

# ChatGPT を用いた e スポーツ専門用語自動解説システムの開発と評価

井上奏太<sup>†1</sup> 梅原旅詩<sup>†1</sup> 川合康央<sup>†1</sup>

**概要**：本研究は、e スポーツのライブストリーミング中に専門用語を自動で解説するシステムの開発と評価を行ったものである。ストリーマーや視聴者に操作を求めることなく、リアルタイムで専門用語の解説を提供することで、ストリーマーのゲームプレイへの集中力を維持しつつ、初心者視聴者の知識向上を支援することを目的とした。評価実験ではシステムの解説精度と視聴者満足度を検証し、提案システムが視聴体験向上に有効であることが示された。

## 1. はじめに

### 1.1 研究背景

近年、esports 業界は大きな盛り上がりを見せている。市場規模は拡大の一途を辿っており、オリンピック e スポーツゲームズも開催されることが決まり、より大きな盛り上がりを見せようとしている。また e スポーツの発展（ゲームの購入やプレイ）には、観戦することが強い影響を与えることが指摘されている[1]。特に、個人によるライブストリーミングは従来の esports イベント放送よりも視聴者にとってアクセスしやすく、双方向のコミュニケーションを可能にすることが確認された[2]。また、ライブストリーミングを魅力的にするための要素の一つとして、リアルタイムでの解説やコメンテーターのコメントが観戦体験を豊かにすることが示された[3]。一方で、現在主流である、ストリーマー 1 人でのライブストリーミングでは、ストリーマーのパフォーマンスが低下することが示されている[4]。主な原因として、ゲームプレイと同時に実況を行うことが、認知的に負荷が高いことが挙げられる。具体的には、プレイ中にゲームの進行を説明したり、視聴者と対話したりすることが集中力を分散させ、ゲーム内での判断力や反応速度が低下する可能性がある。難解な戦術の説明や、視聴者との対話などは、視聴者が求めているコンテンツであるが、初心者に向けた簡単な用語の解説は、ウェブサイトでも学ぶことが可能である。しかし、ウェブサイトを用いて検索などを行う場合、本来視聴したかったライブストリーミングから離れることになり、ユーザー体験を損失する可能性がある。そこで筆者は、ストリーマーが一人で扱うことができる、視聴者に操作を必要とせず、自動で専門用語を解説するシステムがあることが望ましいと考えた。

### 1.2 研究目的

本研究は、ストリーマーが発した専門用語に関する解説文を、自動で生成するシステムを開発し評価する試みである。特に、ストリーミング中にストリーマーに操作を求め

ないこと、視聴者に操作を求めないことの 2 点に焦点を当てて開発を行った。このシステムを用いることで、視聴者へ向けた簡単な用語の説明を自動化し、ストリーマーのゲームへの集中や、視聴した初心者知識のステップアップによるゲームへの参入を促すことを目的としている。

## 2. 先行事例

### 2.1 視聴者の手を介する動的な解説生成に関する研究

下田ら[5]は、視聴者が専用のビューワーを操作し、理解度に合わせた動的な専門用語の解説を閲覧するシステムの提案と評価を行った。専門用語を解説する方法として、YouTube 動画プレイヤーを組み込んだ自作のビューワーを用いた。このビューワーは、動画内での専門用語の発出に伴い、画面に「用語タブ」を表示し、このタブの操作状況を用いて視聴者の理解度を推定する。また、測定された理解度に応じ、用語タブの自動提示の有無や自動提示した際にタブを展開するか畳むかを判定する。評価として、画面に表示された動的な解説は視聴者にとって望ましいものであり、YouTube などの既存ビューワーより専門用語に対する視聴者の理解に寄与すると述べている。また、視聴中に検索エンジン等で不明点を調べるより、ビューワーによる動的な専門用語の解説のほうが視聴者の理解に寄与すると述べている。一方で、問題点として試合観戦が忙しく専門用語の解説やタブの操作が行えなかったことが示された。

### 2.2 動画に付随するコンテンツが視聴者に与える影響に関する研究

梶並[6]は対戦型格闘ゲームを対象にした、動画コンテンツを用いたプレイ分析支援について、2 つの手法を提案した。1 つ目は時系列に応じた戦略的思考の変化をグラフ表現する手法、2 つ目はプレイヤーの着目個所を基本図形の組み合わせで表現する手法である。梶並ら[7]は、これから発展し、対戦型格闘ゲームにおける初心者向け観戦支援を目的とした、キャラクターの位置情報を基に試合状況をアノテーション（注釈）で視覚的に表示するシステムを提案

した。実験では、遠距離、近距離、画面端という3つの典型的な試合状況に応じて、文字や図形を用いたアノテーションを試合映像に重ねることで、攻防の戦略や場所の理解、展開の予測などのしやすさが向上するのではないかと仮説を立てた。検証結果として、文字と図形のアノテーションの提示が攻防のポイントの理解の支援に有用であると示した。一方で画面が煩雑になり、状況理解を妨げる可能性があるとし、観戦者の観戦レベルに応じたアノテーションの提示方法を検討する必要があるとした。

## 2.3 観戦によるeスポーツの発展に関する研究

Maceyら[1]は、eスポーツ観戦およびビデオゲームをプレイする人々194人に構造方程式モデリングを用いて調査を行った。この調査により、観戦意図と購入意図の関係がゲームプレイ意図によって完全に媒介されていることが示された。また、観戦動機とゲームプレイ動機には相関がある可能性があるとし、eスポーツ観戦の満足度を高める要因（例：社会的交流や試合のドラマ性）についての研究が今後重要になると示した。

Jangら[3]は、eスポーツのプレイ経験を持つ598人を対象にオンライン調査を行った。調査では、eスポーツのレクリエーションプレイ、eスポーツコンテンツのライブストリーミング、eスポーツイベント放送（eスポーツ大会や競技の放送）、ストリーマーとの識別、プロプレイヤーとの識別の5つの次元に基づく15項目を評価した。この調査により、eスポーツコンテンツライブストリーミング消費が、eスポーツのプレイ意図とイベント放送消費意図をつなぐ重要な媒介役を果たしていることが示された。

これらの研究により、eスポーツのライブストリーミングなどのコンテンツ提供を強化することが、ゲームプレイや関連商品の購入促進につながる可能性があることが示された。

## 2.4 対戦型格闘ゲーム初心者に対する支援に関する研究

対戦型格闘ゲーム初心者に対する支援では、前述した梶並ら[7]の研究のほかに、初心者へ行動の指針を提示する形で支援する研究が行われた。

酒見ら[8]は、初心者が定着しにくい原因の一つである「読み合い」に焦点を当て、格闘ゲーム初心者に向けた「読み合い」理解支援システムを提案した。読み合いとは、対戦相手の行動を予測し、それに応じて自分の行動を決定する駆け引きのことである。初心者がこの読み合いを学ぶため、自作の簡略化されたゲームとカリキュラムを通じて理解を支援する方法を提案している。提案された「拡張じゃんけんゲーム」は、格闘ゲームの読み合い要素を単純化し、リアルタイム性を排除することで、初心者でも集中して学べるように設計されている。このゲームでは、打撃、投げ、防御の3種類の行動を選び、相手との読み合いを行う。プレイヤーは相手の行動を予測し、行動を決定することで進行する。また、ゲームを通じて学べるよう、ヒントや行動

パターンの提示が行われ、段階的に難易度が上がるカリキュラムも用意されている。

実験結果では、初心者の相手行動予測に対する自己評価が向上したものの、定量的評価では有意差が見られなかったとした。

天川ら[9]は、格闘ゲーム初心者の技術向上を支援するため、試合後の振り返り（感想戦）を支援するシステムを提案した。このシステムでは、初心者によく見られるミスプレイを指摘し、適切なアドバイスや再挑戦の機会を提供することを目的としている。実験では、ミスプレイリストの表示、ミスに対するアドバイスの表示、悪質行為への警告機能について評価が行われた。結果として、ミス検知とアドバイス内容についてはおおむね肯定的な評価、表示の分かりやすさや操作性への評価はやや低いとした。

## 2.5 本研究の位置づけ

下田らや梶並らが動画に付随する動的な解説やアノテーションが視聴者の動画内容の理解に有用であることを示している。一方で、先行研究ではeスポーツ発展の重要な媒介役である、ライブストリーミングに対して解説を付随する手法をとっていない。本研究では、ライブストリーミングに対してリアルタイムで解説を表示するシステムを提案する。

## 3. システムの提案

本章では、ChatGPTを用いたeスポーツ専門用語自動解説システムである「auto-term」について述べる。

また、本研究ではゲームタイトルとして株式会社カプコンがリリースした「ストリートファイター6」を対象とした。分析の対象とするゲームタイトルは、先行研究が行われた対戦型格闘ゲームであり、eスポーツ大会開催の実績があるものとした。よって、JeSuプロライセンス公認タイトル[10]であり、プロリーグである「ストリートファイターリーグ」[11]や「Capcom Pro Tour」[12]が開催されている本作品を採用した。

### 3.1 auto-termの概要

本研究で開発するWebアプリケーションのauto-termについて説明する。本システムの開発はWindows 10 Homeで行うこととし、統合開発環境としてVisual Studio Code、開発言語としてPython、JavaScript、HTML、SCCを用いた。また、主な使用APIやフレームワーク等はFlask、ChatGPT、WebSpeechAPI、SQLiteとなっている（表1）。

表 1 本システムの開発環境及び主な使用 API 等

種別	詳細
OS	Windows 10 Home
統合開発環境	Visual Studio Code
開発言語	Python JavaScript HTML SCC
API	ChatGPT (解説生成) WebSpeechAPI (音声認識)
フレームワーク	Flask
データベース	SQLite

また、auto-term のフローは次のとおりである (図 1)。

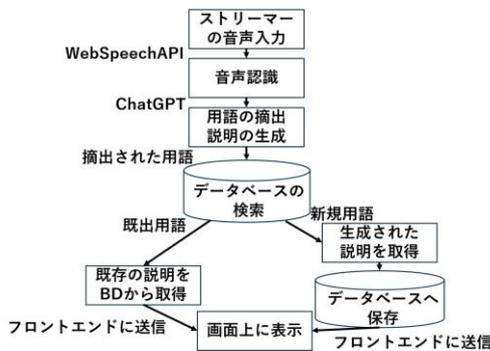


図 1 システムフロー図

音声認識は WebSpeechAPI を用いて行い、認識された文章を次のようなプロンプトに変換する (表 2)。

表 2 用いるプロンプト

以下の[]内の文章に含まれる (ゲーム名) の専門用語を 1 つだけ、1 行で解説してください。複数ある場合は、最も後に出てきたものを優先してください。その際、a= 専門用語、b=解説とすると、「a : b」となるようにしてください。ない場合は、「生成中」と出力してください。  
[ (認識された文章) ]

(ゲーム名) には設定されたゲームの名称が代入され、(認識された文章) には認識された文章が代入される。このプロンプトは設定された秒数ごとに、その間隔で認識されたすべての文章が代入され、ChatGPT に送信される。設定については 3.3 設定の節にて後述する。ChatGPT によって専門用語が認識された場合、認識された用語がデータベースにある場合は保存されている説明を画面に表示する。ない場合は新たにデータベースに登録後、画面に表示する。

また、専門用語が認識されなかった場合は画面の切り替えは行われない。

### 3.2 ホーム画面について

以下の図がホーム画面である (図 2)。

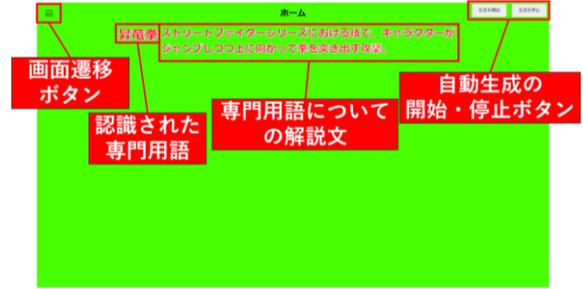


図 2 Web アプリケーションのホーム画面

画面右上のボタンを押すことで、音声認識の開始と停止を行うことができる。専門用語が認識された場合、画面中央に認識された専門用語と解説文が表示される。画面をグリーンバックに設定可能なため、背景透過を用いてストリーミングに組み込むことができる。また、画面左上のボタンを用い、「ホーム」「設定」「データベース」のページに遷移することができる。

### 3.3 設定について

設定では以下の項目を設定することができる (表 3)。

表 3 設定項目

設定項目	詳細
ゲーム名	プロンプトに代入される文字の設定
背景色	ホーム画面の背景色の設定
フォント	以下の 5 つに設定可能 ・ Arial ・ Kosugi Maru ・ Kaisei Decol ・ Noto Sans JP ・ Sawarabi Mincho
フォントサイズ	pt で設定
文字の色	RGB, スポイトで設定
文字の縁の大きさ	pt で設定
文字の縁の色	RGB, スポイトで設定
影の有無	有無を設定
影の色	RGB, スポイトで設定
解説生成の頻度 (秒)	ChatGPT への送信頻度
文字の移動の有無	・ 解説文を横にスクロール ・ 全文を固定して表示する
解説文の横の長さ	画面全体に対する % で設定

このように、利用するストリーマーが自分に合わせた画面を構成できるよう項目を設定した。設定を変更することで、サンプルテキストが更新される (図 3)。

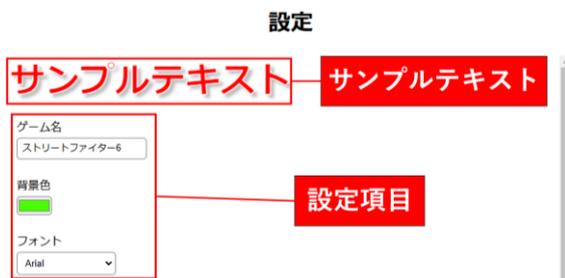


図 3 Web アプリケーションの設定画面

また、文字の移動を「無」に設定し、解説文の横の長さを短くすることで、縦長に表示した際の可読性を上げることができる。

### 3.4 データベースについて

データベースは前述したとおり SQLite を使用した。画面は以下のようになっている (図 4)。



図 4 Web アプリケーションのデータベース画面

データベースに登録された専門用語が一覧表示されており、専門用語をクリックするとデータの編集、削除ボタンでデータの削除、画面上部からデータの追加が行える。生成された解説が間違っており訂正を行いたい場合や、事前にデータを追加しておきたい場合に活用が可能である。

### 3.5 想定したストリーミングでの使用方法

想定した利用方法は、OBS Studio などの配信ソフトウェアを用い、ストリーミングの画面に解説文を合成する方法である。実際に OBS Studio を用い、画面を作成した (図 5)。この図では画面下部に横長で表示されているが、3.3 節で前述したとおり、画面右側のスペースに縦長で表示することも可能である。



図 5 ストリーミングでの使用例

## 4. 評価実験

3 章で提案した auto-term の評価実験を行った。本章では実験結果及び検定結果について述べる。

### 4.1 実験の概要

評価実験の目的は、本システムがどの程度の精度で専門用語を解説できるか、および音声認識によるリアルタイムの解説は視聴者にとって望ましいものであったかの評価である。この実験は大きく 2 つに分かれており、システムの精度を評価する実験と、システムを用いた動画を視聴後アンケートに回答する実験の 2 つである。

### 4.2 システムの精度評価

システムの精度評価実験として、ストリーミングを模したストリートファイター6での対戦を 3 人で 5 時間行った。また auto-term の設定は、ゲーム名「ストリートファイター」、解説生成の頻度「30 秒」とした。また、この評価における専門用語は、以下の項目とする。

- ・ストリートファイターに登場するキャラクター名
- ・ストリートファイターに登場する技名
- ・格闘ゲーム全般における用語 (戦術的に意味がある語)

格闘ゲーム全般における用語については明確な線引きが難しく、筆者の主観が含まれることに注意が必要である。また ChatGPT は時期による回答の変化が考えられる。なお、本実験は 2024 年 11 月 5 日に行った。

実験中に発言した専門用語と、WebSpeechAPI が認識した専門用語をリスト化し、Levenshtein 距離 (挿入語数 + 置換語数 + 削除語数) を測定、発言した語数を用いて標準化し類似度を算出した (表 4)。

表 4 発言した語と認識した語の類似度

Levenshtein 距離 ÷ 発言した語数 (発言した語数 = 1097)	0.522
--	-------

川原[13]によると、人間が見て意味が分かるレベルは 75%~、

発言内容が理解できるレベルは 85%~、放送などの字幕付与に必要なレベルは 95%~である。よって、WebSpeechAPI による専門用語の音声認識は実用的ではないことが示された。

次に WebSpeechAPI が認識した専門用語を、ChatGPT が専門用語として認識した回数と確率について示す (表 5, 表 6)。

表 5 ChatGPT が専門用語を認識した回数

項目	回数
WebSpeechAPI が認識した専門用語 (a)	89 回
ChatGPT が解説した専門用語 (b)	56 回
ChatGPT が解説した専門用語以外の語 (c)	12 回

表 6 ChatGPT が専門用語を認識した確率

専門用語が解説された確率 表 5 より $(b \div a)$	0.629
専門用語以外が解説された確率 表 5 より $(c \div (b+c))$	0.176

表 5 の a は実験中に専門用語を発言したすべての回数ではなく、認識される 30 秒の間に 1 回以上専門用語が認識された回数である。専門用語が解説されるべきタイミングで解説した確率は 62.9%、ChatGPT が専門用語以外の語や意味が通じない語 (実例:リターン型系リサルト 3500, 鳴かないんだよな) を解説した確率は 17.6%となった。このことから、ChatGPT による解説は完全なものではなく、誤った用語を抜き出す可能性があることが示された。

### 4.3 システムのアンケート調査

4.2 章の結果により、音声認識及び ChatGPT による解説の精度が低く、現時点では実用的ではないと判断した。よって、アンケートに用いる動画では音声認識は用いず、動画内で発話される専門用語を人力で抽出し、ChatGPT に解説させた。この解説を用い、本来の動作と同じになるように字幕を付与した動画を作成した。

実験参加者は 20 代の男女 21 人であり、Google Form で作成したアンケートフォームを利用した。実験では、字幕を付与した動画と、字幕を付与していない動画を視聴してもらい、各動画視聴終了後に対応したアンケートに回答してもらった。なお、実験は順序効果を打ち消すため、実験参加者をグループ 1 (字幕なし→字幕あり) とグループ 2 (字幕あり→字幕なし) に分けた。

まず、動画視聴後アンケートの、動画を楽しめたかの結果を示す (図 6)。

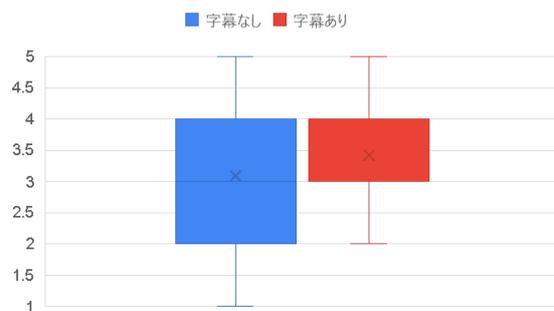


図 6 動画を楽しめましたかの箱ひげ図

字幕なしの平均値は 3.095、標本不偏分散は 0.990、字幕ありの平均値は 3.429、標本不偏分散は 0.557 である。また、ウィルコクソンの符号順位和検定 (有意水準 5%) の結果、p 値は 0.020 となり、2 群の順位平均値に有意差があると認められた。

次にアンケートの、動画内で発言された専門用語が理解できたかの結果を示す (図 7)。

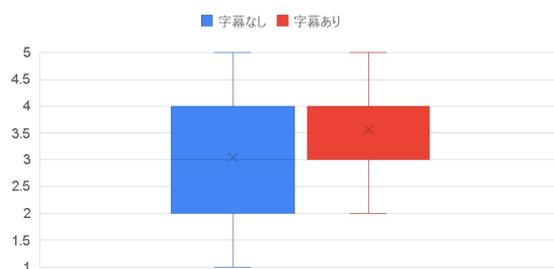


図 7 動画内で発言された専門用語が理解できたかの箱ひげ図

字幕なしの平均値は 3.048、標本不偏分散は 1.848、字幕ありの平均値は 3.571、標本不偏分散は 0.957 である。また、ウィルコクソンの符号順位和検定 (有意水準 5%) の結果、p 値は 0.012 となり、2 群の順位平均値に有意差があると認められた。

次にアンケートの、字幕で専門用語の理解が進んだかについての結果を示す (図 8)。

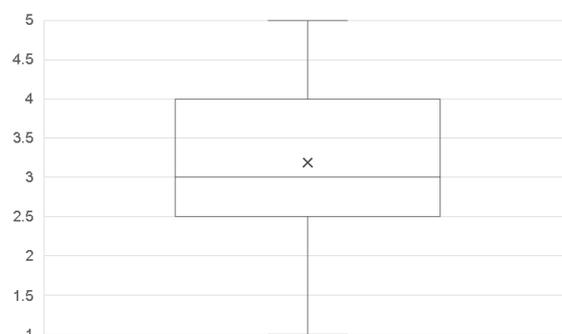


図 8 字幕で専門用語の理解が進んだかの箱ひげ図

平均値は 3.190 であり、実験参加者は概ね字幕により

専門用語の理解が進んだと感じていた。

次にアンケートの、解説されている専門用語はその時必要な情報であったかどうかについての結果を示す (図 9)。

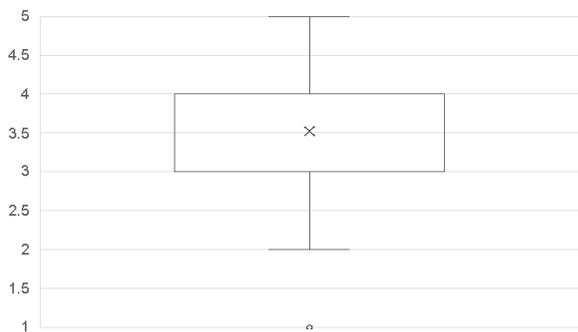


図 9 解説されている専門用語がその時必要な情報であったかの箱ひげ図

平均値は 3.524 であり、実験参加者は概ね必要な情報が表示されたと感じていた。

次にアンケートの、字幕はあった方が楽しめると感じたかの結果を示す (図 10)。

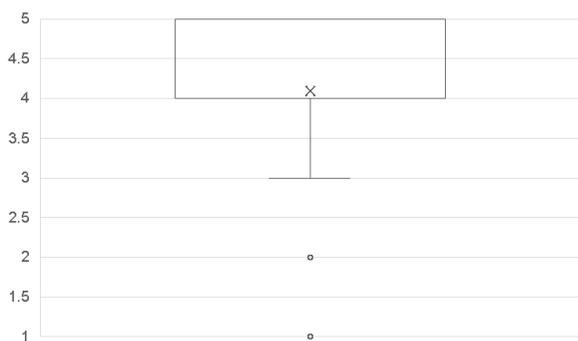


図 10 字幕はあった方が楽しめると感じたかの箱ひげ図

平均値は 4.095 であり、実験参加者は字幕があるほうが楽しめると感じていた。

## 5. まとめ

本研究では、ChatGPT を用いてリアルタイムで専門用語を解説することで、視聴者へ向けた簡単な用語の説明を自動化し、ストリーマーのゲームへの集中や、視聴した初心者知識のステップアップによるゲームへの参入を促すことを目指した。

評価実験の結果、表 4 と表 5 で示したように、本システムは現状実用的ではないことが示された。改善案として、音声認識については別の音声認識 API を用いること、ChatGPT についてはファインチューニング等を用いて専門用語を学習させることが考えられる。

図 6 や図 10 で示した結果からは、本システムが実用的になった場合、動画の楽しさが上がることが示唆された。また、図 7 や図 8、図 9 で示した結果から、字幕によって

専門用語の理解が進むことが示唆された。一方で、解説が出るタイミングが遅いこと、視聴者側のニーズが反映されていないなどの意見が寄せられた。これらは本システムが最後に認識された専門用語を解説するため起こることだと考えられる。下田ら[5]の研究では視聴者の理解度を推定し、解説に利用していたため、このようなアルゴリズムの実装が望ましいと考える。留意点として、下田らの研究では視聴者側に操作を要求しているため、1.1 節で述べた目的に反する点が挙げられる。

今後は前述した改善のうち、音声認識について検証し、専門用語の認識率を上げていきたいと考えている。

**謝辞** 本研究は JSPS 科研費 JP 23K11728 の助成を受けたものです。今回研究をするにあたり、実験に協力していただいた方々、サービスを提供してくださっている企業に深く感謝の意を示し、ここに謝辞とさせていただきます。

## 参考文献

- [1] Macey, J., Tyrväinen, V., Pirkkalainen, H., & Hamari, J.. Does esports spectating influence game consumption?. Behaviour & Information Technology. 2022, vol. 41, no. 1, p. 181-197.
- [2] Wooyoung William Jang, Kevin K. Byon, Baker, T. & Tsuji, Y.. Mediating effect of esports content live streaming in the relationship between esports recreational gameplay and esports event broadcast. Sport Business and Management An International Journal. 2020, p. 89-108.
- [3] Qian, Y., James J. Zhang, Jerred Junqi Wang, Hulland, J.. Beyond the Game: Dimensions of Esports Online Spectator Demand. Communication & Sport. 2019, vol. 8, no. 6, p. 1-27.
- [4] Matsui, A., Anna Sapienza, A., & Ferrara, E.. Does Streaming Esports Affect Players' Behavior and Performance?. Games and Culture. 2019.
- [5] 下田隆介, 兼松祥央, 松吉俊, 三上浩司. ゲーム実況動画における視聴者のビューワー操作に基づく用語理解度推定を利用した動的解説生成. 日本デジタルゲーム学会 年次大会 予稿集, 2024, p. 260-265.
- [6] 梶並知記. プレイ意図を伝えるための動画コンテンツを用いた e-Sports 観戦支援手法の検討. Web インテリジェンスとインタラクション研究会 予稿集, 2013, p. 79-84.
- [7] 梶並知記, 長谷川和也. キャラクタの位置情報に基づいた対戦型格闘ゲームの初心者向け観戦支援システム. 情報処理学会論文誌デジタルコンテンツ, 2018, vol. 6, no. 1, p. 17-27.
- [8] 酒見真, シュエジュウシュエン, 池田 心. 格闘ゲーム初心者のための読み合い理解支援システムの改善. 情報処理学会第 49 回 GI 研究発表会, 2023, vol. 2023-GI-49, no. 6, p. 1-8.
- [9] 天川拓海, 荒川達也. 格闘ゲーム初心者向け感想観戦支援システムの提案. ゲームプログラミングワークショップ 2019 論文集, 2019, p. 109-113.
- [10] “競技タイトル一覧”. <https://jesu.or.jp/contents/titles/>, (参照 2024-12-15).
- [11] “ストリートファイターリーグ - CAPCOM eSports”. <https://sf.esports.capcom.com/sfl/>, (参照 2024-12-15).
- [12] “TOP | Capcom Pro Tour - The Home of Street Fighter Esports”. <https://sf.esports.capcom.com/cpt/jp/>, (参照 2024-12-15).
- [13] 川原達也, 話し言葉の音声認識の進展—議会の会議録作成から講演・講義の字幕付与へ—. メディア教育研究, 2012, vol. 9, no. 1, p. S1-S8.