Synchronization in Games Sound: An Audiovisual Study on Player Experience and Performance

Yu Chen
The Chinese University of Hong Kong, Shenzhen, China
yuchen@link.cuhk.edu.cn

Tian Min
The Chinese University of Hong Kong, Shenzhen, China
116010168@link.cuhk.edu.cn

Juntao Zhao
The Chinese University of Hong Kong, Shenzhen, China
117010399@link.cuhk.edu.cn

Wei Cai
The Chinese University of Hong Kong, Shenzhen, China
Shenzhen Institute of Artificial Intelligence and Robotics for Society Shenzhen, China
caiwei@cuhk.edu.cn

ABSTRACT

The asynchronous perception is one of the critical factors in the game experience. It includes temporal asynchrony in the game elements and interaction. The aesthetic coherence is also a part of the synchronous perception. There are emerging studies on asynchronous perception as the rising of the next generation gaming platform. We conducted two studies in this paper. First, we investigated the sensitivity of human perception to the asynchronous audiovisual and analyzed how the audiovisual content affects the perception. Then, we further explored the in-game asynchronous perception, focusing on the asynchronous sound effect and background music. The result indicates that the ability to distinguish the asynchronous audiovisual is a normal distribution among the people, and the in-game asynchronous sound will reduce the overall performance and experience, especially in the tempi of the background music.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; • **Applied computing → Computer games**.

KEYWORDS

audiovisual, asynchronous, game perception, player experience

ACM Reference Format:
Yu Chen, Tian Min, Juntao Zhao, and Wei Cai. 2022. Synchronization in Games Sound: An Audiovisual Study on Player Experience and Performance. In *2nd edition of the Game Systems Workshop (GameSys '22), June 14, 2022, Athlone, Ireland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3534085.3534342

GameSys '22, June 14, 2022, Athlone, Ireland

ACM ISBN 978-1-4503-9381-2/22/06…$15.00
https://doi.org/10.1145/3534085.3534342

1 INTRODUCTION

The asynchronous perception can reduce multimedia's quality of experience (QoE). As one of the most dynamic media forms, video games will suffer an even heavier QoE reduction. The problem may not seem urgent on classical gaming platforms like PC, consoles, or mobiles, where the audio and video are processed on the local client. However, the gaming platform has been evolving in recent years. With the increasing power of distributed computing and network transmission, cloud gaming has been realized. Cloud gaming has been discussed for years both in the academy and industry. Cai et al. classified cloud gaming and discussed the future of cloud gaming in 2016 [5]. Google released its cloud gaming platform supporting several *3A* games in November 2019 [1]. The players are excited about the coming cloud era. In the foreseeable future, cloud gaming may become the mainstream. Due to the inevitable delay in media data transmission, the asynchronous perception in video games should be seriously considered and it is worth studying the asynchronous perception in video games.

In the game study, the effect of asynchronous perception has already attracted many researchers' interest. In the gaming experience, asynchronous perception can be summarized in two aspects: asynchronous interaction and asynchronous audiovisual. Previous literature focuses on the different in details from these two aspects. Previous efforts investigated the asynchronous thematic of background music (BGM) [16] and the asynchronous sound effects (SFX) with the game difficulty [7]. It's still lack of a systematic study of the concurrent effects of these factors in games. In this work, we discuss both the asynchronous audiovisual and asynchronous in-game interaction through two experiment-driven studies. The first study focuses on audiovisual where we set up a crowdsourcing experiment to investigate the sensitivity of asynchronous audiovisuals and how media content affects this perception. The second study is the in-game asynchronous interaction study. In this study, the time delay, asynchronous tempo, and asynchronous aesthetic music are tested in a series of laboratory experiments. The core contributions of this work can be summarized as: 1) We confirmed the different sensitivity to the asynchronous audiovisual stimuli in the crowd. 2) We gave preliminary evidence of the impact of media content on asynchronous audiovisual perception. 3) We explored various aspects of asynchronous perception that influence the performance and player experience (PX).

The structure of our paper is as follows: In section 2, we present related work on asynchronous perception and the PX. Section 3 describes the research methodologies adopted in studies. The research design and hypothesis of the two studies in our work are described in section 4. Section 5 gives the result and the analysis of the two studies. We summarize the work and discuss the future work direction in section 6.

2 RELATED WORK

2.1 Synchrony perception of game

Video game creates an interactive virtual world for players. Player receives information about the virtual world mainly in visual and audio (Vibration also provides a novel game experience in some game design, especially in the console game, but it is not especially focused in this work.) Players can control avatars or environments in games through a hardware interface (e.g., console, mouse, keyboard). It changes the state of the virtual world and feedback to the player in a close loop form.

Excellent game experience needs a sense of immersion. One of the critical concepts in the gaming immersion experience is the synchrony of perception. It includes the coherence of the virtual world itself, where the aesthetics should be unified. The Spatio-temporal variation among the in-game elements should also accord with the cause and effect of commonsense (Some gaming designs introduce anti-commonsense features as the gameplay is not discussed in this work). The asynchrony of audiovisuals reduces the objective experience and the performance of tasks. Perham and Vizard [13] study the effect of irrelevant sound effects on task performance. It shows that the noisy sound effect makes the cognitive task performance poorer. The study result further confirmed Anshel's point of view in 1978 [4]. However, to the best of our knowledge, no previous study concentrates on the impact of irrelevant sound on video game performance.

2.2 Game sound

Developers and players are paying more and more attention to the game sound, making it one of the most critical components of the game design process. Mainstream video game media and awards in the industry like Imagine Games Network (IGN), The Game Award (TGA), and The Steam Award [2] all highly valued the importance of game sound and have been annually awarded "the best game soundtrack" over the years. The game sound has become more than the aesthetics. It affects the game experience and even becomes a part of the game mechanism.

The game sound can be classified into background music (BGM), sound effects (SFX), and dialogue. In this work, we focus on BGM and SFX. Background music is usually non-narrative, lower than other sound elements. It conveys emotions and cooperates with the story for a better experience[23]. All the game designers and players hope video games can be as immersive as possible.[6] There is much literature that discusses how background music affects the game experience. Compared with background music, few studies focus on the game sound effect, usually short and triggered by players. Amezcua et.al. examine the effects of musical tempo in a video game [3] and combine it with the game's difficult design [7]. Ribeiro et al. [16] study the audiovisual thematic coherence through the methodologies of questionnaire and psychophysiology signal. Although the thematic coherence may not impact the self-report player experience, some psychophysiology features can be found in a close relationship.

2.3 Game experience

Analyzing the personal experience during game playing is a growing and critical issue. Game designers and researchers have widely adopted the scaled questionnaire. This section introduces some practical scales and describes those we adopted in this paper.

The Player Experience of Need Satisfaction (PENS) [17] is a model claimed to evaluate "fun" and player satisfaction that provides heuristic and analytic value. It is a 21-item questionnaire with a 7-point scale for each question. The model contains five aspects of player experience (PX): immersion, relatedness, intuitive controls, competence, and autonomy. It is the first commercialized and widely used scale. Although there are numerous pieces of literature adopting or validating the PENS [9, 10], the academic publications are not allowed to disclose the detail of items in the PENS. Due to the limited usage of the PENS, studies in this work do not adopt the PENS scale.

The Player Experience Inventory (PXI) is designed with a solid theoretical basis. It is based on the 'Mechanics-Dynamics-Aesthetics' (MDA) framework and the Means-End theory.[19] It has been validated that the PXI can systemically reflect the player experience [20]. The PXI contains 33 questions in 10 fields, half in psychology and half in dynamics. Each field's value is the sum point of three questions. The question is 7-point scaled from -3 to 3. The PXI can apply to the study in various game genres, including the musical role-playing game in our study.

Game experience Questionnaire (GEQ) [8] is another frequently used scale. It contains four modules: core, in-game, social presence, and post-game. Although GEQ was widely adopted in a large number of studies, some researchers [10, 11] criticized it as not rigorous and lacking scientific support. Hence, we only use the core module because our study is designed for a single-player flow experience. It is used as a validation for our result drawn from IPX. The core module of GEQ contains 33 questions in seven aspects, where each question is on 5 scales.

3 METHODOLOGY

The methodologies to access the player experience can be divided into two main categories, the objective, and the subjective methods. The objective methods are usually generated during gameplay, such as in-game actions and player physiological signals. The subjective methods are rooted in questionnaires, surveys, and interviews. The state-of-the-art works combine both objective and subjective methods to draw a solid conclusion. In [7], the objective performance is used to evaluate the difficulty of the asynchronous tempo, while the survey is used for the analysis of in-game stress and preference. In [16], physiological signal electromyography is adopted as objective assessment, and several player experience questionnaires are used as subjective evaluation.

Similarly, we adopt both of the methods in order to give rigorous and comprehensive results. For the subjective method, we saved the operations generated by the player and any other in-game element into a detailed log in the JSON file. The log contains all the objects during the gameplay, which can reconstruct the whole playing session. We design a survey requiring subjects to finish during the break time immediately after each session's game experience for the subjective method. The survey contains demographic background questions, the PXI questionnaire, and the GEQ questionnaire. The GEQ is only adopted as supportive information for the result of the PXI analysis.

4 STUDY DESIGN

We focus on the relationship between BGM/SFX and visual stimuli in this work. Video game playing is the players' interaction with the virtual environment created by digital devices. The synchrony perception is composed of interaction between game and player and the virtual world itself. We conducted two studies covering the details of the two components mentioned above. This work aims to provide solid evidence for the importance of synchrony and systematically describe synchrony perception in the PX.

The first study aims at learning the perception of audiovisual synchrony with an emphasis on sensing the disparity. We developed an online system, APIC (Audiovisual Perception Integration Crowdsourcing Platform, to promise the demographic variety and the number of the participant. Given that video games pass information and experience mainly through visual and audio stimuli, we use clips instead of actual games as material in this synchronization test to prompt the subjects to concentrate on audiovisual to the utmost extent. This design also screens out the effect of the different action complexity by gameplay or genre (e.g., real-time strategy games require much higher action per minus than visual novel). The task detail and the material content will be reported in subsections.

The second study explores the audiovisual synchrony of video games from another perspective called In-game Asynchronous Perception Study. This study pays particular attention to the performance and overall PX during game playing. Subjects experience a turn-based role-playing game with their action and experience recorded. The detail of the game will be introduced in the following subsection. The performance and player experience may greatly vary in different environments and gaming devices. Hence, we conducted this experiment offline in laboratory. We recruited volunteers from the university campus. Each subject can take a fifteen-minute practice session to avoid the learning effect and wired actions caused by unexpected game elements. Two formal behavior study sessions are followed. Each behavior study session contains ten minutes of game playing that is slightly different from the practice session. Subjects should complete a survey of the game experience immediately after playing. Subjects are not informed of the change of the game set before the experiment ends.

4.1 Study 1: Explore the Audiovisual Temporal Asynchronous Perception in the Crowd

4.1.1 Recruitment Strategy & Material Selection. We hosted our system on Amazon Mechanical Turk (MTurk) over October 2020. The task was delivered in three successive seven-day sessions. Participants can only give a single attempt in each session to prevent the learning effect. Apart from MTurk, we also sourced a small batch in a Chinese mobile crowdsourcing platform and put an open call for volunteers on the forum to enrich data source diversity. The flow chart of how crowdsourcing workers participate in our experiment is shown in Fig. 1.

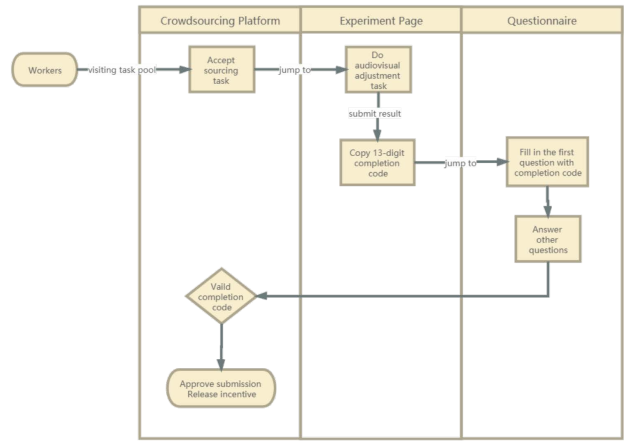

Figure 1: The flow chart for crowdsourcing. The completion code is used to identify and give reward to participants.

We selected 9 clips from 3 genres: *1. Music Video. Contains instrument or vocal clips*; *2. Beats. Contains the clips with a strong linkage between video and audio, or obvious rhythm*; *3. Language. Contains human conversation selected from movies or TV series.* All the clips are selected from the internet or social media. By investigating participants' performance, we tried to cluster them in categories with explainable commonness. Previous efforts[12, 18, 21] proved that perception sensitivity can be reflected by Temporal Integration Windows(TIW) or Point of Subjective Simultaneity(PSS) distribution. Hence, we presumed that the content of audio and visual affects the sensitivity of synchronous perception, and participants would tend to tolerate higher asynchronous values with strong coupling content.

4.1.2 Experiment Detail. Participants will be directed with the guidelines, privacy policies, and experiment purpose before starting the tasks. The system will examine participants' browsers and discourage those who use the mobile client to minimize inevitable system errors. There are two tasks in total: *Task-1: Determine if the audio and video are synchronized, then determine their order; Task-2: Adjust the audio track to fit the video track for three attempts.*

A clip with random content will be shown on the top of the screen with an asynchrony value from 0 to 1.5 seconds, either the video or the audio ahead in uniform distribution. The system's interface is shown in Fig. 2. We divided the web page from top to bottom into media player and operation panel. The media player has been modified to prevent operations in progress. Participants can only start or restart the media by clicking the button below. The content in the operation panel varies for different tasks.

In Task-1, there will be two single-choice questions: *Q1: Are the audio and the video synchronized? (Yes/No)*; *Q2: Are the audio and the video synchronized? (Audio/Video)*

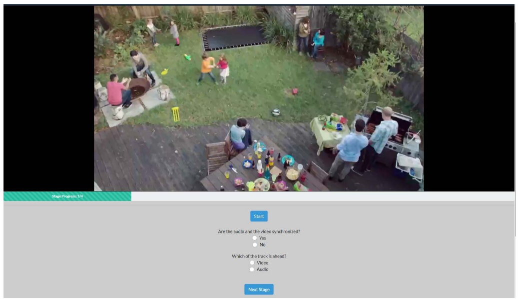

Figure 2: The main interface of the task. Participants should make choices or finish the operation in the gray area.

Figure 3: The interface of Rhythm Dungeon

In Task-2, the participant can rewind or forward the audio track by clicking two buttons where the adjustment step can be tuned by a scroll bar above from 0.1 to 0.9 seconds. The default value is set to 0.3 seconds. We recorded the direction, adjustment step, and timestamp of every operation. Participants can choose *'Done'* if they are confident that the tracks are synchronized. We set a *'Give Up'* button as well, in case the audiovisual is still asynchronous after many attempts.

After submitting the test result, the participant will receive a completion code for identifying the upcoming questionnaire. Apart from essential investigation including age, gender, and native language, we sought answers to the following questions regarding users' perceptions of audiovisual asynchrony:*Q1: Are you confident with your ability to distinguish audiovisual asynchrony?*; *Q2: Does this experiment better reflects your perception of audiovisual asynchrony?*; *Q3: To what degree of audiovisual asynchrony annoying your experience?*

4.2 Study 2: Explore the Asynchronous Perception during the Game Playing

The design of this study is motivated by to explore how audiovisual asynchrony affect the subjective experience and objective performance. Rhythm Dungeon is a mixture of the music game and a rogue-like game [22]. Each round, the player starts with a new character. During the play, the player encounters randomly generated enemies, according to the power of the player's character. The player must press buttons in sync with the drum beats to fight against enemies. The combat is turn-based, and each turn lasts for two bars in terms of music. Each turn, the player can choose to move upward or downward, attack, or charge. Each bar contains four beats, and the player has to press four buttons on each beat to deploy one action. The player should press the keys in harmony with the background music to make a successful move. A visual helper is shown near the bottom of the screen, as shown in figure 3, which indicates the correct timings to press the keys.

Three control experiments are randomly assigned to each behavior study session: The first task is to explore the effect of delay SFX. The delay time is [0, 0.8]s in a step of 0.2s, a total of 5 SFX delays. All other settings are the same as those in the practice session.

The second task is to observe the effect of asynchronous (in some literature, refer to it as incoherent) BGM. In the original game soundtrack, All the BGMs are composed consistently with the aesthetic and beats of the game, but the beats per minus are changed to 86, 112, 168, and 196, which is the 60%, 80%, 120%, and 140% of the original 140 BPM BGM. The sound effect and the rhythm of each turn are still 140 BPM unchanged.

The third task is about aesthetics. The original game's background is an adventure in a dungeon. The visual and sound effects are customized in accord with the theme. In the experiment, we alter the BGM into Japanese light music, Workout music in the gym, and solo metronome sound. All the BGMs keep in 140BPM, which is in line with the SFX.

4.3 Hypotheses

We expect to find evidence to support the hypotheses below.

- **H1. The sensitivity of audiovisual asynchrony among people follows a normal distribution.** The perception sensitivity varies among people. The final asynchronous time value's expected histogram should be a normal distribution with a mean around 0. The result indicates that most people's temporal synchronous perception is approaching perfect synchrony.
- **H2. Content of audiovisual media may affect synchrony perception.** The clips were selected with different genres. The temporal sensitivity of some content may be higher than others. We hypothesize that certain content may improve participants' ability to fix the asynchrony.
- **H3. The performance of the asynchronous gameplay is worse than the original game.** Study 2 is with a set of controlled experiments. We hypothesize that subjects' performance in the asynchronous cases is worse among all cases. The result indicates the importance of synchrony in the player experience.
- **H4. Players may or may not be aware of the asynchrony subjectively.** Although the game experience is reduced in the asynchronous gameplay, not all subjects can be aware of the tiny changes. It is common in other PX matrices, like reducing the resolution of the game graphics.

5 RESULT

5.1 Result of Study 1

5.1.1 Recruitment Result. We recruited 611 participants in the form of three batches (B1:102, B2:406, B3:103). Among these periods, only the first batch is fully completed. The other two batches have a

	MTurk B1	MTurk B2	MTurk B3	Mobile
Assignment	100	1000	200	49
Completion	102	380	103	54
Approved Rate	99	377	101	48
Average Time	3min 54s	4min 23s	4min 57s	/
Duration	10min	10min	30min	60min

Table 1: Summary of crowdsourcing batches. Each batch was collected in 7 days

submission rate of 38% and 51% in seven days. Seven individuals participated in all three periods; 33 individuals took part in two periods. The total sample size after filtering out 5 participants due to incomplete or duplicated submission, or invalid entries, was 606, with a Human Intelligence Task (HIT) approval rate of 99.2%. On the mobile crowdsourcing platform where worker background is much different from MTurk, we collect 48 samples with an approval rate of 88.89%.

Table 1 summarizes the sourcing batches. Because submission of the questionnaire is not mandatory, the number of the questionnaires we retrieved is less than the operation submission, which is 518 respondents from all platforms. Some participants gave positive feedback on the experiment and explained the disparity between the number of completed tasks and the questionnaire: participants claimed they made multiple attempts with different clips to get better performance and submitted the questionnaire once.

5.1.2 Performance Analysis. To analyze **Hypothesis 1**, we design data mining and visualization strategies to analyze the log data submitted by our participants. As the first step, we screen out the poor quality data. The initial time gap between the audio and video tracks generated by the system is between ±1.5s with a uniform distribution. We collect the final time gaps adjusted by participants to find out the Temporal Integration Window figure of the Timeline Adjustment (TA) modality. The distribution is normal as expected, with its mean slightly less than 0. The negative middle point does not in line with the Simultaneous Judgment (SJ) in previous works[14, 15]. The explanation is participants operate an audio track to fit the video track, which means they will focus on the audio first and then find an opportunity to match the video track. We exclude the data with a final gap of more than 5 seconds, which are identified as outliers exceeding the Three-σ Bond. These outliers can be explained by some workers doing meaningless operations or technical problems, including internet fluctuation or unloaded cache.

5.1.3 Content Analysis. To analyze **Hypothesis 2**, we complied the final average Point of Subjective Simultaneity (PSS). In the previous section, we define nine clips into three classes: Music Videos, Beats, and Language. The final time gaps for musical videos are worse than the original gap (1.5s). No significant improvement shows in beats videos. For language videos, all clip's final time gaps improve. All results are statistically significant.

Three clips in the Language class give us the best result. The result indicates the content of the clips has a substantial impact on the synchronization performance. We suspect the reason beneath this outcome is the degree of linkage between image and sound. For example, the clips we chose in Music Video include footage of hands playing the piano quickly or a man playing the violin. Participants can hardly find the trait and pattern to connect the sound with the image. Alternatively, participants can quickly realize and match the voice and speakers in the Language clips, judging whether the audio or video is ahead. Then, precisely locate the soundtrack's position by observing the mouth shape in the clips. Besides, we notice that participants operate when an extraordinary event happens in the clips. For instance, objects are dropping or speakers switching. To sum up, the content of the media determines the amount and intensity of the stimulus, which can affect the synchrony perception or the ability to eliminate the errors further.

5.2 Result of study 2

Thirty-three subjects are recruited from the university students and staff. The participants are claimed as volunteering, but a coupon for a working meal is rewarded as a gift. All subjects took the practice session before the formal sessions, and most of the subjects finished two formal sessions. We abandoned the uncompleted sessions (Some subjects give up, and some are due to technical issues). In total, study two collected 55 valid formal sessions. Table 2 shows the detail of each control variable in the sub-studies.

We processed the objective performance data by extracting the average time error and the correct rate. The average time error is the value of the time gap in each turn. There may be 4 to 6 keypress events in each turn. We calculated the absolute value of the error of pressing time and the correct time for all participants in a session, then took the average for all events in turns. The time error for each turn can be expressed as $E = \frac{1}{n}\sum^{n}(T_{correct} - T_{press})$, where n is the number of events in the specific turn.

Another value we took is the correct rate. Here we define the correct turn: Initially, the number of keypresses is compared with the expected number. Participants' input times is considered invalid if it is less than the expected value. If the participant's input times are accepted, the correctness C of the input is determined by $C = \sum_{i=1}^{n}(t_i - T_i)^2$ Where n denotes the expected number of keypresses, t_i the time when the i-th keypress occurs, and T_i the time when the i-th keypress is expected to take place if C is larger than a threshold H, then the input is valid, and the character that the player controls make the corresponding move. If C is less than the threshold, the input is regarded as a failure. The total correct rate during the playing session can reflect the performance during the game playing. In the following few sections, we report the result of each sub-study.

5.2.1 Study 2-1. Study 2-1 is the temporal asynchronous SFX of the game. Table 2 shows the result of the descriptive statistics of the PXI. The result shows no significant difference for all ten components in the PXI. The value of time delay of SFX does not significantly affect the PX. The subjective performance test also denied the difference between the original game and the delay of the SFX experiment (F=0.855, p=0.505). The subject's comments are only asked about the general comments to the overall PX, and they have not informed the differences between the formal session and the practice session. However, 60% of subjects who experienced the 0.8s delay of SFX complained that the session is temporally asynchronous in SFX, making it hard to control the game. These complaints indicate that subjects are aware of the considerable time delay.

	Study 2-1					Study 2-2					Study 2-3			
	0.2s	0.4s	0.6s	0.8s	p	84BPM	112BPM	168BPM	196BPM	p	soft	metronome	workout	p
Meaning	1/0.53	1.23/0.22	0.25/0.83	1.11/0.47	0.542	1.58/0.79	1.39/0.43	0.8/0.43	0.22/0.48	0.391	0.78/0.51	1.40/0.56	1.00/0.57	0.720
Curiosity	2.33/0.30	2.86/0.14	1.92/0.48	2.56/0.20	0.119	2.92/0.08	2.39/0.26	2.73/0.19	1.22/0.40	**0.004**	2.22/0.47	2.67/0.26	1.93/0.49	0.514
Mastery	1.25/0.67	2.19/0.34	1.08/0.42	1.89/0.28	0.221	1.50/0.44	1.44/0.44	2.27/0.24	0.33/1.20	0.191	1.33/0.58	1.53/0.48	1.27/0.44	0.934
Autonomy	1.42/0.69	2.10/0.33	1.33/0.19	2.39/0.26	0.200	2.25/0.75	1.00/0.53	2.40/0.37	0.44/0.68	0.102	0.78/0.56	2.00/0.28	1.93/0.43	0.130
Immersion	1.42/0.39	1.71/0.37	0.92/0.55	2.39/0.38	0.143	1.92/0.50	1.94/0.29	2.40/0.22	0.22/1.06	**0.050**	1.83/0.34	1.60/0.55	1.40/0.51	0.799
Prog. Fb.	0.33/0.76	1.33/0.54	0.33/0.89	0.56/0.79	0.707	0.58/0.79	1.83/0.27	0.93/0.55	0.78/1.06	0.428	0.56/0.77	1.33/0.54	1.20/0.49	0.763
Av. Appeal	1.58/0.64	1.76/0.59	1.75/0.81	2.11/0.37	0.936	2.75/0.16	1.94/0.64	1.73/0.63	0.22/1.39	0.227	2.28/0.41	2.20/0.39	1.53/0.72	0.554
Challenge	0.67/0.72	1.67/0.40	1.42/0.34	1.67/0.31	0.419	1.42/0.74	1.17/0.24	1.60/0.49	1.44/0.99	0.585	1.56/0.32	0.53/0.95	1.80/0.39	0.321
Control	1.75/0.39	2.33/0.24	1.92/0.48	2.11/0.20	0.576	1.92/0.44	2.00/0.34	2.33/0.28	1.33/0.88	0.550	2.22/0.41	2.29/0.29	1.67/0.46	0.522
Goals	1.75/0.37	2.29/0.20	1.25/0.60	0.83/0.63	0.133	1.75/0.44	1.72/0.30	1.07/0.64	0.56/1.28	0.532	1.44/0.59	0.07/1.02	0.80/0.72	0.469

Table 2: The descriptive statistics result of the study 2. Mean/Standard deviation for each components in the PXI. The p-value of components are shown in the right column.

	(Variable)/ (Number of subjects)				sessions
2-1	0.2s/ 4	0.4s/ 7	0.6s/ 4	0.8s/ 5	21
2-2	84BPM/ 4	112BPM/ 6	168BPM/ 5	196BPM/ 3	18
2-3	Soft/ 6	Metronome/ 5	Workout/ 5		16

Table 3: The summary of the three sub-studies in the study 2. The number of available in each control variable is shown.

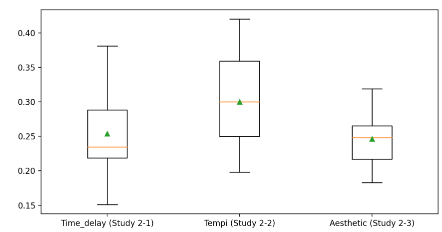

Figure 4: The box plot figure of the average time error of the three sub-studies. The vertical axis is the time in second. Higher error time means worse performance. The performance of the tempo asynchrony is the worst one across the studies

5.2.2 *Study 2-2.* Study 2-2 is the tempo asynchronous of the game. It shows that when the tempo reduces 40% to 84 BPM, the evaluation is significantly different from the other three control variables. Among all components of the PXI, curiosity ($p<0.005$) and immersion ($p=0.05$) are the most significant components. The average time error ($M=0.300$, $SD=0.004$) is much worse than the original game, studies 2-1 and 2-3 ($F=3.313$, $p=0.044$). Nonetheless, the average error ($F=0.430$, $p=0.735$) and the correct rate ($F=0.247$, $p=0.624$) are identical within the different tempo in Study 2-2.

5.2.3 *Study2-3.* Study 2-3 is about the aesthetic asynchronous of the game. Similar to studies 2-1, no significant result in the PXI. The subjects cannot ignore the change in BGM because the genres of the BGMs are considerably diverse. However, only one subject in all sessions complained about the aesthetic of the BGM. The average time error ($F=0.171$, $p=0.684$) and the correct rate ($F=0.974$, $p=0.335$) also have no significant difference within the sub-study.

5.2.4 *Analysis.* We investigated three different cases of asynchronous during game playing. The result indicates that the effect of these asynchronous sets is of varying degrees. From Fig. 8, We can see that the asynchronous tempi reduce the player performance. However, the subjective survey result does not differ across all variables, except the slowest tempi 84BPM. The objective and subjective results support the **H4**, although the performance is reduced due to asynchrony, not all subjects are aware of the reduction and reflect it in the PX questionnaire.

The **H3** cannot get a direct result from the studies. The performances are not statistically distinct worse in studies 2-1 and 2-3. Here we give two potential reasons to explain the result. First, the task load is too light for subjects. The experience time is ten minus each session at a low difficulty level. Elongating the experiment time or increasing the game difficulty may reveal the effect of asynchrony better. Second, game-playing perception highly relies on visual stimuli. The asynchronous sound shows little impact on the performance. Further studies can dig into this problem.

6 CONCLUSION

This work investigates the asynchronous audiovisual in video games. We conduct two experiment-driven studies, including a crowdsourcing test for audiovisual asynchrony perception and in-game asynchrony perception. We investigated this topic comprehensively, including numerous testing data, media content, temporal asynchrony from SFX and BGM, tempi of game sound, and aesthetic of game music. From Study 1, our result confirms that the human ability to distinguish asynchronous audiovisual follows a balanced distribution, and the content of media plays a critical role in determining whether an individual can fix asynchrony in a multimedia clip; In Study 2, the experiment gives us a better understanding in how asynchrony sound affects the performance and gaming experience. We also explored the exciting fact that the asynchronous tempo has the most profound impact on our subjects among all factors. However, the statistical analysis indicates that this influence is limited under our specific testing environment. Future research in this field should seek a universally applicable standard for the scope and influence of human perception of asynchrony through massive experiments in various games and environments to help formulate norms and improve user experience.

REFERENCES

[1] [n.d.]. Googole Stadia. http://https://stadia.google.com/. Accessed 17 Feb. 2021.

[2] [n.d.]. Steam Best Soundtrack Award. https://store.steampowered.com/steamawards#BestSoundtrack. Accessed 17 Feb. 2021.

[3] Claudia Amezcua, Miguel Angel Guevara, and Julieta Ramos-Loyo. 2005. Effects of musical tempi on visual attention ERPs. *International journal of neuroscience* 115, 2 (2005), 193–206.

[4] Mark H Anshel and Dan Q Marisi. 1978. Effect of music and rhythm on physical performance. *Research Quarterly. American Alliance for Health, Physical Education and Recreation* 49, 2 (1978), 109–113.

[5] Wei Cai, Ryan Shea, Chun-Ying Huang, Kuan-Ta Chen, Jiangchuan Liu, Victor CM Leung, and Cheng-Hsin Hsu. 2016. The future of cloud gaming [point of view]. *Proc. IEEE* 104, 4 (2016), 687–691.

[6] Jiulin Zhang Xiaoqing Fu. 2015. The influence of background music of video games on immersion. *Journal of Psychology & Psychotherapy* 5, 4 (2015).

[7] Aline Hufschmitt, Stéphane Cardon, and Éric Jacopin. 2020. Manipulating Player Performance via Music Tempo in Tetris. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*. 146–152.

[8] Wijnand A IJsselsteijn, Yvonne AW de Kort, and Karolien Poels. 2013. The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven* 46, 1 (2013).

[9] Daniel Johnson and John Gardner. 2010. Personality, motivation and video games. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. 276–279.

[10] Daniel Johnson, M John Gardner, and Ryan Perry. 2018. Validation of two game experience scales: the player experience of need satisfaction (PENS) and game experience questionnaire (GEQ). *International Journal of Human-Computer Studies* 118 (2018), 38–46.

[11] Effie L-C Law, Florian Brühlmann, and Elisa D Mekler. 2018. Systematic review and validation of the game experience questionnaire (geq)-implications for citation and reporting practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 257–270.

[12] Joost X Maier, Massimiliano Di Luca, and Uta Noppeney. 2011. Audiovisual asynchrony detection in human speech. *Journal of Experimental Psychology: Human Perception and Performance* 37, 1 (2011), 245.

[13] Nick Perham and Joanne Vizard. 2011. Can preference for background music mediate the irrelevant sound effect? *Applied Cognitive Psychology* 25, 4 (2011), 625–631.

[14] Stian Reimers and Neil Stewart. 2015. Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior research methods* 47, 2 (2015), 309–327.

[15] Stian Reimers and Neil Stewart. 2016. Auditory presentation and synchronization in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods* 48, 3 (2016), 897–908.

[16] Giovanni Ribeiro, Katja Rogers, Maximilian Altmeyer, Thomas Terkildsen, and Lennart E Nacke. 2020. Game Atmosphere: Effects of Audiovisual Thematic Cohesion on Player Experience and Psychophysiology. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 107–119.

[17] Scott Rigby and Richard Ryan. 2007. The Player Experience of Need Satisfaction (PENS). *White Paper* (2007).

[18] Rob LJ Van Eijk, Armin Kohlrausch, James F Juola, and Steven van de Par. 2008. Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. *Perception & psychophysics* 70, 6 (2008), 955–968.

[19] Vero Vanden Abeele, Lennart E Nacke, Elisa D Mekler, and Daniel Johnson. 2016. Design and preliminary validation of the player experience inventory. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. 335–341.

[20] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies* 135 (2020).

[21] Argiro Vatakis, Jordi Navarra, Salvador Soto-Faraco, and Charles Spence. 2008. Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Experimental brain research* 185, 3 (2008), 521–529.

[22] Tengfei Wang, Shuyi Zhang, Xiao Wu, and Wei Cai. 2019. *Rhythm Dungeon: A Blockchain-Based Music Roguelike Game*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3337722.3341836

[23] Sean M Zehnder and Scott D Lipscomb. 2006. The Role of Music in Video Games. (2006).