Educational Innovation Research, 2026, Volume 4, Issue 5
https://bbwpublisher.com/journal/EIR
Online ISSN: 3029-1852
Print ISSN: 3029-1844

Effects of Learners' Perceptual Styles and L2 Proficiency on L2 Subtitle Utilization

Yidan Wang, Yuxuan Li, Thi Bao Ngoc Nguyen

Jiangsu Second Normal University, Nanjing 211200, Jiangsu, China

Author to whom correspondence should be addressed.

Abstract: Taking L2 video learning with subtitles as the research context, this study adopts a 2 (Perceptual Style: Visual/Auditory) × 2 (L2 Proficiency: Intermediate-High/Intermediate-Low) between-subjects experimental design. Combining comprehension tests, eye-tracking, questionnaires and interviews, it systematically investigates how perceptual styles and L2 proficiency jointly affect learners' subtitle learning outcomes, cognitive processing and subjective experiences. The results reveal significant main effects and interaction effects of both variables. Visual learners gain greater advantages in subtitle processing, with an especially prominent compensatory effect at the low-proficiency stage. L2 proficiency exerts a stronger influence on learning performance and cognitive efficiency, and higher proficiency can substantially eliminate gaps caused by perceptual styles. Low-proficiency and auditory learners rely more heavily on subtitles, accompanied by heavier cognitive load and lower processing efficiency. This study uncovers the bidirectional interactive mechanism: proficiency moderates perceptual styles while styles compensate for insufficient proficiency, providing empirical evidence for personalized L2 subtitle teaching design.

Keywords: Perceptual style; L2 proficiency; Eye-tracking; Cognitive load

Online publication: May 26, 2026

1. Introduction

Perceptual style refers to sensory preferences in receiving and processing information. Chang (2011) defines it as "the primary sensory mode individuals adopt to acquire and retain new information" [1]. As a core dimension of learners' cognitive traits, perceptual styles are categorized into visual and auditory types. Learners with distinct perceptual styles demonstrate marked disparities in handling visual and auditory input. Visual learners tend to construct cognitive frameworks via images and written text, whereas auditory learners decode information relying on voice and intonation feedback. As a stable learner characteristic, perceptual style constitutes one internal factor shaping the modality effect.

Existing literature confirms that perceptual styles modulate information processing in multimedia learning, yet few studies examine perceptual differences within L2 subtitle learning scenarios. L2 proficiency

is another core factor affecting subtitle utilization. High-proficiency learners can rapidly integrate subtitle and audio information, expanding vocabulary and analyzing pragmatics through captions. In contrast, low-proficiency learners depend on instant linguistic explanations from subtitles to lower comprehension barriers due to weak linguistic foundations [2].

Prior research mostly explores perceptual style or L2 proficiency in isolation, without integrating both to analyze L2 subtitle learning. Against this gap, the present study establishes a two-dimensional analytical framework of perceptual style and L2 proficiency. Using a mixed-methods design incorporating achievement tests, eye-tracking, questionnaires and interviews, it systematically explores how the two variables jointly shape subtitle processing, and identifies dynamic patterns of subtitle demands across proficiency levels for visual and auditory learners.

2. Research questions, participants and experimental materials

Within multimodal L2 video environments with subtitles, perceptual style and L2 proficiency may interact to determine learning performance. Four core research questions are proposed: (1) Do perceptual styles (visual/auditory) exert significant effects on learners' subtitle-based learning outcomes, eye-movement patterns and subjective perceptions? (2) Does L2 proficiency (intermediate-high/intermediate-low) significantly impact learning performance, eye-tracking indicators and subjective experience when viewing videos with full English subtitles? (3) Is there a significant interaction between perceptual style and L2 proficiency? How do the two factors jointly regulate subtitle processing and learning effects? (4) What internal cognitive mechanisms underpin subtitle use for learners of different perceptual styles and proficiency levels?

A total of 135 sophomores from a normal university in Nanjing were initially recruited. Lynn O'Brien's Learning Channel Preference Survey screened 50 typical visual learners and 52 typical auditory learners. Based on a CET-4 mock exam, 10 intermediate-high learners (550–625 points) and 10 intermediate-low learners (425–500 points) were selected from each style group, forming four experimental groups of 10 participants each: Visual-High, Visual-Low, Auditory-High, Auditory-Low.

A 15-minute PBS NOVA documentary, The Great Flood, was selected, covering flood causes and human disaster prevention strategies. PBS NOVA videos are widely adopted L2 learning materials. Four English teachers verified that its lexical difficulty and thematic familiarity matched participants' linguistic backgrounds.

3. Research instruments

(1) Perceptual Style Scale. Lynn O'Brien's (1990) Learning Channel Preference Survey was adopted, consisting of Visual Subscale and Auditory Subscale (12 items each, 5-point Likert). Higher scores indicate stronger corresponding sensory preference; participants scoring above 4 were classified as visual or auditory learners. Cronbach's $\alpha = 0.72$ (visual subscale), 0.77 (auditory subscale), indicating acceptable reliability and validity.

(2) Comprehension Test. A paper-and-pencil test with 20 multiple-choice items measuring main idea comprehension, detail retention and inferential reasoning (1 point per item, full score = 20). Three experts validated the test; Cronbach's $\alpha = 0.87$. Testing duration: 30 minutes.

(3) Eye-Tracking Equipment. EyeLink 1000 Plus eye tracker with a sampling rate of 1000 Hz, recording

fixation duration, fixation count, subtitle viewing ratio, regression frequency and other metrics. The screen was divided into a subtitle area and an image area for region-of-interest (ROI) analysis.

(4) Questionnaire & Semi-structured Interviews. The questionnaire included 5-point Likert items measuring multimodal learning experience plus an open-ended comment section. 15-minute one-on-one semi-structured interviews explored adaptation to English subtitles, attention allocation across subtitles/audio/images, perceived advantages and drawbacks of captions, and interactions between style and proficiency. All interviews were recorded, transcribed and thematically coded. SPSS 26.0 was used for all statistical analyses [3].

4. Experimental design & procedure

A2 (Visual/Auditory) × 2 (Intermediate-High/Intermediate-Low) between-subjects two-way ANOVA design was adopted, controlling topic familiarity as a confounding variable. Dependent variables: learning performance (comprehension test scores), eye-tracking metrics, subjective experience (cognitive load, learning satisfaction, motivation). Qualitative data were collected via questionnaires and interviews. The whole experiment lasted about 40 minutes per participant in independent language lab cubicles.

Procedure:

(1) Pre-experiment eye-tracking calibration (positional error controlled within 0.5°) plus 3-minute adaptive video training to eliminate equipment interference.

(2) 15-minute viewing of the subtitled documentary while eye movements were recorded in real time.

(3) Immediate 10-minute comprehension test after viewing.

(4) 5-minute questionnaire completion.

(5) Random selection of 3 participants per group (12 total) for post-experiment semi-structured interviews.

5. Results and discussion

5.1. Comprehension test results

Table 1. Descriptive statistics of comprehension test scores (N = 40)

Group	N	Mean (M)	SD	Min	Max
Visual-High	10	16.83	1.25	15	19
Visual-Low	10	12.56	1.38	10	15
Auditory-High	10	15.21	1.41	13	17
Auditory-Low	10	9.72	1.56	7	1

Two-way ANOVA results (**Table 2**): Main effect of perceptual style: $F(1, 36) = 7.82$, $p = 0.025 < 0.05$, $\eta^2 = 0.29$ (medium effect). Main effect of L2 proficiency: $F(1, 36) = 65.39$, $p = 0.011 < 0.05$, $\eta^2 = 0.54$ (large effect, stronger than style). Interaction effect: $F(1, 36) = 6.95$, $p = 0.033 < 0.05$, $\eta^2 = 0.28$ (medium effect).

Table 2. Two-way ANOVA for comprehension scores

Source	SS	df	MS	F	*p*	η^2
Perceptual Style	18.20	1	18.20	7.82	0.025	0.29

Source	SS	df	MS	F	*p*	η^2
L2 Proficiency	152.10	1	152.10	65.39	0.011	0.54
Style × Proficiency	16.20	1	16.20	6.95	0.033	0.28
Error	83.70	36	2.33	—	—	—
Total	270.20	40	—	—	—	—

LSD post-hoc comparisons show: (1) Visual learners (M = 14.65) outperformed auditory learners (M = 12.45, $p = 0.02$); high-proficiency learners (M = 16.08) scored significantly higher than low-proficiency counterparts (M = 11.13, $p < 0.008$). (2) At the low proficiency level: Visual learners earned 2.8 more points than auditory learners ($F(1, 36) = 10.25$, $p = 0.003$), confirming a strong compensatory effect of visual style for weak L2 ability. (3) At high proficiency level: Visual learners only held a 1.6-point insignificant advantage ($F(1, 36) = 3.89$, $p = 0.056$), proving advanced L2 competence neutralizes perceptual gaps. (4) Within the same style, high-proficiency subgroups significantly outperformed low-proficiency ones ($p < 0.003$ for visual; $p < 0.002$ for auditory), demonstrating cumulative proficiency effects.

Visual learners integrate text subtitles and visual scenes seamlessly, converting audio input into visual representations to reduce pure listening load. Auditory learners constantly shift attention between audio and captions, creating cross-modal cognitive conflict. High-proficiency learners automate linguistic decoding, freeing working memory for deep inference; low learners exhaust resources on basic word/sentence parsing. The interaction pattern reveals visual compensation only matters at low levels; advanced linguistic competence diminishes style-based disparities.

5.2. Eye-tracking analysis

Five core eye-movement indicators include total fixation duration, mean fixation length, subtitle viewing percentage, image-subtitle switching frequency, and regression count (**Tables 3 and 4**).

Table 3. Descriptive statistics of eye tracking indicators for each group (mean ± SD)

Total Fixation Time (s)	Average Fixation Time (s)	Subtitle Viewing Ratio (%)	Subtitle-Image Switch Count	Regression Count
682.5 ± 45.2	1.15 ± 0.12	35.2 ± 3.1	42.6 ± 5.8	18.5 ± 4.2
815.3 ± 52.7	1.32 ± 0.15	50.8 ± 4.5	38.9 ± 6.2	32.6 ± 5.7
705.8 ± 48.6	0.92 ± 0.10	39.5 ± 3.8	36.5 ± 5.1	21.3 ± 4.8
926.7 ± 60.3	1.45 ± 0.18	54.2 ± 5.2	58.3 ± 7.5	45.8 ± 6.3

Table 4. Two-way ANOVA for eye-tracking indicators

Metric	Source	F	*p*	η^2
Total Fixation	Style	6.58	0.03	0.07
	Proficiency	48.72	< 0.001	0.57
	Interaction	5.92	0.04	0.06
Mean Fixation	Style	7.25	0.02	0.08
	Proficiency	36.89	< 0.001	0.51
	Interaction	6.33	0.03	0.07

	Style	8.15	0.01	0.09
Subtitle Ratio	Proficiency	52.36	< 0.001	0.59
	Interaction	7.02	0.02	0.08
	Style	4.89	0.05	0.05
Switch Count	Proficiency	10.25	0.003	0.12
	Interaction	15.68	< 0.001	0.18
	Style	7.89	0.02	0.09
Regression Count	Proficiency	60.58	< 0.001	0.63
	Interaction	6.88	0.03	0.08

Key findings comprise: (1) Low-proficiency participants spent longer total and mean fixations, allocated a higher percentage of gaze to subtitles, and produced more regressions, indicating heavy reliance on text for decoding. (2) Auditory learners exhibited longer fixations and more regressions than visual learners due to cross-modal attention shifts. (3) The Auditory-Low group had drastically higher image-subtitle switching frequency, reflecting severe cognitive conflict between preferred audio and mandatory text reading. (4) All eye metrics correlated significantly negatively with comprehension scores (**Table 5**): more fixations, higher subtitle dependence, and frequent switching/regressions predict poorer learning performance.

Table 5. Pearson correlations between eye metrics and test scores

Eye Indicator	r	*p*
Total Fixation Duration	-0.78	< 0.001
Mean Fixation Length	-0.72	< 0.001
Subtitle Viewing Ratio	-0.81	< 0.001
Image-Subtitle Switch Count	-0.65	< 0.001
Regression Count	-0.85	< 0.001

Eye-tracking mechanism summary: Visual learners' gaze first targets subtitles, then expands to scenes; auditory learners prioritize images and read captions reluctantly. Low-proficiency learners fixate excessively on text and ignore visual contextual cues. As L2 proficiency rises, visual learners reduce subtitle reliance, and auditory learners improve multimodal coordination, narrowing style-based gaze differences. The Auditory-Low subgroup's frequent attention switching serves as direct physiological evidence of inefficient information integration.

5.3. Questionnaire & interview qualitative results

The questionnaire contained 12 five-point Likert items (Cronbach's $\alpha = 0.89$) covering attention allocation, subtitle dependence, perceived learning benefits, and cognitive load (**Table 6**).

Table 6. Two-way ANOVA results of eye tracking indicators

Eye Movement Indicator	Source of Variation	F Value	*p* Value	Effect Size (η^2)
	Perceptual Style	6.58	0.03	0.07
Total Fixation Time	L2 Proficiency	48.72	< 0.001	0.57
	Style × Proficiency	5.92	0.04	0.06

	Perceptual Style	7.25	0.02	0.08
Average Fixation Duration	L2 Proficiency	36.89	< 0.001	0.51
	Style × Proficiency	6.33	0.03	0.07
	Perceptual Style	8.15	0.01	0.09
Subtitle Viewing Ratio	L2 Proficiency	52.36	< 0.001	0.59
	Style × Proficiency	7.02	0.02	0.08
	Perceptual Style	4.89	0.05	0.05
Subtitle-Image Switch Count	L2 Proficiency	10.25	0.003	0.12
	Style × Proficiency	15.68	< 0.001	0.18
	Perceptual Style	7.89	0.02	0.09
Regression Count	L2 Proficiency	60.58	< 0.001	0.63
	Style × Proficiency	6.88	0.03	0.08

Interview thematic coding results are: (1) Visual learners regard subtitles as core learning resources, especially low-proficiency visual learners who state captions resolve unknown vocabulary and sentence structures. High-proficiency visual learners use subtitles to verify listening input and analyze pragmatic expressions. (2) Auditory learners treat subtitles as secondary aids. High auditory learners prioritize audio prosody and only check captions for difficult lexis; low auditory learners report constant distraction between sound and text, leading to heavy cognitive burden. (3) High-proficiency learners deploy flexible metacognitive strategies: ignore captions for simple segments and briefly scan subtitles for complex content. Low-proficiency learners lack regulatory strategies and get trapped in disjointed audio-text processing [4].

Qualitative data corroborate quantitative outcomes, showing perceptual styles define learners' functional positioning of subtitles, while L2 proficiency determines strategic flexibility. At low proficiency, visual learners gain the most from captions; with advanced linguistic competence, style gaps diminish as all learners achieve balanced multimodal processing.

5.4. Bidirectional interactive mechanism: Proficiency moderates style, style compensates proficiency

(1) Low-proficiency stage: Visual perceptual style generates prominent compensation. Visual text offsets weak auditory decoding capacity. Auditory learners face persistent cross-modal conflict, lowering comprehension efficiency.

(2) High-proficiency stage: Automatic linguistic processing frees cognitive resources for multimodal integration, which neutralizes style disparities. Auditory learners narrow performance gaps against visual counterparts.

(3) Long-term trend: Improved L2 proficiency partially overrides innate sensory biases, yet style preferences persist in strategy selection—visual learners leverage subtitles for lexical/syntactic analysis, auditory learners prioritize audio pragmatic input[5].

This dual-interaction framework provides actionable guidelines for differentiated subtitle teaching design.

6. Conclusion

Visual learners show stronger preference and processing efficiency for subtitled videos than auditory learners, thanks to integrated visual dual-input (text + images) matching their innate cognitive bias. L2 proficiency positively correlates with flexible, autonomous subtitle use; advanced learners employ captions for semantic comparison, while beginners rely entirely on text decoding. Significant interaction exists between perceptual style and L2 proficiency: visual styles effectively compensate linguistic deficiencies at low levels, yet proficiency advancement weakens style-based gaps. Although high L2 competence mitigates perceptual disparities, sensory preferences continuously shape learners' subtitle utilization strategies. The bidirectional interplay of internal cognitive traits and linguistic ability underpins multimodal L2 video learning, delivering empirical support for personalized subtitle-assisted instruction.

Funding

Philosophy and Social Science Research Project for Universities by Jiangsu Provincial Department of Education: A Study on Influencing Factors and Optimized Modes of Subtitle Use Among Chinese Foreign Language Learners (Project No.: 2021SJA0599)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Chang CC, Tseng KH, Tseng JS, 2011, Is Single or Dual Channel with Different English Proficiencies Better for English Listening Comprehension, Cognitive Load and Attitude in Ubiquitous Learning Environment? Computers & Education, 57(4): 2313–2321.

[2] Chen YR, Liu YT, Todd A, 2018, Transient but Effective? Captioning and Adolescent EFL Learners' Spoken Vocabulary Acquisition. English Teaching & Learning, 42(1): 25–56.

[3] Al Qatavraeh S, Eltahir ME, Alsalhi NR, 2019, The Effect of Blended Learning on the Achievement of HDE Students in the Methods of Teaching Arabic Language Course and Their Attitudes Towards Its Use. Education and Information Technologies, 25: 2101–2127.

[4] Lightbown PM, Spada N, 2015, How Languages Are Learned (4th ed.). Oxford University Press, London.

[5] Sagarra N, 2017, Longitudinal Effects of Working Memory on L2 Grammar and Reading Abilities. Second Language Research, 33(3): 341–363.