IEEE CoG 2025 Auxiliary Paper

Measuring Time Delay Tolerance in Third-Person Live Commentary for Super Smash Bros. Ultimate

Ryosuke Matsushita[1], Ryosuke Sakai[1], Koki Fukuda[1], Shinnosuke Takamichi[1,2], Kota Iura[2],Yuki Saito[2], Graham Neubig[3], Katsuhito Sudou[4], Hiroya Takamura[5], Tatsuya Ishigaki[5]

(1: Keio University, Japan, 2: The University of Tokyo, Japan,
3: Carnegie Mellon University, USA, 4: Nara Women's University, Japan, 5: AIST, Japan)

Takamichi Laboratory
慶應義塾大学 高道研究室

Research background

Benefits of automatic commentary generation

- Audio commentary is what makes **the match exciting**, by not the player.
- Audio commentary makes it possible to **help viewers better understand the video** [Ishigaki 21].
- However, providing commentary **requires knowledge and skill**.
  - In fact, there are a lot of match videos without any commentary.
- Automatic commentary generation from the game video.
  - Real-time capability[Ishigaki 21]
  - Multilingual support[Zhou 22]
  - No need for the commentator's knowledge or skill[Ishigaki 21]

Automatic commentary generation

Input：video without commentary audio

Output：video with commentary audio

Do this in real time!

“A vegetable hits Daisy!”

This is no sound.

This video is not generated automatically.

Delay in commentary audio

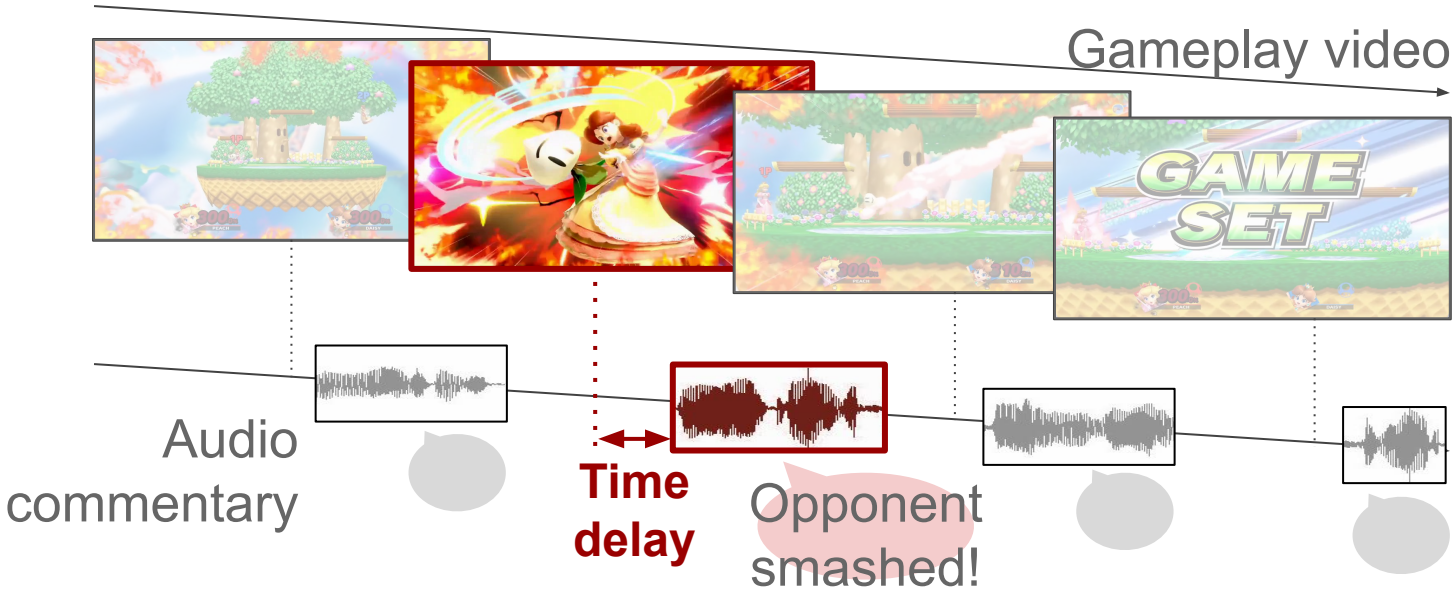
- However, when generating commentary, understanding the video and synthesizing the audio take time, **resulting in a delay**.
- To produce natural commentary audio,
  it is important to **keep the delay within an acceptable time**.

e.g.) Significant audio delay

The commentary is mistimed.

The scope of this study

- Using gameplay videos of Super Smash Bros. Ultimate, we measure **how much delay in commentary can be tolerated**.



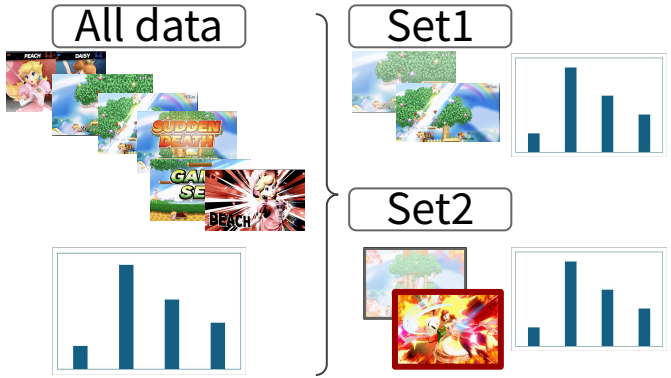
How much delay in commentary is acceptable to viewers?

Proposed method

Proposed method

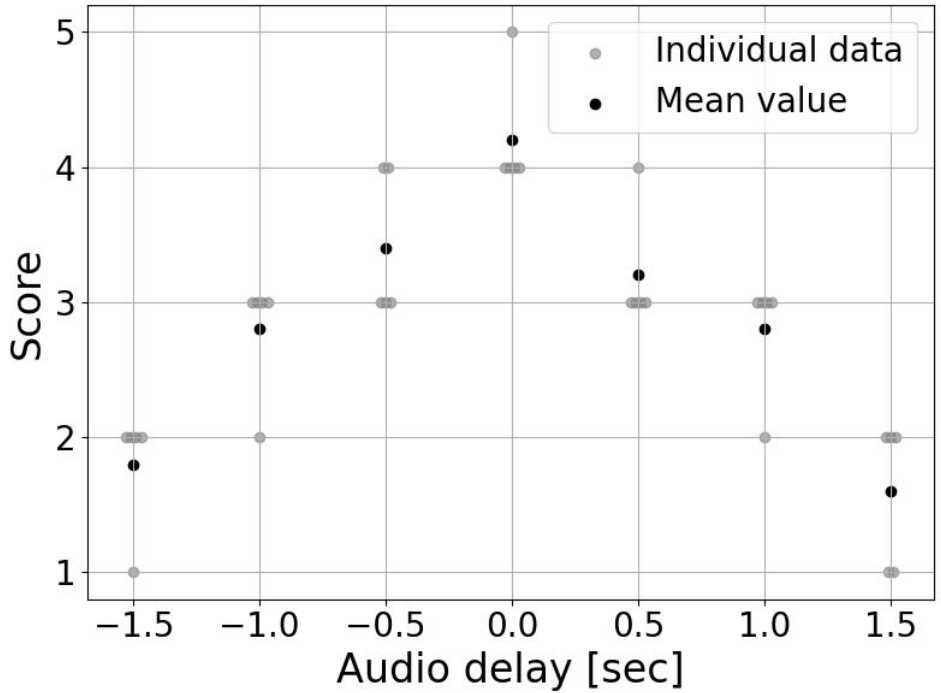
- **Creation of video stimulus**
  - Add **positive** / **negative** delay to the audio.
  - Include the preceding utterances.

- **Construction of video stimulus sets**
  - Using paired data {gameplay video, commentary audio, topic tag}.
    - Topic tags are labels (e.g. Item) describing the commentary content.
  - Include an equal number of samples for each delay condition.



Evaluation of time delay by participants

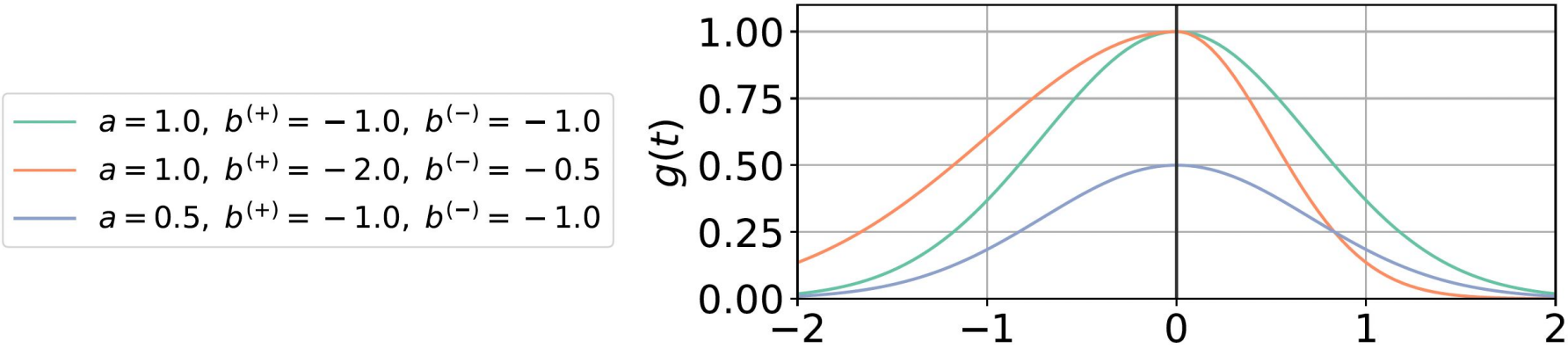
- Evaluated by the constant method[Watson 90] using a 5-point scale.
- The average evaluation score is calculated for each delay condition.



The regression curve used in this study

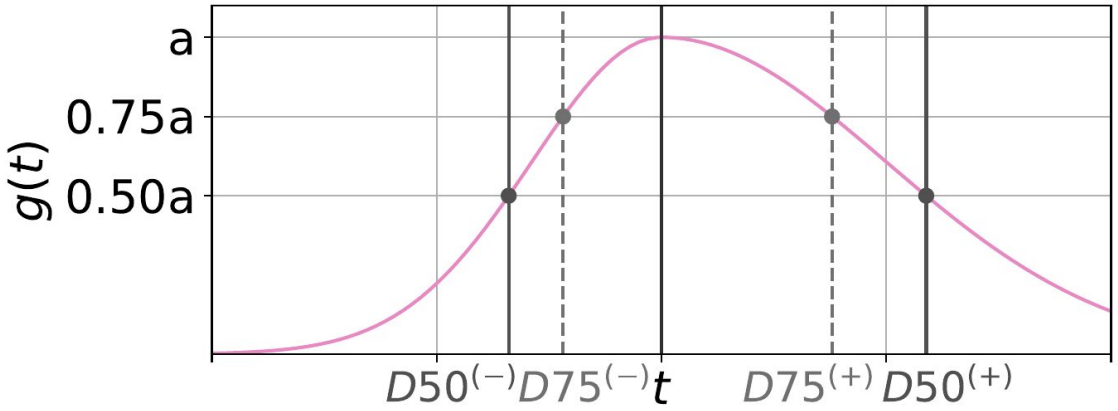
- A regression curve fits to the scores.
  - Unnormalized Gaussian func. with each-side variances, centered at 0.

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



Measurement of acceptable delay time

- The allowable delay time is defined as the time when the regression curve reaches **75%** and **50%** of its maximum.
  - The study[Hyodo 24] defined the threshold as 50% of its maximum.
  - To examine less perceived discomfort, we also use 75%.

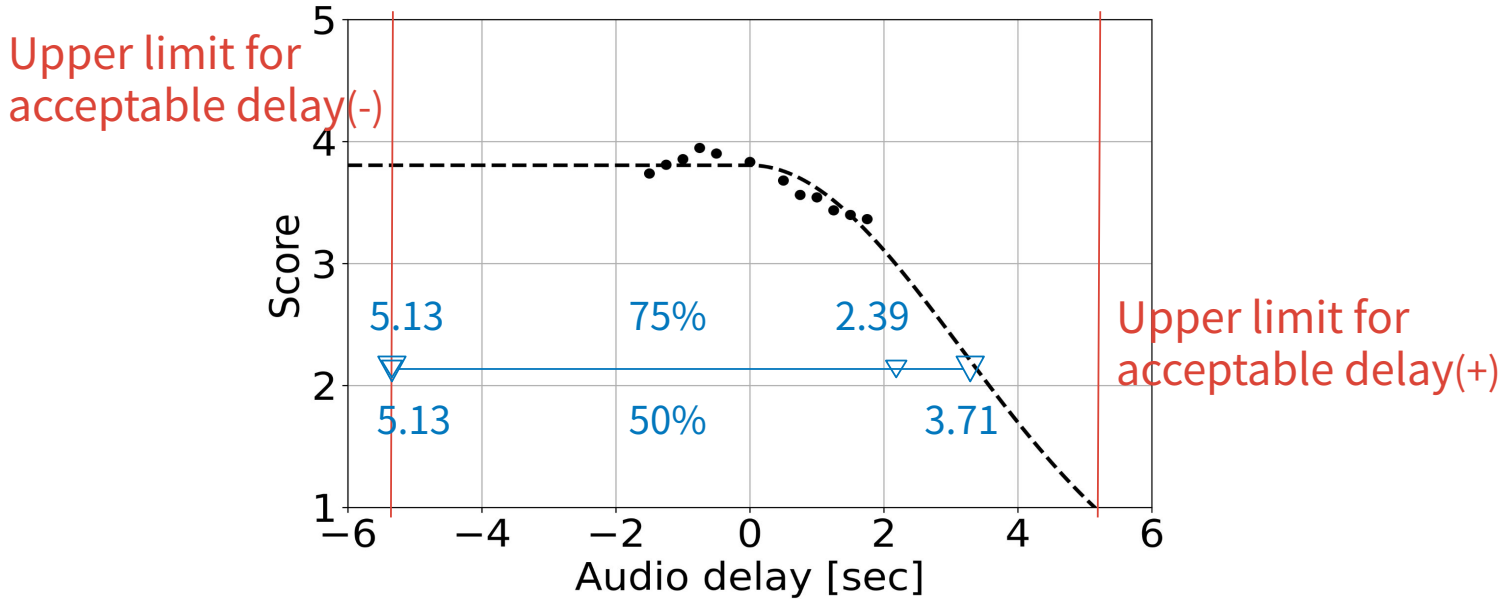


Experiments

Experimental conditions

Used corpus	SMASH corpus[Saito 20]
Number of participants	727 (hired by crowdsourcing)
Viewer attributes	Experience (Gameplay, Viewing commentary), etc. Details in the paper
Topic tag	Result, Scene, Item, Fighter
Delay time	-1.50 to +1.75 sec with 0.25 increments
Max duration of video	20 sec
Upper limit for acceptable delay	5.13 sec

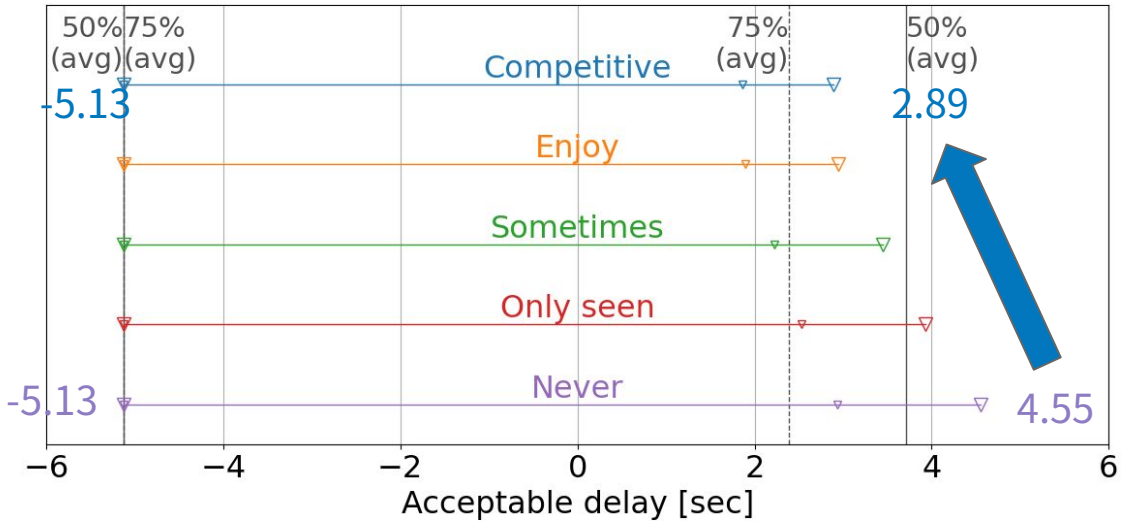
Experimental results: overall



- For positive delays, the delays is 3.71 seconds at the 50% threshold and 2.39 seconds at the 75% threshold.

Experimental results: gaming experience

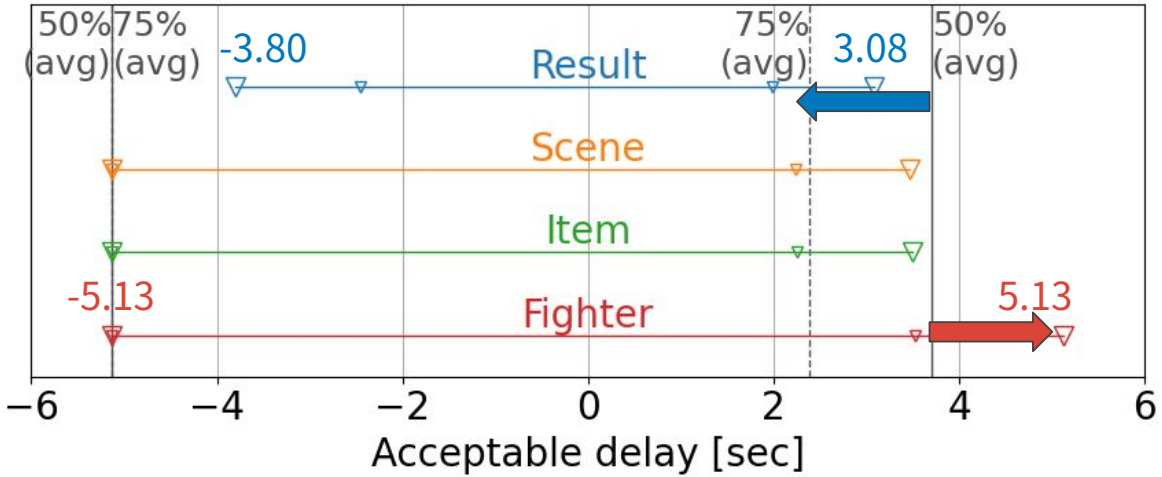
- We investigated whether gameing experience affects the acceptable.



- The acceptable delay time tends to be **shorter for participants with more gameplay experience**.

Experimental results: topic tag

- We investigated whether topic tag affects the acceptable.



- The acceptable delay time is extremely short for the "**Result**" topic and extremely long for the "**Fighter**" topic.

Conclusion

- Objective:
  - To measure the acceptable delay time in commentary audio.
- Proposed method:
  - Participants evaluate gameplay with delayed commentary audio, and the acceptable delay time is measured based on their responses.
- Evaluation results:
  - Acceptable delay time varies depending on participant attributes.
  - It also varies depending on video content, such as topic tags.
- Future work:
  - Automatic evaluation of commentary audio.

- 動画自体は日本語のもので良いが，日本語を知らない人でもわかるような工夫（英語はある程度）
- キーワードになるものをスライドに載せとくと，本番忘れた時にも何とかなる
-