

三人称ゲーム実況音声に対する時間遅延許容量の測定

松下 嶺佑[†] 阪井 瞭介[†] 福田 航希[†] 高道慎之介^{†,††} 井浦 昂太^{††}

齋藤 佑樹^{††} ニュービッググラム^{†††} 須藤 克仁^{††††} 高村 大也^{†††††} 石垣 達也^{†††††}

[†] 慶應義塾大学

^{††} 東京大学

^{†††} カーネギーメロン大学

^{††††} 奈良女子大学

^{†††††} 国立研究開発法人産業技術総合研究所

E-mail: [†]{ryosuke.jp66,shinnosuke __ takamichi}@keio.jp

あらまし 本研究では、三人称ゲーム実況の遅延の許容時間を測定する。三人称ゲーム実況とは、ゲームプレイ動画の視聴者に状況の理解を手助けすることや、盛り上がりを届けるために、プレイヤー以外の人物が実況することである。これを自動化し、ゲームプレイ動画から実況音声を実動合成する際には、動画理解や音声言語合成による時間遅延が生じる。自動実況が視聴者にとって自然であるためには、許容される時間内に遅延を抑える必要がある。そこで本研究では、三人称ゲーム実況の時間遅延の許容量を測定する方法論を提案する。実験的評価では、大乱闘スマッシュブラザーズ SPECIAL を用いて、許容時間と、視聴者属性やゲームプレイ内容の影響を調査する。

Ryosuke MATSUSHITA[†], Ryosuke SAKAI[†], Koki FUKUDA[†], Shinnosuke TAKAMICHI^{†,††}, Kota IURA^{††}, Yuki SAITO^{††}, Graham NEUBIG^{†††}, Katsuhito SUDOH^{††††}, Hiroya TAKAMURA^{†††††}, and Tatsuya ISHIGAKI^{†††††}

[†] Keio University

^{††} The University of Tokyo

^{†††} Carnegie Mellon University

^{††††} Nara Women's University

^{†††††} National Institute of Advanced Industrial Science and Technology (AIST)

E-mail: [†]{ryosuke.jp66,shinnosuke __ takamichi}@keio.jp

1. はじめに

三人称ゲーム実況とは、プレイヤーや観客に対して第三者がそのゲーム内容を解説・補足しながら進行状況を伝える形式のゲーム実況である。この形式では、実況者がゲーム画面やプレイヤーの行動を観察し、それを言語的に表現することで、視聴者にゲームの状況やプレイヤーの意図をわかりやすく伝える役割を担う [1]。

実況は視聴者に対して臨場感を提供し、プレイや試合の状況をわかりやすく伝えるだけでなく、映像そのものの価値を高める重要な役割を果たしている [2]。一方で、実況の付いていない動画に対して、実況を付与するには専門的なスキルが必要となるため、スポーツ映像やビデオゲーム映像の多くには、実況が付与されていない現状がある [2]。

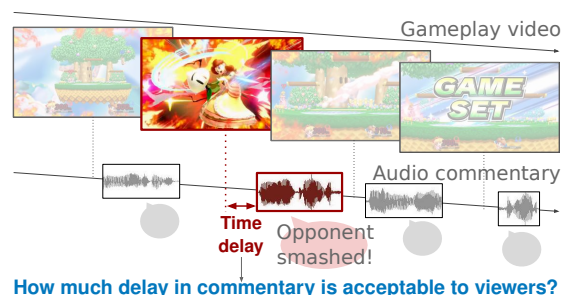


図1 本研究の概要。実況音声の遅延許容量を測定する。

こうした背景から、近年では、実況を自動生成する技術として、試合の状況を分析し、適切な実況を自動生成する手法が提案されている [2]~[6]。これらの研究は、映像コンテンツの付加価値を高めるための有効なアプローチとして注目を集めている。

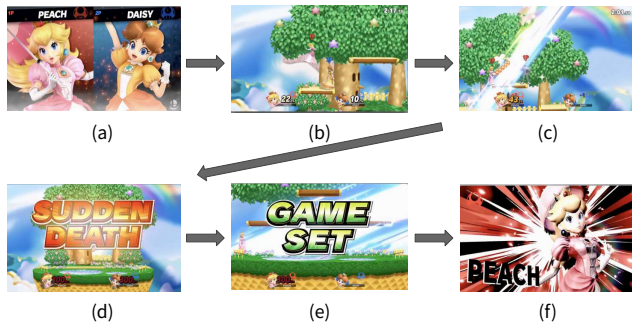


図2 ゲームプレイ動画の流れ。

る[2],[6]。しかし、実況を自動生成する場合、動画理解や実況音声の合成に時間を要するため、実況音声に遅延が生じる。自然な実況音声視聴者に届けるには、許容される時間内に遅延を抑える必要がある。

そのため、本研究では、実況自動生成の前段階として、三人称ゲーム実況の遅延の許容時間を測定する方法論を提案する。具体的には、図1で示すように、ゲームのプレイ動画をを用い、視聴者が実況の時間ずれに対してどの程度の違和感を覚えるかを評価する実験を実施する。大乱闘スマッシュブラザーズ SPECIALを用いた実験結果より、許容時間と、視聴者属性やゲームプレイ内容の影響を調査する。本研究の成果として得られる実況の許容時間は、実況自動生成の評価での利用が期待される。

2. 関連研究

2.1 大乱闘スマッシュブラザーズ SPECIAL

大乱闘スマッシュブラザーズ SPECIAL^(注1)のゲームルールについて紹介する。ゲームプレイヤーが使用するキャラクターの攻撃などで相手のファイターにより多くのダメージを与え、ステージの場外に飛ばすことでスコアを稼ぐことができる[7]。また、ステージ上にランダムに出現するアイテムを活用することで、プレイヤーは戦況をより有利に運ぶことができる。

次に、ゲームプレイ動画の特徴を説明する。図2に示したように、いくつかの場面が存在する。序盤(a)は登場するキャラクターの紹介を行い、中盤(b)(c)では対戦が行われる。制限時間内のスコアが同点の場合は、サドンデス(d)が行われる。その後試合終了(e)が示され、終盤(f)では結果発表が行われる。

2.2 実況自動生成

実況を自動生成する際、発話生成と、発話タイミングの識別の2つのサブタスクを行う必要がある[2]。前者は、サッカー[3],[4],[8],[9]、クリケット[5]、バスケットボール[10]、野球[11]といったスポーツの発話生成から、将棋[12],[13]、チェス[14]といったボードゲーム[15]~[18]のようなeSportsと多岐にわたるジャンルで行われてきた。これらの発話生成タスクには、実況対象の専門知識が必要となる[4],[11],[15]。しかし、従来の研究では、これらの発話生成タスクのみ行なっているものが多い。

発話タイミングの識別には、実況が適切なタイミングで行われているかを判断する[2],[4],[11]。動画理解や実況音声の合成

に時間を要するため、実況音声に遅延が生じる。自然な実況音声視聴者に届けるには、許容される時間内に遅延を抑える必要がある。本研究では、実況自動生成の中でも、発話タイミングの着目に重きを置いた研究を行う。

2.3 音声やテキストの遅延による主観評価

音声やテキストの遅延による主観評価を行った先行研究[19]~[22]では、いずれも5件法の恒常法[23]によって行われている。5件法の恒常法とは、被験者が提示された項目に対して、あらかじめ定められた5段階の評価基準に基づき、自身の主観的な判断を行う。各評価項目における回答の一貫性を測定するため、分析において分散や標準偏差を活用することが可能である。本研究でもこの方法を採用する。

2.4 視聴者属性による影響

被験者属性によって、刺激に対する反応時間が異なることが示されている[24],[25]。

- **年齢**：加齢による影響の多くが一般的な処理速度の低下であることが確認されている[24]。
- **経験の有無**：楽器演奏経験について、経験のあるほうがリズムに合わせて動作する能力が高い[25]。

ゲーム実況音声の視聴について反応時間(許容遅延時間)に言及する直接的な研究はないが、上記の関連研究から属性依存が予想されるため、本研究においても視聴者属性を収集する。

3. 提案手法

三人称ゲーム実況音声に対する許容遅延時間を測定する方法を提案する。提案法では、時間同期したゲームプレイ映像と三人称ゲーム実況音声のデータセットを使用する。この三人称ゲーム実況音声には、発話ごとに発話開始時刻、発話トピックタグが付与されているものとする。発話トピックタグとは当該発話の実況内容に関するラベルである。例えば、後述する本研究の実験では、対戦風景、キャラクター、対戦結果などを指す。このデータセットに基づいて、まず、各被験者に提示する動画刺激セットを作成する。次に、多数の被験者に動画刺激を提示し、その時間遅延の知覚の程度を回答させる。最終的に、複数被験者の結果を集計することで、許容遅延時間を測定する。なお、本研究における遅延時間は負の値、すなわち、基準時刻よりも早いタイミングで実況が開始される設定を含むことに注意されたい。これにより、実況音声ゲームプレイ映像の内容を先行して提示する場合における許容量も評価対象とする。

3.1 動画刺激の作成

データセットから、各発話の時間遅延を評価させるための動画セグメントを作成する。各動画セグメントは、評価対象発話およびその直前 N 発話からなる実況音声と、それに対応するゲームプレイ映像からなる。評価対象発話だけでなく直前 N 発話を含めるのは、ゲームプレイの文脈や状況を評価者に理解させるためである。 N は、動画セグメントの時間長が、事前に定める閾値(動画刺激の最大時間長)を超えない範囲で最大の値に設定する。この実況音声に対し時間遅延を付与する。時間遅延量の候補を事前に定め、動画セグメントと時間遅延量候補の全組み合わせについて、時間遅延付き動画セグメント(動画刺

(注1) : https://www.smashbros.com/ja_JP/

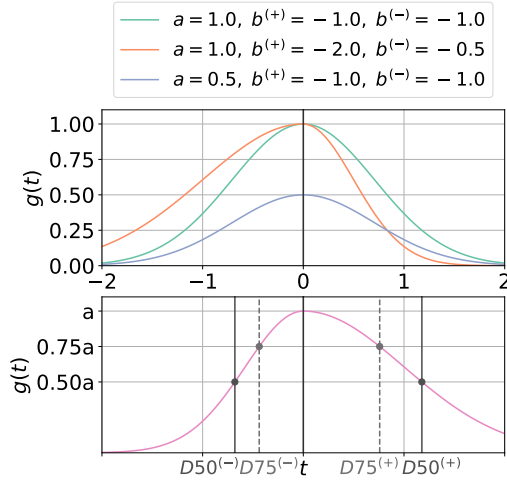


図3 回帰曲線(上)と許容遅延時間(下)の例。

激)を作成する。時間遅延量候補には、正負値および0を含むものとする。0は基準値であり、データセットに含まれる実況音声に時間遅延を付与しないことに相当する。

3.2 各被験者に提示する刺激セットの作成

各被験者に提示する動画刺激セットを作成する。セットを作成する際には、被験者とセットの組み合わせによる意図しないバイアスが評価に影響しないよう、各被験者について均質なセットを作成する必要がある。本研究では、以下の2点についてセットの均質を保つ。

- **時間遅延量**: 極端な例として、あるセットの動画刺激の時間遅延が全て正に大きいとする。その場合、そのセットを評価する被験者は、実況の大きく遅延した動画ばかりを評価することになり、正しい評価値を付与できない恐れがある。そのため、各セットには、3.1節にて定めた時間遅延量の各候補を、それぞれ同数だけ含ませる。

- **発話トピック**: 例えば、突発したゲームイベントには即座の実況が求められる一方、ゲーム内容に関係しない雑談では大きい遅延でも許容されると予想される。そのため、特定のトピックに偏らないセットを被験者に提示すべきである。この実現には、発話トピックの出現頻度について、データセット全体の確率分布と各動画刺激セットの確率分布をそれぞれ計算する。すべての動画刺激セットについて、確率分布間のKullback-Leibler情報量が事前に定める閾値以下になるよう、各動画刺激セットを作成する。この作成には貪欲法[26]を用いる。後述するように、評価は5件法の恒常法[23]に基づくため、セット内の動画刺激の順番はランダムとする。

3.3 被験者による時間遅延評価

各被験者は動画刺激を視聴し、ゲームプレイ映像と実況音声の時間ずれについて1から5までの5段階MOS評価を行う。すべての動画刺激を評価した後、被験者属性に関するアンケートに回答する。

3.4 許容時間の測定

3.3節の評価値を遅延時間ごとに平均し、以下の回帰曲線でフィッティングする。

$$g(t) = \begin{cases} a \cdot \exp(b^{(+)} \cdot t^2) & (t \geq 0) \\ a \cdot \exp(b^{(-)} \cdot t^2) & (t < 0) \end{cases} \quad (1)$$

t は遅延時間である。この式の概形を図3上に示す。この関数は、 $t=0$ を中心として左右で精度パラメータの異なる、正規化されていないガウス関数である。 $a \geq 0$ はスケールパラメータ、 $b^{(-)} < 0$, $b^{(+)} < 0$ はそれぞれ精度パラメータの負値である。

本論文では、回帰曲線の値が最大値の75%と50%になる秒数を、許容遅延時間と定義する(図3下)。許容遅延時間を $D50^{(-)}$, $D75^{(-)}$, $D75^{(+)}$, $D50^{(+)}$ ($^{(+)}$, $^{(-)}$ はそれぞれ、負と正の遅延に対応)と定義すると、それぞれ以下の式で計算される。

$$D50^{(-)} = \sqrt{\frac{\ln 0.5}{b^{(-)}}}, D75^{(-)} = \sqrt{\frac{\ln 0.75}{b^{(-)}}} \quad (2)$$

$$D50^{(+)} = \sqrt{\frac{\ln 0.5}{b^{(+)}}}, D75^{(+)} = \sqrt{\frac{\ln 0.75}{b^{(+)}}} \quad (3)$$

4. 実験的評価

4.1 実験条件

対戦アクションゲームの実況音声を対象として、時間遅延許容量を測定した。被験者は、クラウドソーシングプラットフォームであるランサーズ^(注2)にて募集した727人である。各被験者には240円を支払った。

4.1.1 動画刺激と刺激セットの作成

動画刺激は、SMASH コーパス[7]に含まれる1vs1の対戦形式の対戦動画32本、約1.3時間を用いて作成した。当該コーパスは、大乱闘スマッシュブラザーズ SPECIALのゲーム動画、実況音声、トピックタグ、発話開始時刻情報から成る。対戦ルールは制限時間150secの時間制乱闘であり、ゲーム動画は、序盤のキャラクタ紹介、中盤の150secの対戦、終盤の対戦結果紹介から成る。本論文では以下のトピックタグを評価対象とした。

- **Result**: 対戦結果に関する実況
- **Scene**: 対戦風景に関する実況
- **Item**: 対戦中に登場するアイテムに関する実況
- **Fighter**: 対戦キャラクタに関する実況

雑談トピックの発話は、その発話内容が対戦と直接的に関係しないため、許容時間の推定に無関係として評価対象から除外した。実況は2人の男性が担当し、各動画はそのうちどちらか1人が実況した。音声のサンプリング周波数は16kHz、動画のフレームレートは30fpsである。

動画刺激の最大時間長は20secとした。この数値は、視聴者が対戦状況を把握できるよう事前調査より決定した。時間遅延量候補は、-1.50secから1.75secまで0.25sec刻みの12通りとした。最終的に5,016個の動画刺激を作成した。各被験者に提示する動画刺激数は24動画とした。Kullback-Leibler情報量の閾値は0.1とした。回帰曲線の $b^{(-)}$, $b^{(+)}$ が0に近い値をとるとき、式(2),(3)の許容遅延時間は非常に大きい値になる。しかしながら、動画の説明を含む実況音声があるような値をとることは不自然である。そこで本研究では、データセットの発話開始時刻の平均的な時間間隔を求め、その値を許容遅延時間

(注2): <https://www.lancers.jp>

表 1 被験者属性のアンケート。全ての項目が単一選択。

ID	Content
01	年齢
02	性別
03	当該ゲームを知っているか否か
04	当該ゲームの経験歴の程度
05	当該ゲームの三人称実況動画視聴の程度
06	当該ゲームの一人称実況動画視聴の程度

表 2 質問 02：年齢

10s	20s	30s	40s	50s	60s	70s<	Total
1	59	221	272	131	37	6	727

表 3 質問 04：当該ゲームの経験歴の程度

Competitive	Enjoy	Sometimes	Only seen	Never	Total
10	46	305	235	131	727

表 4 質問 05：三人称実況動画視聴の程度

Often	Sometimes	Never	Total
34	328	365	727

表 5 質問 06：一人称実況動画視聴の程度

Often	Sometimes	Never	Total
45	390	292	727

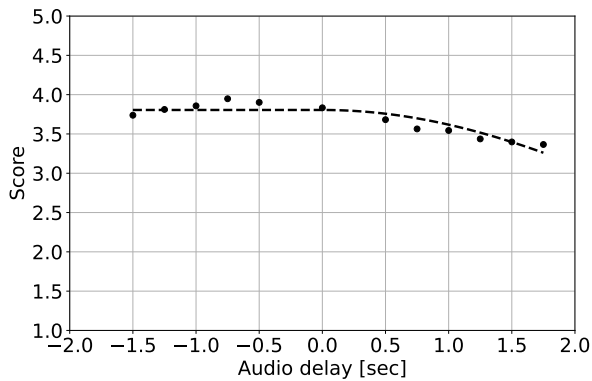


図 4 全被験者の評価結果の平均と回帰曲線。

の上限値とした。本実験のデータセットでは 5.13 sec であった。

4.1.2 被験者属性

表 1 に示す被験者属性アンケートを実施した。表 2 から表 5 に、各属性の被験者数を示す。

4.1.3 調査内容

本節では、以下の内容について調査する。

- **平均的な時間遅延許容量 (4.2 節)：**実況音声の平均的な時間遅延許容量を測定する。
- **被験者属性の影響 (4.3 節)：**時間遅延許容量が、表 1 に示した被験者属性によって変化するかを調査する。
- **動画内容や対戦展開の影響 (4.4 節, 4.5 節)：**時間遅延許容量が、ゲーム内時刻や対戦盛り上がりなどの要素によって変化するかを調査する。

4.2 全評価データの傾向

まず、評価データの全体傾向をつかむために、すべての評価結果から回帰曲線と許容時間を求めた。図 4 に結果を示す。まず、負の遅延時間では回帰曲線がほぼ一定である。これに対する仮説として、基準時刻のとり方が考えられる。本実験では、遅延 0 sec を人間実況者の発話開始時刻としている。人間実況

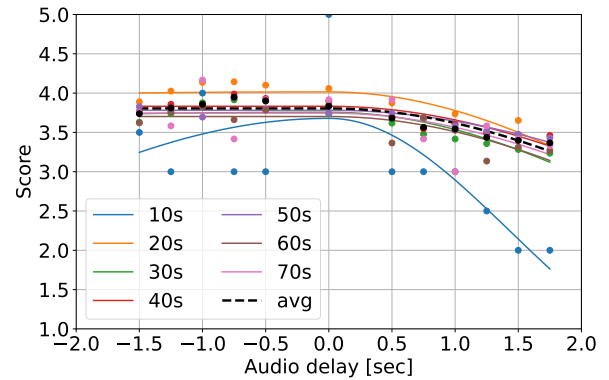


図 5 年齢別の評価結果の平均と回帰曲線。

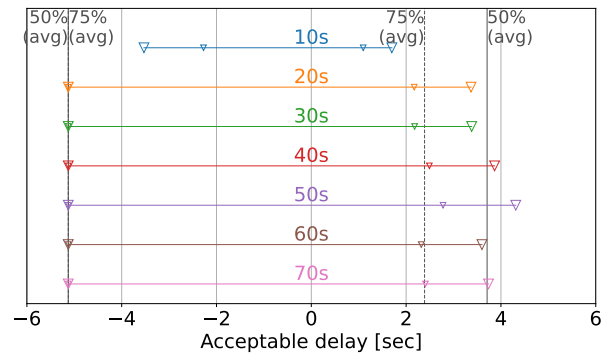


図 6 年齢別の許容遅延時間。小さい三角は $D75^{(\pm)}$ 、大きい三角は $D50^{(\pm)}$ を表す。

者の実況音声そのものが動画に対して遅延しているため、その実況に対し負の遅延時間を付与しても、一定の遅延時間範囲においては実況の自然さに影響しなかったと考えられる。この仮説に対する回答には、人間実況者の実況音声の遅延を計測する必要があるが、この点については今後の予定とする。なお、回帰曲線がほぼ一定になる現象については、その要因を後述する。

許容遅延時間は $D50^{(-)} = 5.13, D75^{(-)} = 5.13, D75^{(+)} = 2.39, D50^{(+)} = 3.71$ sec となった。これらの値は、動画内容や視聴者属性に依存しない、平均的な許容遅延時間として利用できる。

4.3 視聴者属性の影響

4.3.1 年齢の影響

被験者の年齢が許容遅延時間に与える影響を調査するために、年齢ごとに回帰曲線を求め許容遅延時間を求めた。図 5 と図 6 はそれぞれ回帰曲線と許容遅延時間である。図 5 に示されるように、年齢が増加するにつれ、正の遅延に対する曲線の減衰が緩やかになることがわかる。この傾向は、図 6 においても確認でき、若い被験者ほど、許容遅延時間が短い傾向にある。この要因として、年齢のほかに、後述するゲームプレイ経験との関連が考えられる。ゲームプレイ経験の深い Competitive または Enjoy を選んだ割合は、20 代の 0.22 から 50 代の 0.02 にかけて、年齢が増すごとに単調に減少する。この経験の深さが影響した可能性がある。

4.3.2 ゲームプレイ経験の影響

ゲームプレイ経験が許容遅延時間に与える影響を調査した。図 7 は回帰曲線、図 8 は許容遅延時間である。ゲームプレイ経

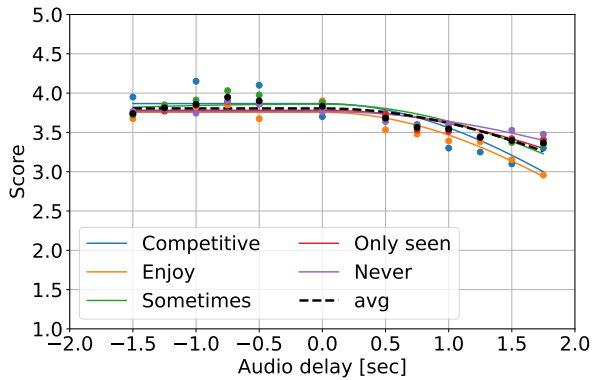


図7 ゲームプレイ経験別の評価結果の平均と回帰曲線。

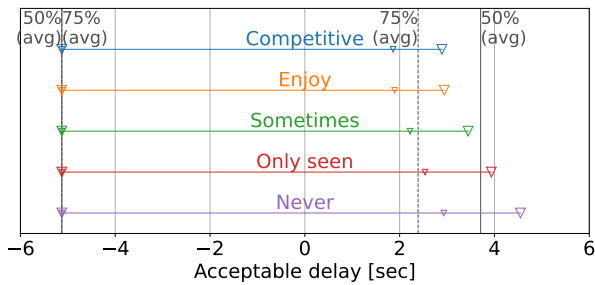


図8 ゲームプレイ経験別の許容遅延時間。

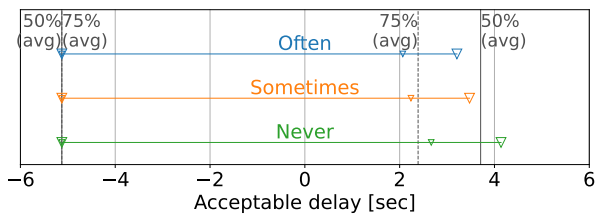


図9 一人称実況の視聴経験別の許容遅延時間。

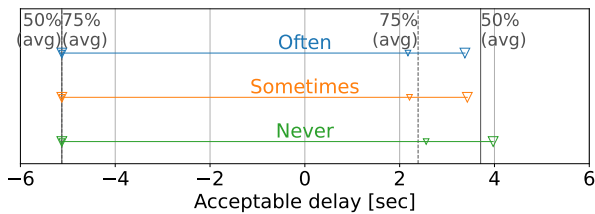


図10 三人称実況の視聴経験別の許容遅延時間。

験が深くなるほど許容遅延時間が短くなる傾向にあることがわかる。この傾向は、経験のある方が刺激に対する反応時間が早くなるという既存研究[25]と整合する。D50⁽⁺⁾を比較すると、Competitiveが2.89 sec, Neverが4.55 secと、最大1.66 secの開きがある。

4.3.3 ゲーム実況視聴歴別の評価

図9と図10のそれぞれに、一人称実況の視聴経験、三人称実況の視聴経験ごとの許容遅延時間を示す。4.3.2節の結果と同傾向であり、一人称と三人称を問わず視聴経験が深いほど許容遅延時間が短くなる（OftenではD75⁽⁺⁾ = 2.07, 2.18 secと、全体平均を0.32, 0.21 sec上回っている）。以上より、自身によるゲームプレイ経験が深い場合と、他者によるゲームプレイを実況付きで視聴する経験が深い場合の両方で、視聴者の求める遅

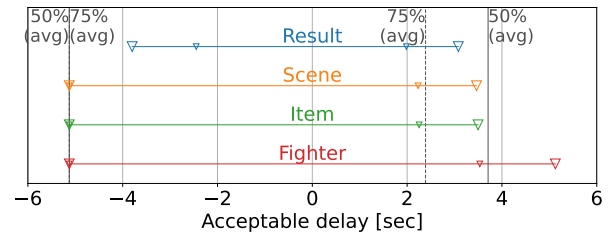


図11 発話トピック別の許容遅延時間。

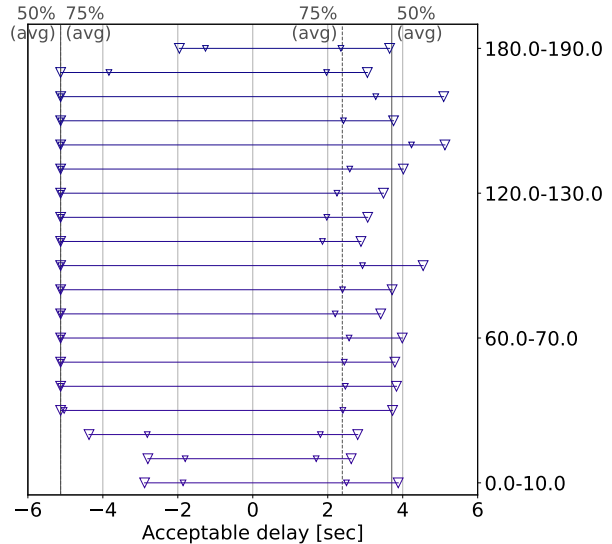


図12 ゲーム内時刻別の許容遅延時間。

延時間が短いと言える。

4.4 動画内容の影響

4.4.1 発話トピックの影響

動画内容のうち、実況の発話トピックが許容遅延時間に与える影響を調査した。調査方法は4.3節と同様であり、ここでは発話トピックごとにスコアを平均している。図11に発話トピックごとの許容遅延時間を示す。

Resultトピックの許容遅延時間が際立って短いことがわかる。ResultトピックはD50⁽⁺⁾ = 3.08 secであり、全体平均よりも0.63 sec程度短い。これは、Resultトピックでは動画最終盤に対戦結果を伝えるため、結果が動画に明示されたのち即座の音声実況が求められるためだと考えられる。また、Resultトピックでは負の遅延時間に対しても許容遅延時間が極端に短い。これは、負の遅延時間を与えることで、動画で結果が明示されるよりも前に音声実況が開始されてしまい、非因果的な実況が発生したものだと考えられる。

Fighterトピックは、他のトピックに比べ極端に長い許容遅延時間を有する。これは、Fighterトピックでは、対戦風景ではなく対戦キャラクタ自体を解説することが多く、即座の実況を要請しないためだと考えられる。

4.4.2 ゲーム内時刻の影響

動画内容のうち、ゲーム内時刻が許容遅延時間に与える影響を調査した。図12にゲーム内時刻ごとの許容遅延時間を示す。

動画開始直後（図2の(a)）では、負の許容遅延時間が際立って短いことがわかる。これは、動画開始直後の発話内容が、ファ

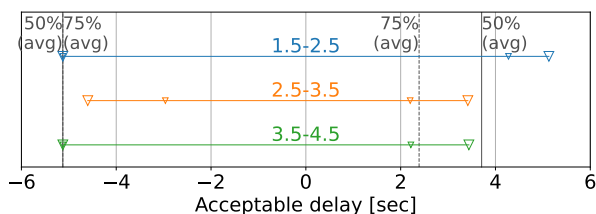


図 13 実況の盛り上がりスコア別の遅延時間。

イター紹介のものが多くに関係すると考えられる。実際に、10.0 sec までに開始した発話の中で、発話トピックが雑談のものを除くと、7 文中 6 文が、発話内容がファイター紹介であった。これは、負の遅延時間を与えることで、動画でファイターが登場するよりも前に音声実況が開始されてしまい、非因果的な実況が発生したものだと考えられる。

次に、動画の中盤(図 2 の (b)-(d))では、正の許容遅延時間が短いことがわかる。これは、必殺技を打つ場面や、撃墜が発生した割合が高かったことが要因であると考えられる。実際に、70.0 sec から 90.0 sec までに開始した発話の中で、発話トピックが雑談のものを除くと、57 文中 28 文が、これに類する内容の実況であった。

動画終盤(図 2 の (e)(f))では、許容遅延時間が際立って短いことがわかる。これは 4.4.1 節でも扱ったが、Result のトピックタグの割合が多いからであると考えられる。全データの発話トピックタグのうち、Result は 15 個あったが、160.0 sec 以降に開始した発話の中で、Result は 12 個あった。

4.5 実況の盛り上がりの影響

実況の盛り上がりが許容遅延時間に与える影響を調査した。盛り上りの定量化には既存文献[27]で付与されたスコアを用いた。これは、実況音声聞いてその盛り上がり度を 1-5 の 5 段階で付与したものである。図 13 に実況盛り上がりスコアごとの許容遅延時間を示す。

盛り上がりが中・高程度の実況では、許容される遅延時間が短いことがわかる。この原因としては、盛り上がりが高いところでは、プレイヤーの重要なアクションやゲーム内の劇的な展開が発生することが多いため、被験者はゲーム内の展開に対して即時的な反応を期待するためであると考えられる。

5. ま と め

本研究では、大乱闘スマッシュブラザーズ SPECIAL の実況の遅延の許容時間を推定することを行った。そこから、被験者の個人属性、動画内容、実況者の属性によって許容される遅延時間が異なるといった知見を得られた。

謝辞：本研究には、内閣府が実施する「研究開発成果の社会実装への橋渡しプログラム (BRIDGE) /AI × ロボット・サービス分野の実践的グローバル研究」により得られた成果が含まれている。また、JSPS 科研費 21H04900、JST 創発的研究支援事業 JPMJFR226V の支援を受けて実施した。

文 献

- [1] 山本浩, スポーツアナウンサー——実況の真髄, ser. 岩波新書. 東京: 岩波書店, 2021.
- [2] T. Ishigaki et al., “Generating racing game commentary from vision, language, and structured data,” in *Proc. INLG*, Aug. 2021, pp. 103–113.

- [3] J. Rao et al., “Matchtime: Towards automatic soccer game commentary generation,” in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024.
- [4] Y. Taniguchi et al., “Generating live soccer-match commentary from play data,” in *Proc. AAAI*, 2019.
- [5] K. Shashank Yadav et al., “Automated cricket commentary generation for videos,” in *Computer Vision and Image Processing*, 2024, pp. 432–444.
- [6] E. Marrese-Taylor et al., “Open-domain video commentary generation,” in *Proc. EMNLP*, Y. Goldberg et al., Eds., Dec. 2022, pp. 7326–7339.
- [7] Y. Saito et al., “SMASH corpus: A spontaneous speech corpus recording third-person audio commentaries on gameplay,” in *Proc. LREC*, May 2020, pp. 6571–6577.
- [8] R. Pasunuru, M. Bansal, “Game-based video-context dialogue,” in *Proc. EMNLP*, Oct.-Nov. 2018.
- [9] A. Suglia et al., “Going for GOAL: A resource for grounded football commentaries,” in *CIKM* 23, 2022.
- [10] R. Puduppully, M. Lapata, “Data-to-text generation with macro planning,” *Trans. of ACL*, 2021.
- [11] B. J. Kim, Y. S. Choi, “Automatic baseball commentary generation using deep learning,” in *Proc. ACM SAC*, 2020, p. 1056–1065.
- [12] S. Mori et al., “A Japanese chess commentary corpus,” in *Proc. LREC*, May 2016, pp. 1415–1420.
- [13] H. Kameko et al., “Learning a game commentary generator with grounded move expressions,” in *2015 IEEE Conference on Computational Intelligence and Games (CIG)*, 2015.
- [14] A. Sadikov et al., “Automated chess tutor,” in *Computers and Games*, H. J. van den Herik et al., Eds., 2007, pp. 13–25.
- [15] Z. Sun et al., “Graspsnooker: Automatic chinese commentary generation for snooker videos,” in *Proc. IJCAI*, 7 2019, pp. 6569–6571.
- [16] Z. Wang, N. Yoshinaga, “Commentary generation from data records of multiplayer strategy esports game,” in *Proc. NAACL*, Jun. 2024, pp. 263–271.
- [17] C. Nimpattanavong et al., “Am I fighting well? fighting game commentary generation with ChatGPT,” in *Proc. IAIT*, 2023.
- [18] 下田隆介 et al., “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*, vol. 14, pp. 260–265, 2024.
- [19] 小森智康 et al., “ネット配信動画のための適切な字幕表示タイミングの検討,” *聴覚研究会資料*, vol. 51, no. 6, pp. 255–260, 10 2021.
- [20] 西宏之, 大平洋一郎, “コネクションレス音声通信を対象とした遅延と雑音の主観評価,” *電子情報通信学会技術研究報告. SP, 音声*, vol. 96, no. 566, pp. 1–6, 03 1997.
- [21] 栗田孝昭 et al., “音声遅延と映像遅延が通信品質に及ぼすインタラクション効果,” *NTT R & D / NTT 先端技術総合研究所 企画編集*, vol. 42, no. 10, p. p1207–1214, 1993.
- [22] 小川修太, 大盛善啓, “K-020 リアルタイム字幕放送における字幕の提示タイミングずれに対する補正方法の主観評価,” *情報科学技術フォーラム講演論文集*, vol. 8, no. 3, pp. 567–568, 08 2009.
- [23] 黒田剛士, 蓮尾絵美, “早わかり心理物理学実験 (やさしい解説),” *日本音響学会誌*, vol. 69, no. 12, pp. 632–637, 2013.
- [24] T. Salthouse, “The processing-speed theory of adult age differences in cognition,” *Psychological review*, vol. 103, pp. 403–28, 08 1996.
- [25] B. Repp, “Sensorimotor synchronization: A review of the tapping literature,” *Psychonomic bulletin & review*, vol. 12, pp. 969–92, 01 2006.
- [26] 相澤彰子, “スキーマ貪欲な集団型探索アルゴリズムの構成,” in *全国大会講演論文集*, vol. 第 48 回, no. 人工知能及び認知科学, mar 1994, pp. 225–226.
- [27] 井浦昂太 et al., “盛り上がり制御可能な対戦ゲーム実況解説音声合成モデルの検討,” *第 155 回音声言語情報処理研究発表会*, 2025, (発表予定).