Effects of Visual Stimuli on Perceived Sound Volume in Virtual Reality Spaces*

Yuki Matsuda[1,2,3][0000-0002-3135-4915], Toma Kobayashi[2], Hiroki Watanabe[4][0000-0002-6854-4448], and Keiichi Yasumoto[2,3][0000-0003-1579-3237]

[1] Okayama University, Okayama 700-0082, Japan yukimat@okayama-u.ac.jp
[2] Nara Institute of Science and Technology, Nara 630-0101, Japan
{kobayashi.toma.ks5, yasumoto}@is.naist.jp
[3] RIKEN Center for Advanced Intelligence Project (AIP), Tokyo 103-0027, Japan
[4] Future University Hakodate, Hokkaido 041-8655, Japan hwata@fun.ac.jp

Abstract. With the proliferation of affordable and high-performance virtual reality (VR) devices, VR content such as games and the metaverse is becoming increasingly widespread. In VR environments, users experience various sensory stimuli, primarily through visual and auditory cues. However, subjective perception of these stimuli varies based on user context. Existing studies have shown that auditory perception can be influenced by visual stimuli, however, most of them have focused on congruent audiovisual stimuli, leaving the effects of non-congruent pairings unexplored. This study investigates how visual stimuli, specifically color and crowdedness, influence perceived sound volume in VR. In the experiment that participants experienced VR environments with different room colors while listening to test tones, the results showed that warm colors led to higher perceived volume at low sound levels. Also, in the experiment that participants viewed VR scenes with varying crowd densities while hearing announcements, less crowded environments resulted in higher perceived sound volume. These findings suggest that visual context impacts auditory perception, providing insights for optimizing hearable devices and enhancing VR auditory experiences.

Keywords: Virtual Reality · Subjective sound volume · Visual stimuli

1 Introduction

With the advent of inexpensive and high-performance virtual reality (VR) devices, VR content such as VR games and metaverse are rapidly becoming more widespread. In VR spaces, users can experience various sensations and stimuli such as speed or falling, with most systems relying on visual and auditory cues. However, previous studies have reported that subjective perceptions of these stimuli can vary depending on the user's circumstances. For instance, research

* This study was partly supported by JST PRESTO Grant Number JPMJPR2039, JPMJPR2138 and JSPS KAKENHI Grant Number JP24K20763.

has shown that visual stimuli presented via smart glasses can alter users' perception of time passage [10], and the playback speed of videos viewed on smart glasses can affect the speed at which users perform tasks afterward [6]. Moreover, social factors such as the COVID-19 pandemic have been reported to influence perceived crowding in certain spaces [7]. Regarding auditory stimuli, subjective sound volume can also vary based on the user's context. For example, the perceived sound volume can be influenced by the preceding sound volume [15]. Also, the visual effects can alter the perceived sound volume or impressions of the sound even with sounds of the same physical intensity [9]. However, in these studies, the visual and auditory stimuli used were typically congruent (e.g., car sounds with car visuals), and the effects of non-congruent pairings (e.g., park visuals with voice guidance) have not been explored yet. Watanabe et al. found that quieter environments, such as parks, lead to higher perceived sound volume compared to busier environments like crowded spaces [13]. However, because the distinction between visually "busy" and "calm" stimuli in that study was somewhat ambiguous, further evaluation using simplified and generalized stimuli is needed. By investigating the generalized impact of visual stimuli on perceived sound volume, it could inform improvements in hearable devices, such as earphones, by not only adjusting volume based on environmental noise but also considering visual stimuli, thereby providing optimal auditory experiences.

From this background, this study investigates the effect of visual stimuli on perceived sound volume in VR environments, focusing on simplified stimuli of color and crowdedness. In the experiment focusing on "color," visual stimuli (VR room with various colors) and auditory stimuli (test tones with randomly varied volumes) were presented to the participants through a head mounted display (HMD). The experiment is conducted with 20 participants aged 20–30 years, and we compare the perceived sound volume in the room with warm and cool colors. The results indicated significant differences in perceived sound volume when actual sound levels were low, with warm colors making it harder to perceive volume decreases (sounds felt louder) compared to cool colors. In the experiment focusing on "crowdedness," 16 participants aged 20–30 years were shown four-type 360-degree videos in VR (two scenes and two crowdedness levels) along with an announcement sound whose volume varied. The statistical analysis was conducted to analyze the effects of the environment category, crowdedness level, and sound volume level on perceived sound volume. As the results, main effects were found for both the environment category and sound volume level, with the city being perceived as louder than the cafeteria. Furthermore, the analysis of simple main effects revealed that when the sound volume was high, crowdedness had a significant effect, with less crowdedess leading to higher perceived sound volume compared to more crowdedness.

2 Related Work

This section outlines existing research related to the effects of auditory and visual stimuli on perceived sound volume.

The phenomenon of perceived sound volume decreasing with sustained auditory stimuli has been studied extensively over time [4]. Yataka et al. proposed a system that dynamically adjusts the volume output from earphones based on environmental conditions, such as surrounding noise or ongoing conversations [15]. Similarly, Yataka et al. leveraged the influence of previously heard sound volume on perceived volume to develop a mechanism that ensures consistent perceived volume when delivering auditory information from a system [16].

Several prior studies suggest that visual and auditory information can mutually influence each other. Sakuma et al. investigated whether changes in visual information affect perceived sound volume [9]. Using a head-mounted display (HMD) with a webcam, they implemented a system that modifies visual effects at user-specified locations to control perceived sound volume. Abe et al. studied the impact of visual information on the perception of environmental sounds [1]. Their evaluation experiments, in which visual stimuli were presented alongside auditory stimuli, showed that visual additions affected perception across factors other than sound brightness.

Tokunaga et al. reported that subjective sound volume changes when participants hear only automobile noise compared to when they hear the noise alongside visual information [12]. Similarly, Fujimoto et al. showed that evaluations of sound intensity, noisiness, and disturbance differed when participants listened to actual traffic noise on-site, noise alone in a laboratory, and noise combined with visuals in a laboratory setting [5]. Szychowska et al. confirmed that evaluations of environmental noise differ depending on whether auditory stimuli, visual stimuli, or both are presented [11]. Bangjun et al. examined the impact of visual confirmation of noise sources, reporting that visible noise sources increase discomfort in identical noise environments [3]. Miyakawa et al. investigated the impressions of 62 types of environmental sounds presented with and without visual information [8]. Similarly, Asakura demonstrated that the subjective impressions of urban environmental sounds differ depending on the presence and type of visual information [2]. Their findings indicated that perceived sound volume increases when visual stimuli such as waterfalls or airplanes are present, compared to conditions without visual stimuli.

These studies explored the influence of visual information on auditory perception and confirmed that visual and auditory information are interrelated. However, these works primarily investigated scenarios where visual and auditory stimuli corresponded. The current research, by contrast, examines the effects of non-corresponding visual and auditory stimuli.

Building on prior work, the authors investigated how visual stimuli influence perceived sound volume. They found that perceived sound volume tends to be higher for less stimulating visuals, such as park scenes, compared to more stimulating visuals, such as crowds [13]. However, the criteria for evaluating the intensity of visual stimuli were ambiguous. Consequently, they concluded that further evaluation using simplified and generalized stimuli is necessary.

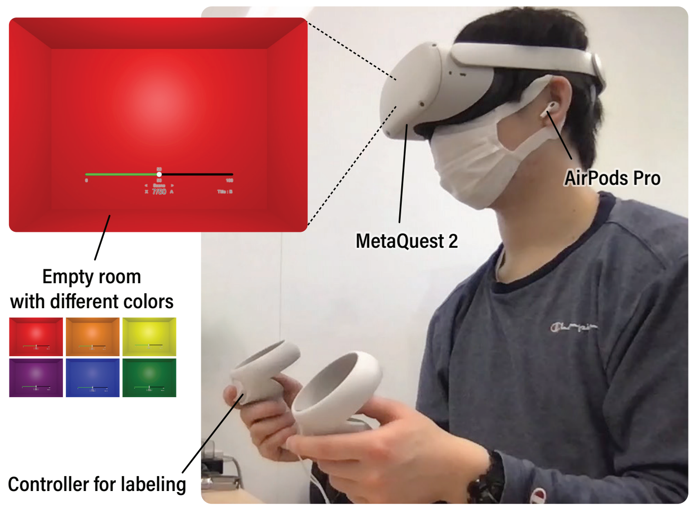

Fig. 1. Experimental setup

3 Impact of Colors on Perceived Sound Volume

This study aims to investigate how simple visual stimuli in a VR environment affect perceived sound volume. To simplify the stimuli compared to the setup used by Watanabe et al. [13], which involved visual stimuli such as city and park videos and auditory stimuli such as news audio, we focus on “color” as the visual stimulus and “test tones” as the auditory stimulus. In this paper, we examine whether differences in visual stimuli, specifically changes in the “color” of a room in a VR space, influence perceived sound volume.

3.1 Experimental Conditions

The experimental setup is shown in Fig. 1. Participants wore an immersive HMD (MetaQuest2[5]) for visual stimuli and used earphones (Apple AirPods Pro[6]) for auditory stimuli. Visual stimuli consisted of a room created in Unity (Fig. 2) within a VR environment. The room measured 8m (length) $\times$ 4m (width) $\times$ 3m (height), and its wall color alternated between white, red, orange, yellow, purple, blue, and green. The auditory stimuli were defined as 1 kHz sine waves (test tones), with volumes varying within a $\pm$6dB range to present different stimuli.

3.2 Experimental Procedure

Using the setup described above, the following procedure was used to collect data. In this experiment, the reference auditory stimulus was defined as volume

[5] https://www.meta.com/jp/quest/products/quest-2/
[6] https://www.apple.com/airpods-pro/

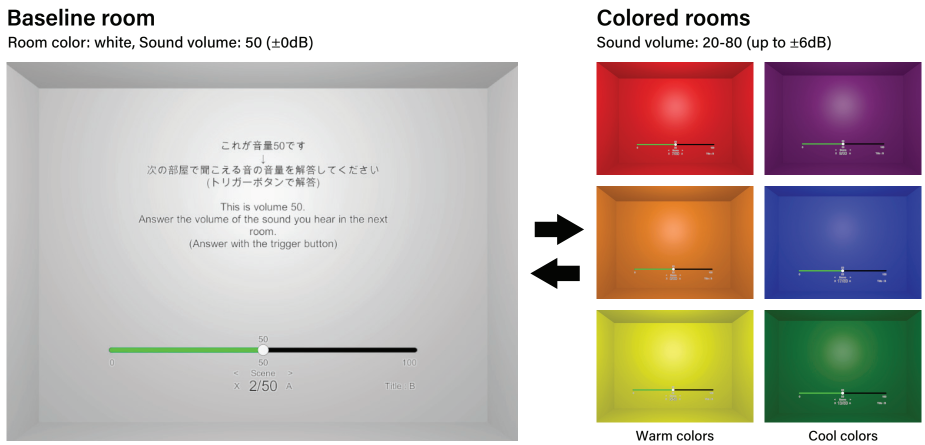

Fig. 2. Display screen in the VR environment

50 on a scale of 0–100. First, as a baseline, the room in the HMD displayed white walls while presenting the reference sound (volume 50). Subsequently, the HMD displayed a colored room (randomly selected from red, orange, yellow, purple, blue, or green) while presenting auditory stimuli with randomly adjusted volumes within ±6dB (set to 20 for -6dB and 80 for +6dB). Finally, participants rated how the auditory stimulus compared to the reference sound using a scale from 0 to 100. For example, if a participant felt the stimulus was slightly louder than the reference sound (volume 50), they might respond with 70. This sequence constituted one trial. After each trial, the HMD returned to the baseline white room with the reference sound, and the next trial began. The process was repeated 24 times (6 colors × 4 repetitions). Visual and auditory stimulus combinations and their presentation order were randomized for each participant. The experiment included 20 participants aged in their 20s and 30s.

3.3 Experimental Results

Participants rated the perceived sound volume (hereafter V_{subj}) relative to the presented volume (hereafter V_{true}). To analyze this, the difference V_{diff} between the presented and perceived volumes was calculated as shown in Fig. 3 and the equation below:

$$V_{\text{diff}} = \begin{cases} V_{\text{subj}} - V_{\text{true}} \ (V_{\text{true}} \geq 50) \\ V_{\text{true}} - V_{\text{subj}} \ (V_{\text{true}} < 50) \end{cases} \tag{1}$$

Analysis Using All Data The differences V_{diff} were categorized by visual stimuli (room colors) and are shown in Fig. 4. A Shapiro-Wilk test (significance level 5%) indicated that the dataset did not follow a normal distribution

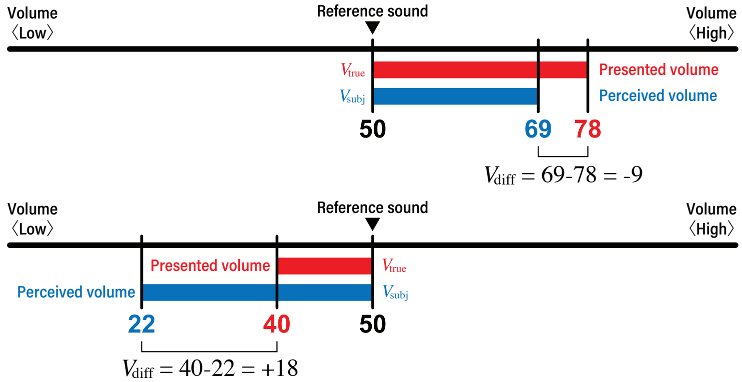

Fig. 3. Calculation of the difference V_{diff} between presented and perceived volumes

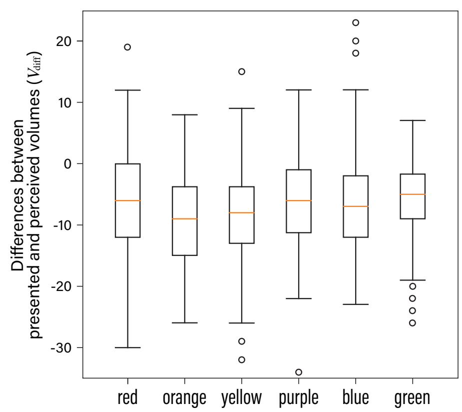

Fig. 4. Differences between presented and perceived volumes (V_{diff}).

($p = 0.010, p < 0.05$). Therefore, a Kruskal-Wallis test was conducted to determine whether the visual stimuli significantly influenced perceived sound volume, yielding $p = 0.034$ ($p < 0.05$), indicating a significant effect. However, post-hoc Bonferroni tests revealed no significant differences between specific color pairs, likely due to the reduced statistical power from multiple comparisons.

Volume-Dependent Analysis To assess whether the influence of visual stimuli depends on volume, data were grouped into three categories based on V_{true}: low (20–40), medium (40–60), and high (60–80). The result is shown in Fig. 5 (a), (b), and (c). Shapiro-Wilk tests confirmed that none of the subsets followed a normal distribution ($p < 0.05$). Kruskal-Wallis tests showed significant effects of visual stimuli on perceived sound volume only in the low-volume group ($p = 0.01, p < 0.05$), but not in the medium or high-volume groups ($p > 0.05$). This result supports findings from prior studies [13].

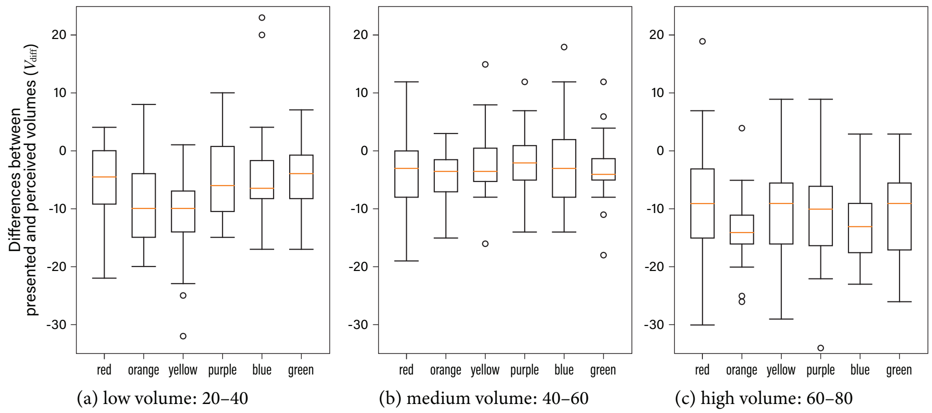

Fig. 5. Differences between presented and perceived volumes (V_{diff}) for six room colors.

Analysis by Color Categories Finally, we investigated the influence of "color temperature" (warm vs. cool colors) on perceived sound volume. Warm colors (red, orange, yellow) and cool colors (purple, blue, green) were grouped into categories, and the differences V_{diff} were analyzed for each volume group (Fig. 6 (a), (b), and (c)). Shapiro-Wilk tests confirmed that none of these datasets followed a normal distribution ($p < 0.05$). A Mann-Whitney U test (significance level 5%) revealed a significant effect of color temperature only in the low-volume group ($p = 0.01, p < 0.05$), where warm colors were associated with reduced perceived sound volume changes compared to cool colors.

This result suggests that when the actual sound volume is low, warm colors may reduce the perception of sound volume changes more effectively than cool colors. This contrasts with previous findings by Watanabe et al. [13], which showed that more stimulating visuals (e.g., urban scenes) were associated with reduced perceived sound volume compared to less stimulating visuals (e.g., park scenes). This discrepancy highlights the need for further studies investigating other factors, such as motion intensity in the visual stimuli, which were not considered in the present research.

4 Impact of Crowdedness on Perceived Sound Volume

Watanabe et al. [13] suggested that less stimulating visuals, such as parks, result in higher perceived sound volumes compared to more stimulating visuals, such as urban scenes. However, "more stimulating visuals" is a vague concept requiring further clarification. Factors such as "crowdedness" or "density of buildings" could contribute to this perception. Therefore, this study focuses on the impact of "crowds" as a specific aspect of visual stimulation in VR environments and investigates how differences in visual stimuli related to crowds affect parceived sound volume.

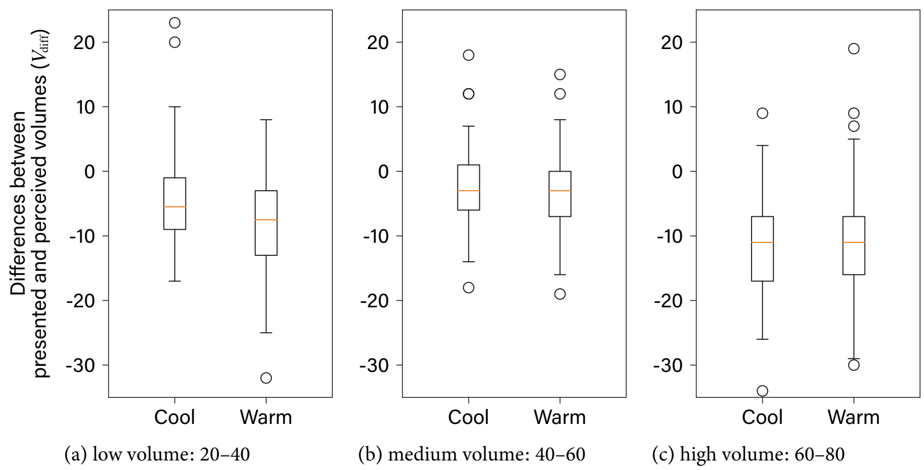

Fig. 6. Differences between presented and perceived volumes (V_{diff}) for cool and warm colors.

4.1 Experimental Conditions

As in the color experiment, the visual stimuli were presented using an immersive HMD, and auditory stimuli were delivered through earphones. Four types of 360-degree videos were prepared: two scenes (a city and a cafeteria) each with two conditions (high and low crowdedness), as shown in Fig. 7, 8, 9, and 10. The city scene used footage from Shinjuku Station, while the cafeteria scene was filmed at the university cafeteria of NAIST (Nara Institute of Science and Technology). Auditory stimuli consisted of announcement audio files with three volume levels (+3, 0, −3 dB) as described in Table 1.

4.2 Experimental Procedure

The procedure followed these steps: First, the reference sound (volume 30, 50, 70) was played against a black screen to normalize participant responses. Next, one of the 360-degree videos (Fig. 7, 8, 9, and 10) was randomly displayed, along with a randomly selected announcement audio at one of the three volume levels. Participants then rated the perceived volume relative to the reference volume on a 0–100 scale. This process was repeated 24 times (4 videos $\times$ 6 trials), with randomized video and audio combinations for each participant. Sixteen participants in their 20s and 30s (both male and female) participated.

4.3 Experimental Results

A three-way ANOVA was conducted to analyze the effects of environmental factors (city, cafeteria), crowdedness (high, low), and stimulus volume (high=70,

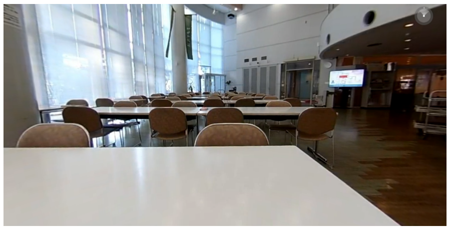

Fig. 7. Cafeteria (low crowdedness)

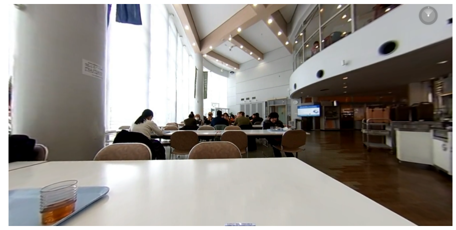

Fig. 8. Cafeteria (high crowdedness)

Fig. 9. City (low crowdedness)

Fig. 10. City (low crowdedness)

Table 1. Announcement audio used in the experiment

Reference sound	**Stimulus sound**
This is volume 30.	Thank you for attending today.
This is volume 50.	Please refrain from taking photos or videos during the performance.
This is volume 70.	We appreciate your cooperation in silencing your mobile devices.

Table 2. Results of the three-way ANOVA

Factor Combination	Sum of Squares	df	F-value	p-value
Environment	648	1.0	11.0	0.001 *
Crowdedness	0.211	1.0	0.004	0.952
Stimulus Volume	5890	2.0	49.8	0.000 *
Environment * Crowdedness	22.5	1.0	0.381	0.538
Environment * Volume	75.7	2.0	0.640	0.528
Crowdedness * Volume	705	2.0	5.95	0.003 *
Environment * Crowdedness * Volume	20.7	2.0	0.175	0.839

medium=50, low=30) on perceived sound volume. The difference (V_{diff}) between presented volume (V_{true}) and perceived volume (V_{subj}) was plotted in Fig. 11.

The results of the three-way ANOVA are summarized in Table 2, and the interaction plot is shown in Fig. 12. The analysis revealed significant main effects for the environmental factor and the stimulus volume factor ($p < 0.05$). Examining the interaction plot, it was observed that across all samples, the difference between the presented and perceived volumes was smaller for the city environment compared to the cafeteria, meaning that sound was perceived as relatively louder in the city.

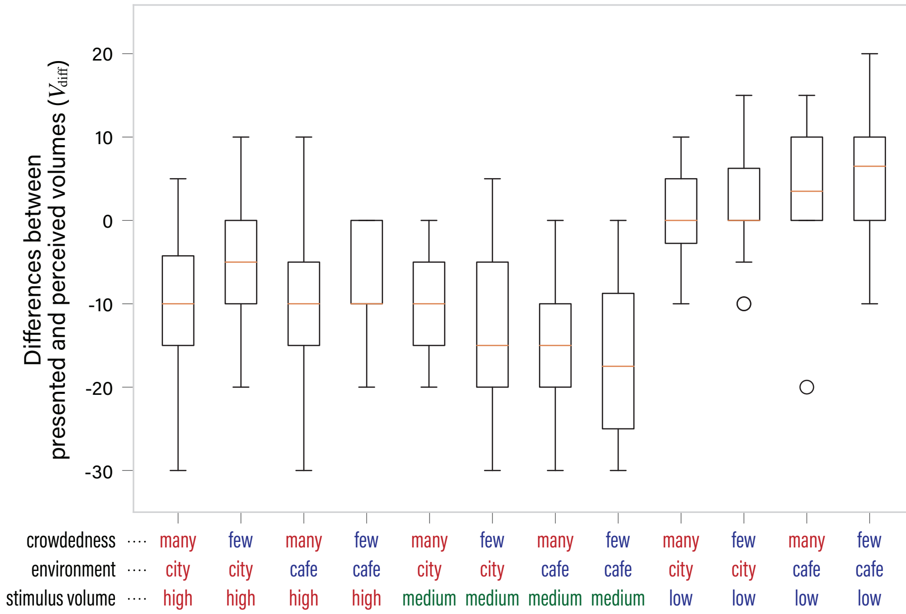

Fig. 11. Differences between presented and perceived volumes (V_{diff}) by three factors

Table 3. Simple main effects analysis for crowdedness and stimulus volume

Factor	Sub-group	Sum of Squares	df	F-value	p-value
Crowdedness	Crowded (many)	3280	2.0	23.9	0.000 *
	Empty (few)	3320	2.0	31.8	0.000 *
Stimulus Volume	High volume (70)	481	1.0	7.09	0.009 *
	Medium volume (50)	130	1.0	1.82	0.180
	Low volume (30)	94.5	1.0	2.28	0.133

In addition, a significant first-order interaction was observed between crowdedness and stimulus volume ($p < 0.05$). A simple main effects analysis (Table 3) was conducted to further investigate this interaction.

For crowdedness, the main effect of stimulus volume was observed for both high- and low-crowd conditions. This indicates that changes in stimulus volume significantly affect the difference between the presented and perceived sound volumes. However, the sequential presentation of reference sounds ($30 \rightarrow 50 \rightarrow 70$) may introduced order bias in the results. Future experiments should consider randomized or counterbalanced presentation orders for the reference sounds.

For stimulus volume, the main effect of crowdedness was observed only for volume 70. The interaction plot (Fig. 12) shows that, under high stimulus volume, sound was perceived as louder in the low-crowd condition compared to the high-crowd condition. This suggests that lower crowdedness amplifies the perceived loudness of sound under high-volume conditions.

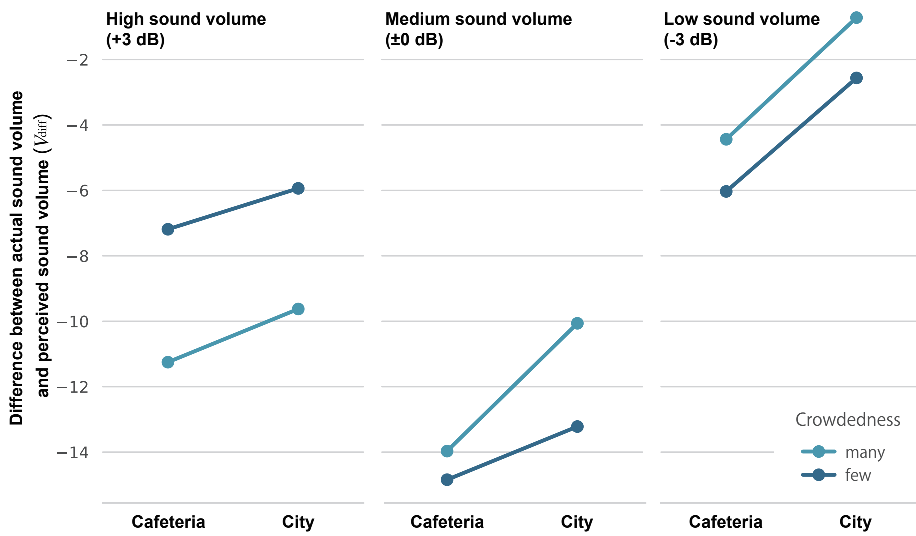

Fig. 12. Interaction plot (Environment, Crowdedness, and Stimulus volume)

5 Discussion

Conditions Where Visual Stimuli Affect Perceived Sound Volume This study found that for "color," cool colors were associated with a greater perceived decrease in sound volume compared to warm colors when the actual sound volume was low. For "crowds," sound was perceived as louder in low-crowd conditions compared to high-crowd conditions when the actual sound volume was high. Based on these findings, the conditions where sound is perceived as loudest could be "warm colors with low crowdedness," while the conditions where sound is perceived as softest could be "cool colors with high crowdedness." However, since this conclusion integrates results from two separate experiments, additional experiments under combined conditions are needed to validate the findings.

Perceived Sound Volume in Different Environments Watanabe et al. [13] reported that less stimulating visuals, such as parks, result in higher perceived sound volumes compared to more stimulating visuals, such as urban scenes. This study revealed that sound was perceived as louder in city environments compared to cafeterias. Combining these results suggests the following order of perceived sound volume: "cafeteria $<$ city $<$ park." Furthermore, as cafeterias are enclosed spaces while cities and parks are open spaces, the findings imply that perceived sound volume increases in open spaces and decreases in enclosed spaces.

Improvements to the Experimental Design In the color experiment, only a reference sound with a volume of 50 was presented, making it challenging to evaluate differences beyond this volume. To address this, the crowdedness experiment presented reference sounds sequentially at volumes 30, 50, and 70. However, this sequential approach may have introduced order bias. Future experiments should randomize or counterbalance the presentation order of reference sounds. Additionally, providing a broader range of reference sounds, such as 20, 40, 60, and 80, may improve accuracy.

Brightness Levels of Colors The colors used in the experiment were defined by their RGB values as follows: red (255, 0, 0), orange (255, 165, 0), yellow (255, 255, 0), green (0, 128, 0), blue (0, 0, 255), and purple (128, 0, 128). Converted to HSV, the brightness (V) values were as follows: red (100), orange (100), yellow (100), green (50), blue (100), and purple (50). Brightness affects room illumination, suggesting that future experiments should consider brightness as a variable when designing the experimental setup.

Audio Levels of Original Videos In the crowdedness experiment, the audio levels of the videos were measured at 32 ± 5 dB. Although the reference sound was fixed at 35 dB, the actual audio levels in the video environments were not consistent. While the video audio was muted during the experiment, future studies should examine the perceived sound volume of the video audio itself.

Color and Perceived Time Related research has examined the relationship between color and perceived time [14]. In one study, participants counted three minutes while exposed to different lighting colors (red, orange, yellow, green, blue, purple) and raised a placard when they felt three minutes had passed. Results showed that yellow, orange, blue, purple, red, and green influenced perceived time in descending order. This suggests that "cool colors make perceived time slower." While the present study found that "cool colors make perceived sound softer," these results together may indicate that "cool colors dull sensory perception."

6 Conclusion

VR environments can present various sensory stimuli, such as speed and free fall, primarily through visual and auditory information. However, subjective impressions like perceived sound volume vary based on the VR environment. Therefore, sound volume adjustments should consider the surrounding conditions. To clarify how visual stimuli affect perceived sound volume, this study focused on simplified visual stimuli, "color" and "crowdedness," and investigated their effects in a VR environment.

Future research should investigate other visual factors, such as distance, light and shadow, and movement intensity, while considering individual attributes (e.g., gender, age) and activities (e.g., motion) that may influence perception in VR environments.

References

1. Abe, K., Ozawa, K., Suzuki, Y., Sone, T.: The Influence of Visual Information on Perception of Environmental Sounds. Journal of the Acoustical Society of Japan **56**(12), 793–804 (2000). https://doi.org/10.20697/jasj.56.12_793
2. Asakura, T.: Effect of visual stimuli on subjective evaluation of sound. Journal of the Acoustical Society of Japan **74**(12), 655–662 (2018). https://doi.org/10.20697/jasj.74.12_655
3. Bangjun, Z., Lili, S., Guoqing, D.: The influence of the visibility of the source on the subjective annoyance due to its noise. Applied Acoustics **64**(12), 1205–1215 (2003). https://doi.org/10.1016/S0003-682X(03)00074-4
4. Charron, S., Botte, M.: Frequency selectivity in loudness adaptation and auditory fatigue. The Journal of the Acoustical Society of America **83**(1), 178–187 (1988). https://doi.org/10.1121/1.396443
5. Fujimoto, K., Sakata, N., Yada, M.: A comparison of annoyance of road traffic noise evaluated in the field with that in the laboratory. Journal of Architecture, Planning and Environmental Engineering **364**, 10–19 (1986). https://doi.org/10.3130/aijax.364.0_10
6. Hasegawa, E., Isoyama, N., Monteiro, D.V., Sakata, N., Kiyokawa, K.: The Effects of Speed-Modulated Visual Stimuli Seen through Smart Glasses on Work Efficiency after Viewing. Sensors **22**(6) (2022). https://doi.org/10.3390/s22062272
7. Matsuda, Y.: IoPT: A Concept of Internet of Perception-aware Things. In: The 12th International Conference on the Internet of Things. pp. 201–204. IoT '22 (2022). https://doi.org/10.1145/3567445.3571108
8. Miyakawa, M., Suzuki, S.i., Aono, S., Takagi, K.: The effects of visual information on the impressions of environmental sounds. Journal of the Acoustical Society of Japan **56**(6), 427–436 (2000). https://doi.org/10.20697/jasj.56.6_427
9. Sakuma, I., Terada, T., Tsukamoto, M.: Design and Implementation of a Subjective-Volume Control System using Visual Effects. In: Entertainment Computing Symposium. pp. 357–364 (2015)
10. Shimizu, T., Futami, K., Terada, T., Tsukamoto, M.: In-Clock Manipulator: Information-Presentation Method for Manipulating Subjective Time using Wearable Devices. In: Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia. pp. 223–230. MUM '17 (2017). https://doi.org/10.1145/3152832.3152873
11. Szychowska, M., Hafke-Dys, H., Preis, A., Kociński, J., Kleka, P.: The influence of audio-visual interactions on the annoyance ratings for wind turbines. Applied Acoustics **129**, 190–203 (2018). https://doi.org/10.1016/j.apacoust.2017.08.003
12. Tokunaga, Y., Terashima, T., Ishikawa, A.: Influence of visual information on subjective evaluation of road traffic noise. In: INTER-NOISE and NOISE-CON Congress and Conference Proceedings. InterNoise '14, vol. 249:4, pp. 3607–3611 (2014)
13. Watanabe, H., Matsuda, Y.: Investigation of the effect of visual information on subjective-volume change of audio information presentation. In: Multimedia, Distributed, Cooperative, and Mobile Symposium. pp. 1318–1322. DICOMO '22 (2022)
14. Yamashita, M.: Effects of color space on human temporal experience. Journal of the Color Science Association of Japan **44**(3), 25–28 (2020). https://doi.org/10.15048/jcsaj.44.3__25

15. Yataka, S., Tanaka, K., Terada, T., Tsukamoto, M.: A context-aware audio presentation method in wearable computing. In: Proceedings of the 2011 ACM Symposium on Applied Computing. p. 405–412. SAC '11 (2011). https://doi.org/10.1145/1982185.1982274
16. Yataka, S., Terada, T., Tsukamoto, M.: An Audio Information Presentation Method Considering Subjective Volume for Wearable Computing. In: IPSJ SIG Technical Reports. vol. 2011-UBI-29:38, pp. 1–8 (2011)