

Article

Not peer-reviewed version

Visual Harmony Between Avatar Appearance and On-Avatar Text: Effects on Self-Expression Fit and Interpersonal Perception in Social VR

[Yang Guang](#)*, Sho Sakurai, Takuya Nojima, Koichi Hirota

Posted Date: 2 December 2025

doi: 10.20944/preprints202512.0242.v1

Keywords: metaverse; social VR; avatar; text design; harmony; self-expression fit; interpersonal perception; Proteus effect; hyperpersonal communication; VRChat



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Visual Harmony Between Avatar Appearance and On-Avatar Text: Effects on Self-Expression Fit and Interpersonal Perception in Social VR

Yang Guang , Sho Sakurai , Takuya Nojima  and Koichi Hirota 

Graduate School of Informatics and Engineering, University of Electro-Communications, Chofu, Tokyo 182-8585, Japan

* Correspondence: yangguang@vogue.is.uec.ac.jp

Abstract

In social virtual reality (VR) and metaverse platforms, users express identity through both avatar appearance and on-avatar textual cues, such as speech balloons. However, little is known about how the harmony between these cues influences self-representation and social impressions. We propose that when avatar appearance and text design, including color, font, and tone, are consistent, users experience stronger self-expression fit and elicit greater interpersonal affinity. A within-subject study ($N = 21$) in VRChat manipulated social context, color harmony between avatar hair and text, and style or content consistency between tone and font. Questionnaires provided composite indices for perceived congruence, self-expression fit, and affinity. Analyses included repeated-measures ANOVA, linear mixed-effects models, and mediation tests. Results showed that congruent pairings increased both self-expression fit and affinity compared to mismatches, with mediation analyses suggesting that self-expression fit partially carried the effect. These findings integrate theories of avatar influence and computer-mediated communication into a framework for metaverse design, highlighting the value of consistent avatar and text styling.

Keywords: metaverse; social VR; avatar; text design; harmony; self-expression fit; interpersonal perception; Proteus effect; hyperpersonal communication; VRChat

1. Introduction

Communication in online environments has shifted from text-based exchanges to embodied interactions using avatars in social VR and the metaverse. Avatars act as digital bodies, while speech balloons and text provide an additional channel for identity cues. Prior research has shown that avatar appearance influences self-perception and the impressions formed by others, a phenomenon often called the Proteus effect [1]. Similarly, research on computer-mediated communication has demonstrated that consistency across communication channels amplifies impressions, as explained by the hyperpersonal model [2]. Despite these findings, the combined effect of avatar appearance and on-avatar text design has not been systematically examined.

Recent commercial social VR platforms, including VRChat, Cluster, and Rec Room, now attract millions of users. These platforms are used not only for entertainment but also for education, professional meetings, and collaboration. In such contexts, the clarity and naturalness of self-representation become essential. Previous studies in human–computer interaction suggest that inconsistencies across modalities disrupt presence and rapport, while coherence across channels supports trust and smooth interaction. The present study positions avatar and text harmony as a key component of self-presentation, arguing that the ability to maintain a coherent identity is fundamental for sustainable communication in the metaverse.

Beyond entertainment and social activities, the metaverse is increasingly used for industrial applications such as remote collaboration, medical training, and educational instruction. For instance,

enterprises deploy VR platforms for distributed teamwork, while universities adopt immersive classrooms that allow students to engage through avatars. In such domains, coherent identity projection is not merely aesthetic but instrumental to credibility, efficiency, and inclusivity (e.g., [3,4]).

Furthermore, self-presentation theory emphasizes that individuals strategically manage impressions in order to achieve social goals. When visual and textual signals conflict, this impression management is disrupted, potentially leading to reduced trust and less willingness to collaborate. Therefore, understanding avatar-text harmony is not only a theoretical challenge but also a practical necessity for ensuring effective interaction in professional metaverse contexts (e.g., [2]).

In addition, theoretical perspectives from psychology offer further motivation. Gestalt principles of perception highlight that humans interpret stimuli holistically, and fluency theory proposes that stimuli which are easier to process are evaluated more positively [5–7]. Taken together, prior evidence suggests a principled bridge from theory to hypotheses. Proteus-effect findings imply that visual appearance can shape self-perception, color and typography studies imply that textual styling conveys warmth/competence, and multimodal-coherence and processing-fluency accounts imply that cross-channel consistency enhances authenticity and interpersonal judgments. Therefore, we derive: H1 (harmony elevates self-expression fit and affinity), H2 (the effect on affinity is mediated by self-expression fit), and H3 (effects increase with immersion where cue coupling tightens).

This work tests three hypotheses stated verbatim in Section 3.

1.1. Contribution and Scope

We formalize a theory-driven account that links cross-modal harmony to self-expression fit and interpersonal affinity; operationalize harmony with jointly controlled color, font, and tone; evaluate effects across immersion levels with reliability/sphericity/effect-size and model-based robustness; and derive metaverse interface implications while restricting claims to the tested settings (within-subject, $N = 21$).

2. Related Work

2.1. Avatar-Mediated Behavior and Impressions

Avatars serve as transformed self-representations that can alter user behavior and shape how others respond. Studies have demonstrated that variations in avatar appearance, such as height, attractiveness, and clothing, can shift dominance, confidence, and perceived warmth. These effects show that the visual channel is powerful for signaling identity traits and is interpreted rapidly by observers. This pattern accords with the Proteus effect and related evidence that avatar appearance shapes anthropomorphism, credibility, and attraction [1,8–12].

2.2. Textual Expressivity in Online Communication

Textual features, including font, length, and positivity, also influence impression formation. The hyperpersonal account emphasizes that senders manage cues to achieve desired impressions, especially in environments that allow careful self-presentation. In metaverse settings, on-avatar text functions as a coupled presentation channel that can either reinforce or contradict visual signals from avatars. Work in color psychology and typography also demonstrates that seemingly subtle choices affect perceived warmth, competence, and trust [2,13–17].

2.3. Multimodal Coherence

Research on multimodal perception highlights that coherence across channels supports presence and reduces cognitive dissonance. When avatar appearance and text style align in meaning, the combination is processed more fluently and judged more positively; deliberate mismatches may draw attention but can undermine comfort and authenticity [5,18,19].

2.4. Metaverse Identity Work

Recent studies document how users explore gender and identity in VRChat [4] and how organizations value avatar customization [3]. These developments underscore the practical importance of coherent avatar and text design for self-presentation and social interaction in immersive spaces.

2.5. Gap and Positioning

Prior work has rarely treated avatar and on-avatar text as a unified semiotic package in immersive settings. The present work positions harmony as the bridge that unifies these channels in social VR, links it to perceived authenticity and affinity, and specifies conditions under which harmony matters most.

3. Theory and Hypotheses

Figure 1 depicts the hypothesized structure in this study. Solid arrows represent causal paths from avatar–text harmony to self-expression fit and from self-expression fit to interpersonal affinity. Dashed lines represent the moderating role of immersion (2D < VR < social VR). Because moderation does not imply an additional causal direction, dashed lines are drawn without arrowheads.

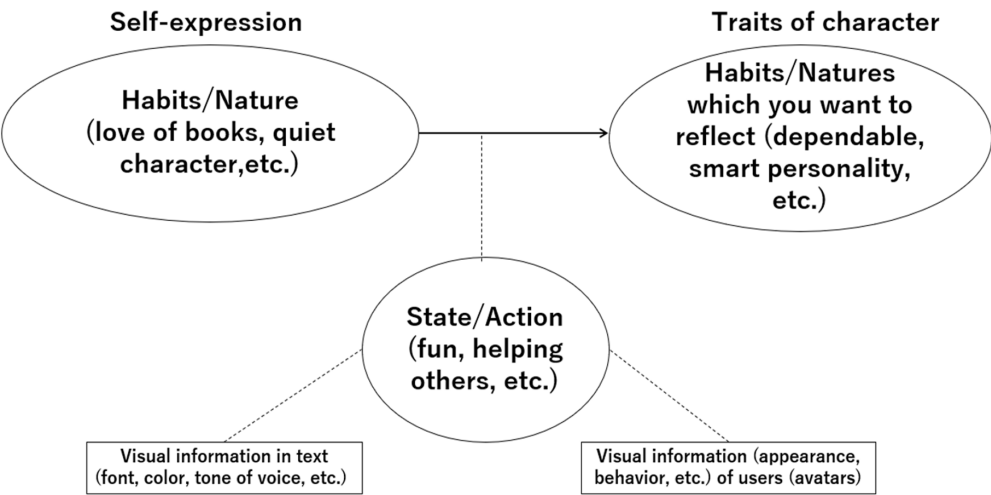


Figure 1. Conceptual model illustrating causal and moderating relations among variables. Solid arrows denote hypothesized causal paths (Harmony → Self-expression fit → Interpersonal affinity). Dashed lines denote the moderation by immersion (2D → VR → Social VR); they do not indicate causal direction but highlight that immersion strengthens the overall pathway. Dashed lines indicate moderation by immersion (2D < VR < social VR). Dashed lines intentionally have no arrowheads because moderation does not imply additional causal direction.

We posit that avatar–text harmony—defined as the congruence of color, font, and tone with avatar appearance—enhances self-expression fit, which in turn increases interpersonal affinity. Perceived congruence is treated as a proximal check that the harmony manipulation is noticed, whereas self-expression fit captures the deeper identity-relevant appraisal that we expect to drive affinity. We further predict a graded moderation by immersion (2D < VR < social VR), consistent with embodiment and telepresence accounts. Higher immersion increases embodiment and telepresence, which tighten sensorimotor coupling and the integration of visual and textual cues; as a result, cross-modal congruence becomes more diagnostic of identity and more consequential for judgments.

H1. Harmonious compared to mismatched avatar–text combinations produce higher self-expression fit and higher interpersonal affinity.

H2. The effect of avatar–text harmony on affinity is mediated by self-expression fit.

H3. Effects strengthen with immersion, with greater embodiment and tighter cue coupling in social VR compared with VR and 2D.

4. Materials and Methods

4.1. Overview of Approach

We test H1–H3 in a within-subject design from a fixed third-person vantage in a private VRChat world (standardized lighting/camera; free chat disabled). After each scene, participants rate perceived congruence, self-expression fit, and interpersonal affinity. Text-only stimuli are included to isolate textual design from avatar appearance.

4.2. Participants and Ethics

Twenty-one adults from a university community took part in the study with informed consent. The protocol was approved by the ethics board of the University of Electro-Communications (approval no. H25019). All participants reported normal or corrected vision and basic familiarity with VR.

4.3. Apparatus and Platform

The study used a Meta Quest 2 headset tethered to a PC [20]. A private VRChat world was created with controlled lighting and fixed viewing angles to standardize experience across participants. Speech balloons with predesigned fonts and colors were displayed above avatars. Chat input was disabled to prevent uncontrolled variation.

4.4. Sessions and Immersion Levels

We operationalized immersion at three levels: (i) a 2D desktop session (non-embodied viewing on a standard monitor), (ii) a VR session (single-user embodied viewing via HMD), and (iii) a social VR session (multi-user, in-world viewing in a private VRChat world). Scene content, camera angle, and timing were matched as closely as possible across sessions; order was counterbalanced at the block level.

4.5. Task

Across all conditions, participants always evaluated the presenter-side avatar (or its textual substitute in text-only conditions) as the single focal target.

On each trial, participants observed a scene in which a single target avatar (the presenter avatar standing in front of the poster) was shown from a fixed third-person viewpoint in a private VRChat world. Even in scenes where another avatar appeared for contextual purposes, instructions clarified that participants must evaluate only the presenter-side avatar. Participants never role-played this avatar; instead, they evaluated the target avatar strictly as observers.

After each scene, participants rated three constructs using 5-point Likert scales: (1) perceived harmony (congruence between the text balloon and the target avatar when text was present), (2) self-expression fit (two items; whether the target avatar represented an authentic and suitable identity), and (3) interpersonal affinity (six items: willingness to talk, friendliness, trust, warmth, positive impression, and satisfaction).

Two practice trials using non-experimental screenshots ensured comprehension. Trials were randomized with breaks. Free chat, gestures, and motion were disabled, and avatar poses were fixed to avoid additional cues.

4.6. Stimuli and Manipulations

We designed warm vs. cool manipulations following prior associations in color and typographic form, and validated all stimuli via manipulation checks before testing main effects. To avoid stereotype assumptions, cultural and individual differences are acknowledged in the Limitations; claims are restricted to the tested setting. Avatar clothing, facial expression, pose, and camera angle were held constant across appearance conditions; only hair color varied to manipulate warm vs. cool appearance. Stimulus screenshots removed extraneous text (e.g., in-world labels) to avoid unintended cues.

Fonts (added). We used Rounded M+ 1c for the warm style and Roboto Condensed for the cool style, with size/letter-spacing/line-height/placement/contrast held constant; see Tables 1 and 2 for licensing and theoretical justification.

Table 1. Stimulus font and justification (Warm; added).

Field	Detail
Font family	Rounded M+ 1c
Mapping	Warm
License / source	SIL Open Font License / M+ Fonts
Stimulus control	Size, letter-spacing, line-height, placement, and contrast were held constant across conditions.
Rationale	Rounded curvature is linked to warmth/approachability and lower “hardness/rigidity” impressions in typography studies [13–15,17].

Table 2. Stimulus font and justification (Cool; added).

Field	Detail
Font family	Roboto Condensed
Mapping	Cool
License / source	Apache 2.0 / Google Fonts
Stimulus control	Same controls as Table 1 (size, tracking, line-height, placement, and contrast fixed).
Rationale	Angular/condensed forms are associated with competence/professionalism and cooler impressions in prior work [13–15,17].

4.6.1. Fonts and Typographic Mapping

To instantiate the “warm” (rounded) vs. “cool” (angular/condensed) text styles, we used Rounded M+ 1c for warm and Roboto Condensed for cool. These choices align with typographic findings that link rounded curvature and higher perceived warmth/approachability vs. angularity/condensation and higher perceived competence/professionalism. To avoid confounds beyond warm/cool, font size, letter-spacing, line height, placement, and contrast ratios were held constant across conditions. The exact font names, licenses, and versions are listed in Tables 1–2.

As Table 3 shows, twelve conditions were organized into three independent groups (Harmony, Presence, Text-content).

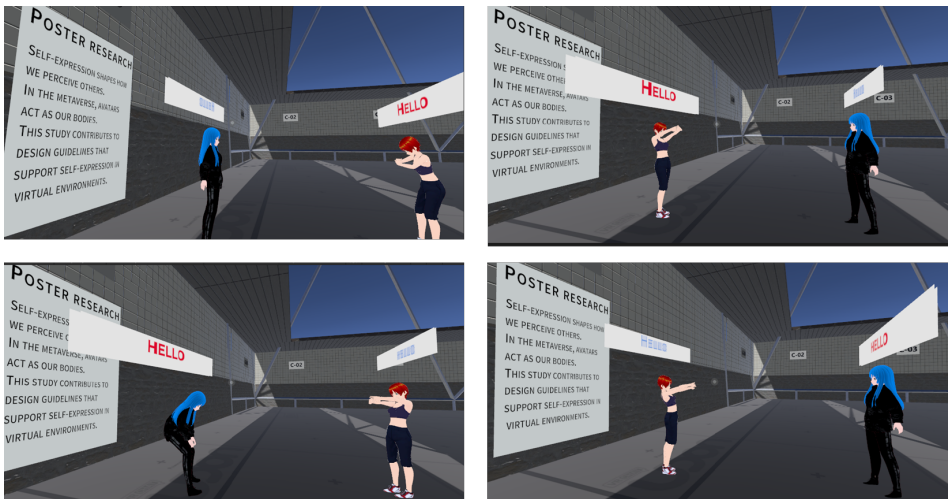


Figure 2. Experimental conditions in the metaverse study (Harmony group). Top-left: congruent warm avatar with warm text; Top-right: congruent cool avatar with cool text; Bottom-left: mismatched cool avatar with warm text; Bottom-right: mismatched warm avatar with cool text. (The in-world text is displayed in Japanese for ecological validity; the conceptual contrasts are warm vs. cool.)

Table 3. Twelve experimental conditions structured as 4 × 3 groups.

Group	Condition	Description
Harmony (4)	HC-warm	Congruent warm avatar + warm text (red/orange hair; warm-colored rounded font, cheerful wording)
	HC-cool	Congruent cool avatar + cool text (blue hair; cool-colored angular font, neutral wording)
	HM-warm	Mismatched warm avatar + cool text (red/orange hair; cool-colored angular font, neutral wording)
	HM-cool	Mismatched cool avatar + warm text (blue hair; warm-colored rounded font, cheerful wording)
Presence (4)	P1	Warm avatar only (no text balloon)
	P2	Cool avatar only (no text balloon)
	P3	Warm + cool avatars together, no text
	P4	Warm avatar facing cool avatar, no text
Text-content (4)	T1	Cheerful tone × Warm style (cheerful wording, warm-colored rounded font)
	T2	Cheerful tone × Cool style (cheerful wording, cool-colored angular font)
	T3	Neutral tone × Warm style (neutral wording, warm-colored rounded font)
	T4	Neutral tone × Cool style (neutral wording, cool-colored angular font)

Note. The twelve conditions were evaluated as three independent groups (Harmony, Presence, Text-content) to limit participant burden; a full 4 × 4 × 4 factorial (64 cells) was not implemented. The Text-content group preselected four tone × style combinations from a pilot for discriminability and balanced burden; remaining combinations are deferred to follow-up studies.

4.6.2. Avatar-Only Control (Figure 3)

To estimate impressions attributable to appearance alone, we used an avatar-only control with no text balloons. Screenshots were captured under the same lighting and camera settings as the main stimuli.

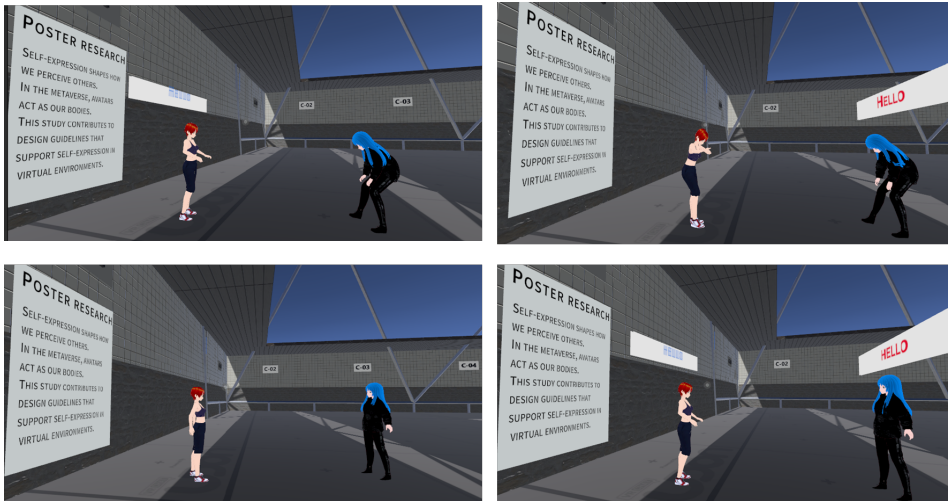


Figure 3. Control conditions without text balloons (avatar-only baselines). These screens isolate appearance-driven impressions for comparison with avatar–text combinations.

4.6.3. Text-Only Emphasis (Figure 4)

To isolate the contribution of textual design independent of avatar appearance, we included a text-only emphasis condition. As illustrated in Figure 4, the text element (color, font style, and tone) was presented without the target avatar, and participants evaluated the same constructs with respect to this text stimulus after viewing each scene.

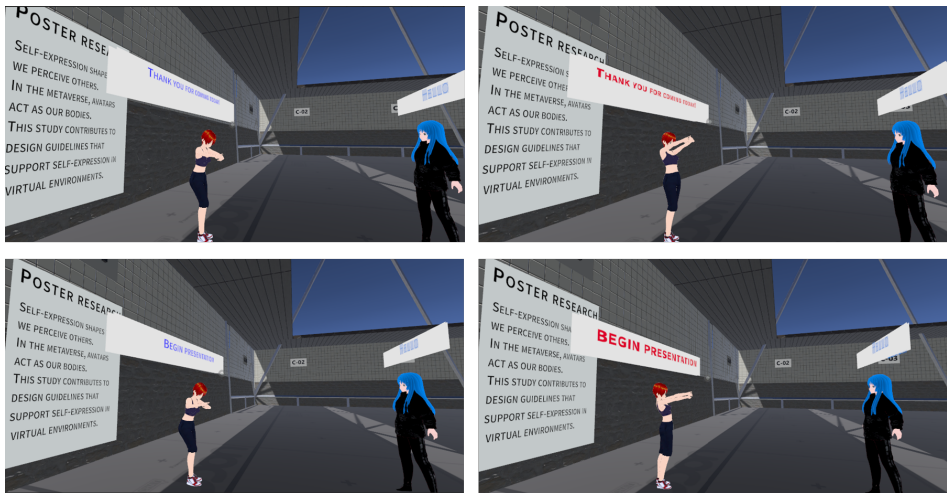


Figure 4. Text-only emphasis stimuli isolating the contribution of text design (tone × style) without avatar appearance.

4.7. Measures

We targeted three constructs: perceived congruence (1 item), self-expression fit (2 items), and interpersonal affinity (6 items). All item wordings (English translations; original Japanese in Supplementary Materials) are listed in Table 4. Composites were formed by averaging items on 5-point Likert scales; reliability was typically $\alpha \geq .85$. Cronbach’s α is reported for internal consistency and is distinct from hypothesis testing. Note that, while Table 4 describes the constructs in terms of an “avatar–text pair” for brevity, in the Presence (avatar-only) conditions participants applied the same questions to the target avatar alone, and in the Text-content (text-only) conditions they applied them to the text stimulus alone.

Table 4. Survey items for each construct (English translation).

Construct	Items
Perceived congruence	“The text matches the avatar.” (1 item)
Self-expression fit	(1) “This avatar-text pair feels like my self-representation.” (2) “This avatar-text pair supports my personality expression.”
Interpersonal affinity	(1) “I would be willing to talk with this person.” (2) “This person feels friendly.” (3) “This person feels trustworthy.” (4) “This person gives a positive impression.” (5) “This person feels warm.” (6) “I am satisfied with the interaction.”

4.8. Procedure

Each scene was displayed for 10–15 seconds. Trials were presented in randomized order, with short breaks between sessions to reduce fatigue. Two practice trials using non-experimental screenshots preceded the main block. Participants completed all assigned conditions within a single session.

4.9. Evaluation Framework and Statistical Plan

Primary analyses used repeated-measures ANOVA (Mauchly’s test; Greenhouse–Geisser corrections) with planned contrasts and effect sizes (partial η^2 , d_z). Complementary linear mixed-effects models included random intercepts for participants (and random slopes for Harmony where identifiable). Mediation (harmony → self-expression fit → affinity) was examined with bootstrap confidence intervals; indirect effects were considered significant if 95% CIs excluded zero. Exclusion rules, diagnostics, and sensitivity checks (e.g., removal of low-attention trials, subject-wise z-normalization, and one-item-drop analyses) are reported to avoid duplication elsewhere. Cronbach’s α indexes internal

consistency and is distinct from hypothesis testing. The analysis plan and reporting follow internal statistical guidance to ensure appropriate corrections and effect-size interpretation.

5. Results

5.1. Descriptive Statistics

Across participants, composite reliability was high for affinity (Cronbach’s $\alpha = .88$) and acceptable for self-expression fit ($\alpha = .86$). Means indicated a clear advantage for congruent conditions. For self-expression fit, the average rating was higher for congruent pairs ($M = 3.54$, $SD = 0.68$) than for mismatched pairs ($M = 2.57$, $SD = 0.75$). Affinity ratings were also higher for congruent pairs ($M = 3.67$) than mismatched pairs ($M = 3.37$). Distributions were unimodal and showed no severe skew.

As visualized in Figure 5, both self-expression fit and interpersonal affinity were higher in congruent than mismatched conditions, with non-overlapping 95% confidence intervals, consistent with the ANOVA results. As summarized in Table 5, descriptive statistics confirmed that congruent conditions yielded higher mean ratings than mismatched ones.

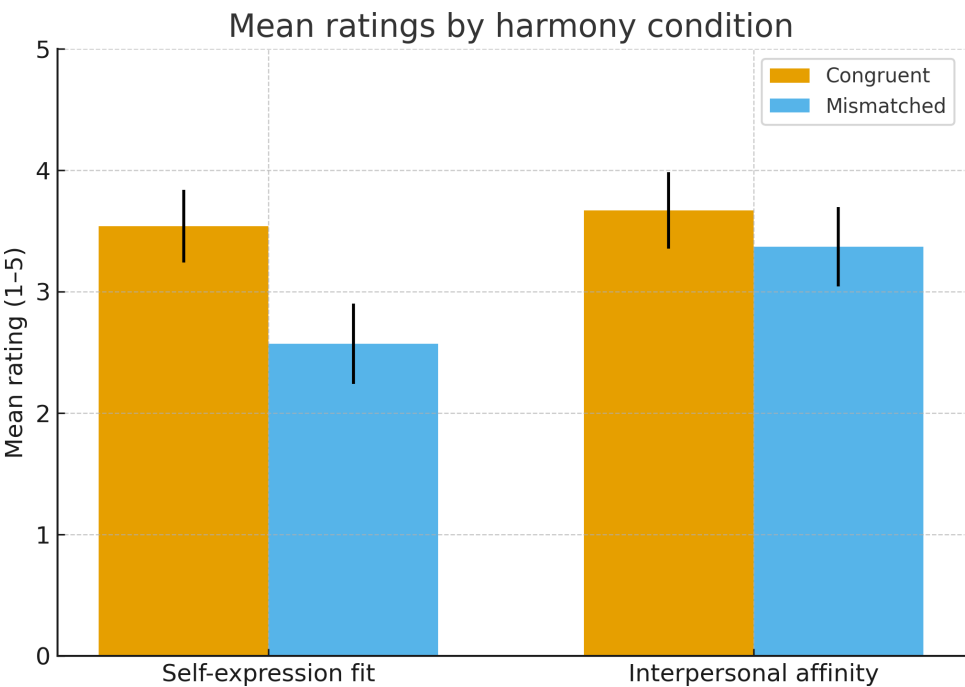


Figure 5. Mean ratings of self-expression fit and interpersonal affinity by harmony condition. Error bars represent 95% confidence intervals. Congruent conditions yielded significantly higher scores than mismatched conditions, supporting H1.

Table 5. Descriptive statistics for key outcomes by Harmony condition (collapsed across sessions). Means, standard deviations, and 95% CIs are reported.

Outcome	Condition	Mean	SD	95% CI
Self-expression fit	Congruent	3.54	0.68	[3.24, 3.84]
Self-expression fit	Mismatched	2.57	0.75	[2.24, 2.90]
Interpersonal affinity	Congruent	3.67	0.70	[3.36, 3.99]
Interpersonal affinity	Mismatched	3.37	0.73	[3.05, 3.70]

Notes. Means and SDs are on 5-point scales. Self-expression fit is the average of two identity-resonance items. Affinity averages six interpersonal items. 95% CIs are bias-corrected via participant-level resampling.

5.2. Manipulation Checks

Color and tone manipulations were recognized as intended. Ratings of congruence were significantly higher in matched conditions compared to mismatched ones, confirming that the manipulation was effective. For tone, cheerful wording was judged significantly more positive than neutral wording, ($F(1,20) = 162.25, p < .001, \eta^2 = .74$).

5.3. ANOVA Results

Repeated-measures ANOVA revealed a main effect of Harmony on self-expression fit, ($F(1,20) = 50.31, p < .001, \eta^2 = .72$), indicating that congruent conditions supported stronger self-expression. A smaller but significant effect of Session suggested that immersion enhanced self-fit. The interaction between Harmony and Session trended in the predicted direction, with harmony effects largest in the metaverse condition.

For affinity, the main effect of Harmony was also significant, ($F(1,20) = 15.84, p < .001, \eta^2 = .44$). Session effects were weaker but again consistent with stronger effects under immersive conditions.

Table 6. Repeated-measures ANOVA summaries. Greenhouse–Geisser corrections applied where required.

Effect (DV)	df	F	p	partial η^2
Harmony (Self-fit)	1, 20	50.31	< .001	.72
Session (Self-fit)	2, 40	4.10	.025	.17
Harmony $\times$ Session (Self-fit)	2, 40	2.45	.099	.11
Harmony (Affinity)	1, 20	15.84	< .001	.44
Session (Affinity)	2, 40	2.18	.126	.10
Harmony $\times$ Session (Affinity)	2, 40	1.95	.156	.09

Notes. All means use five-point scales. Where sphericity was violated, Greenhouse–Geisser corrections were applied.

5.4. Linear Mixed-Effects Models

Mixed-effects models with random intercepts for participants confirmed the main effects of Harmony on both self-expression fit and affinity. Inclusion of random slopes for Harmony did not qualitatively change conclusions.

5.5. Mediation Analysis

Regression-based mediation analyses tested whether self-expression fit explained the link between Harmony and Affinity. As summarized in Figure 6, Harmony significantly predicted affinity in the total model ($\beta = 0.97, p < .001$). When self-expression fit was added, the direct effect of Harmony became nonsignificant ($\beta = -0.10, p = .37$), while the path from self-expression fit to affinity remained significant ($\beta = 0.41, p < .001$). Bootstrapped confidence intervals for the indirect effect excluded zero, indicating that self-expression fit fully carried the effect of Harmony on affinity. This mediation pattern confirms H2.

Table 7. Mediation summary with standardized coefficients. The indirect effect ($a \times b$) was significant by bootstrap CI.

Path	Estimate	p
Total effect (Harmony $\rightarrow$ Affinity), c	0.97	< .001
Harmony $\rightarrow$ Self-fit, a	1.02	< .001
Self-fit $\rightarrow$ Affinity (controlling Harmony), b	0.41	< .001
Direct effect (Harmony $\rightarrow$ Affinity), c'	-0.10	.37
Indirect effect, $a \times b$	0.42	95% CI excludes 0

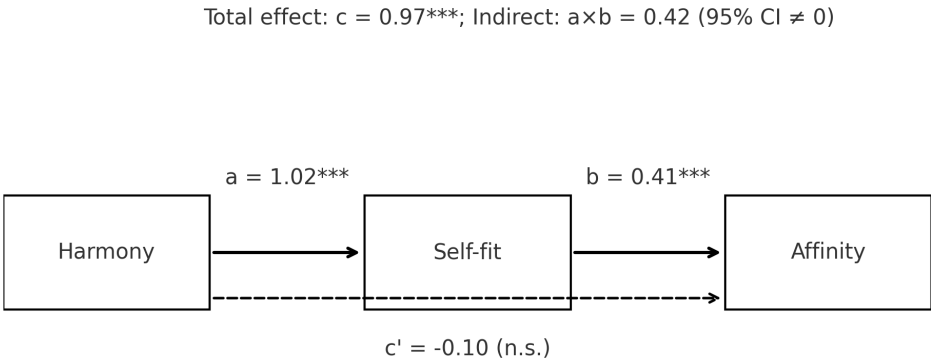


Figure 6. Mediation model depicting the relationship between harmony, self-expression fit, and interpersonal affinity. Solid arrows indicate significant standardized paths ($a = 1.02$; $b = 0.41$; $p < .001$), and the dashed arrow indicates a non-significant direct path ($c' = -0.10$). The total effect was $c = 0.97$ and the indirect effect was significant (bootstrap 95% CI excluded zero), confirming H2.

5.6. Exploratory Analyses

Exploratory comparisons showed that warm–warm avatar–text pairs produced the highest affinity scores (around 4.2 on a **5-point** scale), while reserved–reserved pairs yielded the lowest (around 1.9). Cross-pair conditions fell in between. Additional analyses of attention, memory, and preference subscales revealed higher attention ratings than memory or preference, ($F(2, 40) = 28.95$, $p < .001$, $\eta^2 = .59$).

5.7. Robustness and Additional Checks

Results were stable after excluding low-attention trials and after subject-wise z-normalization. Alternative composite definitions that dropped one affinity item at a time did not change the pattern of significance. Mixed-effects models with random slopes for Harmony, where identifiable, yielded qualitatively similar estimates to the ANOVA effects.

6. Discussion

The results confirm H1: Harmony between avatar and text enhances both self-expression fit and affinity. These effects align with the Proteus effect and with the hyperpersonal model, which highlights the amplification of impressions through coherent multimodal cues. Mediation analyses supported H2 by showing that self-expression fit explained the effect of harmony on affinity. The observed trend that harmony effects were stronger in immersive contexts partially supports H3.

From a theoretical perspective, these findings extend personality impression research by demonstrating that text cues embedded in avatars function similarly to visual appearance cues [6,21]. Harmony may work by increasing perceptual fluency, which enhances judgments of authenticity. Moreover, the results suggest that avatar–text harmony can be seen as a special case of Gestalt principles, where the whole impression exceeds the sum of parts [22]. Presence-based models in virtual environments [18,19] also predict stronger consequences under higher immersion.

From a practical design standpoint, platforms could implement adaptive systems that harmonize text style with avatar appearance (e.g., palette-aware defaults and accessibility-aware contrast), with user controls to lock/relax harmony for emphasis.

7. Limitations and Future Work

The sample size ($N = 21$) was modest and limited to university participants. While sufficient for within-subject medium effects, larger and more diverse samples are needed.

Manipulations focused on color, font, and tone; expanding to voice prosody, gesture, and movement would broaden the test of harmony. Longitudinal studies could assess persistence and adaptation.

In addition, our mediation analysis modeled self-expression fit as the sole mediator between harmony and affinity. A richer structural model in which perceived congruence is inserted as an additional mediator (harmony → perceived congruence → self-expression fit → affinity) is theoretically plausible but would require larger samples and more targeted measurement; exploring this extended pathway is an important direction for future work.

Contribution Recap

This article (i) formalizes a theory-driven account linking avatar–text harmony to interpersonal outcomes via self-expression fit, (ii) isolates harmony with jointly controlled color, font, and tone, (iii) demonstrates robust effects across immersion levels using ANOVA, mixed-effects models, and mediation with bootstrap inference, and (iv) translates findings into design guidance for metaverse interfaces.

8. Conclusions

Harmony between avatar appearance and on-avatar text significantly enhances both self-expression fit and interpersonal affinity in social VR. Congruence supports users in feeling authentic and fosters positive impressions among others.

Supplementary Materials: The following supporting information can be downloaded: Interview_Results_Summary_EN.xlsx, Post_Experiment_Congruence_Evaluation_EN.csv, and Avatar_Attention_Questionnaire_EN.csv. All original Japanese items and responses are available from the corresponding author upon reasonable request.

Author Contributions: Conceptualization, G.Y.; methodology, G.Y.; formal analysis, G.Y.; investigation, G.Y.; writing—original draft preparation, G.Y.; writing—review and editing, G.Y., T.N., K.H., S.S.; visualization, G.Y.; supervision, T.N., K.H., S.S.; project administration, G.Y.; funding acquisition, G.Y.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Electro-Communications (H25019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Anonymized data and analysis scripts are available from the authors upon request.

Acknowledgments: We thank colleagues and student participants at the University of Electro-Communications for their support with pilot testing and study logistics.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yee, N.; Bailenson, J.N. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research* **2007**, *33*, 271–290. <https://doi.org/10.1111/j.1468-2958.2007.00299.x>.
2. Walther, J.B. Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal. *Communication Research* **1996**, *23*, 3–43. <https://doi.org/10.1177/009365096023001001>.
3. Lim, C.; Ratan, R.A.; Foxman, M.; Meshi, D.; Liu, H.; Hales, G.E.; Lei, Y.S. An Avatar's Worth in the Metaverse Workplace: Assessing Predictors of Avatar Customization Valuation. *Computers in Human Behavior* **2024**, *158*, 108309. <https://doi.org/10.1016/j.chb.2024.108309>.
4. Zhang, J.; Juvrud, J. Gender Expression and Gender Identity in Virtual Reality: Avatars, Role-Adoption, and Social Interaction in VRChat. *Frontiers in Virtual Reality* **2024**, *5*, 1305758. <https://doi.org/10.3389/frvir.2024.1305758>.
5. Asch, S.E. Forming impressions of personality. *The Journal of Abnormal and Social Psychology* **1946**, *41*, 258–290. <https://doi.org/10.1037/h0055756>.
6. Cui, Y.; Wan, Z.; Xia, Q.; Feng, Y.; Gu, W.; Yang, L.; Zhou, Z. Fairy Tale Situation Test for Implicit Theories of Personality. *American Journal of Applied Psychology* **2021**, *9*, 172–181. <https://doi.org/10.11648/j.ajap.20200906.15>.

7. Elliot, A.; Maier, M.; Moller, A.; Friedman, R.; Meinhardt, J. Color and Psychological Functioning: The Effect of Red on Performance Attainment. *Journal of experimental psychology. General* **2007**, *136*, 154–68. <https://doi.org/10.1037/0096-3445.136.1.154>.
8. Nowak, K.L.; Rauh, C. The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. *Journal of Computer-Mediated Communication* **2005**, *11*, 153–178. <https://doi.org/10.1111/j.1083-6101.2006.tb00308.x>.
9. Kang, H.; Kim, H.K. My avatar and the affirmed self: Psychological and persuasive implications of avatar customization. *Computers in Human Behavior* **2020**, *112*, 106446. <https://doi.org/10.1016/j.chb.2020.106446>.
10. Aseeri, S.; Interrante, V. The influence of avatar representation on interpersonal communication in virtual social environments. *IEEE Transactions on Visualization and Computer Graphics* **2021**, *27*, 2608–2617.
11. Hobbs, M.F.D.J.M.D.J. Designing avatars for social interactions. *Animating Expressive Characters for Social Interaction* **2008**, pp. 195–211.
12. Koda, T.; Ishida, T.; Rehm, M.; et al. Avatar culture: cross-cultural evaluations of avatar facial expressions. *AI & Society* **2009**, *24*, 237–250.
13. Hall, J.A.; Horgan, T.G.; Murphy, N.A. Nonverbal Communication. *Annual Review of Psychology* **2019**, *70*, 271–294. <https://doi.org/10.1146/annurev-psych-010418-103145>.
14. Hepperle, D.; Purps, C.; Deuchler, J.; Bruder, G. Aspects of visual avatar appearance: self-representation, display type, and uncanny valley. *The Visual Computer* **2022**, *38*, 1227–1244. <https://doi.org/10.1007/s00371-021-02168-0>.
15. Ekdahl, D.; Osler, L. Expressive Avatars: Vitality in Virtual Worlds. *Philosophy & Technology* **2023**, *36*, 1–20. <https://doi.org/10.1007/s13347-022-00590-1>.
16. T., H.J.; Dunham, P.J. Impression formation in computer-mediated communication revisited: An analysis of the breadth and intensity of impressions. *Communication Research* **2001**, *28*, 325–347.
17. Shaikh, A.D.; Chaparro, B.S.; Fox, D. Perception of Fonts: Perceived Personality Traits and Uses. In Proceedings of the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2006, Vol. 50, pp. 1277–1281. <https://doi.org/10.1177/154193120605001807>.
18. Slater, M. Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments. *Philosophical Transactions of the Royal Society B* **2009**, *364*, 3549–3557. <https://doi.org/10.1098/rstb.2009.0138>.
19. Biocca, F. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication* **1997**, *3*, JCMC324. <https://doi.org/10.1111/j.1083-6101.1997.tb00070.x>.
20. Meta. Meta Platforms Inc: Oculus Release Notes. <https://developer.oculus.com/>, 2021.
21. Tshivhase, M. On the possibility of authentic self-expression. *Communicatio* **2015**, *41*, 374–387. <https://doi.org/10.1080/02500167.2015.1093329>.
22. Wertheimer, M. Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung* **1923**, *4*, 301–350. English translations available in anthologies.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.