Check for updates

How Hangul Subtitle Typography Shapes YouTube Viewing Experiences: Effects Across Video Genre and Format

Beomseok Lee[1], Hyosun Kwon[2*]

[1]Department of Convergence Design, Graduate School of Techno Design, PhD Student, Kookmin University, Seoul, Korea

[2]Department of Industrial Design, Associate Professor, Kookmin University, Seoul, Korea

Abstract

Background Subtitles are widely used for YouTube videos and can shape viewer engagement. While typeface effects on perception are well documented in static media, subtitle typography remains underexplored in dynamic video contexts, especially for non-Latin scripts. Hangul's syllable-block architecture differs from the linear organization of Latin scripts, which limits the direct transferability of findings derived from Latin typography to Hangul. Empirical evidence remains limited regarding how evaluations of Hangul subtitle typefaces vary by video format and genre.

Methods A mixed-factorial experiment with 71 participants (short-form n = 36; long-form n = 35) examined four Hangul typeface categories (sans-serif, serif, display, script) across three genres (informational, entertainment, daily life) of both short- and long-form YouTube videos. Participants rated each condition on a 10-item questionnaire measuring readability, visual cohesion, attractiveness, and functional suitability using seven-point Likert scales. Two-way repeated-measures analysis of variance (ANOVA) with Bonferroni corrections was used to analyze genre × typeface interaction effects.

Results Larger genre × typeface interaction effect sizes were observed under long-form conditions for readability (η^2_p = 0.167), visual cohesion (η^2_p = 0.296), and attractiveness (η^2_p = 0.235). Sans-serif and serif typefaces scored high on readability but significantly lower on visual attractiveness, indicating a functional–aesthetic dissociation. Script typefaces demonstrated the highest context sensitivity, performing well in the daily life genre but receiving lower cohesion and attractiveness ratings in long-form entertainment.

Conclusions Subtitle typefaces in Korean YouTube content functioned as a context-dependent design variable whose evaluation varied by genre and format, making them more than a stylistic choice. These findings support context-dependent subtitle evaluation, while differences in participant groups and procedures limit duration-only interpretation. Dynamic, context-aware subtitle design is advanced through the proposal of an adaptive strategy using sans-serif and serif typefaces as stable baselines with script typefaces as contextual accents for tonal transitions.

Keywords Subtitle Typography, Hangul Typefaces, Korean Subtitles, Video, YouTube

*Corresponding author: Hyosun Kwon (hyosun.kwon@gmail.com)

Citation: Lee, B., & Kwon, H. (2026). How Hangul Subtitle Typography Shapes YouTube Viewing Experiences: Effects Across Video Genre and Format. *Archives of Design Research, 39*(4), 197-213.

http://dx.doi.org/10.15187/adr.2026.11.39.4.197

Received : May. 21. 2026 ; **Reviewed :** Sep. 14. 2026 ; **Accepted :** Sep. 14. 2026
pISSN 1226-8046 **eISSN** 2288-2987

Copyright : This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

1. Introduction

Online video platforms like YouTube provide multimedia environments where moving imagery, speech, sound, and on-screen text combine to create complex multimodal viewing experiences. Within this context, subtitles have evolved beyond their conventional function as translation or accessibility features. Subtitles increasingly operate as creative, postproduction elements that shape viewer engagement and comprehension (Sasamoto et al., 2017, 2021; Sasamoto, 2014).

The impact of typeface on readability and semantic perception has been extensively documented for static media (Fox et al., 2007; Brumberger, 2003). However, research on subtitle typography in video remains limited. Existing studies have primarily emphasized the legibility of standardized, translation-oriented subtitles, which are typically placed at the bottom-center of the screen (Bisson et al., 2014; Liao et al., 2021). This approach tends to treat typeface as a secondary variable rather than an active design parameter. However, this assumption is increasingly misaligned with contemporary online video editing practices that integrate subtitles as primary editorial elements (Sasamoto et al., 2017; Sasamoto et al., 2021).

Furthermore, short- and long-form content present different typographic considerations. For example, short-form videos may employ temporally dynamic typography as an expressive element (Lee et al., 2006), which is often coordinated with pacing and sound effects (Sasamoto et al., 2021). Meanwhile, long-form content requires sustained legibility over extended viewing periods (Bisson et al., 2014; Liao et al., 2021). Despite the ubiquity of these practices in online video editing, there is little empirical evidence regarding how subtitle typography varies across formats and genres.

Typeface effects in videos remain especially underexplored in non-Latin-script environments. For example, Hangul syllables are organized as compact, visually unified blocks where consonants and vowels integrate into a single perceptual unit (Pae, 2011; Yoo et al., 2025). Unlike Latin letters' linear sequencing, Hangul's modular structure means that a syllable's visual form varies depending on the selection and spatial arrangement of its component graphemes (Pae, 2011; Yoo et al., 2025). Meanwhile, Korean screen media has developed a distinct subtitle culture in which on-screen Hangul text functions as an integral narrative and visual element (Lee, 2019; Park, 2009). These factors motivate an empirical investigation of Hangul typeface selection in audiovisual meaning-making, especially as this context has received substantially less attention compared to Latin-script typography.

The present study therefore focused on Hangul subtitle use in Korean-language YouTube videos. It examined how viewer evaluations of four Hangul typeface categories (sans-serif, serif, display, and script) varied by format and genre. These included perceptions of readability, visual cohesion, and attractiveness for the four typeface categories across short- and long-form videos in three genres: informational, entertainment, and daily life (or "vlogging"). This paper thus both provides empirically grounded typographic heuristics for video creators and contributes to a more nuanced understanding of how typeface selection shapes viewer evaluations of Hangul-based multimodal media.

2. Background

2. 1. Subtitle Typography in Video Content

In video contexts, subtitles are inherently temporal phenomena that appear for limited durations while competing for viewer attention against moving imagery, speech, music, and audio effects (Bisson et al., 2014; Liao et al., 2021). Subtitle processing thus constitutes a complex, multimodal activity that requires the coordination of visual scanning and semantic comprehension during a continuous audiovisual stream, which makes it fundamentally different from conventional linear reading (Lång et al., 2021; Liao et al., 2021). Previous studies approached subtitles as presentation constraints to be optimized, focusing on temporal segmentation (Lee et al., 2025), spatial placement and occlusion management (Al Amin et al., 2022; Kurzhals et al., 2020), and gaze distribution dynamics (Kurzhals et al., 2017; Kurzhals et al., 2020). Scholars have also compared professional and automatic subtitles to assess caption quality (Kim et al., 2023). However, reliable quality metrics that reflect real-world viewing conditions remain elusive (Kim et al., 2023; Arroyo Chavez et al., 2024).

Typography and human-computer interaction research has demonstrated that typeface significantly influences affective perception, including tone, semantic appropriateness, and perceived personality (Shaikh et al., 2006; Evans et al., 2004; Brumberger, 2003). Individual variance in typographic perception furthermore suggests that adaptive or context-sensitive design interventions may further enhance user experience (Wallace et al., 2022; Kadner et al., 2021). As noted above, much of this work has focused on Latin-script media. To establish the script-specific context for the present paper, it is necessary to first examine the orthographic architecture of Hangul and its implications for typographic design.

2. 2. Hangul's Unique Structure and Its Typographic Implications

Hangul, created by King Sejong (r. 1418–1450) in 1443 and promulgated in 1446 in *Hunminjeongeum (The Correct Sounds for the Instruction of the People)*, is a systematically designed writing system created explicitly to enhance literacy in Korean citizens (Pae, 2011). Hangul's most distinctive architectural principle is its use of syllable blocks as the primary orthographic unit (Pae, 2011; Yoo et al., 2025). Unlike Latin scripts, where individual letters are sequenced horizontally along a single baseline, Hangul organizes consonant and vowel graphemes within compact syllabic blocks in which they occupy both horizontal and vertical positions (Yoo et al., 2025; Taylor, 1980). Figure 1 illustrates this architecture using the syllable sam (삼: [sam]) in sans-serif, serif, display, and script typefaces, demonstrating how the individual consonant (ㅅ: [s], ㅁ: [m]) and vowel (ㅏ: [a]) graphemes merge to form a single perceptual unit.

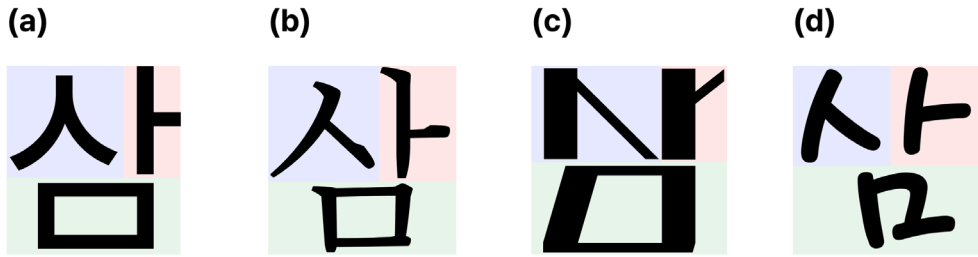

Figure 1 Hangul Syllable Block Structure by Typeface Category. The syllable shows the integration of consonant (blue and green) and vowel (red) graphemes within a single block, displayed in (a) sans-serif, (b) serif, (c) display, and (d) script typeface examples.

This syllable-block structure has unique implications for typeface design. In Latin typography, individual letterforms maintain consistent proportions and positioning relative to the fixed baseline and cap height. Hangul meanwhile requires designers to manage variable internal spatial arrangements. Hangul syllable blocks feature different internal configurations depending on how their component graphemes are spatially combined. In Korean typography, these are commonly classified as horizontal-combination forms (e.g., 가, ga), vertical-combination forms (e.g., 고, go), and mixed-combination forms (e.g., 과, gwa; Lee, 2012). Variations in stroke thickness, secondary stroke length, or letter separation influence the entire block's visual density and the typeface's perceived characteristics. When combined with temporal constraints and competing audiovisual stimuli (Bisson et al., 2014; Liao et al., 2021), these intra-block variations may impact how viewers perceive subtitle typefaces. However, the present study treated this as theoretical motivation rather than a directly tested mechanism.

2. 3. Hangul Typographic Context for On-Screen Subtitles

Korean screen media has developed distinctive subtitle conventions shaped both by Hangul's structural properties and cultural viewing practices. These practices reflect both technical adaptation and cultural convention in which on-screen Hangul text can function as an editorial design element. This editorial integration of subtitles has direct implications for typeface semantics. In short-form content, typographic variation can serve as a pacing and emphasis device, synchronized with sound design and visual cuts. In long-form content, however, evaluations of the same typeface may vary with genre and narrative tone under sustained viewing conditions. To our knowledge, no prior work has systematically examined how evaluations of subtitle typefaces vary jointly across genre and video format in Korean video contexts, nor how design parameters established in legibility research relate to viewer ratings of readability, visual cohesion, and attractiveness across these contexts.

2. 4. Research Gap

Previous studies have clarified the functional constraints required for subtitle legibility. However, a critical gap remains regarding how viewers evaluate typeface as an active editorial variable in Korean YouTube content, particularly as video format and genre context vary. By empirically examining evaluations of Hangul typeface categories across video format and genre, this study developed evidence-based guidance for content creators while extending typeface research to a Hangul-based multimodal viewing context.

3. Method

3. 1. Participants

The final analysis included data from 71 Korean participants. The short-form condition included 36 participants (18 males, 18 females; age range = 22–39 years, M = 25.9, SD = 3.0), and the long-form condition included 35 participants (20 males, 15 females; age range = 24–33 years, M = 25.9, SD = 2.2). Participants were recruited through social media posts, personal networks, and referrals. The study was approved by the institutional review board (KMU-202511-HR-525), and informed consent was obtained prior to the study.

3. 2. Stimuli Design

This study employed publicly available YouTube videos as stimuli. The original subtitles were masked and replaced with the newly designed subtitles on opaque rectangular backgrounds. For each genre, the researchers created four subtitle conditions using sans-serif, serif, display, and script typefaces (Figure 2a). The placement and coloring were matched to each video's editing style (Figures 2b, c). Six videos were used, with two per genre (i.e., informational, entertainment, daily life). Three were short-form (under 20 s) and three were long-form (approximately 4 min). Stimulus screenshots and source URLs are provided in Appendix A.4.

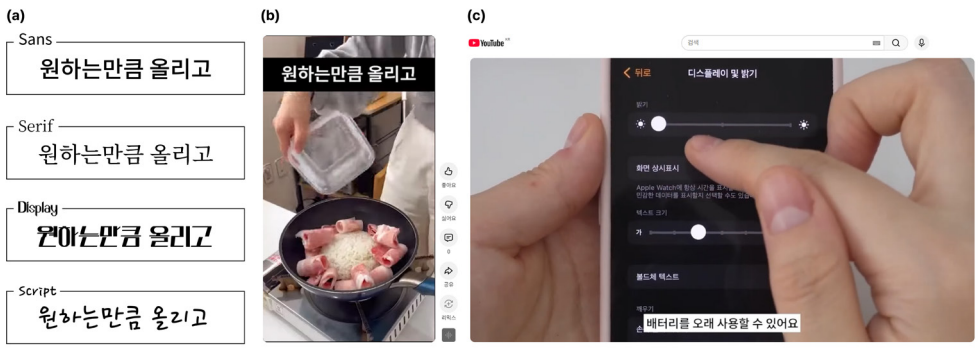

Figure 2 Stimuli and Typefaces. (a) four typeface categories used in the study (sans-serif, serif, display, and script); (b) short-form, daily life video example; (c) long-form informational video example.

The four subtitle conditions each used a single representative typeface. The sans-serif typeface was Noto Sans (400), the serif typeface was Noto Serif (500), the display typeface was 116 Watermelon (400), and the script typeface was Hakgyoansim Mindeulleholssi (400). These respectively correspond to the categories of *minburi, buri,* display, and handwriting that are used in current Korean type practice. Noto Sans and Noto Serif were selected as neutral, text-oriented representatives, the handwriting face for its regular, upright forms, and the display face as a distinctive, commonly used Korean display typeface. All were freely available. Apparent size was optically adjusted. A weight of 500 was used for Noto Serif because weight 400 appeared lighter compared to the other typefaces.

This four-way scheme was adopted to reflect the range of style choices available to creators when creating subtitles, from neutral to expressive, and because it corresponds to how typefaces are grouped in mainstream font services and editing tools. It does not represent a formal taxonomy of Hangul type, nor does it mean that Latin serif and sans-serif share a lineage with Korean *buri* and *minburi*, which describe Hangul-specific terminal conventions rather than a shared historical lineage with Latin categories.

The short-form stimuli featured a pharmacist providing food-related advice (informational), a humorous clip from a Korean online creator (entertainment), and a cooking creator briefly introducing a personal recipe (daily life). The long-form stimuli covered Apple Watch battery management tips (informational), a segment from a Korean comedy program (entertainment), and a honeymoon travel vlog by a Korean travel creator (daily life). The videos were selected to clearly distinguish between genres and because they either featured topics likely to be familiar to or engaging for participants or content from widely recognized programs and creators. All videos used stable, nonmoving subtitle layouts to minimize the potential variation introduced by animated or dispersed subtitle placement.

3. 3. Procedure

In the mixed factorial design, video format (short vs. long) was the between-subjects factor, and genre (three) and typeface category (four) were the within-subjects factors. Each participant therefore evaluated 12 typeface–genre combinations regardless of their assigned condition (see Appendix). The stimuli presentation was also adapted to accommodate the structural differences between formats. In the short-form condition (Figure 2b), the participants viewed the same clip four times per genre, with the subtitle typeface varying in each iteration. In the long-form condition (Figure 2c), each video was segmented into four distinct sequential parts (one per typeface) of approximately 60 s each to limit total viewing time and reduce repetition effects. Presenting a complete four-minute video four times would have required approximately 16 minutes of viewing per genre (48 minutes total), and repeating the same one-minute segment four times would have introduced substantial repetition. Using sequential segments allowed the four typeface conditions to share the same topic, narrative context, speakers, and editing style without repeating identical content.

Typeface-to-segment assignments were counterbalanced across participants using a Latin square design, so no typeface was fixed to a particular segment position or content section. After viewing each typeface condition, the participants completed a 10-item questionnaire using seven-point Likert scales (1 = strongly disagree, 7 = strongly agree). The short-form condition involved approximately three minutes of total viewing time and was conducted anonymously online. The long-form condition involved approximately 12 minutes of total viewing time and was conducted primarily in person or via supervised video call to support sustained attention and completion of the procedure.

3. 4. Questionnaire Design

The questionnaire items were informed by three complementary theoretical traditions. The readability construct (Q1–Q3) operationalized the functional quality dimension of typeface perception (Bartram, 1982; Shaikh et al., 2006) and was adapted for temporal viewing conditions (Bisson et al., 2014; Liao et al., 2021). The visual cohesion construct (Q4–Q6) drew on the concept of typeface appropriateness, or the perceived congruence

between typeface and its communicative context (Brumberger, 2003), and extended it to audiovisual congruence by incorporating questions of multimodal fit with auditory elements (Sasamoto et al., 2017, 2021). The attractiveness construct (Q7–Q9) integrated semantic quality assessment (Evans et al., 2004) with the hedonic–pragmatic quality distinction established in user experience research (Hassenzahl, 2004; O'Donovan et al., 2014). Finally, Q10 served as a global functional suitability measure reflecting holistic evaluative judgment (Brumberger, 2003; Hassenzahl, 2004). The questionnaire was administered in Korean. The full questionnaire is provided in Appendix A.1.

4. Findings

Composite scores were computed to assess user perception. These demonstrated high internal consistency across both the short- and long-form conditions in readability (Q1–Q3; short $\alpha = 0.90$, long $\alpha = 0.90$), visual cohesion (Q4–Q6; short $\alpha = 0.90$, long $\alpha = 0.94$), and attractiveness (Q7–Q9; short $\alpha = 0.85$, long $\alpha = 0.88$). Item Q10 was analyzed separately to measure functional suitability.

4. 1. Overall User Perception and Interaction Effects

Overall, the sans-serif and serif typefaces obtained consistently high ratings, whereas the display typefaces were generally rated lower. The script typefaces exhibited the highest variance, with evaluations fluctuating significantly depending on the context. To examine whether the genre–typeface relationship differed within each format, separate two-way repeated-measures ANOVAs with Bonferroni corrections were performed. The magnitude of the interaction effect sizes (η^2_p) was prioritized to interpret the practical impact of each typeface across genres in the two different formats.

Genre × typeface interactions were observed in both formats but were notably stronger in the long-form condition (Table 1). Item-level analyses confirmed this divergence. The short-form condition showed significant interactions for visual cohesion (Q4, Q5), attractiveness (Q7), and functional suitability (Q10). The long-form condition produced substantially larger genre × typeface interaction effects across nearly all measures except Q2 (Figure 3; Appendix A.2). The interaction effect sizes (η^2_p) were greater in the long-form condition for nine of the ten items, with Q2 as the exception. This indicates stronger context dependence for most measures in the long-form experimental configuration.

Table 1 Composite-Level Genre × Typeface Interactions. *Significant at the Bonferroni-corrected threshold of $\alpha = 0.05/3 \approx 0.0167$ for three composite-level interaction tests within each format. G = genre; T = typeface.

Factors	Readability			Visual cohesion			Attractiveness		
	F	p	η^2_p	F	p	η^2_p	F	p	η^2_p
G×T (short)	3.239	〈0.0167*	0.085	3.898	〈0.0167*	0.100	2.226	0.042	0.060
G×T (long)	6.796	〈0.0167*	**0.167**	14.27	〈0.0167*	**0.296**	10.47	〈0.0167*	**0.235**

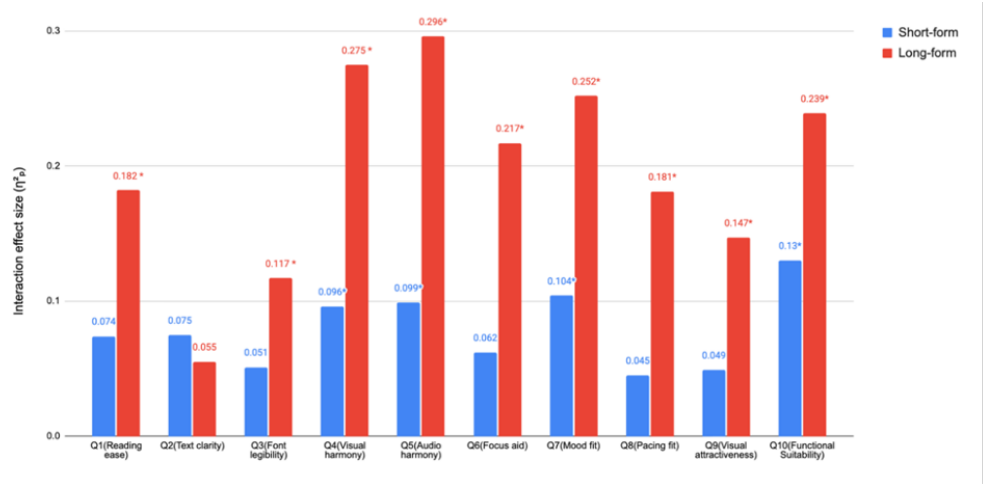

Figure 3 Interaction Effect Sizes for Genre × Typeface Interactions. *p 〈 0.005.

4. 2. Functional Dominance against Aesthetic Limitations

The sans-serif and serif typefaces tended to receive lower ratings for visual attractiveness (Q9) than the other metrics in all six format × genre conditions (Figure 4). To statistically verify this pattern, pairwise comparisons were performed between Q9 and all other items (Q1–8, Q10) within each condition while applying a Bonferroni correction threshold (α = 0.05/9). The readability items (Q1–3) were significantly higher than Q9 for all conditions. However, the differences between Q9 and other items were not consistently significant. Because all mean differences were positive, readability consistently exceeded visual attractiveness for sans-serif and serif typefaces across every condition.

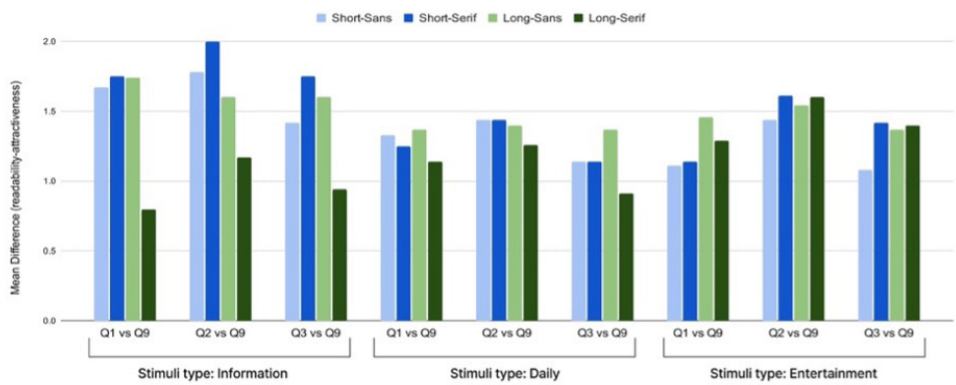

Figure 4 Mean Score Differences: Readability versus Attractiveness. Only comparisons significant at the Bonferroni-corrected level (α = 0.0056) are reported.

The long-form serif in the informational genre (dark green) showed notably compressed gaps (0.80–1.17). Short-form videos using the serif typeface (medium blue) demonstrated the largest gaps, particularly in Q2–Q9 in the same genre (2.00). This indicates that text clarity was rated much higher than visual appeal in that genre. The readability comparisons (Q1–Q3 versus Q9) are presented in Appendix A.3.

4. 3. Genre Sensitivity and Differences by Format

The script typeface exhibited significant genre sensitivity in both short- and long-form videos. It received its highest ratings in the daily life genre. In the short-form condition, ratings for this genre were higher than those for both the informational and entertainment genres in terms of readability and visual cohesion.

In the long-form condition, the daily life and informational genres were rated significantly higher than the entertainment genre in both visual cohesion and attractiveness (Table 2, top). Differences were also observed between conditions within genres (Table 2, bottom). In the entertainment genre, the script typefaces were rated significantly lower for cohesion and attractiveness in the long-form than the short-form condition. Conversely, they scored higher for cohesion in the long-form condition for the informational genre. Notably, the ratings remained robustly high across both conditions for the daily life genre, which suggests a consistent perceived fit within the tested stimuli.

Table 2 Bonferroni–Corrected Comparisons for Script Typeface. Top: pairwise genre comparisons within each format. Bottom: between–format comparisons within each genre. Only comparisons significant at the Bonferroni–corrected level (α = 0.0167) are reported. Daily = daily life; Ent = entertainment; Info = informational.

Format	Variable	Direction	Δ	t
Short	Readability	Daily 〉 Info	0.91	3.70
	Readability	Daily 〉 Ent	0.56	2.55
	Cohesion	Daily 〉 Info	1.31	4.41
	Cohesion	Daily 〉 Ent	0.81	3.30
	Attractiveness	Ent 〉 Info	0.57	2.58
	Attractiveness	Daily 〉 Info	1.08	4.43
Long	Readability	Daily 〉 Ent	0.61	3.90
	Cohesion	Info 〉 Ent	2.07	6.18
	Cohesion	Daily 〉 Ent	2.10	8.01
	Attractiveness	Info 〉 Ent	1.68	5.43
	Attractiveness	Daily 〉 Ent	1.70	6.93

Genre	Variable	Direction	Δ	t
Info	Cohesion	Long 〉 Short	1.09	3.10
Ent	Cohesion	Short 〉 Long	1.49	4.33
Ent	Attractiveness	Short 〉 Long	1.66	5.10

5. Design Implications

5. 1. Adaptive Typography and Typographic Hierarchy

The findings of this study indicate that the relationship between typefaces and genres was more pronounced under the long-form experimental condition. This suggests that subtitle typography does not need to be static. Rather, subtitle typography can be treated as a dynamic layer that adapts to narrative shifts involving mood, speaker, focus, or audiovisual style.

In the long-form stimuli, highly stylized typefaces received more context-dependent evaluations. Creators should therefore avoid consistently using a highly stylized typeface for

the complete duration of the video, especially if the content or tone fluctuates. In the short-form condition, typeface evaluations were generally less context-dependent, which may indicate room for greater stylistic flexibility. Typeface variation in long-form videos may be most useful at structural transitions to reflect changes in topic, tone, or scene. A practical approach would be establishing a consistent baseline typeface and deploying alternative styles only when distinct contextual shifts occur. This approach can be understood through the concept of typographic hierarchy.

The findings indicate a distinction between baseline and accent typefaces. For example, sans-serif and serif typefaces can serve as practical typographic baselines for the majority of a video's runtime based on their consistently high readability ratings. Because script typefaces exhibited high context-dependent variance, they may function best as contextual accents. Their visual character may align especially well with emotional moments in narrative or conversation. However, because lower cohesion and attractiveness ratings were observed for the script typeface in the long-form entertainment condition, a hybrid strategy is preferable. This approach aims to balance stylistic expression with readability.

6. Limitations and Future Work

This study had several limitations. First, given the hybrid nature of YouTube content (e.g., "edutainment"), rigid genre classification remains an approximation. Furthermore, the four typeface conditions follow categories used in Korean type practice rather than a formal Hangul typographic classification. For example, the square-frame (*nemotteul*) and nonsquare-frame (*tal-nemotteul*) distinction is specific to Hangul and closely related to the syllable-block structure discussed in Section 2.2, but it was not treated as a controlled variable in the present study. Future work should manipulate this axis as a separate factor.

The script and display typefaces used in this study represent only a small subset of the vast stylistic diversity available in these families, especially compared with the more standardized sans-serif and serif groups. Furthermore, only one video was used per genre–format combination, so video-level properties may have contributed to the observed effects. The short- and long-form conditions also used different presentation procedures, administration settings, and nonoverlapping participant groups. Differences between conditions therefore cannot be attributed solely to video duration. Although typeface-to-segment assignments in the long-form condition were counterbalanced using a Latin square, residual segment-level variation was possible. Finally, the sample was limited to adults in their twenties and thirties (22–39 years in the short-form condition; 24–33 years in the long-form condition), which limited generalization both to younger and older viewers.

Future research can address these limitations by employing granular genre taxonomies, expanded typeface datasets, multiple videos per genre–format cell, and more closely matched procedures across formats. Haraguchi and Otani (2023) developed a font recommendation system for banner design, but extending such systems to video content is challenging because of the simultaneous cognitive demands of processing motion, audio, and timing. Leveraging larger datasets spanning diverse content styles would help further model perceived readability and attractiveness as functions of typographic attributes (O'Donovan et al., 2014).

Building on this, the researchers plan to develop artificial intelligence-based curation tools that automatically suggest subtitle typefaces tailored to a video's tone and viewing conditions.

7. Conclusion

Korean YouTube videos offer a unique environment for studying subtitle typography, as they are characterized by the ubiquity of on-screen text and the distinct visual density of Hangul. The present investigation into Hangul subtitles showed that typeface selection is a functional variable whose evaluation varies across viewing contexts. The findings of this study indicate that while sans-serif and serif typefaces provided a reliable baseline for readability, stylized script typefaces were highly context-dependent and received more favorable aesthetic evaluations in specific contexts, particularly in the tested daily-life stimuli.

This study extended typeface-appropriateness research from static documents to dynamic Hangul subtitle environments by demonstrating that subtitle evaluations varied by genre and experimental format. Differences between format conditions should be interpreted considering the accompanying differences in participant groups and procedures, rather than being attributed solely to video duration. The findings also support a baseline–accent strategy in which relatively neutral sans-serif and serif typefaces serve as stable defaults and expressive script typefaces are applied selectively when they fit the scene and genre. These contributions advance context-aware subtitle design grounded in measured perceptions of readability, visual cohesion, attractiveness, and functional suitability.

References

1. Al Amin, A., Hassan, S., Lee, S., & Huenerfauth, M. (2022). Watch it, don't imagine it: Creating a better caption-occlusion metric by collecting more ecologically valid judgments from DHH viewers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)* (pp. 1–14). ACM. https://doi.org/10.1145/3491102.3517681
2. Arroyo Chavez, M., Feanny, M., Seita, M., Thompson, B., Delk, K., Officer, S., Glasser, A., Kushalnagar, R., & Vogler, C. (2024). How users experience closed captions on live television: Quality metrics remain a challenge. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)* (pp. 1–16). ACM. https://doi.org/10.1145/3613904.3641988
3. Bartram, D. (1982). The perception of semantic quality in type: Differences between designers and non-designers. *Information Design Journal, 3*(1), 38–50. https://doi.org/10.1075/idj.3.1.04bar
4. Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 399–418. https://doi.org/10.1017/S0142716412000434
5. Brumberger, E. R. (2003). The rhetoric of typography: The awareness and impact of typeface appropriateness. *Technical Communication, 50*(2), 224–231.
6. Evans, M. B., McBride, A. A., Queen, M., Thayer, A., & Spyridakis, J. H. (2004). The effect of style and typography on perceptions of document tone. In *Proceedings of the 2004 IEEE International Professional Communication Conference (IPCC '04)* (pp. 300–303). IEEE. https://doi.org/10.1109/IPCC.2004.1375314
7. Fox, D., Shaikh, A. D., & Chaparro, B. S. (2007). The effect of typeface on the perception of onscreen documents. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 51*(5), 464–468. https://doi.org/10.1177/154193120705100508

8. Haraguchi, D., & Otani, M. (2023). Coarse-to-fine font recommendation for banner designs. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23 Companion)* (pp. 148–150). ACM. https://doi.org/10.1145/3581754.3584156
9. Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human–Computer Interaction, 19*(4), 319–349. https://doi.org/10.1207/s15327051hci1904_2
10. Kadner, F., Keller, Y., & Rothkopf, C. (2021). AdaptiFont: Increasing individuals' reading speed with a generative font model and Bayesian optimization. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* (pp. 1–11). ACM. https://doi.org/10.1145/3411764.3445140
11. Kim, H., Tao, Y., Liu, C., Zhang, Y., & Li, Y. (2023). Comparing the impact of professional and automatic closed captions on video-watching experience. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)* (pp. 1–6). ACM. https://doi.org/10.1145/3544549.3585634
12. Kurzhals, K., Cetinkaya, E., Hu, Y., Wang, W., & Weiskopf, D. (2017). Close to the action: Eye-tracking evaluation of speaker-following subtitles. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)* (pp. 6559–6568). ACM. https://doi.org/10.1145/3025453.3025772
13. Kurzhals, K., Göbel, F., Angerbauer, K., Sedlmair, M., & Raubal, M. (2020). A view on the viewer: Gaze-adaptive captions for videos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)* (pp. 1–12). ACM. https://doi.org/10.1145/3313831.3376266
14. Lång, J., Vrzakova, H., & Mehtätalo, L. (2021). Modelling gaze behaviour in subtitle processing: The effect of structural and lexical properties. *Journal of Audiovisual Translation, 4*(1), 71–95. https://doi.org/10.47476/jat.v4i1.2021.104
15. Lee, D., Choi, J., & Noh, J. (2025). OptiSub: Optimizing video subtitle presentation for varied display and font sizes via speech pause-driven chunking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–12). ACM. https://doi.org/10.1145/3706598.3714199
16. Lee, H. (2019). Konglish? Or multilingual practice? An investigation of subtitles in Korean TV shows. *The Korean Language in America, 23*(1), 27–52. https://doi.org/10.5325/korelangamer.23.1.0027
17. Lee, J., Jun, S., Forlizzi, J., & Hudson, S. E. (2006). Using kinetic typography to convey emotion in text-based interpersonal communication. In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)* (pp. 41–49). ACM. https://doi.org/10.1145/1142405.1142414
18. Lee, Y.-J. (2012). 효율적인 한글 활자디자인을 위한 대표 글자 연구: 글자의 균형과 비례를 중심으로 [Research on representative letterforms for developing effective Hangul fonts: With focus on balance & proportion]. *LetterSeed, 4*(1), 1187–1201.
19. Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435. https://doi.org/10.1080/10888438.2020.1823986
20. O'Donovan, P., Lībeks, J., Agarwala, A., & Hertzmann, A. (2014). Exploratory font selection using crowdsourced attributes. *ACM Transactions on Graphics, 33*(4), Article 92. https://doi.org/10.1145/2601097.2601110
21. Pae, H. K. (2011). Is Korean a syllabic alphabet or an alphabetic syllabary? *Writing Systems Research, 3*(2), 103–115. https://doi.org/10.1093/wsr/wsr002
22. Park, J. S.-Y. (2009). Regimenting languages on Korean television: Subtitles and institutional authority. *Text & Talk, 29*(5), 547–570. https://doi.org/10.1515/TEXT.2009.029
23. Sasamoto, R. (2014). Impact caption as a highlighting device: Attempts at viewer manipulation on TV. *Discourse, Context & Media, 6*, 1–10. https://doi.org/10.1016/j.dcm.2014.03.003
24. Sasamoto, R., O'Hagan, M., & Doherty, S. (2017). Telop, affect, and media design: A multimodal analysis of Japanese TV programs. *Television & New Media, 18*(5), 427–440. https://doi.org/10.1177/1527476416677099

25. Sasamoto, R., Doherty, S., & O'Hagan, M. (2021). The "hookability" of multimodal impact captions: A mixed-methods exploratory study of Japanese TV viewers. *Translation, Cognition & Behavior, 4*(2), 253–280. https://doi.org/10.1075/tcb.00055.sas
26. Shaikh, A. D., Chaparro, B. S., & Fox, D. (2006). Perception of fonts: Perceived personality traits and uses. *Usability News, 8*(1), 1–6.
27. Taylor, I. (1980). The Korean writing system: An alphabet? A syllabary? a logography? In *P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), Processing of visible language* (Vol. 2, pp. 67–82). Springer. https://doi.org/10.1007/978-1-4684-1068-6_5
28. Wallace, S., Bylinskii, Z., Dobres, J., Kerr, B., Berlow, S., Treitman, R., Kumawat, N., Arpin, K., Miller, D. B., Huang, J., & Sawyer, B. D. (2022). Towards individuated reading experiences: Different fonts increase reading speed for different individuals. *ACM Transactions on Computer-Human Interaction, 29*(4), Article 38. https://doi.org/10.1145/3502222
29. Yoo, S.-A., Jeong, Y. J., & Joo, S. J. (2025). The flexible encoding of syllable positions in Korean Hangul. *Scientific Reports, 15*(1), Article 42987. https://doi.org/10.1038/s41598-025-26940-y

APPENDIX

A.1 Survey Questionnaire

Participants evaluated subtitle typography using the following ten items.
All items were rated on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree).

Readability
Q1. The subtitles were easy to read and quickly noticeable.
Q2. The subtitle text was clearly distinguishable from the video background.
Q3. The design of the subtitle typeface (e.g., weight, stroke contrast, spacing) was appropriate for reading.

Cohesion
Q4. The subtitle typeface matched well with the visual style of the video (e.g., imagery and color).
Q5. The overall feel of the subtitles blended naturally with the speaker's voice and background music without feeling out of place.
Q6. The subtitle design helped viewers focus on the video content without distracting their attention.

Attractiveness
Q7. The subtitle typeface was well aligned with the overall tone of the video (e.g., humorous, serious, emotional).
Q8. Considering the video length and pacing (short-form vs. long-form), the subtitle style felt appropriate.
Q9. The design of the subtitle typeface itself was visually appealing.

Overall Effect
Q10. The subtitle style enhanced the video's original purpose (entertainment: made the video more enjoyable; information: helped convey the content more clearly at a glance; daily life: increased a sense of relatability).

A.2

Table A1 Bonferroni-corrected item-level genre × typeface interactions (Q1–Q10)

Factors	Q1 (Reading ease)			Q2 (Text clarity)			Q3 (Font legibility)			Q4 (Visual harmony)			Q5 (Audio harmony)		
	F	p	η^2_p	F	p	η^2_p	F	p	η^2_p	F	p	η^2_p	F	p	η^2_p
G × T (short)	2.797	0.012	0.074	2.825	0.012	0.075	1.874	0.087	0.051	3.734	<.005*	0.096	3.853	<.005*	0.099
G × T (long)	7.557	<.005*	**0.182**	1.991	0.068	0.055	4.523	<.005*	**0.117**	12.93	<.005*	**0.275**	14.30	<.005*	**0.296**
Factors	Q6 (Focus aid)			Q7 (Mood fit)			Q8 (Pacing fit)			Q9 (Visual attractiveness)			Q10 (Functional Suitability)		
	F	p	η^2_p	F	p	η^2_p	F	p	η^2_p	F	p	η^2_p	F	p	η^2_p
G × T (short)	2.325	0.034	0.062	4.060	<.005*	0.104	1.662	0.132	0.045	1.822	0.096	0.049	5.225	<.005*	0.130
G × T (long)	9.406	<.005*	**0.217**	11.45	<.005*	**0.252**	7.525	<.005*	**0.181**	5.844	<.005*	**0.147**	10.68	<.005*	**0.239**

*Significant at the Bonferroni-corrected threshold of α = .05/10 = .005 for ten item-level interaction tests (Q1–Q10) within each format.

A.3

Results of pairwise comparisons between Q9 and the readability items (Q1–Q3) across all 12 conditions defined by format (short, long), genre (information, entertainment, daily), and typeface (sans, serif). Bonferroni correction was applied to account for multiple comparisons, with the significance threshold set at $\alpha = .05/9 \approx .0056$.

Format	Genre	Typeface	Comparison	Mean_Q9	Mean_Other	Δ	t	p
Short	Info	Sans	Q1 vs Q9	4.69	6.36	**1.67**	7.07	<0.0056
			Q2 vs Q9		6.47	**1.78**	6.80	
			Q3 vs Q9		6.11	**1.42**	7.20	
		Serif	Q1 vs Q9	4.22	5.97	**1.75**	6.83	<0.0056
			Q2 vs Q9		6.22	**2.00**	8.04	
			Q3 vs Q9		5.97	**1.75**	8.40	
	Entertainment	Sans	Q1 vs Q9	5.17	6.28	**1.11**	5.39	<0.0056
			Q2 vs Q9		6.61	**1.44**	6.27	
			Q3 vs Q9		6.25	**1.08**	4.33	
		Serif	Q1 vs Q9	4.78	5.92	**1.14**	3.78	<0.0056
			Q2 vs Q9		6.39	**1.61**	6.01	
			Q3 vs Q9		6.19	**1.42**	4.78	
	Daily	Sans	Q1 vs Q9	4.86	6.19	**1.33**	4.99	<0.0056
			Q2 vs Q9		6.31	**1.44**	5.92	
			Q3 vs Q9		6.00	**1.14**	4.82	
		Serif	Q1 vs Q9	4.83	6.08	**1.25**	5.43	<0.0056
			Q2 vs Q9		6.28	**1.44**	6.37	
			Q3 vs Q9		5.97	**1.14**	5.70	
Long	Info	Sans	Q1 vs Q9	4.91	6.66	**1.74**	5.93	<0.0056
			Q2 vs Q9		6.51	**1.60**	4.87	
			Q3 vs Q9		6.51	**1.60**	5.07	
		Serif	Q1 vs Q9	4.69	5.49	**0.80**	3.41	<0.0056
			Q2 vs Q9		5.86	**1.17**	4.67	
			Q3 vs Q9		5.63	**0.94**	3.32	
	Entertainment	Sans	Q1 vs Q9	5.11	6.57	**1.46**	6.80	<0.0056
			Q2 vs Q9		6.66	**1.54**	7.33	
			Q3 vs Q9		6.49	**1.37**	6.68	
		Serif	Q1 vs Q9	4.89	6.17	**1.29**	5.97	<0.0056
			Q2 vs Q9		6.49	**1.60**	6.67	
			Q3 vs Q9		6.29	**1.40**	5.68	
	Daily	Sans	Q1 vs Q9	5.17	6.54	**1.37**	6.31	<0.0056
			Q2 vs Q9		6.57	**1.40**	6.21	
			Q3 vs Q9		6.54	**1.37**	6.43	
		Serif	Q1 vs Q9	5.14	6.29	**1.14**	5.45	<0.0056
			Q2 vs Q9		6.40	**1.26**	6.09	
			Q3 vs Q9		6.06	**0.91**	4.05	

A.4

Screenshots of the visual stimuli used in the experiment. The stimuli were organized into six format–genre combinations: short-form information, short-form entertainment, short-form daily, long-form information, long-form entertainment, and long-form daily. In each figure, panels (a), (b), (c), and (d) show the sans, serif, display, and script subtitle conditions, respectively.

Video sources:

Short-form information: https://youtube.com/shorts/RSmjn3-DZaA

Short-form entertainment: https://youtube.com/shorts/0CopOMG9AxA

Short-form daily life: https://youtube.com/shorts/eU6VoHNUTlM

Long-form information: https://www.youtube.com/watch?v=x-yDmepnxss

Long-form entertainment: https://www.youtube.com/watch?v=x_N8Z86Y-6s

Long-form daily life: https://www.youtube.com/watch?v=PUXO5o42KPE

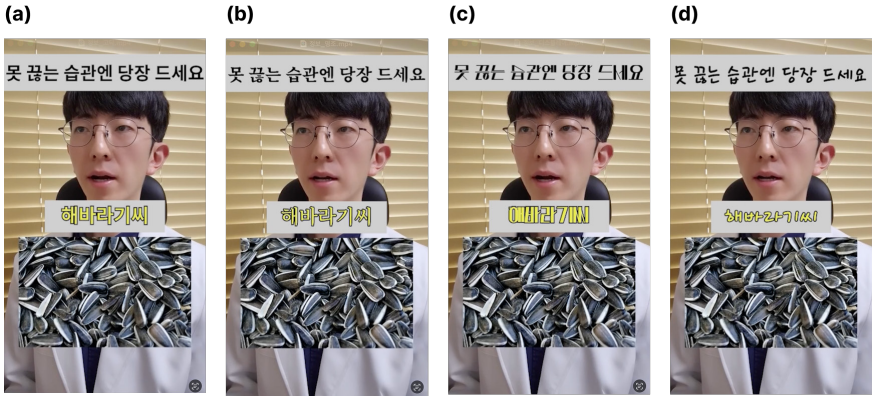

Figure A1 Screenshots of short-form information stimuli.

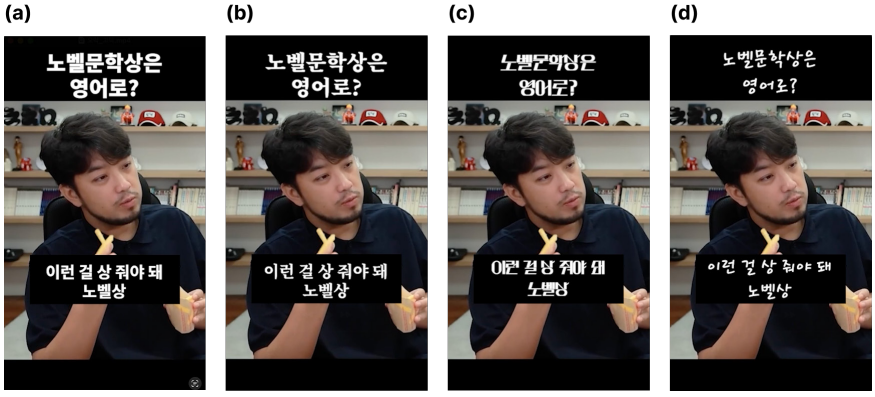

Figure A2 Screenshots of short-form entertainment stimuli.

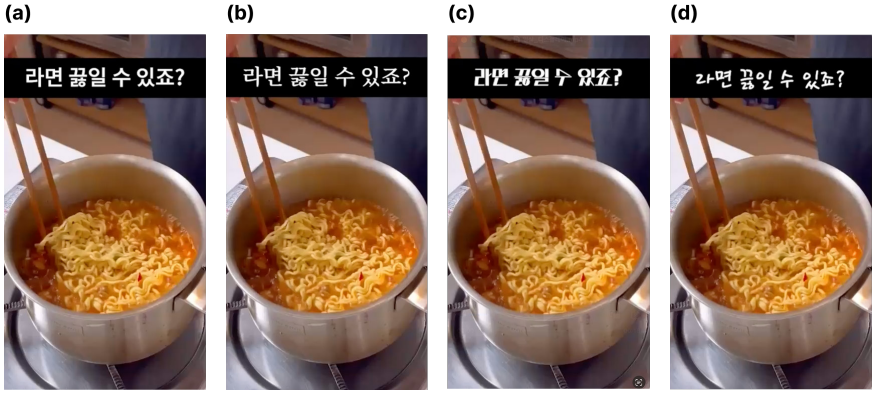

Figure A3 Screenshots of short-form daily stimuli.

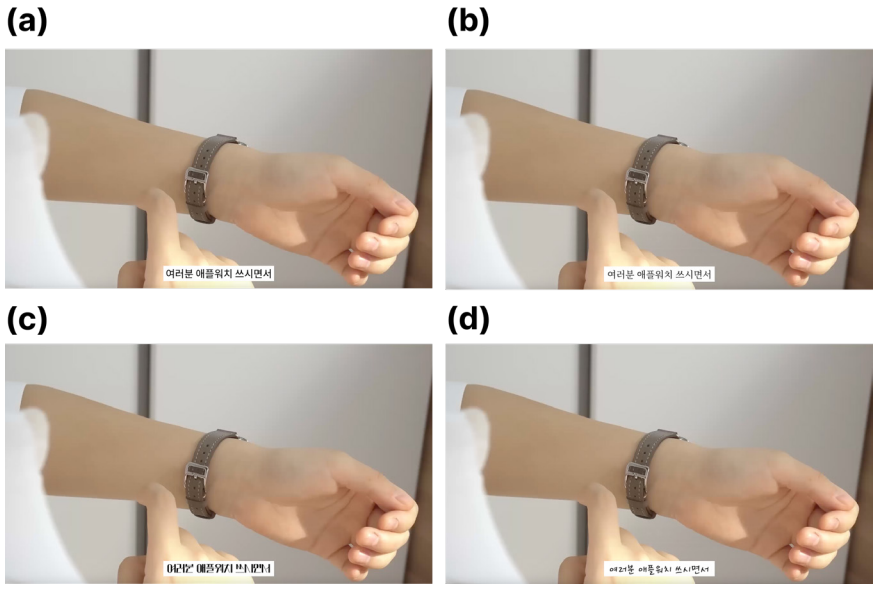

Figure A4 Screenshots of long-form information stimuli.

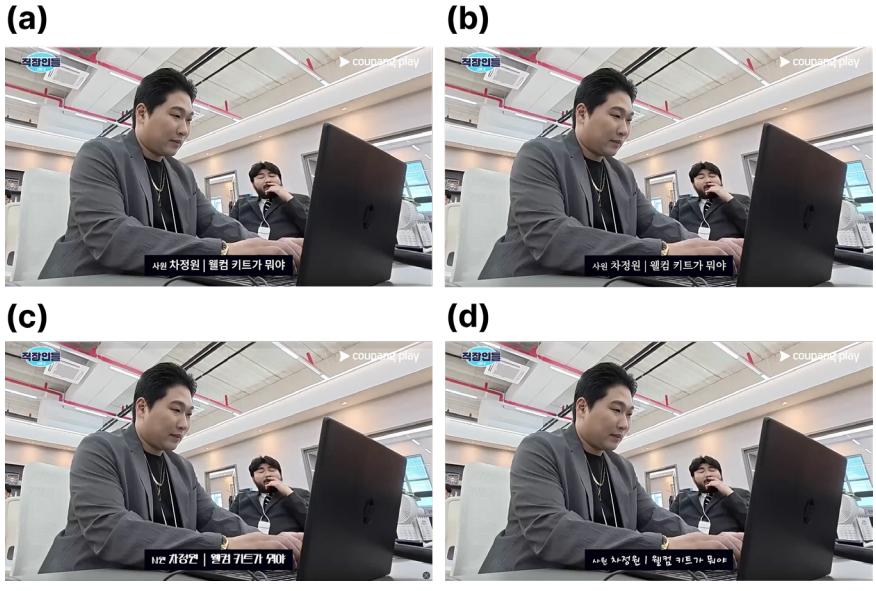

Figure A5 Screenshots of long-form entertainment stimuli.

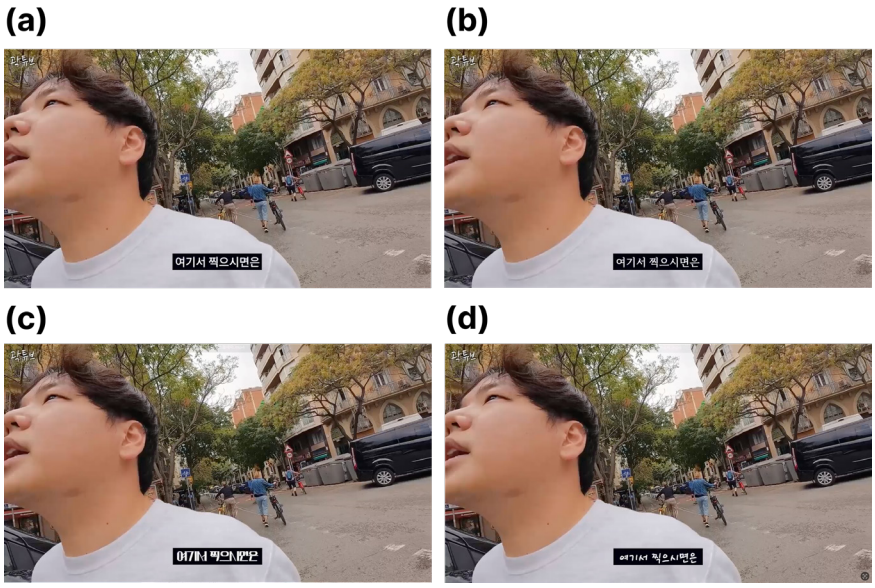

Figure A6 Screenshots of long–form daily stimuli.