケートを用いて調査する。

2. 関連研究

Yu らは、字幕の存在によって視聴の仕方がどのように変化するかを分析している [2]。Yu らの実験では、英語音声に対して、中国語字幕、英語と中国語の2か国語字幕、字幕なしの3種類の字幕タイプを用意し、コンテンツとしてシーン情報が多いもの（HS）とシーン情報が少ないもの（LS）、各3分を用意し、被験者実験を行っている。被験者実験では、視聴者の視線情報を取得しながらコンテンツを視聴させ、視聴直後に、シー

ン情報と内容情報に関するテストを実施している。

それによると、LSでのシーン情報に関するテストは、HSでのそれに比べて低下するという結果が得られた。ここから、LSではシーン情報が少ないことから、字幕を読むことに視覚が活用されたと考えられる。その結果として当然ながら、LSでの内容情報のスコアは、HSの場合よりも高いことも確認されたとされている。しかしながら、Yu らの取り組みでは、字幕の存在の、シーン情報の取得への影響は詳述されておらず、「（シーン情報の取得のために字幕の存在が）必ずしもいつも有益であるわけではない」と述べるに留まっている。

本研究は、Yu らの取り組みを参考に以下の2点を拡張することで、コンテンツ視聴時の視聴覚の役割分担をより詳しく分析することを狙うものである。第1は吹き替え映像の導入である。Yu らの取り組みは、すべて英語音声を使っており、音声のみから内容情報を取得することが困難な状況であったと推測される。ここに、吹き替え映像を加えることで、母国語における日常に近い状況を評価対象として加えることが可能になる。第2は、コンテンツジャンルの多様化である。Yu らの取り組みは、シーン情報の多いコンテンツ（HS）と少ないコンテンツ（LS）と述べるに留まっており具体的にどのようなコンテンツを用いたのか定かではない。今回、実験に際し、技術解説、主に子どもが視聴対象となる映画、主に大人が視聴対象となる映画の3種類を用いることで、より詳細な分析を可能とする。

3. 実験

吹き替え・字幕による視聴行動の変化とシーン情報および内容情報の理解度を把握するために実験を行う。まず、実験を実施するにあたっての準備を以下に示す。

3.1 コンテンツ準備

本実験では、4種類のコンテンツを用意した。第1のコンテンツは技術解説コンテンツである。これは、「What AI is – and isn't」というタイトルのTED Talkより一部を抜粋したものである。これは2名による対話コンテンツであり、オリジナルは英語である。字幕はボランティアにより付与されたものが存在している。ただし、字幕版と吹き替え版との違いをなくしつつ、シーン情報として評価しやすい挙動を埋め込むために、以下の変換を加えた。まず、対話を行う2名の動きを、Animaze by FaceRig¹を用いてアライグマおよびネコのアバターに置き換えた。次に、瞬きやしっぽの動き、舌を出すなど特徴的な動作の頻度を多く設定した。これにより、吹き替え版、字幕版でシーン情報を全く変えることなく、シーン情報に含まれる情報を増やしたコンテンツを作成することができた。図1に実際に作成したコンテンツのスナップショットを示す。このようにネコとアライグマが横に並んで対話しているコンテンツとなっている。図2に、対話シーンに字幕を追加した例を示す。ネコによる発言には水色字幕、アライグマによる発言にはオレンジ色の字幕を表示した。

残りの3種類のコンテンツは、吹き替え版、字幕版が備わる

(注1)：Holotech Studios Inc.



図1 animaze を用いた吹き替え映像の一部



図2 animaze を用いた字幕映像の一部

映画コンテンツを利用した。最も子ども向けと思われるものは「くまのプーさん」であり、複雑な内容情報は含まないと考えている。一方で、最も内容情報が難しいと考えられるものが「ガーディアンズ・オブ・ギャラクシー」である。これはマーベル社の SF アクションであり対象とする年齢も「くまのプーさん」より高く、より複雑な内容情報を含む。そして、その中間をイメージした「ミスターインクレディブル・ファミリー」を選定した。各々 1.5 から 2.5 分程度の部分を切り出して実験に使用した。以降、「ミスターインクレディブル・ファミリー」を第 2 コンテンツ、「くまのプーさん」を第 3 コンテンツ、「ガーディアンズ・オブ・ギャラクシー」を第 4 コンテンツという。

3.2 視聴環境および視聴後テスト

視聴および視聴後テストを円滑に実施するために、Google Forms を活用した。Google Forms では YouTube 動画の再生が可能となっている。動画再生画面および視聴後テスト画面の例を図 3、4 に示す。なお、各々のコンテンツについて、シーン情報と内容情報とを問うテストを行っているため質問はコンテンツ毎に異なるが、問題のサンプルを示す。これは、技術解説コンテンツでの例である。

問 1 動画の中で右側にいた動物に気になる動きは何かありましたか？

問 2 問 1 に対してどこが気になりましたか？

問 3 動画の中で左側にいた動物に気になる動きはありましたか？

問 4 問 3 に対してどこが気になりましたか？

問 5 動画の内容を簡単に説明してください。(難しい場合はキーワードを記入してください)

ここで、問 1、問 3 がシーン情報に関する問であり、問 5 が内



図3 動画視聴画面

字幕

[字幕] 問 1 動画の中で右側にいた動物に気になる動きは何かありましたか？

☐ はい

☐ いいえ

問 2 どこが気になりましたか？
問 1 で「はい」を選んだ方は回答してください。
複数ある場合は複数記述してください。

回答を入力

[字幕] 問 3 動画の中で左側にいた動物に気になる動きは何かありましたか？

☐ はい

☐ いいえ

問 4 どこが気になりましたか？
問 3 で「はい」を選んだ方は回答してください。
複数ある場合は複数記述してください。

回答を入力

問 5 動画の内容を簡単に説明してください。
難しい場合はキーワードを記入してください。

回答を入力

図4 視聴後テスト画面

容情報に関する出題である。詳細はコンテンツ毎に異なるが、いずれのコンテンツについても、シーン情報に関する問、内容情報に関する問を含めた複数の設問を用意している。

なお、視聴に用いるディスプレイは、なるべくノート PC、デスクトップ PC を用い、スマートフォンやタブレットなど画面が小さいものを利用しないように要望を行ったがそれ以上の制約は与えていない。アンケートに基づいてデバイスの確認は行っており、タブレット・スマートフォンでの回答はないことを確認しているが正しく回答しているか、さらにノート PC の場合の画面サイズなどは不明である。

3.3 被験者

実験は日本語を母国語とする 20 歳から 23 歳を被験者の対象

とし、第1コンテンツでは43名、第2コンテンツでは28名、第3コンテンツでは28名、第4コンテンツでは19名で実施した。

3.4 アイトラッキング

一部の被験者に対して、視聴中の視線情報を取得するためアイトラッキングを実施した。アイトラッキングには、Pupil Labos 社製の Pupil Core を用いた。一般的なメガネ式のアイトラッカである。本実験の際には、43-inch の液晶ディスプレイを使用し、被験者にはディスプレイ全体を見ることができる距離に着座させた。

3.5 実験手順

各被験者は、吹き替え条件もしくは字幕条件のいずれか1条件でコンテンツを視聴し、その後続くテストに回答した。映像を繰り返し視聴することは許していない。ただし、回答にあたっての制限時間は設けていない。

3.6 結果

3.6.1 技術解説コンテンツ

図5と6にシーン情報に関する質問の結果を示す。これはシーン情報のひとつと考えられる「キャラクタに気になる動きがあったか?」という問に対する回答であり、図5は右側(アライグマ)。図6が左側(ネコ)についての回答である。また、両図において、左が字幕でコンテンツ視聴した群の結果であり、右が吹き替えでコンテンツ視聴した群における結果である。図5, 6より、左右のキャラクタで傾向に大きな違いが出たことが分かる。

図5に示したアライグマのキャラクタについては、吹き替え版視聴群の方が、シーン情報に対して注意が向いていたことが伺える。シーン情報の具体例に関する調査においても、吹き替え版視聴群では、舌を出す、ウィンクなど幅広い表情表出を確認できていることが分かった。

一方、図6から、ネコのキャラクタにおいては字幕動画と吹き替え動画とで気になる動きの有無に差がないことが分かる。この原因については考察において後述する。また、内容理解に関しては、字幕・吹き替えのどちらの映像を視聴した被験者も適切なキーワードを回答しており、理解度において大きな差異はなかったものと判断している。

3.6.2 映画コンテンツ

第2コンテンツを視聴した被験者における内容情報に関する問の正解率とシーン情報に関する問の正解率を図7に示す。同様に、第3コンテンツ、第4コンテンツの結果を、それぞれ図8、および図9に示す。

まず、各コンテンツにおける吹き替えと字幕での比較を述べる。内容情報に関しては、吹き替えでも字幕でも適切に取得可能であり、同等のスコアが期待される。今回の実験の範囲においては、ほぼ同等であるものの、字幕版において若干のスコアの向上が見られた。一方、シーン情報は、字幕を読むことでシーン情報の取得が困難となることを想像していた。しかしながら、結果は、想像に反し、字幕版において大きなスコア上昇がみられた。

次に、コンテンツ難度による変化について確認する。著者ら

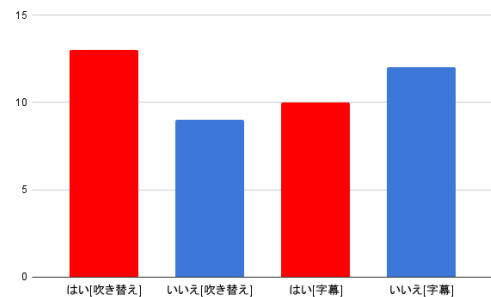


図5 第1コンテンツのアライグマに関する回答の集計結果 (左: 吹き替え 右: 字幕)

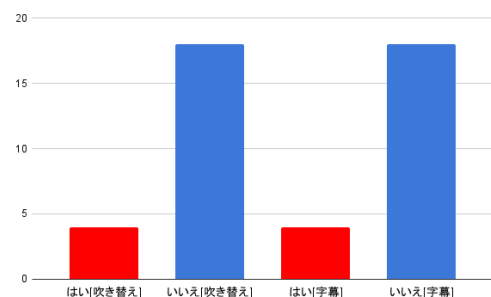


図6 第1コンテンツのネコに関する回答の集計結果 (左: 吹き替え 右: 字幕)

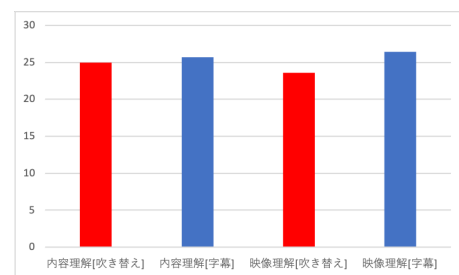


図7 第2コンテンツのスコアの結果

(左: 内容理解の比較 右: シーン理解の比較 赤: 吹き替え 青: 字幕)

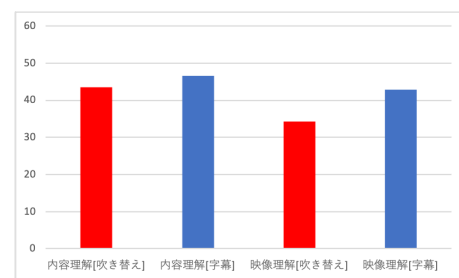


図8 第3コンテンツのスコアの結果

(左: 内容理解の比較 右: シーン理解の比較 赤: 吹き替え 青: 字幕)

は、コンテンツの内容が高度になると字幕を読み理解することに要する時間が長くなり、シーン情報の取得に割くことができる時間が短くなる可能性を考えていた。そこで、コンテンツの難易度を、第2, 第3, 第4と高めた。しかしながら結果は、難易度が高まるにつれて、字幕版のスコア上昇が顕著となった。

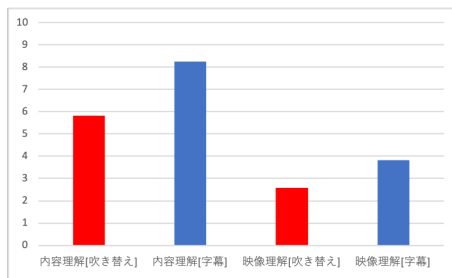


図9 第4コンテンツのスコアの結果
(左:内容理解の比較 右:シーン理解の比較 赤:吹き替え 青:字幕)

4. 考 察

4.1 アイトラッカーを用いた検証

結果を詳しく調査するために、アイトラッキングにより、吹き替え、字幕の双方における視線の動きの違いについて検証を行った。図10、11に第1コンテンツに対するアイトラッキングの結果を示す。図10から、視線がほとんどずっと字幕上に停滞していることが分かる。一方で、吹き替えでは図11から分かる通り、常にアライグマの口元に視線が向いていたことが確認された。

これには2つの要因が考えられる。第1に、対談であることに被験者が気づき、内容情報が重要と判断し、自然と内容情報にフォーカスするような視聴方法を選択した可能性がある。これにより画面内の様々な部分を確認するような視聴の仕方をしなかった可能性がある。第2に、発話者に注視が集まった可能性である。このコンテンツでは、大部分の時間を、アライグマが発話している。そのため吹き替え版視聴者は、発話者であるアライグマの口元を注視していたものと思われる。どちらもコンテンツ視聴時の自然な挙動であると考えられ、双方の影響から吹き替え版において、ネコに関するシーン情報に注目が向かなかったと推測している。

別の例として、第2コンテンツの1シーンを図12と図13に示す。なお、コンテンツ権利の観点から、図12と図13は、実際の視線提示画像をトレースしたものとなっている。当該シーンでは、3つのオブジェクトが存在し、字幕は中央下部に提示されている。字幕版では、字幕を含めて、画面上の主要なオブジェクトに対して視線が向いているのに対して、吹き替え版においては、字幕部分には視線が向いていないことが見て取れる。字幕以外に視線が向く場所は大きな違いがないことも見て取れる。

ここで注目したいのは、字幕の短さと字幕への滞留時間の短さである。しっかりと制作された映画コンテンツにおいて字幕は、専門家によりカットの時間長毎の文字数など厳しい制約の下で適切に作成されている。そのため、時間的制約の下であっても、十分にシーン情報を取得するだけの余裕があったものと考えられる。

4.2 文字情報として得られることによる定着効果

本実験では、内容情報についても、シーン情報についても、字幕版においてスコアが上昇する傾向が得られた。ここについ



図10 アイトラッキングを用いた字幕版第1コンテンツ



図11 アイトラッキングを用いた吹き替え版第1コンテンツ



図12 アイトラッキングを用いた吹き替え版第2コンテンツ

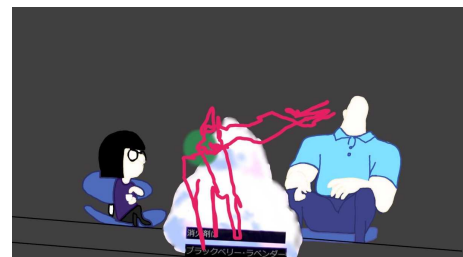


図13 アイトラッキングを用いた吹き替え版第2コンテンツ

ては正確な判断が可能な情報がそろっているとは言えない。しかしながら、実験を通じた主観として、文字情報として情報を得ることによる定着への効果があると考えている。シーンの中に画像として出ているものが目に入っただけでは次の瞬間に忘れてしまうものが、言語として取得することで記憶しやすくなる。吹き替え音声においても、聞くことよりも、字幕文字を読むことの方が記憶しやすくなる可能性があるのではないかと考えている。

5. ま と め

本研究では、同一内容の字幕および吹き替えコンテンツを用

意し、内容情報およびシーン情報の理解について調査を行った。字幕においては、内容情報を入手するために視覚を利用してしまいうために、シーン情報の把握レベルが低下するという仮説のもの調査を行った。しかしながら、映画コンテンツにおいては、そのような傾向はみられず、むしろ字幕が提示されていることにより記憶の定着が進み、質問に対する正解率という観点ではスコアの向上が見られた。

一方で、当初想定通りとなったコンテンツもあった。それは、技術解説コンテンツであった。これには2つの要因があると思われる。第1が技術内容の紹介がメインのコンテンツであり視聴時にシーン情報の取得が不要という意識が自然と働いたこと。第2が、字幕の付与がプロではなくボランティアによる翻訳であったため、視聴者に負荷のかからない字幕という観点では適切な字幕ではなかった可能性がある。

しばしば、映画を字幕で見るか、吹き替えで見るかという論争が行われるが、今回の検討の範囲においては、どちらも十分に内容理解もシーン理解も可能であると考えられる。ただし、TED Talk や、YouTube など近年しばしばみられる、ボランティアなどによる字幕付与や自動翻訳においては、字幕付与のエキスパートのものと違い適切な文字数での字幕提示ではない可能性がある。そのようなケースでは、シーン情報の見落としのリスクが増大する可能性が高い。

今後より広範なコンテンツジャンル、また、バラエティ番組などにおけるテロップなどについても内容理解に対する効果を検証していきたい。

文 献

- [1] T.-Y. Lin et al., Microsoft COCO: Common objects in context, In Proceedings of the European Conference on Computer Vision (ECCV), pages 740–755, 2014.
- [2] Yana Yu et al., Influence of Bilingual Subtitles on Attention Allocation and Information Acquisition during Film Viewing, 日本認知心理学会第18回大会, 2020.
- [3] 江草ら, 視線計測装置を用いた吹き出し型字幕提示法の視線移動量低減効果に関する有効性の評価, ヒューマンインタフェース学会論文誌, 21 巻, 4 号, pp.381-390, 2019.
- [4] Z. Cao et al., Openpose: Realtime multi-person 2D pose estimation using part affinity fields, IEEE T.PAMI, 2019.