

2025年3月 音声研究会

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¹, 阪井 瞭介¹, 福田 航希¹, 高道 慎之介^{1,2}, 井浦 昂太²,
齋藤 佑樹², ニュービッグ グラム³, 須藤 克仁⁴, 高村 大也⁵, 石垣 達也⁵

(1: 慶大, 2: 東大, 3: CMU, 4: 奈良女子大, 5: 産総研)



研究背景

実況自動生成の利点

- 実況をつけることによって、**試合映像の価値を高める**ことや、**臨場感を提供することができる** [Ishigaki 21].
 - 視聴者の動画理解を助ける.
- しかし、実況を行うには、**知識やスキルが必要**となる.
 - 実際、実況の付与されていない試合映像は、非常に多い.
- そのため、実況を自動で生成する技術が必要とされている.
 - リアルタイム性
 - 多言語対応
 - 実況者の知識やスキルが不要

研究背景

- スポーツやゲームの試合映像の価値を高めるために、
実況を自動生成する技術が開発されている [Ishigaki 21] .

入力：実況をつけたい動画

実況自動生成

出力：実況のついた動画



実況音声の遅延

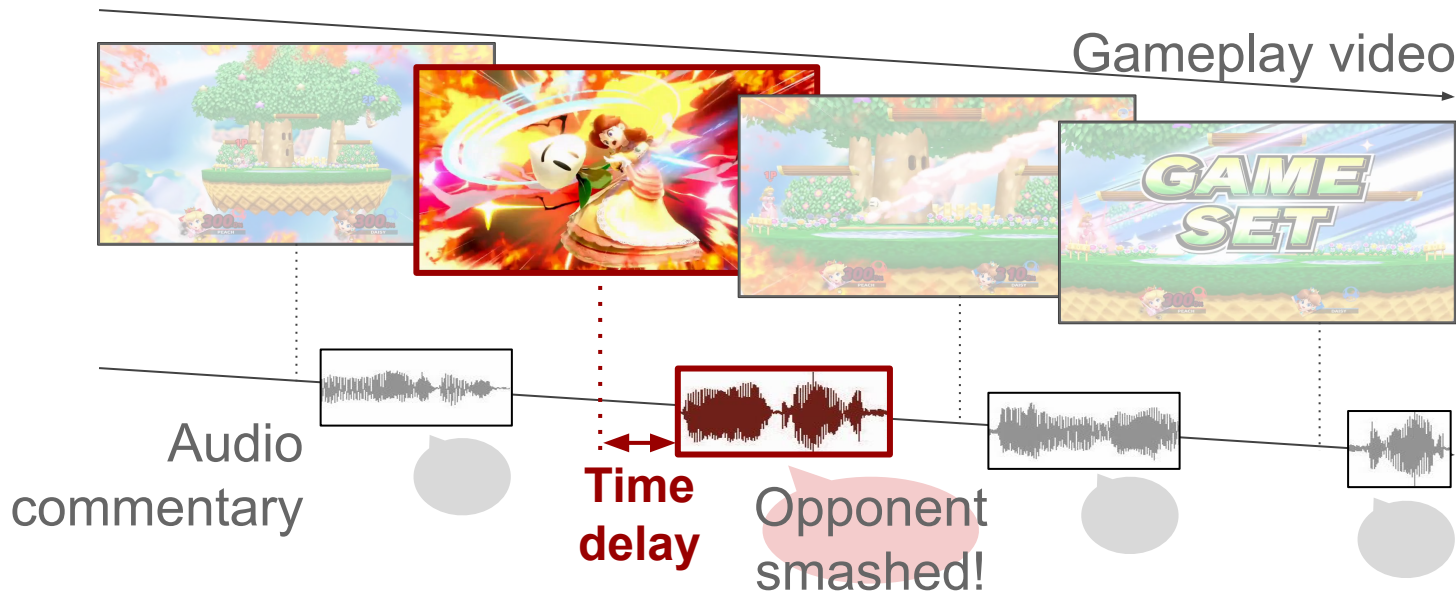
- 一方で、実況を自動生成する場合、動画理解や実況音声の合成に時間を要するため、実況音声に遅延が生じる。
- 自然な実況音声生成のために、**許容される時間**内に遅延を抑えたい。

音声の遅延が大きい例



本研究の立ち位置

- 大乱闘スマッシュブラザーズSPECIALのプレイ動画を用い、実況の遅延時間について、どれくらい許容できるかを測定する。



How much delay in commentary is acceptable to viewers?

関連研究

大乱闘スマッシュブラザーズSPECIAL¹について

- プレイヤーは相手をステージの外に飛ばすとスコアを稼げる [Saito 20] .
- アイテムを活用すると、有利に進めることができる [Saito 20] .



キャラクター紹介



対戦風景



スコア獲得



サドンデス



対戦終了



試合結果

関連研究：5件法の恒常法と視聴者属性による影響

- **遅延時間の測定法：5件法の恒常法** [Kuroda 13]
 - 時間遅延させたメディアをランダムに提示し，5段階で主観評価．
 - 5段階の評価基準は，メディアに対してどれだけ違和感を覚えたか．

▶ **この方法で許容遅延時間の測定をする．**

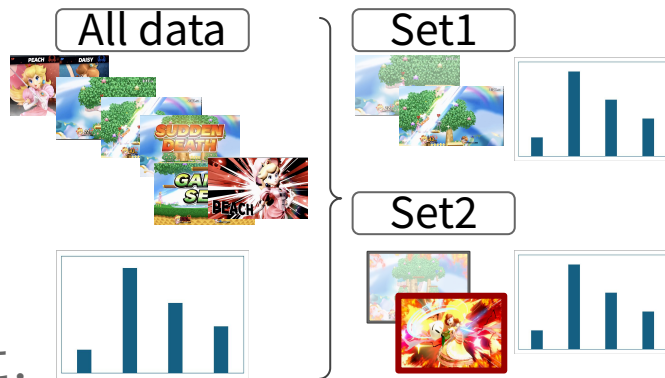
- **視聴者属性**
 - **年齢** [Salthouse 96]
 - 加齢によって，一般的な処理速度が低下することが確認されている．
 - **経験の有無** [Repp 06]
 - 楽器演奏経験がある方が，リズムに合わせて動作する能力が高い．

▶ **本研究でも視聴者属性を集めることにする．**

提案手法

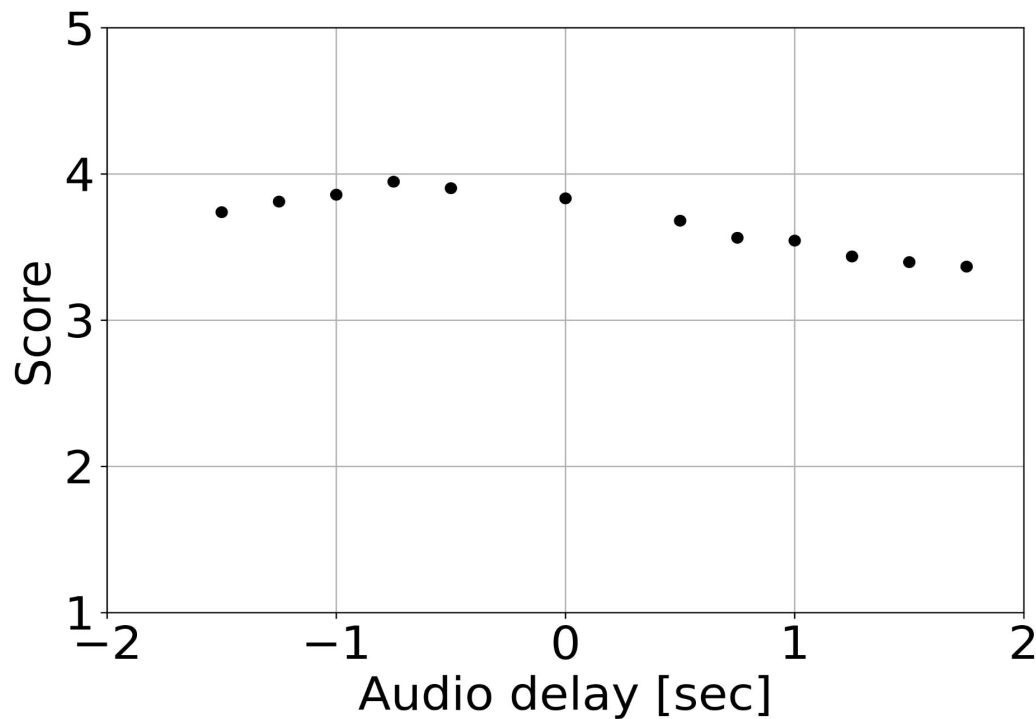
提案手法

- {ゲームプレイ動画，実況音声，**トピックタグ**}の対データを使用。
 - トピックタグとは，実況内容に関するラベルを指す。
- 被験者に示す動画刺激セットの作成
 - **動画刺激の作成**
 - 音声に遅延を付ける（早い or 遅い）。
 - 着目させる発話以前の発話も含める。
 - **動画刺激セットの作成**
 - 各遅延量を同数含める。
 - トピックに偏りが出ないように，トピックの出現頻度について，各セットと全体のKLダイバージェンスの値を閾値以下に。



被験者による時間遅延評価

- 5件法の恒常法を用いて、被験者が評価を行う。
- その結果を元に各遅延量において、評価スコアの平均値を算出する。

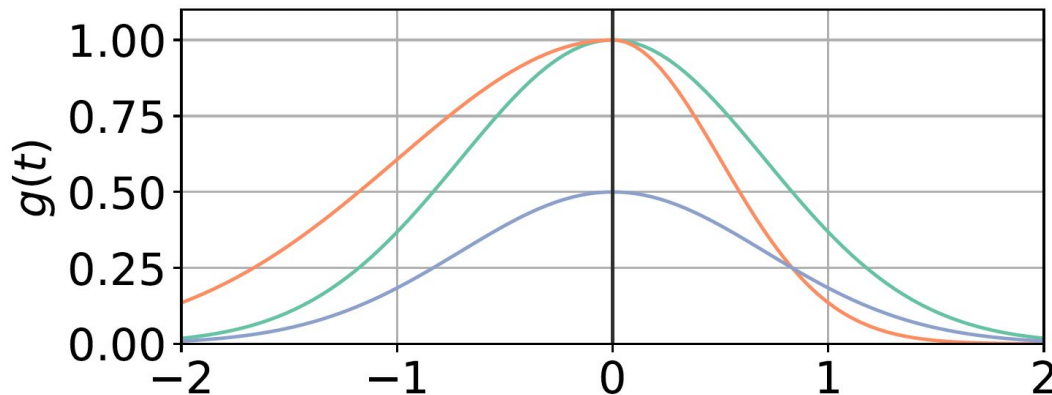


本実験で用いる回帰曲線

- 各点に対して、以下の数式によって回帰曲線を作成する。
 - 0を中心として左右で精度の異なる、正規化されていないガウス関数

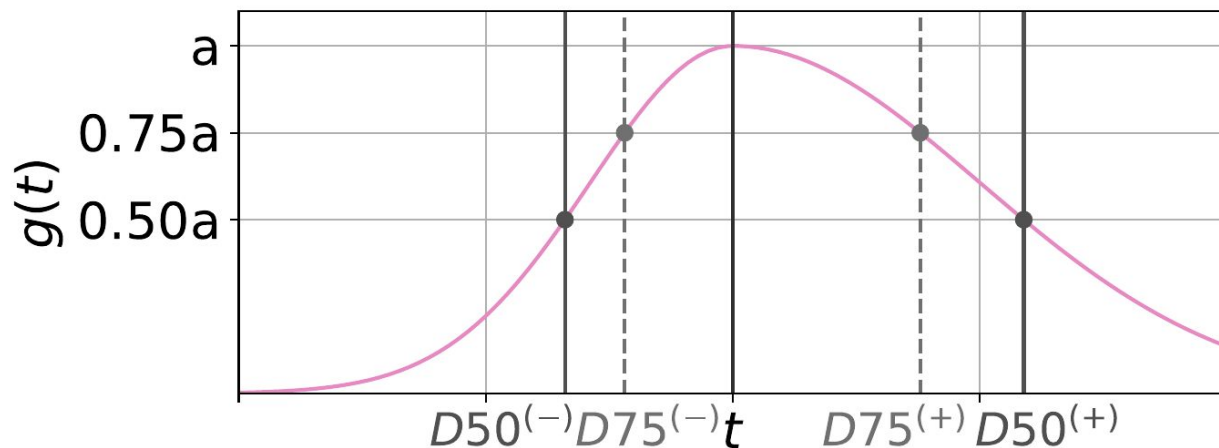
$$g(t) = \begin{cases} a \cdot \exp(b^{(+)} \cdot t^2) & (t \geq 0) \\ a \cdot \exp(b^{(-)} \cdot t^2) & (t < 0) \end{cases}$$

- $a = 1.0, b^{(+)} = -1.0, b^{(-)} = -1.0$
- $a = 1.0, b^{(+)} = -2.0, b^{(-)} = -0.5$
- $a = 0.5, b^{(+)} = -1.0, b^{(-)} = -1.0$



許容遅延時間の測定

- **最大値の 75 % と 50 % になる秒数を許容遅延時間とする。**
 - 知覚刺激の閾値を関数の最大値の50%にした研究^[Hyodo 24]を参考にする。
 - 違和感をさらに覚えにくい許容遅延時間を，回帰曲線の最大値と50%の半分である，最大値の 75 %に設定する。



実験

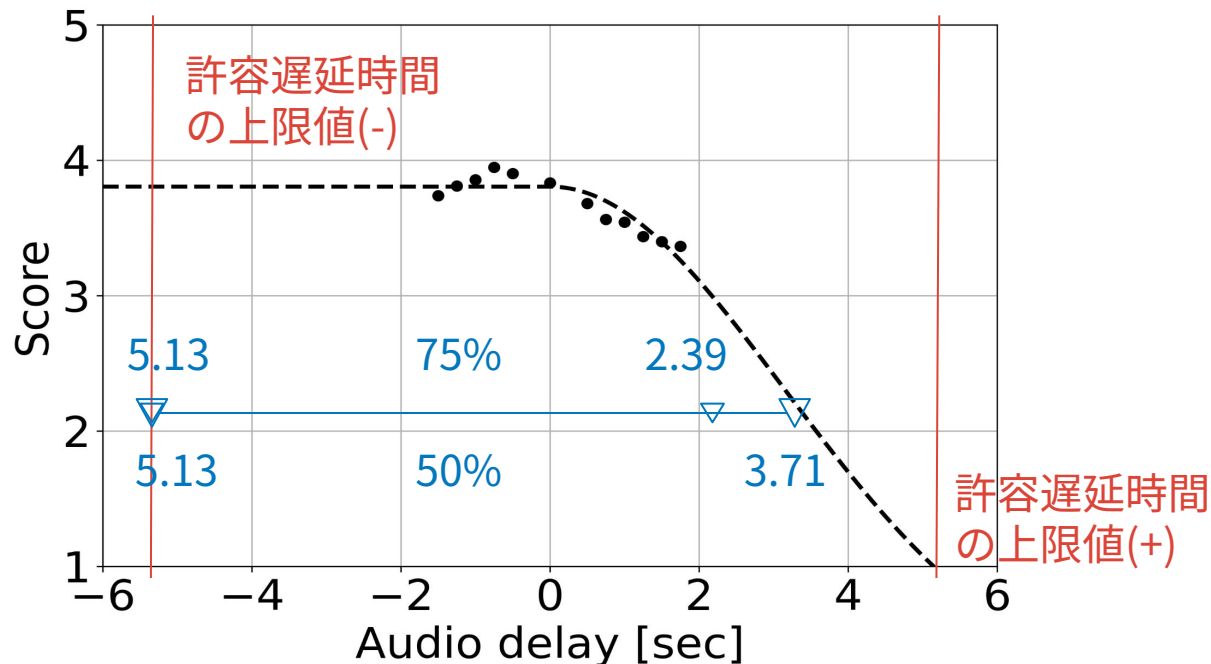
実験条件

使用したコーパス	SMASHコーパス[Saito 20]
被験者人数	727人
被験者属性	年齢, ゲームプレイ経験, 性別, 実況の視聴経験
トピックタグ	Result (試合結果), Scene (対戦風景) Item (アイテム), Fighter (キャラクタ)
時間遅延量	-1.50 sec ~ + 1.75 secまでの0.25 sec刻みの12通り
1動画の最大時間長	20 sec
許容遅延時間の上限値	5.13 sec

算出した許容遅延時間が
大きすぎると不自然

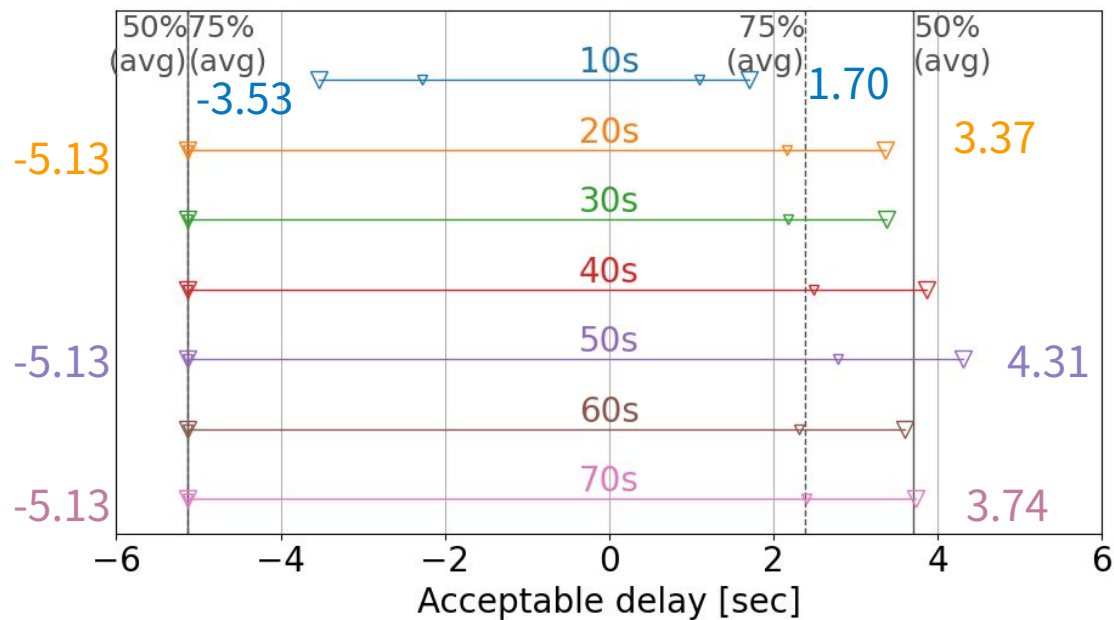
元データセットの
1発話の平均時間に設定

実験結果：全体傾向



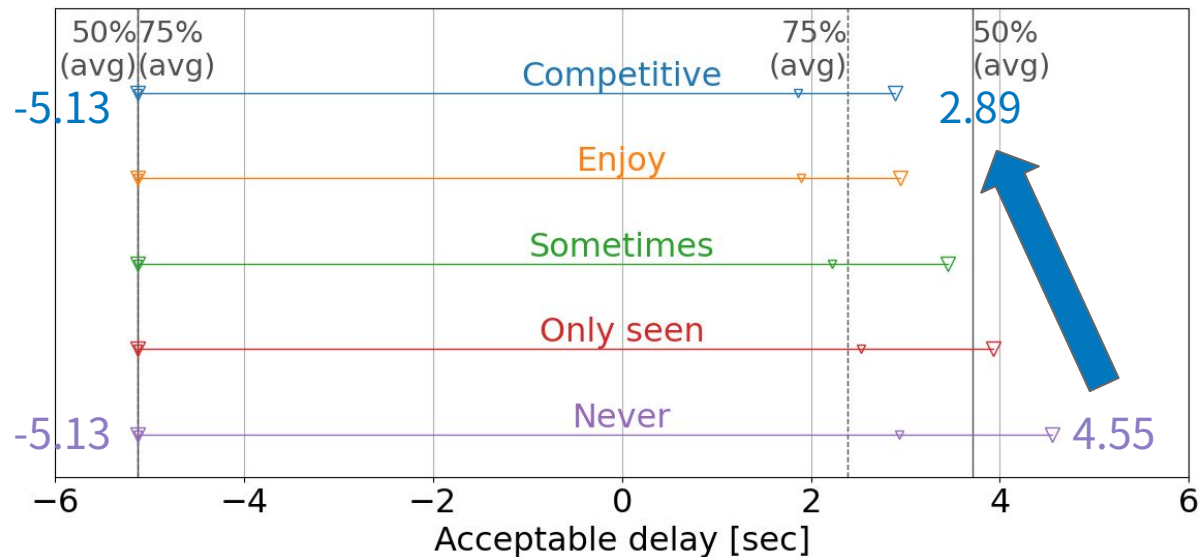
- 負の遅延時間のところでは、評価スコアに影響は無い。
 - 実況者の発話開始時刻を遅延0秒としたが、イベント発生から実況までに遅延が生じるため、一定の遅延時間内では実況は自然に感じた。

実験結果：年齢別



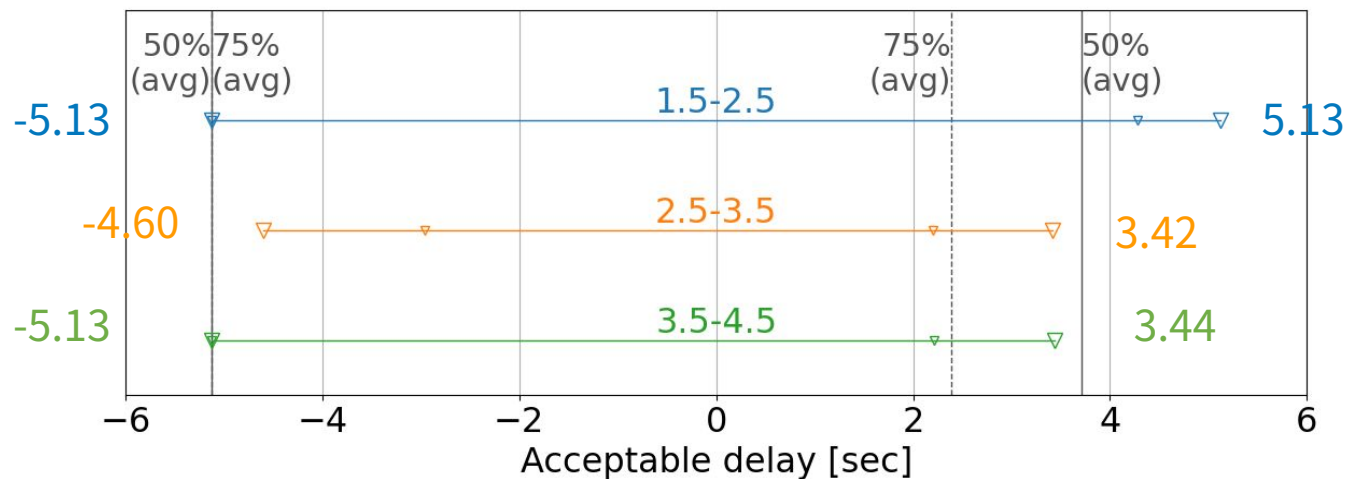
- 許容遅延時間は、年齢が若い方が短い傾向にある。
 - 加齢による処理速度の低下 [Salthouse 96] や、若年層のゲームプレイ経験が関係すると考えられる。

実験結果：ゲームプレイ経験別



- 許容遅延時間はゲームプレイ経験が深くなるほど短い。
 - 経験のある方が反応時間が早くなるという関連研究 [Repp 06] に一致するため、被験者属性による許容遅延時間の差はあると考えられる。
 - 視聴者層に合わせて、実況自動生成をする必要がある。

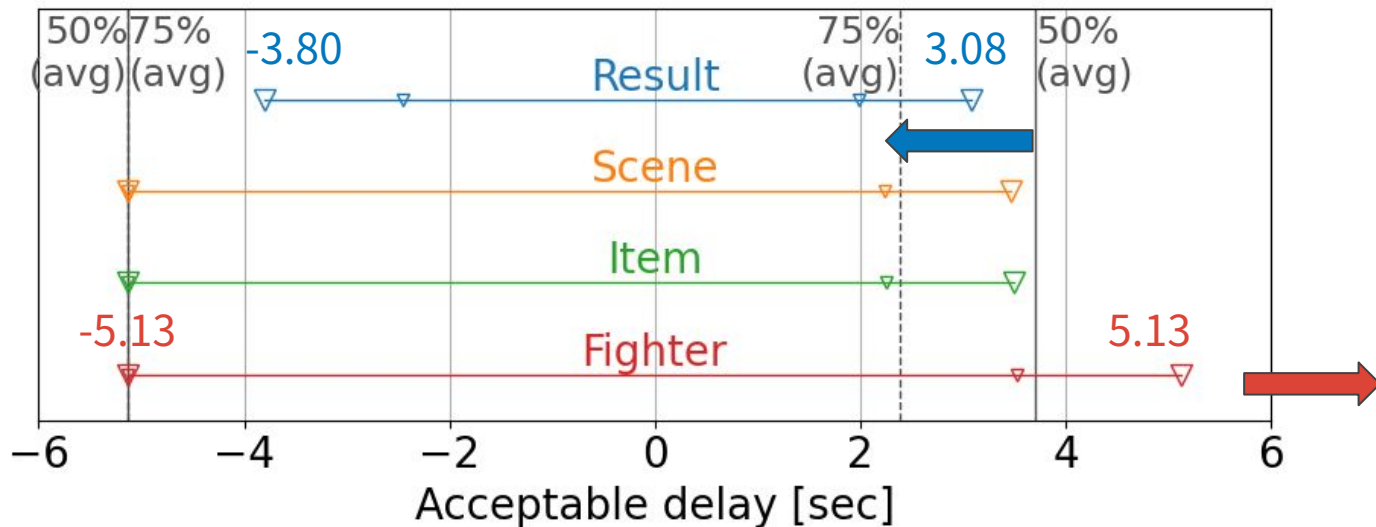
実験結果：実況の盛り上がり別



盛り上がりスコアは [Iura 2025] を参照

- 盛り上がりの中・高程度の実況では、許容される遅延時間が短い
 - 盛り上がりが高いところでは、重要なアクションが発生するため、ゲーム内の展開に対して、即座の実況を望むのではないか。

実験結果：トピックタグ



- 許容遅延時間は**Result**では極端に短く，**Fighter**では極端に長い。
 - Resultは，視聴者の関心が向くので，許容遅延時間が短い。
 - Fighterは，内容が予測がしやすいので，許容遅延時間が長い。

実際の実況 (Result)

遅延0秒



遅延+1.75秒



- Resultは、遅延時間が大きくなると違和感を覚えやすい。

その他の分析と結果

- ゲーム実況視聴経験別
 - ゲーム実況視聴経験歴が長いほど許容時間が短い.
- その他の被験者属性
 - 許容時間に対する影響が小さい.
- ゲーム内時刻
 - 序盤と終盤で，許容時間が短い.

まとめ

- 目的：
 - 実況音声の許容遅延時間の測定.
- 提案手法：
 - 被験者に音声遅延のついた実況動画を評価させ，許容遅延時間を測定.
- 評価結果：
 - 被験者属性によって，許容遅延時間が異なる.
 - トピックタグなどの動画内容によっても許容遅延時間が異なる.
- 今後の予定：
 - 実況音声の自動評価.