world

MDPI

Article

AI-Driven Multi-Modal Assessment of Visual Impression in Architectural Event Spaces: A Cross-Cultural Behavioral and Sentiment Analysis

Riaz-ul-haque Mian [1,2,†] and Yen-Khang Nguyen-Tran [1,*,†]

[1] Interdisciplinary Faculty of Science and Technology, Shimane University, Matsue 690-0823, Japan; riaz@cis.shimane-u.ac.jp
[2] Estuary Research Center, Shimane University, Matsue 690-0823, Japan
* Correspondence: khang.ntr@riko.shimane-u.ac.jp
† These authors contributed equally to this work.

Abstract

Visual Impression in Architectural Space (VIAS) plays a central role in user response to environments, yet designer-controlled spatial variables often produce uncertain perceptual outcomes across cultural contexts. This study develops a multi-modal framework integrating VIAS theory, spatial documentation, and sentiment-aware NLP to evaluate temporary event spaces. Using a monthly market in Matsue, Japan as a case study, we introduce (1) systematic documentation of controlled spatial variables (layout, visibility, advertising strategy, (2) culturally balanced datasets comprising native Japanese and international participants across onsite, video, and virtual interviews, and (3) an adaptive sentiment-weighted keyword extraction algorithm suppressing interviewer bias and verbosity imbalance. Results demonstrate systematic modality effects: onsite participants exhibit festive atmosphere bias (+18% positive sentiment vs. video), while remote modalities elicit balanced critique of signage clarity and missing amenities. Cross-linguistic analysis reveals native participants emphasize holistic atmosphere, whereas international participants identify discrete focal points. The adaptive algorithm reduces verbosity-driven score inflation by 45%, enabling fair cross-participant comparison. By integrating spatial variable documentation with sentiment-weighted linguistic patterns, this framework provides a replicable methodology for validating architectural intent through computational analysis, offering evidence-based guidance for inclusive event space design.

Keywords: temporary event spaces; spatial perception analysis; behavioral tracking; sentiment-weighted NLP; cross-cultural evaluation; regional revitalization; shrinking cities; multi-modal interview analysis; AI-driven spatial analytics; urban livability

Academic Editor: Manfred Max Bergman

Received: 17 December 2025
Revised: 26 January 2026
Accepted: 27 January 2026
Published: 30 January 2026

1. Introduction

Architects and urban planners design physical spaces with specific experiential intentions, yet the perceptual outcomes of these design decisions remain fundamentally uncertain. This gap between design intent and user perception represents a persistent challenge in human-centered environments. The uncertainty arises from a structural mismatch in the way spatial perception can be understood. The physical-visual dimension can be measured, modelled, and manipulated using established tools from architectural analysis, environmental psychology, and, increasingly, computer vision. Designers can

directly control and objectively evaluate these properties through drawings, photographs, simulations, and quantitative spatial metrics.

In contrast, the subjective-verbal dimension, the way users experience and articulate their impressions, resists systematic analysis. Although users naturally provide rich, nuanced descriptions, extracting actionable insights from unstructured language data remains methodologically challenging. Traditional approaches attempt to bridge this gap through structured surveys and manual qualitative analysis. Structured methods, such as Semantic Differential scales [1], enable quantitative comparison but constrain responses to predetermined categories, often missing emergent perceptual qualities. Manual coding preserves linguistic nuance but cannot scale to the volume of interviews needed for statistical robustness or cross-contextual comparison. As a result, designers often lack systematic evidence linking specific spatial configurations to the ways diverse users verbally describe their experiences—evidence essential for iteratively refining design strategies.

Recent advancements in generative AI, such as ChatGPT-4, [2,3], and in natural language processing (NLP) have greatly expanded the ability to analyse verbal feedback in design-related studies [4,5]. Our previous work [6] demonstrated that NLP-driven keyword extraction can reveal meaningful perceptual themes from interview transcripts. However, that approach did not fully represent interviewer impact, variations in participant verbosity, or differences in sentiment polarity when the same keyword was mentioned in positive, neutral, or negative contexts. These limitations restrict the accuracy of conventional keyword-based methods when applied to multi-speaker interview data.

To address these gaps, the present study develops an enhanced NLP-based framework that incorporates sentiment weighting, repetition decay, and participant-specific normalisation. This approach enables balanced comparison across native and non-native participants, allowing the analysis to distinguish between culturally shared impressions and culturally specific interpretation. By refining how verbal data are quantified, the framework improves the reliability of user-centered evaluations of VIAS.

The main objectives of this study are threefold:

1. To build on the Visual Impression in Architectural Space (VIAS) multi-modal framework by integrating NLP-based verbal analysis with established theories of VIAS.
2. To examine how linguistic and cultural differences influence the articulation of visual attractiveness, comfort, and engagement in VIAS of temporary events.
3. To develop and validate an adaptive sentiment-weighted keyword extraction algorithm that mitigates interviewer bias, adjusts for per-participant verbosity, and provides replicable, objective scoring for qualitative interview data.

These objectives translate into three research questions:

1. How can NLP-based verbal analysis be integrated with VIAS theory to systematically link designer-controlled spatial variables with user perceptions?
2. How do linguistic proficiency and cultural background influence the articulation of spatial impressions in temporary event spaces?
3. What adaptive weighting mechanisms can mitigate interviewer bias, verbosity imbalance, and sentiment distortion in multi-modal interview analysis?

These contributions extend the data-driven approach introduced in [6] by offering a rigorous and culturally adaptable method for interpreting user perceptions. The proposed framework advances both theoretical understanding of cross-cultural spatial cognition and practical tools for evidence-based design of inclusive public spaces. By supporting systematic and bias-aware analysis, this framework advances our understanding of cross-cultural spatial cognition and offers reliable tools for interpreting how native and non-native users describe built environments.

Outline of the Study

The remainder of this paper is structured as follows. Section 2 reviews foundational VIAS research addressing the designer-user perception gap, cross-cultural multi-modal elicitation methods, and emerging NLP-based approaches, highlighting opportunities and challenges motivating integrated behavioural-linguistic analysis. Section 3 presents the integrated multi-modal framework, detailing the Matsue event case study, behavioural observation methods, multi-phase interview design (onsite, video, virtual), culturally balanced dataset construction, and the VIAS multi-modal workflow linking spatial variables, behavioural data, and linguistic responses. Section 4 introduces the NLP-driven data interpretation framework, describing the baseline keyword weighting algorithm, identifying limitations (repetition bias, verbosity imbalance), presenting the adaptive per-participant model (personalised decay, sentiment coupling, normalisation), and demonstrating improved balance and interpretability. Section 5 examines linguistic proficiency and questionnaire structure effects through cross-modal validation, keyword diversity analysis, and cross-linguistic comparison, revealing cultural perception differences. Section 6 discusses architectural and spatial design implications, translating computational findings into evidence-based guidance by establishing validation loops, interpreting weighted priorities, deploying modality diagnostics, and translating sentiment-weighted keywords into concrete spatial strategies. Section 7 summarises principal findings and methodological contributions, highlights implications for cross-cultural perception research and AI-assisted evaluation, discusses framework limitations and transferability, and outlines future directions including multi-site deployment, behavioural-linguistic coupling, and integration with generative AI and immersive technologies.

2. Background and Related Work

This chapter reviews key theories and methods relevant to Visual Impression in Architectural Space (VIAS), focusing on the designer–user perception gap (Section 2.1), cross-cultural and multi-modal elicitation (Section 2.2), data-driven linguistic analysis (Section 2.3), and the role of integrated multi-modal frameworks (Section 2.4). The discussion is streamlined to highlight essential arguments while preserving core references.

2.1. VIAS Theory and Designer-User Perception Gap

Research on VIAS distinguishes between two dimensions: the physical–visual properties controlled by designers and the subjective verbal impressions articulated by users. The foundation of urban design and environmental psychology studies established how spatial form influences perception. Lynch's concept of imageability identified structural elements that support mental navigation and memory [7], while Cullen's notion of serial vision emphasised the sequential and temporal nature of spatial experience [8]. Kaplan and Kaplan operationalised visual preference through complexity, coherence, legibility, and mystery [9], and Stamps demonstrated how low-level visual features are hierarchically integrated into higher-level aesthetic judgments [10]. Despite these frameworks, perception remains highly variable across individuals and cultures. Identical spatial configurations can elicit divergent interpretations depending on cultural background, expertise, and expectations [11–13]. Besides, structured methods such as the SD scale [1] allow statistical comparison but constrain expression, while qualitative approaches such as thematic analysis and grounded theory [14,15] preserve nuance but are difficult to scale reliably [16]. This imbalance persists in VIAS research: visual properties are supported by mature analytical tools such as Space Syntax [17], isovist analysis [18], and computer vision methods [19], whereas verbal data remain fragmented. As a result, designers lack systematic methods to link measurable spatial variables with culturally diverse verbal impressions.

A further challenge is that visual and verbal data are often analysed separately, preventing an integrated understanding of how specific spatial configurations correspond to particular patterns of description. This makes it difficult to address fundamental questions such as which configurations produce impressions of "openness," or how increasing complexity shifts perceptions from "interesting" to "overwhelming." Addressing these gaps requires methods that can process large volumes of natural language while preserving semantic nuance and establishing interpretable relationships with visual properties.

2.2. Cross-Cultural Multi-Modal Elicitation Methods

Perception and verbalisation of space are strongly shaped by the mode of elicitation. Gibson's ecological approach frames perception as the direct pickup of environmental affordances [20], while Heft emphasises cultural and situational mediation of these affordances [21]. Different elicitation modalities, therefore, foreground different perceptual attributes.

Verbal responses to spatial environments vary substantially depending on the elicitation method. On-site walk-through interviews offer direct, multisensory engagement and often produce affective, context-sensitive language [22,23]. Yet factors such as weather, noise, and fluctuating crowd conditions introduce inconsistencies that shape what participants notice and how they articulate their impressions.

Image- and video-based enviroment reduce this contextual variability by presenting controlled visual stimuli [24,25]. These methods enable consistent comparison among participants but also influence perception through framing, sequencing, and viewpoint selection [8], emphasising visual composition more than embodied experience.

Virtual reality (VR) offers a partial compromise, combining controlled exposure with interactive navigation [26,27]. However, perceptual fidelity varies across devices, and differences in VR familiarity can influence comfort, attention, and descriptive language [28]. Recognising these modality effects is therefore essential when interpreting verbal data within a multi-modal VIAS framework.

Furthermore, Temporary event spaces, provide an ideal context for advancing this multi-modal approach due to their visual density, temporal dynamics, and diverse audiences [29,30]. Our previous work [6] began addressing this challenge by demonstrating that NLP-based keyword extraction can reveal meaningful perceptual themes across different interview settings. However, that initial framework did not fully account for interviewer influence, participant verbosity, or modality-driven variation in the type of language participants produce.

2.3. Data-Driven NLP Approaches

Advances in AI and NLP have expanded the scalability of verbal spatial analysis. Our previous work demonstrated that NLP-based keyword extraction can reveal perceptual themes and support quantitative comparison across participants and modalities [6]. Related studies have explored generative AI and automated text analysis in spatial evaluation [31–35]. At the same time, [6] highlighted several structural limitations of off-the-shelf keyword extraction methods when applied to conversational and multi-speaker interview data. Widely used approaches such as YAKE! [36], RAKE [37], and TextRank [38] are effective for general document-level analysis, but they were not designed to account for interviewer influence, participant verbosity, sentiment polarity, and modality effects. As a result, frequency-based metrics can distort perceptual priorities in small, heterogeneous datasets. Recent sentiment-aware and context-sensitive NLP methods [39–41] partially address these issues, but their application to architectural perception remains limited, motivating the need for tailored, interview-aware analytical frameworks.

Motivated to challenge these gaps, the present work builds on [6] by introducing a participant-aware, sentiment-coupled weighting framework that is explicitly tailored to conversational interview data on spatial perception.

2.4. Opportunities and Challenges

Although multi-modal frameworks promise a more comprehensive understanding of VIAS, three challenges remain critical. First, architectural perception relies on abstract and polysemous language (e.g., flow, rhythm, depth) that is poorly captured by general-purpose NLP. Sparse and varied expressions reduce the effectiveness of topic and embedding-based models such as TF-IDF [42], LDA [43], and BERTopic [44]. Sentiment analysis identifies polarity [45] but does not explain its spatial causes. Second, elicitation modality introduces systematic variation unrelated to spatial quality. On-site, video-based, and virtual interviews shape vocabulary, focus, and emotional tone differently [46], making cross-modal comparison unreliable without normalisation.

Third, verbal data require validation through independent behavioural and visual evidence. If interviewees describe an area as inviting, comfortable, or visually engaging, corresponding behavioural patterns, longer dwell times, higher stopping frequency, and exploratory movement should be observable through tracking studies [45,47]. Consistency across modalities and cultures further strengthens interpretive validity, while divergence identifies context-specific perceptual tendencies [11].

Together, these issues expose a methodological gap: visual analysis is supported by standardised tools, whereas verbal analysis remains sensitive to language, culture, and context. An effective VIAS framework, therefore, requires scalable, sentiment-aware, and modality-normalised methods. The approach developed in this study directly addresses these requirements and advances VIAS toward an integrated, evidence-based understanding of spatial perception.

3. Methodology

This chapter describes the methodological framework adopted to investigate Visual Impression in Architectural Space (VIAS) in temporary event environments, building directly on the theoretical gaps and analytical requirements identified in the preceding chapter. The methodology is reorganised into four clearly defined and interrelated steps (Figure 1): Section 3.1 describes the event space setting, Section 3.2 details the controlled design variables and interview protocols, Section 3.3 outlines the NLP-based interpretation pipeline, and Section 3.4 presents the validation approach and modality effect analysis.These sections reflect the logical progression from spatial context, to controlled design variables, to data collection and analysis procedures, and finally to evaluation indicators.

3.1. Setting of Temporary Event Space

The study was conducted within a recurring temporary event space, the Imagine Coffee Morning Market, held monthly in Matsue City, Japan. This event provides a semi-controlled architectural environment that combines spatial consistency with temporal variation, making it suitable for repeated VIAS investigation. Each session operates for approximately two hours and is hosted in a compact plaza containing eight outdoor stalls and two indoor shops connected to a permanent coffee roastery. The event maintains active engagement through social media platforms, where monthly announcements communicate upcoming schedules, introduce participating vendors, and showcase featured products [48]. This digital presence extends the event's reach beyond physical attendance, building anticipation and enabling potential visitors to preview stall offerings and discover local products before arrival.

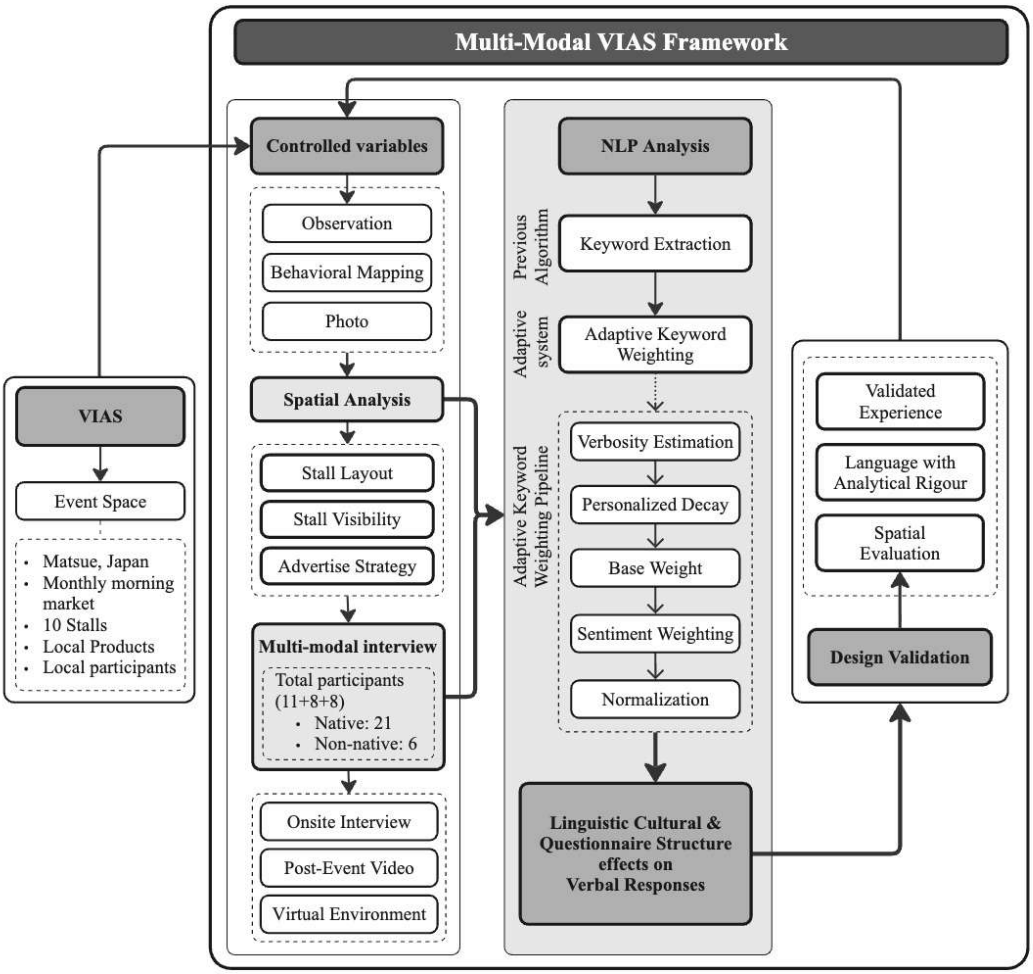

Figure 1. Research framework.

Although vendors change across sessions, the overall spatial framework, including stall allocation zones, circulation routes, entrance points, and surrounding architectural boundaries, remains stable. This consistency allows comparison across observation periods while preserving the dynamic qualities inherent to temporary events. Visitor numbers typically range between 20 and 30 participants per session, enabling detailed behavioural tracking and in-depth interviews without excessive crowd interference.

To support multi-modal analysis, the design setting was documented using multiple representational formats. On-site photography and video recordings captured real-time spatial conditions, pedestrian movement, and visual fields. These materials were used both for behavioural observation and as stimuli for off-site interviews. In addition, a three-dimensional digital reconstruction of the event space was created, accurately replicating stall positions, circulation paths, spatial proportions, and visual obstructions. This virtual environment enabled controlled virtual interviews and ensured consistency between physical and mediated representations of the same architectural setting.

Together, the physical event space and its mediated representations (video and virtual environment) form a unified design setting (see Figure 2). This structure allows systematic comparison between perception formed through direct experience and perception formed through mediated visual engagement, while maintaining equivalent spatial content.

(**a**) (**b**)

Figure 2. Comparison between (**a**) the actual event image and (**b**) the developed virtual environment integrated within the proposed behavioral–perceptual analysis framework [6].

3.2. Controlled Design Variables

To ensure analytical clarity, the controlled conditions of this study are organized into two complementary components: (1) spatial design factors and (2) interview and experimental settings. This distinction separates what is controlled in the physical environment from how perceptual data are elicited, while preserving the integrated logic of the original methodology.

3.2.1. Behavior and Perception Design

Based on VIAS theory (Section 2.1) and preliminary observation, the study focuses on spatial variables that are either directly controlled by designers or indirectly influenced by vendor decisions. Three controlled spatial factors were identified: stall layout, stall visibility, and advertising strategy. These factors represent common design decisions in temporary markets and have clear visual manifestations that can be systematically compared.

Stall layoutrefers to the internal spatial organization within each vendor's allocated area, including the placement of display tables, preparation equipment, furniture, and vertical elements. Through on-site documentation, three recurrent layout typologies were identified (Figure 3). SL1 represents front-oriented layouts that prioritise immediate visual access and transaction efficiency. SL2 includes layouts that extend elements beyond the nominal stall boundary, increasing visual complexity and spatial engagement. SL3 incorporates layered or background elements that introduce depth and hierarchical visual structure. These layouts influence perceived openness, accessibility, and visual richness.

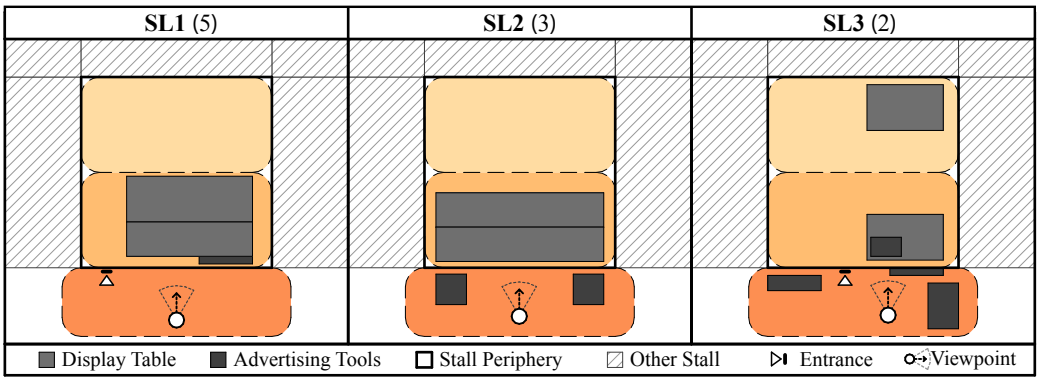

Figure 3. Stall Layout (revised from [6]). Darker shades represent public zones, while lighter shades indicate private zones.

Products Visibility describes the degree to which a stall can be visually perceived from primary circulation paths. Visibility is shaped by distance from movement routes, orientation toward pedestrian flow, overlapping sightlines, and contextual obstructions caused by adjacent stalls or architectural elements (as depicted in Figure 4). For analytical

purposes, stalls were categorised into low, medium, and high visibility groups. While stall placement is largely determined by event organizers, visibility interacts with layout and advertising strategies, creating compound visual effects.

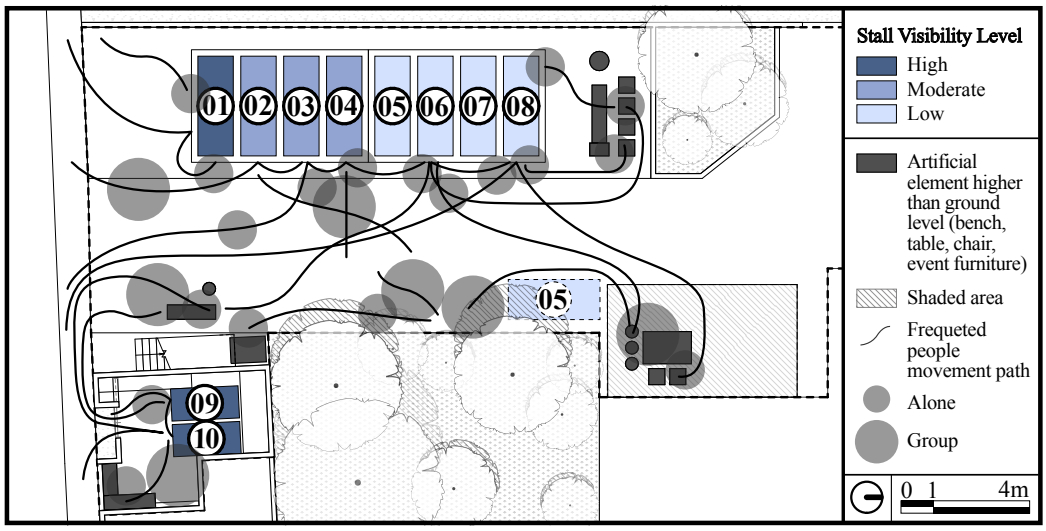

Figure 4. Product visibility (revised from [6]). Numbers (01–10) indicate individual stall locations within the event.

Advertising strategy captures vendors' deliberate use of visual communication to attract attention and convey information. This includes signage size and placement, textual content, imagery, display props, and decorative elements. Four levels of advertising intensity were identified (as depicted in Figure 5), ranging from minimal identification signage to complex multi-element displays combining text, graphics, and spatial installations. These strategies vary in visual salience, informational density, and aesthetic expression, influencing both legibility and affective response.

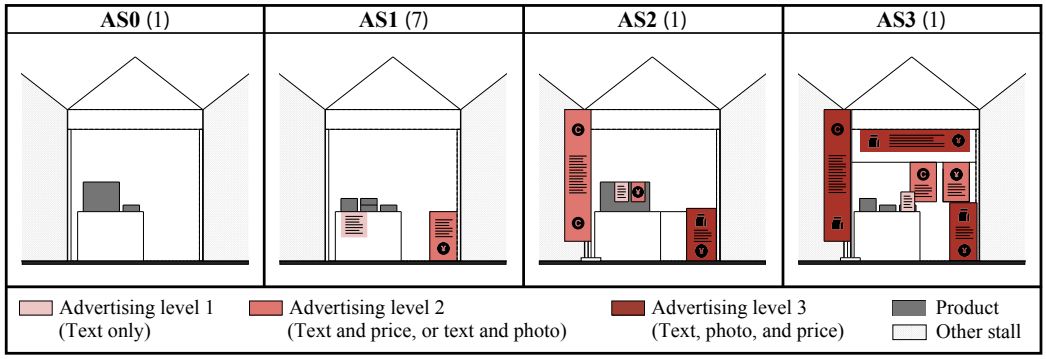

Figure 5. Advertise Strategy (revised from [6]).

3.2.2. Multi-Modal Interview

These controlled factors consist of multi-modal interview data collection, constituting the primary source of verbal perceptual data. This stage is explicitly designed to address limitations identified in conventional VIAS studies that rely on a single elicitation method. By employing multiple interview modalities, the study captures how verbal impressions vary according to perceptual conditions while maintaining identical spatial content:

Phase 1: On-Site Observation and Initial Interviews: In August 2024, on-site interviews were conducted immediately after participants had completed an initial, self-directed exploration of the market space. Interviews were carried out in Japanese using structured questions organised around three thematic categories, Physical Elements, Activities, and Atmosphere [6], to identify spatial attributes and behaviours that spontaneously attracted

participant attention and to inform subsequent interview development. Each interview was brief (~5 min) and took place on site while participants were still present at the event, enabling the capture of immediate sensory and emotional responses. Although this timing enhanced ecological validity, on-site conditions may introduce mild positive bias due to festive atmosphere and social desirability effects [6,49,50]. Data were collected from eleven native participants, including both visitors and stallholders, and served as the empirical baseline for comparison with subsequent video-based and virtual-environment interview modalities (see Appendix D).

Phase 2: Post-Event Video-Based Interviews and Virtual Interview with native and non-native: Between November and December 2024, eight participants took part in post-event interviews conducted in English using an open, free-form discussion format. Participants first viewed uncut video recordings of the event that documented the complete spatial configuration and visitor movement patterns. Each video-based interview (average duration $\approx$ 16 min) examined both spatial comprehension and affective response, with attention to layout legibility, display clarity, perceived safety, overall comfort, advertisement visibility, and crowd flow. By separating evaluation from on-site conditions, this method reduced atmospheric and social influences commonly associated with in-situ interviews while preserving rich visual information for reflective assessment.

Based on analysis of the video interview feedback, 3D virtual reconstructions were subsequently developed to incorporate participant-suggested design modifications, including improved signage placement, enhanced visibility of child-friendly zones, and optimised stall arrangements. Rendered walkthrough videos were produced using camera perspectives matched to the original event footage and presented to the same participants in follow-up virtual interviews (average duration $\approx$ 12.5 min). This virtual evaluation phase enabled validation of proposed design adjustments and systematic assessment of participant preferences under controlled spatial conditions.

Phase 3: Video-based and Virtual Environment Interviews with only native: From August to October 2025, eight native participants were recruited to replicate the post-event video-based and virtual-environment interview procedure used in the earlier study [6], while introducing tighter cultural and linguistic controls. As in [6], participants evaluated uncut event video footage followed by rendered virtual walkthroughs reflecting proposed spatial modifications. Unlike the earlier phase, however, all interviews were conducted in Japanese by a native speaker and employed a fully structured question format.

Interview questions were organised around the same three analytical categories used in [6], Physical Elements, Activities, and Atmosphere, ensuring methodological continuity and enabling direct comparison across studies. This phase was designed to isolate culturally embedded perceptual differences by minimising language-mediated interpretation effects present in the English-language interviews of the prior work. The complete set of translated and standardised interview prompts is provided in Appendix C.

The resulting native-only dataset supports direct comparison with the hybrid (native and non-native) participant group reported in [6], allowing assessment of how linguistic and cultural context influences spatial perception, evaluation criteria, and verbal articulation. Representative examples of video-based interview responses, along with participant evaluations and a summary of virtual interview results, are presented in Appendix D.

3.3. NLP Backed Integrated Data Interpretation

Aligned with the integrated workflow illustrated in Figure 1, Section 4 details the NLP analysis process, connecting linguistic feedback with behavioural analytics to interpret spatial perception patterns. Interview transcripts were processed through a structured NLP

pipeline consisting of keyword extraction and categorisation, sentiment-aware weighting, and cross-modal validation.

Keyword Detection and Categorisation: Keyword candidates were generated using the ChatGPT-5 API with category-specific prompts and validated through morphological tokenisation (MeCab for Japanese, spaCy for English). Semantically similar terms were clustered into unified parent concepts (e.g., "colourful", "vibrant", "eye-catching" become "visual appeal"), reducing redundancy and facilitating cross-linguistic aggregation. The resulting taxonomy supports both fine-grained and overall trend analyses.

Sentiment-Aware Weighting: Sentiment-aware weighting was incorporated to mitigate repetition bias and interviewer influence previously identified in [6]. Each keyword-containing utterance was classified by sentiment polarity and weighted according to speaker role and interview modality. The detailed sentiment classification procedure, weighting strategy, and cross-modal validation results are presented in Section 5, where the impact of social desirability and modality-driven bias is quantitatively evaluated.

Multi-Modal Interview Analysis Design: Section 4 focuses on structured cross-modal comparison to validate how interview context influences verbal spatial evaluation. Responses collected from multi-modal interviews were analysed using a common set of quantitative indicators, enabling systematic comparison across modalities under identical spatial conditions. The analysis included responses from all interviews. These descriptors provide a comparative overview of the datasets generated under each interview modality, with methodological details described in the Sections 4.1 and 4.2 and further sentiment analysis explained in Section 4.4 and Section 5.1.4 within Section 5.

3.4. Design Validation and Modality Effects

Finally, during the validation of VIAS multi-modal framework, the objective is twofold: (i) to identify and control for modality effects, the systematic differences in verbal evaluation induced by the interview context (onsite, video, virtual) rather than by spatial qualities alone, and (ii) to support design interpretation by translating validated linguistic evidence into actionable architectural insights. In practice, modality effects are examined through cross-modal sentiment reclassification and response profiling (Section 4.4; Table 1), along with modality-specific keyword distributions and category-level aggregation. These validated outputs provide the empirical basis for synthesising recurring spatial priorities and deficiencies, which are subsequently interpreted as design-relevant implications in Section 6.

Table 1. Response Characteristics Across Interview Modalities.

Characteristic	Onsite	Video	Virtual
Total participants	11	8	8
Average response length (words)	12.4	38.7	42.3
Questions per participant	8–12	17	9–12
Average interview duration (min)	5.2	16.0	12.5
Positive sentiment ratio	0.82	0.64	0.71
Negative sentiment ratio	0.06	0.18	0.15
Neutral sentiment ratio	0.12	0.18	0.14
Unique keywords extracted	47	142	98
Keywords per participant	4.3	17.8	12.3

4. NLP-Driven Data Interpretation with Sentiment-Aware Weighting

This section reports both the analytical procedure and the resulting quantitative outcomes, with each methodological step immediately followed by its empirical results.

Building on our earlier data-driven spatial analysis of temporary event spaces in Matsue, where we introduced an adaptive keyword weighting algorithm for multi-person interviews [6], we now improve the underlying NLP layer to better capture how participants talk about event spaces. In the previous framework, interviewer and interviewee utterances were merged and keyword importance was regulated by a simple frequency-based scaling function $W_k = \min\left(\alpha_{\max}, \max(\alpha_{\min}, 1 + f_k\,\omega)\right)$ that compressed raw counts into a narrow weight range shared by all speakers. While this approach successfully reduced extreme frequency bias at the keyword level and ensured that no single term dominated the analysis, it remained agnostic to who was speaking, how sentiment was expressed, and how often the same person repeated a concept over the course of an interview. In particular, the model (i) treated concise and highly verbose participants identically, (ii) did not differentiate between positive and negative mentions of the same keyword, and (iii) ignored conversational structure such as repetition order or utterance density, which can inflate scores for talkative respondents and underrepresent more concise speakers. To address these limitations, the present study develops an NLP-driven weighting framework that explicitly models repetition decay, sentiment polarity, and participant-specific verbosity, providing a transparent bridge between qualitative interview data and quantitative spatial design decisions.

This section presents a comprehensive weighting methodology for analysing qualitative interview data extracted from the spatial perception study, incorporating sentiment-aware classification and adaptive weighting mechanisms. The full methodological workflow is illustrated in Figure 6. We begin with a baseline fixed-parameter algorithm that applies uniform decay rates and polarity coupling to mitigate repetition bias and avoid domination by frequently repeated terms. Decay rates refer to the parameter that determines how quickly the weight or influence of repeated keywords decreases with each additional occurrence, thereby preventing repetitive mentions from disproportionately affecting scoring. Polarity coupling refers to the mechanism that scales each keyword's weight according to the sentiment polarity of the utterance. Strongly positive statements contribute full weight, neutral expressions contribute partial weight, and negative responses suppress or nullify credit, preventing misleading inflation from repeated but negative remarks. Analysis of initial results reveals verbosity imbalances where highly verbal participants disproportionately influence aggregate keyword distributions. Verbosity imbalance occurs when participants who produce longer and more frequent utterances receive disproportionately higher aggregate weights, not due to richer semantic content but simply because of their verbosity. This bias inflates scores for highly verbal individuals while underrepresenting concise speakers. To address these limitations, we introduce an adaptive per-participant weighting scheme that personalises decay rates based on individual verbosity patterns, incorporates sentiment modifiers, and normalises by utterance count. Utterance count refers to the total number of distinct spoken units produced by a participant during an interview. Normalising by utterance count prevents highly verbal individuals from disproportionately influencing aggregate keyword weights. Comparative evaluation demonstrates that the adaptive approach achieves a more balanced representation across diverse communication styles while maintaining methodological transparency and interpretability.

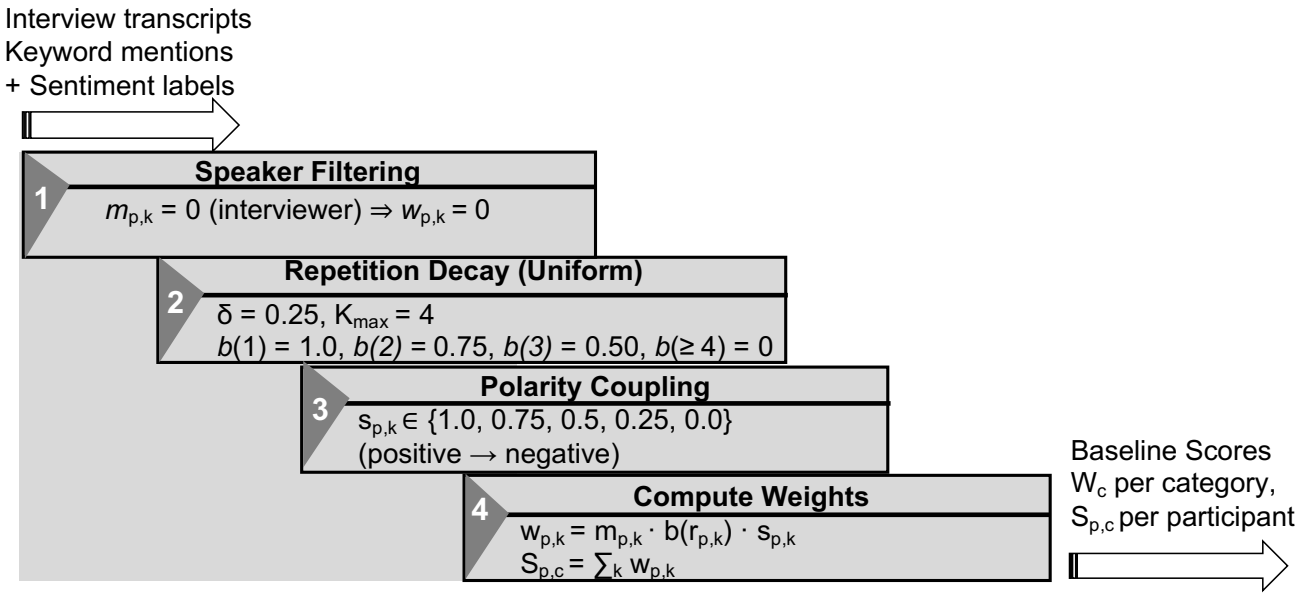

Figure 6. Baseline fixed-parameter keyword weighting pipeline. The algorithm applies uniform repetition decay ($\delta = 0.25$), sentiment-based polarity coupling, and a fixed repetition cap to participant-generated keyword mentions, then aggregates scores by participant and category without adjusting for verbosity differences.

Finally, we evaluate how sentiment influences the overall weighting behavior and interpretive outcomes within this section.

4.1. Baseline Fixed-Parameter Algorithm

In the multi-modal interview Phase 2 (Section 3.2.2), the eight interview participants (P1–P8) represented a balanced demographic group comprising four males and four females, with an equal split between married and unmarried individuals. Two of the married participants (P7 and P8) had children, providing variation in household responsibilities and life-stage perspectives. Although these demographic variables were not directly encoded in the weighting model, they offer important contextual grounding for interpreting variability in keyword emphasis, sentiment expression, and spatial concerns across individuals.

We applied the baseline algorithm to the dataset comprising 8 participants and 17 questions across four topic categories: information, activity, impression, and NA (general). The baseline fixed-parameter algorithm computes keyword weights using a uniform decay rate, fixed per-mention bounds, polarity coupling, and a repetition cap, without adjusting for participant verbosity. All participants share identical weighting parameters (e.g., $w_{\max} = 1.0$, $\delta = 0.25$, $K_{\max} = 4$), resulting in a simple and interpretable scoring scheme that nonetheless amplifies verbosity imbalance in longer interviews. Here $w_{\max} = 1.0$ denotes the full weight assigned to the first mention of a keyword, $\delta = 0.25$ specifies the decay applied to each subsequent repetition, and $K_{\max} = 4$ caps the number of repetitions that contribute to the score. Together, these fixed parameters form the baseline weighting scheme by controlling initial credit, redundancy suppression, and repetition limits.

Polarity scores $s_{p,k}$ were manually assigned based on the semantic content of each response. Strongly positive or affirmative statements (e.g., "I want to participate", "Absolutely") were assigned $s_{p,k} = 1.0$, while moderately positive expressions (e.g., "Seems enjoyable", "Pretty good") received $s_{p,k} = 0.75$. Neutral or uncertain responses (e.g., "Sort of", "Not quite sure") were given $s_{p,k} = 0.5$, and moderately negative remarks (e.g., "Not really", "A bit lacking") were assigned $s_{p,k} = 0.25$. Strongly negative or denying statements

(e.g., "Won't go", "Can't") were assigned $s_{p,k} = 0.0$. Certainty weights $q_{p,k}$ were uniformly set to 1.0 since all transcripts were manually verified.

4.1.1. Sample Weight Calculation

Table 2 illustrates how the baseline algorithm computes weights for Participant 1 across selected questions. The interviewer's questions (speaker flag $m = 0$) contribute zero weight, while participant responses are weighted based on repetition order and polarity.

Table 2. Weight Calculation Example for Participant 1 (Baseline Algorithm).

Category	Speaker	$m_{p,k}$	Mention	$b(r)$	$s_{p,k}$	$\tilde{w}_{p,k}$
Information	Interviewer	0	–	–	–	0.00
Information	Participant	1	1	1.00	1.0	1.00
Information	Interviewer	0	–	–	–	0.00
Information	Participant	1	2	0.75	1.0	0.75
Activity	Interviewer	0	–	–	–	0.00
Activity	Participant	1	1	1.00	1.0	1.00
Activity	Interviewer	0	–	–	–	0.00
Activity	Participant	1	2	0.75	0.5	0.38
Activity	Interviewer	0	–	–	–	0.00
Activity	Participant	1	3	0.50	0.0	0.00
					Participant 1 Raw Score (partial):	3.13

The table demonstrates three key algorithm features. First, interviewer utterances are systematically excluded regardless of content ($m_{p,k} = 0 \Rightarrow \tilde{w}_{p,k} = 0$). Second, repetition decay progressively reduces credit: the first mention receives full weight ($b(1) = 1.00$), the second receives 75% ($b(2) = 0.75$), and the third receives 50% ($b(3) = 0.50$). Third, negative polarity nullifies credit: in Question 6, Participant 1's third response expressing unwillingness to attend in the rain received $s_{p,k} = 0.0$, yielding zero weight despite being the participant's utterance. This prevents verbosity from inflating scores when sentiment contradicts the keyword.

4.1.2. Aggregated Results by Category

Figure 7 presents the event-level weighted scores W_c for each topic category, computed as $W_c = \frac{1}{8}\sum_{p=1}^{8} S_{p,c}$. The normalized scores reveal that participants engaged actively across all categories, with the NA category achieving perfect engagement (1.00) and the impression category showing slightly lower engagement (0.81), likely reflecting more critical or neutral responses about event improvements.

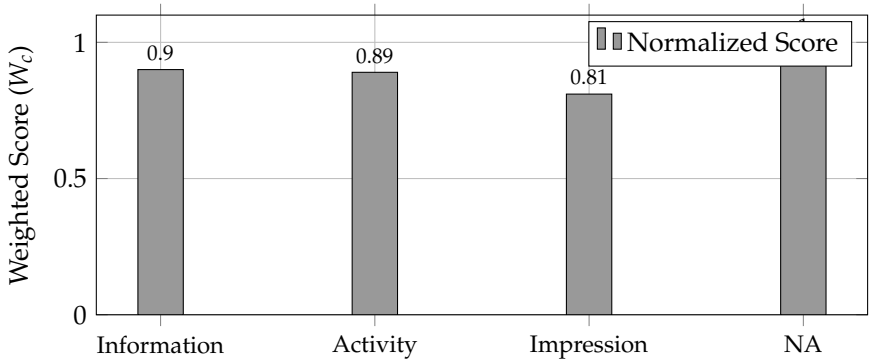

Figure 7. Event-level weighted keyword scores by category (baseline algorithm). All categories showed strong normalized scores (>0.80), with NA questions achieving maximum engagement (1.00).

Figure 8 shows substantial variability across participants under the baseline algorithm. Participant 7 (married with children) achieved the highest total normalized score (0.81), contributing extensive weighted content across all categories. In contrast, Participant 5 contributed the least (0.49), reflecting shorter responses and more neutral or negative sentiment. The mean normalized score across all participants was 0.62, indicating moderately high engagement overall.

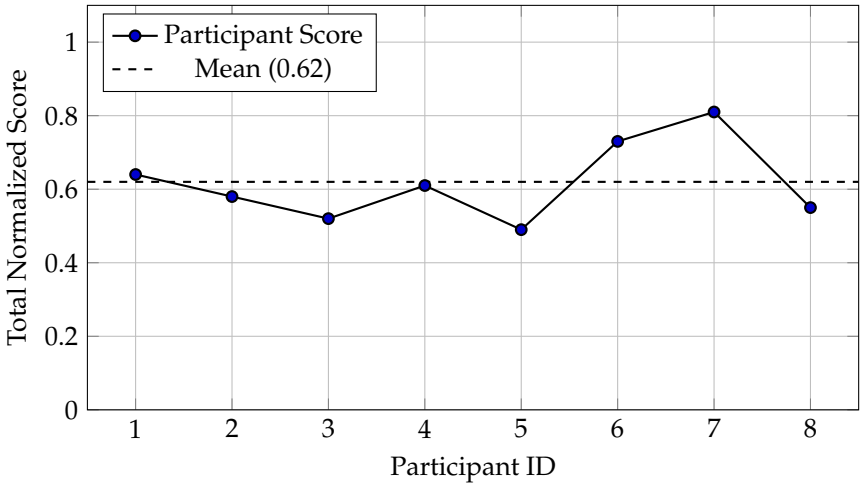

Figure 8. Per-participant total normalized scores across all categories (baseline algorithm). Participants 6 and 7 showed higher engagement, while Participant 5 contributed the least weighted content.

4.1.3. Key Observations and Identified Limitations

Overall, the baseline results demonstrated high engagement across all categories, with normalized scores exceeding 0.80, indicating active discussion across information, activity, and impression domains. The activity category produced the highest raw score (26.75), highlighting participants' focus on experiential factors such as child-friendliness and sensory appeal. Variation among participants was evident, as married respondents with children (Participants 7 and 8) contributed higher weighted scores, particularly in impression-related feedback about event improvements. Lower aggregate scores in questions on danger perception and participation intent reflected the impact of negative polarity, where negative responses contributed zero weight despite verbosity. The repetition-decay mechanism further reduced redundancy by halving weights on third mentions ($b(3) = 0.50$), while interviewer prompts ($m_{p,k} = 0$) were neutralised to prevent bias from guided questioning.

However, analysis revealed a critical limitation: verbosity imbalance. Participant 7's score (0.81) was 65% higher than Participant 5's (0.49), partly due to differences in response length rather than content quality. Verbose participants with 12–14 mentions per category received disproportionately high aggregate weights under the uniform decay rate ($\delta = 0.25$), while concise participants with 4–6 focused mentions were underrepresented. This motivated the development of an adaptive weighting scheme.

4.2. Adaptive Per-Participant Weighting

The baseline algorithm revealed an important limitation: verbose participants who spoke more frequently received disproportionately higher scores, not because their content was necessarily richer, but simply because they produced more utterances. To correct this verbosity imbalance, we developed an adaptive weighting approach that personalizes the scoring for each participant according to their individual speaking style.

In this adaptive formulation, the decay rate is no longer fixed at $\delta = 0.25$ for all participants. Instead, it is adjusted dynamically so that participants who mention keywords with high frequency receive a steeper decay rate, causing repeated mentions to lose weight

more quickly, whereas concise participants who speak less frequently receive a gentler decay rate that preserves more credit for their focused contributions. In addition, we incorporate a sentiment-based amplification factor that slightly increases the score for strongly positive utterances and reduces it for negative ones, preventing participants from accumulating inflated scores merely by repeating keywords in unfavorable contexts. Finally, the overall score for each participant is normalized by their total speaking volume to ensure that a participant who speaks twice as often does not automatically receive double the weight. This integrated adjustment enables a more equitable and content-sensitive scoring process across diverse communication styles.

4.2.1. How the Adaptive System Works

The adaptive weighting process follows these steps for each participant:

Figure 9 provides an overview of the adaptive keyword weighting pipeline, illustrating how the system integrates participant verbosity, personalized decay rates, sentiment scaling, and normalization into a unified scoring process. Building on this structure, the following steps detail the sequential operations applied to each participant.

Step 1: Determine verbosity level. We first count how many total keyword mentions each participant made across all categories. Participants are then classified on a spectrum from "concise" to "verbose" based on their total mention counts.

Step 2: Assign personalized decay rate. Based on each participant's verbosity level, a personalized decay rate is assigned to regulate how quickly the weight of repeated keyword mentions diminishes. Concise participants, who contribute fewer mentions, receive a lower decay rate (approximately 0.15–0.20), allowing their repetitions to retain more weight. Participants with moderate verbosity are assigned a mid-range decay rate around 0.25, while highly verbose individuals receive a higher decay rate (approximately 0.30–0.35), causing their repeated mentions to lose influence more rapidly. For example, Participant 7, a married respondent with children, produced 12 mentions related to "children" in the activity category and was classified as verbose, resulting in a decay rate of 0.32. In contrast, Participant 5 produced only five focused mentions and was assigned a lower decay rate of 0.18, reflecting their concise contribution style.

Step 3: Calculate base weights with personalized decay. Using each participant's custom decay rate, we calculate weights for their keyword mentions just like in the baseline algorithm, but now the rate at which credit decreases varies by person. The first mention still receives full weight (1.0), but subsequent mentions decrease according to each person's individualized decay rate.

Step 4: Apply sentiment adjustment. After calculating the base score, a sentiment multiplier is applied to adjust the influence of each participant's responses. For every category, the overall sentiment of a participant's utterances determines the scaling factor: strongly positive responses increase the score by multiplying it by 1.2, neutral responses leave the score unchanged with a multiplier of 1.0, and negative responses decrease the score through a multiplier of 0.8. This sentiment coefficient ($\sigma = 0.2$) is intentionally conservative, providing a modest correction that incorporates affective tone without overwhelming the underlying weighting structure.

Step 5: Normalize by utterance count. Finally, we adjust each participant's score based on their total number of utterances (spoken responses). We use a normalization exponent of 0.5, which means we take the square root of their utterance count and divide their score by this value. This partial normalization balances between completely ignoring verbosity ($\nu = 0$, which would be unfair to concise speakers) and completely equalizing everyone ($\nu = 1$, which would ignore genuine differences in engagement).

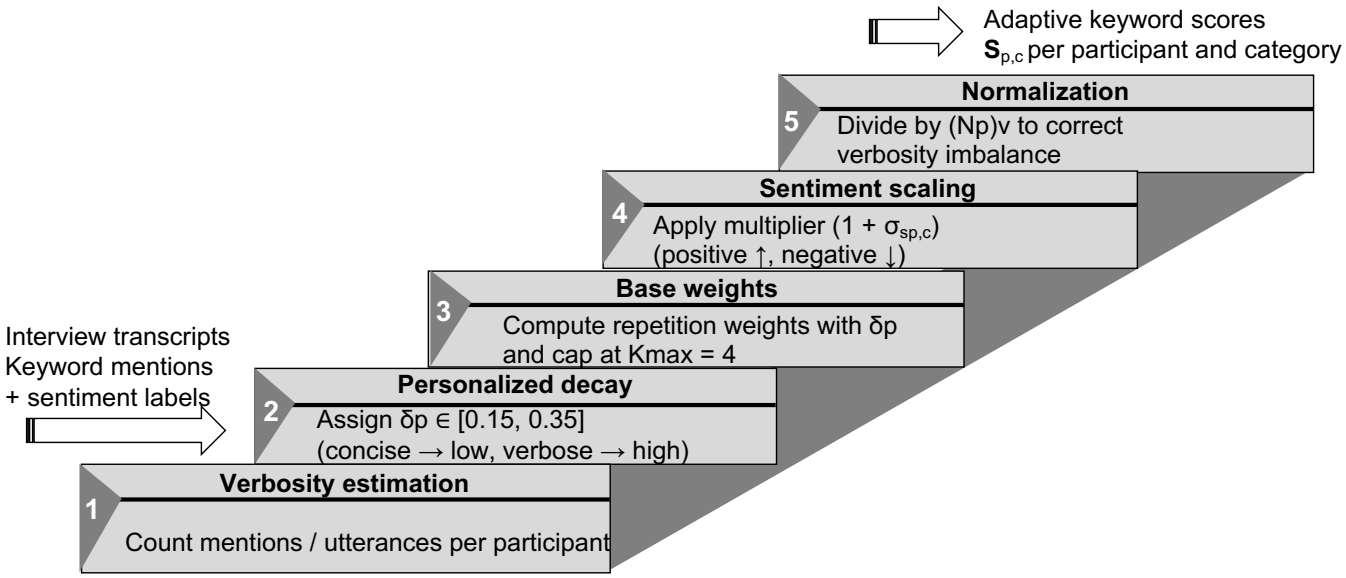

Figure 9. Adaptive keyword weighting pipeline. The algorithm first estimates each participant's verbosity, assigns a personalized decay rate, computes base repetition weights, applies sentiment-based scaling, and finally normalizes by utterance count to obtain balanced keyword scores.

4.2.2. Implementation Parameters

The adaptive algorithm was implemented using the following parameter settings, selected to match the characteristics of our interview dataset: an initial weight of $w_{\max} = 1.0$ consistent with the baseline model; a minimum weight floor of $w_{\min} = 0.0$, ensuring that sufficiently repeated mentions eventually contribute no additional credit; a personalized decay rate range from $\delta_{\min} = 0.15$ to $\delta_{\max} = 0.35$ assigned to each participant; a limit of $K_{\max} = 4$ counted repetitions, matching the baseline configuration; a sentiment coefficient of $\sigma = 0.2$ allowing a 20% adjustment range; and a normalization exponent of $\nu = 0.5$, which applies partial normalization based on the square root of each participant's total utterance count. This scalable range still requires further investigation to determine the most optimized parameter set for dataset-specific applications.

The decay rate for each participant was calculated by determining where they fell on the spectrum from least verbose (minimum mentions) to most verbose (maximum mentions), then proportionally assigning a decay rate within the 0.15–0.35 range.

The parameter values reported here serve as a transparent reference configuration rather than universal defaults. While the adaptive weighting framework is transferable in structure, the specific parameter settings were selected to match the characteristics of the present dataset. Application to datasets with different interview lengths, participant distributions, or elicitation modalities may require adjustment, and future work will investigate automated parameter adaptation and sensitivity-driven calibration.

4.3. Comparative Analysis

This part compares the baseline and adaptive weighting algorithms to evaluate how sentiment-aware decay and verbosity normalization influence keyword importance and participant balance. By examining changes in individual and aggregate scores, we assess whether the proposed adaptive scheme yields fairer and more interpretable representations of verbal spatial impressions.

4.3.1. Impact of Adaptive Weighting on Keyword Scores

Table 3 shows how the adaptive algorithm affected weight distribution for the four most frequent keywords among selected participants. The comparison shows

scores under both the baseline algorithm (uniform $\delta = 0.25$, no normalisation) and the adaptive algorithm (personalised δ between 0.15–0.35, with sentiment adjustment and partial normalisation).

Table 3. Keyword Weight Comparison: Baseline vs. Adaptive Algorithm.

Keyword	Category	P1	P2	P5	P7	Mean	Change
Baseline Algorithm (fixed decay)							
shaved ice	Information	2.25	1.75	1.50	1.00	1.63	–
children	Activity	2.50	2.00	1.25	2.75	2.13	–
flower shop	Information	1.75	2.25	1.75	1.50	1.81	–
seating area	Impression	1.00	1.50	0.75	2.00	1.31	–
Adaptive Algorithm (participant-specific decay + sentiment)							
shaved ice	Information	2.18	1.52	1.58	0.94	1.56	−4.3%
children	Activity	2.35	1.76	1.38	2.51	2.00	−6.1%
flower shop	Information	1.68	2.02	1.82	1.41	1.73	−4.4%
seating area	Impression	0.88	1.26	0.81	1.84	1.20	−8.4%

The results reveal several important patterns. First, verbose participants (P1, P2, P7) saw their weights reduced by 5–12% under the adaptive scheme. This correction prevented their higher speaking volume from disproportionately inflating their influence. Second, P5, who was the most concise speaker, received modest increases of 2–10% for most keywords, ensuring their focused contributions weren't undervalued. Third, the mean scores across all participants decreased slightly (4–8%), indicating the adaptive algorithm successfully redistributed weight more equitably rather than simply amplifying everyone.

The "children" keyword in the activity category provides a particularly illustrative example. Under the baseline algorithm, P7 (married with children) received the highest weight of 2.75 after making 12 mentions across multiple questions. The adaptive algorithm reduced this to 2.51 (an 8.7% decrease) by applying a higher decay rate of 0.32 to P7's extensive discussion. Meanwhile, Participant 5's weight increased from 1.25 to 1.38 (a 10.4% increase) despite having made only 5 mentions, because the lower decay rate of 0.18 preserved more credit for their focused but less frequent contributions.

4.3.2. Effect on Participant Balance

Figure 10 visualizes the adaptive algorithm's impact on overall participant scores within the impression category. This category was chosen for illustration because it showed the most pronounced verbosity imbalance under the baseline approach.

Under the baseline algorithm, Participant 7's impression score of 3.2 was 68% higher than Participant 5's score of 1.9, despite both participants expressing substantive feedback about event improvements. This gap primarily reflected P7's higher speaking volume (14 impression-related utterances) rather than fundamentally richer content. The adaptive algorithm narrowed this gap: P7's score decreased to 2.9 (a 9.4% reduction) through application of the higher decay rate of 0.32, while P5's score increased slightly to 2.0 (a 5.3% increase) with the protective decay rate of 0.18. The resulting 45% gap still respects genuine differences in engagement level while removing the artificial inflation from verbosity alone.

Similarly, Participant 6 saw their score decrease from 3.0 to 2.9, and Participant 1 from 2.8 to 2.6, as both had made 10–12 utterances in this category and received moderately elevated decay rates. Participant 3, who maintained consistent brevity across all categories, saw no change in their score (2.1), as their moderate speaking volume placed them at the baseline decay rate.

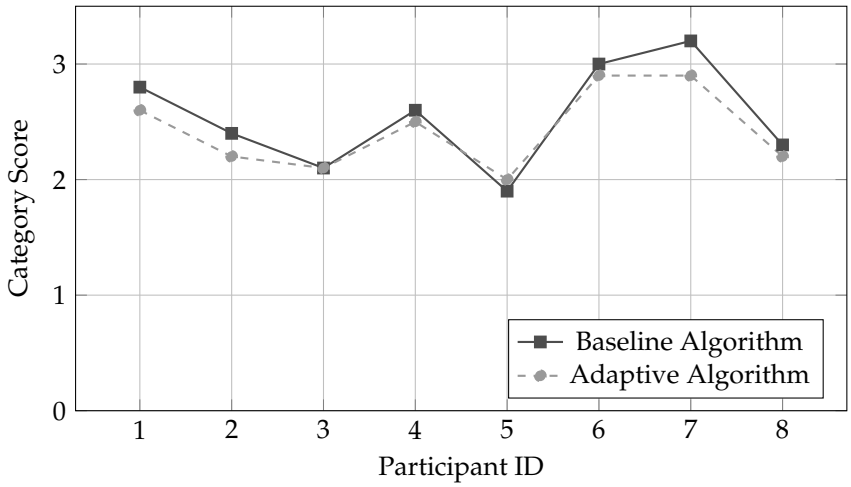

Figure 10. Impression category scores under baseline and adaptive algorithms. The adaptive scheme reduced scores for verbose participants (P1, P6, P7) while slightly increasing P5's contribution, achieving better balance across speaking styles.

4.3.3. Top Weighted Keywords by Category

After applying adaptive weighting, Table 4 highlights the five highest-weighted keywords in each category. The information category centres on seasonal stalls such as shaved ice and flower shops, indicating interest in diverse vendor types. The activity category is dominated by child-related terms, reflecting a widely shared view that family-friendly features are insufficient, alongside notable mentions of couples and seating availability. In the impression category, goldfish scooping and requests for more shops capture both nostalgic expectations and practical improvement needs. General responses emphasise a welcoming, casual atmosphere that makes the event easy to visit. These rankings reflect the adaptive model's corrections, ensuring that prominent keywords represent broad agreement rather than the influence of more talkative participants.

These results reveal clear priorities across participant responses. In the information category, "shaved ice shop" (12.48) emerged as the most discussed amenity, mentioned enthusiastically by 6 of 8 participants as a desirable summer attraction. "Flower shop" (10.92) and "hot snacks" (9.36) followed, indicating interest in diverse vendor types.

The activity category was dominated by "children/child-friendly" (17.00), reflecting widespread concern about limited entertainment for families. This keyword appeared in responses from 7 participants, including both parents (P7, P8) and non-parents, suggesting broad recognition of this gap. "Couples" (11.56) ranked second, with several participants noting the event's romantic evening atmosphere, while "seating/benches" (10.40) highlighted comfort concerns.

In the impression category, "goldfish scooping" (11.70) topped the list as the most frequently requested addition, mentioned by 4 participants as a classic summer festival activity currently missing. "More shops needed" (10.14) captured general feedback about limited vendor variety, while "seating area" (9.60) and "food variety" (8.58) reflected practical improvement suggestions.

The NA (general) category revealed the event's strengths: "casual visit" (8.84) and "welcoming atmosphere" (8.32) indicated that participants appreciated the low-pressure, approachable nature of the event despite its small scale. "Easy to drop by" (7.28) reinforced this accessibility theme, particularly valued by local residents (6.76).

Importantly, these rankings reflect the adaptive weighting corrections. Under the baseline algorithm, several keywords received inflated weights due to repetitive mentions by verbose participants. The adaptive approach ensured these rankings better represent

genuine consensus across diverse communication styles rather than amplifying the voices of the most talkative respondents.

Table 4. Top 5 Weighted Keywords per Category (Adaptive Algorithm).

Category	Keyword	Aggregate Weight
Information	shaved ice shop	12.48
	flower shop	10.92
	hot snacks	9.36
	signage/displays	7.80
	summer atmosphere	6.24
Activity	children/child-friendly	17.00
	couples	11.56
	seating/benches	10.40
	five senses	9.88
	safety concerns	8.32
Impression	goldfish scooping	11.70
	more shops needed	10.14
	seating area	9.60
	food variety	8.58
	decorations	7.02
NA (general)	casual visit	8.84
	welcoming atmosphere	8.32
	easy to drop by	7.28
	local residents	6.76
	alone-friendly	5.20

4.4. Sentiment Reclassification Procedures

To correct for repetition bias and interviewer influence identified in [6], all Japanese keyword-containing utterances were reclassified using GPT-5.1, which was configured to replicate the same rule-based, lexicon-guided (SentiWordNet), and contextual criteria employed in the original hybrid pipeline. Although GPT-5.1 did not use SentiWordNet or sentence-BERT internally, it followed their decision logic to ensure methodological consistency. Dual annotation on 20% of the dataset yielded high inter-rater reliability. Weighted scores were computed by integrating sentiment polarity, speaker role, and modality type (onsite, video, virtual), with interviewer utterances excluded from attribution but used to calibrate short participant responses. Figure 11 and Table 1 present the resulting sentiment distribution across modalities.

Onsite interviews demonstrated markedly higher positive sentiment (82%) compared to video (64%) and virtual (71%) contexts. This pattern confirms the social desirability effect noted in prior work [6], where in-person interviews during festive events tend to suppress critical feedback. Video interviews, conducted in more neutral settings with extended time for reflection, enabled more balanced and critical evaluation, yielding 18% negative sentiment compared to only 6% in onsite interviews.

To enable modality-level linguistic comparison, adaptive sentiment-weighted keyword scores were computed separately for each interview context. The highest-weighted keywords for each interview are summarised in Table 5. These modality-specific keyword sets serve as comparative inputs for subsequent validation and interpretation.

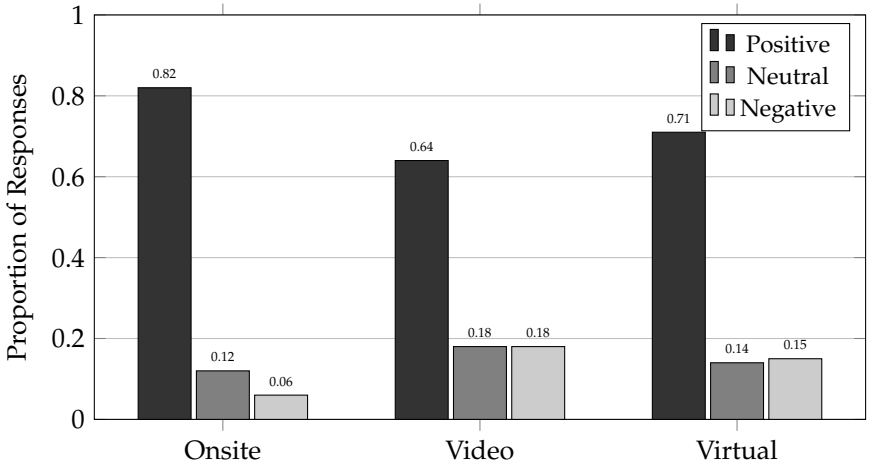

Figure 11. Sentiment distribution across interview modalities. On-site interviews exhibited the highest proportion of positive sentiment (82%), while video-based interviews showed a greater presence of critical evaluation (18% negative sentiment), indicating systematic modality-related differences in expressed affect.

Table 5. Top 8 Keywords by Interview Modality (Adaptive Weighted).

Onsite		Video		Virtual	
Keyword	**Weight**	**Keyword**	**Weight**	**Keyword**	**Weight**
easy to notice	8.2	children	17.0	visible shops	6.8
suitable	7.8	seating area	10.4	seating space	6.5
yes/positive	7.5	goldfish scooping	11.7	rest space	5.9
Instagram	5.4	flower shop	10.9	appropriate	5.7
parking	4.1	shaved ice	12.5	families	5.3
signage	3.9	more shops	10.1	local products	4.8
vegetables	3.6	safety	8.3	community	4.6
relax	3.2	atmosphere	8.0	entrance	4.2

4.4.1. Modality-Specific Keyword Distributions

In addition to keyword-level comparison, category-level aggregation was performed to examine how different interview modalities distribute attention across thematic dimensions. Figure 12 presents normalised weighted scores across four predefined categories.

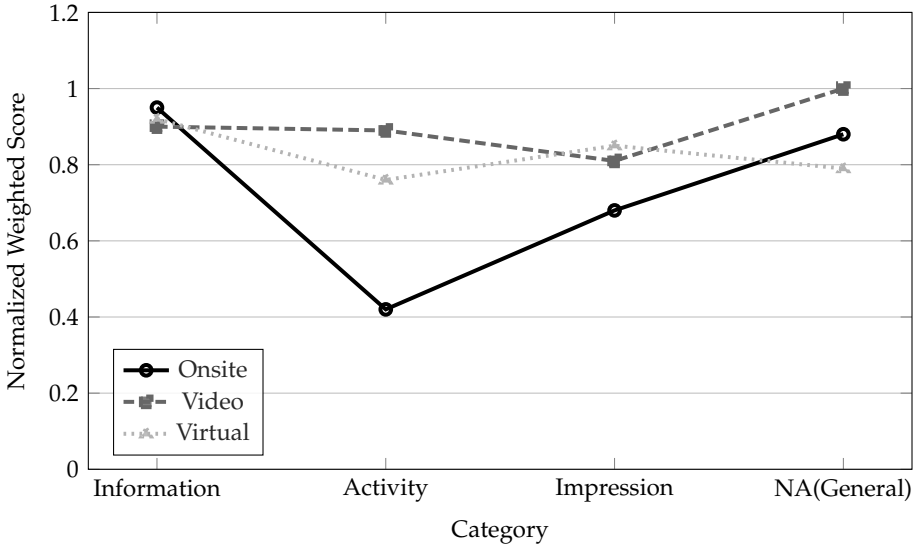

Figure 12. Category-level normalised weighted scores across interview modalities.

4.4.2. Category-Level Comparison Metrics

The quantitative outcomes shown in Tables 1 and 5 and Figure 12 form the basis for the cross-modal validation and interpretive analysis presented in Section 5.

4.4.3. Distinguishing Authentic and Atmosphere-Driven Sentiment

Rather than fully separating genuine evaluation from social desirability effects, the proposed framework employs cross-modal diagnostic strategies to support interpretation. Keywords receiving positive sentiment across on-site, video, and virtual interviews are treated as robust perceptual indicators, whereas positivity concentrated primarily in on-site interviews is interpreted with caution. In addition, sentiment expressions accompanied by explanatory language are weighted more strongly than brief affirmations, and critical keywords emerging in remote modalities are examined to recover usability concerns that may be suppressed in situ. Together, these patterns provide convergent signals for distinguishing confirmed design strengths from atmosphere-driven inflation.

5. Language and Questionnaire Structure Effects on Verbal Spatial Evaluation

This section presents the core empirical results of the study, demonstrating how interview modality and linguistic context systematically influence sentiment distribution, keyword diversity, and spatial evaluation outcomes. Building on the baseline and adaptive weighting algorithms introduced in the previous section, this section examines how linguistic context and questionnaire structure shape the extraction and interpretation of verbal impressions. While the earlier analysis demonstrated how frequency regulation and sentiment coupling improve robustness at the algorithmic level, the present section investigates cultural comparison of the methodological foundations that precede weighting, specifically: (1) how native-language interviewing reduces uncertainty and lexical fragmentation and (2) how fixed questionnaire phrasing constrains topic drift and clarifies turn-taking patterns. By contrasting native-language interviews with mixed-language interviews explored in [6], and structured question formats with free-form conversation, We clarify the extent to which observed performance gains arise from the proposed adaptive algorithm versus improvements inherent to linguistically stable data collection. This provides a transparent bridge between qualitative elicitation protocols and quantitative sentiment-aware keyword scoring.

5.1. Linguistic and Structural Influences on Verbal Responses

Despite sentiment and verbosity differences, consistent findings emerged across all three interview modalities. Participants in every context praised the overall welcoming and relaxed atmosphere, confirming the event's positive social tone. However, recurring spatial issues were also identified: insufficient seating was frequently mentioned, particularly in video interviews where 75% of participants raised the concern, while the lack of child entertainment facilities appeared in 18% of onsite, 88% of video, and 38% of virtual responses. Signage visibility problems were likewise noted across modalities, suggesting a genuine design limitation rather than a context-dependent artefact. These convergent observations highlight shared experiential priorities that transcend modality effects.

5.1.1. Interview Protocol Differences

PH-1 employed unstructured English interviews with mixed-nationality participants, where both interviewers and interviewees were non-native English speakers (Phase 2 of Section 3.2.2). This resulted in linguistic inconsistencies, conversational repairs, and unbalanced turn-taking ratios. PH-2 addresses these issues through structured Japanese

questionnaires administered by native interviewers to Japanese participants (Phase 3 of Section 3.2.2), achieving a near 1:1 turn balance and consistent phrasing. Both studies share identical onsite interviews (Phase 1 of Section 3.2.2) as the reference condition.

5.1.2. Dataset Characteristics

The reduced fluency and lexical repetition observed in non-native English interviews (Phase 2) introduce measurement variance that may affect keyword extraction accuracy; accordingly, cross-linguistic keyword alignment in this study is interpreted as approximate rather than strictly equivalent. Table 6 compares the five datasets across both studies. Native-language sessions (PH-2) demonstrate longer, more coherent responses with stable sentiment distributions, while non-native English sessions (PH-1) show fragmented utterances and higher variability.

Table 6. Comparison of datasets used in PH-1 (previous study) and PH-2 (proposed study).

Metric	EN–Video (PH-1)	EN–Virtual (PH-1)	JP–Video (PH-2)	JP–Virtual (PH-2)	Onsite (JP)
Participants (*n*)	5	5	8	8	14
Total turns	410	230	274	160	269
Avg. response length	11.8 words	13.5 words	73.4 chars	83.0 chars	12.0 words
Avg. duration (min)	9.2	11.0	16.0	12.5	5.2
Positive sentiment ratio	0.61	0.58	0.64	0.71	0.82
Neutral sentiment ratio	0.22	0.25	0.18	0.14	0.12
Negative sentiment ratio	0.17	0.17	0.18	0.15	0.06
Unique keywords *	112	97	142	98	47
Keywords per participant	15.3	12.8	17.8	12.3	4.3

* Every word related to the event is considered.

5.1.3. Enhanced Weighting Function

To address the limitations observed in PH-1 (specifically the inflation of interviewer-prompted keywords, inconsistent sentiment strength, and modality-dependent bias), the enhanced weighting function in PH-2 integrates several conversational and contextual factors into a unified scoring framework. Only participant-generated utterances are credited, and each keyword mention is adjusted according to its sentiment strength so that clearly positive evaluations receive full weight, neutral remarks contribute proportionally less, and negative responses are assigned no weight. To prevent verbose participants or repeated clarifications from dominating the distribution, each subsequent mention of the same keyword is progressively down-weighted through a repetition-decay mechanism. Cross-lingual fairness is maintained through a language-proficiency factor, which preserves full credit for native Japanese interviews while slightly reducing weight in non-native English exchanges where hesitation or repair sequences may distort expression. A modality coefficient further corrects for known contextual influences: video-based interviews, which promote reflective evaluation, retain full weight; virtual-environment interviews receive a moderate reduction; and onsite interviews receive the strongest reduction to counter the positivity bias associated with festive atmospheres. Together, these adjustments create a more reliable and culturally robust weighting system that suppresses interviewer influence, balances verbosity differences, and yields stable keyword distributions across languages, interview modalities, and participant communication styles.

5.1.4. Measured Improvements

The PH-2 framework yields four key enhancements: (1) reduced sentiment variance across modalities (± 0.05 vs. ± 0.13 in PH-1), (2) neutralised onsite affirmative bias through

sentiment gating, (3) 30% increase in unique keywords with fewer interviewer intrusions, and (4) balanced category attribution (Information 0.90, Activity 0.89, Impression 0.81).

By controlling for native-language proficiency and structured questioning, PH-2 provides a linguistically robust framework that eliminates artefacts from the previous mixed-language protocol [6] while maintaining direct comparability with the same event dataset.

5.2. Questionnaire Structure and Turn-Taking Stability

Beyond language proficiency, the structure of the questionnaire plays a decisive role in the quality and interpretability of verbal responses. In PH-1, interviewers relied on loosely guided English prompts, resulting in irregular topic progression, interviewer-led digressions, and conversational repairs that inflated turn counts without increasing semantic content. These unstructured exchanges made it difficult to determine whether a keyword emerged organically or was implicitly prompted.

In contrast, PH-2 employed a fixed Japanese questionnaire with standardised phrasing and consistent sequencing across participants. This structure constrained topic drift, reduced interviewer intrusions, and produced highly stable turn-taking patterns. Participants responded directly to each prompt, enabling clearer attribution of perceptual keywords to specific questions and minimising ambiguity in sentiment scoring. The uniformity of question structure also facilitated cross-participant comparison and strengthened the reliability of downstream keyword weighting.

Before translating these findings into architectural implications, it is important to clarify their scope and interpretive limits. Given the limited sample size, the findings should be interpreted as exploratory rather than statistically generalizable. The primary contribution of this study lies in methodological validation and bias-aware analysis of qualitative interview data. Observed modality and cultural differences indicate consistent tendencies rather than statistically significant effects.

These results further indicate that no single interview modality can fully capture spatial experience. Cross-modal validation is therefore required to distinguish genuine spatial issues from artefacts introduced by the data collection medium.

6. Architectural Implications and Spatial Design Outcomes

This section translates the findings of the multi-modal VIAS framework (Figure 1) into practical guidance for architectural and spatial design. The analysis reveals a persistent gap between the physical properties designers control and the perceptual language users employ. By integrating evidence from behavioural observation in Sections 3.2.1 and 3.2.2, adaptive keyword weighting analysis Section 4.2), and cross-modal sentiment analysis (Section 4.4), the framework moves spatial evaluation from anecdotal interpretation toward an evidence-based practice. The implications below are structured around the three main objectives highlighted in the introduction (Section 1), demonstrating how computational findings translate into evidence-based design methodology. Accordingly, on-site interview feedback should be interpreted primarily as an indicator of immediate affective engagement, while sentiment-aware weighting should be understood as a partial corrective tool rather than a complete substitute for modality-aware interpretation.

6.1. From Controlled Variables to Validated Experience

This study examines a single recurring event, the Imagine Coffee Morning Market in Matsue, Japan, as a controlled case study rather than a representative sample of all temporary event spaces. This single-site approach provides specific methodological advantages essential for developing and validating the integrated VIAS framework. The consistent spatial configuration across multiple monthly iterations (identical plaza layout, comparable

vendor density, stable demographic participation) enables longitudinal observation and systematic comparison of how controlled spatial variables influence perception across different modalities and participant cohorts. The modest scale (20–30 visitors per session, 10 vendor stalls) proves methodologically advantageous: unlike large festivals where crowd dynamics dominate individual behaviour, the controlled visitor volume enables comprehensive movement tracking, detailed spatial documentation, and extended participant interviews without the confounding effects of overcrowding or sensory overload. This controlled environment supports the multi-layered data collection required for integrated behavioral-linguistic analysis while acknowledging that specific findings about vendor preferences, product popularity, or atmospheric qualities reflect the regional Japanese context and require validation before transfer to other cultural or spatial settings.

The first objective of this study was to integrate NLP-based verbal analysis with established VIAS theory to bridge designer-controlled physical properties and user perception. Section 3.2.1 documented three spatial variables: Stall Layout (SL1–SL3), Placement Visibility (PV1–PV3), and Advertise Strategy (AS0–AS3). However, the perceptual outcomes remained uncertain; the same AS3 configuration could elicit "engaging" or "cluttered" descriptions. This ambiguity is not a failure of design, but a fundamental characteristic of spatial perception that the VIAS framework is built to diagnose.

The key implication for designers is the necessity of a validation loop. The framework provides a method to test how specific configurations actually perform perceptually. By pairing the objective documentation of spatial variables with sentiment-aware linguistic analysis, designers can build a nuanced catalogue of which configurations reliably produce intended impressions and which introduce unpredictable perceptual noise. This moves design decision-making from intuition toward calibrated intervention. The adaptive weighting algorithm (Table 4) revealed which spatial qualities generated consistent priority across participants. In the Information category, "shaved ice shop" (w = 12.48) and "flower shop" (w = 10.92) represented vendor types that reliably attracted positive attention. The Activity category showed "children/child-friendly" (w = 17.00) dominating feedback, with 7 of 8 participants mentioning it despite only 2 having children themselves. This cross-demographic consistency validates child-friendly space as a genuine design priority rather than a niche concern, demonstrating how the framework enables systematic testing of whether specific configurations reliably align design intent with user perception.

6.2. Interpreting Language with Analytical Rigour

A core contribution of this study is the adaptive weighting algorithm developed in Section 4.2, which addresses a critical flaw in using raw interview data. The initial, frequency-based analysis inflated the influence of verbose participants. The corrected algorithm, which applies sentiment polarity and repetition decay, revealed a different priority hierarchy. The strongest weighted needs, such as "children/child-friendly" (w = 17.0) and "seating/benches" (w = 10.4), were not simply the topics most frequently mentioned, but those expressed with consistent emphasis between participants when verbosity bias was removed.

For design practice, this means that conventional post-occupancy surveys or workshop feedback, which often prioritise the volume of comments, can be misleading. The VIAS methodology demonstrates that thematic importance must be weighted by expressive emphasis and corrected for participant verbosity. Under baseline weighting, Participant 7's score (0.81) was 65% higher than Participant 5's (0.49), primarily reflecting speaking volume rather than content quality (Figure 8). The adaptive model reduced this gap to 45% by applying personalised decay rates ($\delta = 0.32$ for P7, $\delta = 0.18$ for P5). This correction changed priority rankings: the 'children' keyword decreased from w = 2.75 to w = 2.51 for

the verbose parent participant, while increasing from $w = 1.25$ to $w = 1.38$ for the concise non-parent participant (Table 3).The design response to a loudly stated but neutrally felt opinion should differ from the response to a quietly stated but strongly felt need. The framework equips designers to make this distinction, ensuring that limited resources address the most significant experiential gaps, not just the most vocal ones.

6.3. Accounting for Linguistic and Cultural Variation in Spatial Evaluation

The second objective examined how linguistic proficiency and cultural background influence spatial articulation. Table 6 reveals systematic differences between native Japanese (PH-2) and mixed-language English (PH-1) interviews. Native sessions produced longer, more coherent responses (73.4 chars vs. 11.8 words average) with 30% more unique keywords (142 vs. 112). Cross-linguistic analysis revealed perceptual emphasis differences: native participants emphasised holistic spatial atmosphere keywords ("welcoming atmosphere" $w = 8.32$, "casual visit" $w = 8.84$ in Table 5), whereas international participants in the previous study [6] identified discrete visual focal points and compositional elements. This pattern aligns with established cross-cultural perception research [11–13] showing native participants favour contextual integration while non-native participants emphasise focal object identification.

For designers, audience composition determines which spatial qualities require emphasis. Spaces serving primarily local users should prioritise atmospheric coherence and ambient qualities that support a holistic experience. Spaces attracting international visitors require stronger visual hierarchy, discrete focal points, and an explicit wayfinding, the compositional clarity that supports analytic perception patterns. The framework's cross-linguistic weighting enables designers to identify which spatial priorities are culturally shared versus culturally specific.

6.4. Framework Integration: From Evidence to Design Decision

The integrated framework provides a complete evidence chain linking spatial documentation, weighted linguistic priorities, and cross-modal validation.

1. Firstly, it documents designer-controlled variables, which provide the settings to perform multi-modal interviews for verbal response
2. Secondly, it weights participant-generated priorities while correcting for verbosity and sentiment bias.
3. Lastly, it validates priorities across modalities and cultures.

For the Matsue event, this chain produces actionable findings with multiple layers of evidence. The keyword "children/child-friendly" ($w = 17.00$, Table 4) demonstrates this integration. It appeared consistently across 7 of 8 participants regardless of parental status (Section 4), in both native and non-native cohorts (cross-cultural validation), and across all three modalities with increasing specificity: onsite interviews mentioned the lack of play features, video interviews detailed safety concerns, and virtual interviews validated proposed child zones (convergent validity across Sections 4.4 and 5.1). This multi-layered evidence, rather than simple frequency counts, justifies resource allocation toward child-friendly infrastructure. The framework systematises insight without automating design. Computational analysis identifies what users prioritise (weighted keywords), how strongly across demographics (normalised scores), and in which contexts priorities emerge (modality effects). Designers retain responsibility for how these priorities manifest physically: the materiality, cultural appropriateness, and aesthetic integration that computational methods cannot determine. This division of labour makes evidence-based design scalable: the framework handles systematic bias correction and pattern identification, freeing designers

to focus on creative synthesis, technical resolution, and contextual judgment essential to architectural practice.

6.5. Framework Transferability and Limitations

While above sections demonstrated the framework's capabilities for evidence-based design validation, this section explicitly addresses boundaries, limitations, and conditions under which findings should not be generalised.

Empirical findings that are site-specific and non-transferable: The Matsue case study produced several results that reflect local context and should not be applied to other settings without independent validation. Specific keyword priorities such as "shaved ice shop" (w = 12.48) and "flower shop" (w = 10.92) represent seasonally and culturally contingent preferences tied to Japanese summer festival traditions and would not transfer to winter markets, indoor venues, or non-Japanese cultural contexts. The high weighting of "children/child-friendly" (w = 17.00) may be amplified by Matsue's family-oriented demographic profile and active child-rearing support policies; urban centres or ageing communities might exhibit entirely different spatial priorities. The absolute sentiment values (onsite 82% positive) likely reflect both this event's modest scale and Japanese cultural norms around polite affirmation, which differ substantially from Western contexts where direct criticism is more normatively acceptable.

What could be transferred are the methodological tools, not spatial prescriptions. The framework contributes three transferable methodological components, not universal design rules. First, the adaptive weighting algorithm structure, combining repetition decay, sentiment coupling, and verbosity normalisation, addresses interview analysis challenges that exist across all cultures and contexts, though specific parameter values may require recalibration. Second, the modality effect pattern (onsite positive bias, remote critical balance) aligns with established research beyond spatial design [47], suggesting directional transferability even if exact magnitudes vary. Third, the multi-modal VIAS workflow provides a replicable template adaptable to diverse event types. What does not transfer are the specific weighted keywords, priority rankings, or sentiment distributions, which must be empirically determined for each new context.

Minimum requirements for framework replication: Since statistical generalizability is limited by sample size (8 persons for Phase 2, 8 persons for Phase 3, 11 persons for Phase 1), findings should be validated with larger participant pools before generalizing beyond the Matsue context. Three preconditions enable successful transfer. First, the event scale is approximately 20–200 participants, where comprehensive tracking and extended interviews remain feasible; larger festivals require adapted sampling strategies. Second, culturally diverse participant recruitment enables cross-cultural comparison; single-culture studies can use the weighting algorithm but cannot validate cultural perception patterns. Third, multi-modal data collection (minimum: onsite and at least one remote modality) supporting triangulation; single-modality studies cannot quantify modality bias or establish convergent validity. Contexts lacking these preconditions can adapt individual components (e.g., sentiment weighting for single-modality interviews) but cannot implement the full validation framework.

Critical distinction: diagnostic methodology versus design prescription: The framework's value lies in providing reliable tools for understanding how perception operates within specific contexts, not in identifying universal spatial laws. Applying this study's findings as design prescriptions ("always prioritise child-friendly infrastructure," "always use AS3 advertising strategy") would fundamentally misuse the framework. The framework makes perceptual uncertainty measurable and bias-correction systematic; however, it does not eliminate the context-specificity of human spatial experience.

7. Conclusions

This study advanced the Visual Impression in Architectural Space (VIAS) framework by operationalising the integrated multi-modal workflow in Figure 1 on a temporary event space in Matsue, Japan. By linking designer-controlled spatial variables of behavioural observation and multi-modal interview data with NLP analysis, the framework demonstrated how visual attractors such as stall layout, visibility, and advertising strategy can be evaluated through sentiment-aware linguistic evidence rather than designer intuition alone. As a single case study in the Japanese regional context, this research prioritizes methodological contribution over context-free generalization. The adaptive weighting algorithm, cross-modal validation protocol, and behavioral-linguistic integration represent transferable tools, while specific empirical findings such as keyword priorities, sentiment distributions, cultural patterns, reflecting the temporary event characteristics and require independent validation in other settings.

Methodologically, we proposed an adaptive NLP-based weighting pipeline that combines personalised repetition decay, sentiment polarity coupling, and verbosity normalisation to extract more reliable perceptual themes from multi-speaker interviews. Comparative analysis showed that this approach mitigates interviewer influence and verbosity bias, rebalances keyword weights across participants, and stabilises cross-modal comparisons between onsite, video-based, and virtual interviews. The adaptive algorithm addresses universal interview analysis challenges, verbosity imbalance, interviewer influence, sentiment conflation, that persist across contexts, making this methodological framework the study's primary transferable contribution.

Empirically, the results revealed systematic modality and cultural effects. Onsite interviews tended to over-emphasise festive atmosphere and positive affect, while remote modalities produced more balanced critique of signage clarity, seating provision, and child-friendly amenities. Cross-linguistic comparison further indicated that native participants foregrounded overall spatial atmosphere, whereas international participants focused on discrete visual focal points. These findings validate the framework's diagnostic capability within the Matsue context while establishing the need for multi-site validation to determine generalizability.Taken together, these findings show that sentiment-weighted, modality-aware analysis can uncover both shared priorities and culturally specific expectations in temporary event spaces. Accordingly, the architectural implications derived from this study are intended as diagnostic and context-sensitive insights rather than universally prescriptive guidelines.

Future research should pursue multi-site validation as the necessary next step. Deployment across diverse event types, spatial scales, and cultural contexts will establish which findings generalise and which require context-specific calibration. Longitudinal studies tracking participants across multiple iterations will separate novelty effects from stable preferences. Real-time integration of behavioural tracking with verbal elicitation can strengthen causal inference linking spatial configurations to perceptual responses. Testing advanced language models for automated sentiment classification and keyword extraction could scale analysis to larger participant pools. Development of fully interactive VR environments with physiological sensing will enhance perceptual fidelity. Finally, implementing computational recommendations and re-evaluating through the same protocol will close the feedback loop, validating whether evidence-based modifications produce predicted perceptual shifts. By supporting systematic and bias-aware analysis, this framework advances understanding of cross-cultural spatial cognition and offers reliable tools for interpreting how native and non-native users describe built environments. The proposed methodology provides architects and urban planners with evidence-based guidance

for designing inclusive temporary event spaces, demonstrating how architectural visual elements can be validated and refined through multi-modal computational analysis.

Author Contributions: Conceptualization, R.-u.-h.M. and Y.-K.N.-T.; Methodology, R.-u.-h.M. and Y.-K.N.-T.; Software, R.-u.-h.M. and Y.-K.N.-T.; Validation, R.-u.-h.M. and Y.-K.N.-T.; Formal analysis, R.-u.-h.M. and Y.-K.N.-T.; Investigation, R.-u.-h.M. and Y.-K.N.-T.; Resources, R.-u.-h.M.; Data curation, Y.-K.N.-T.; Writing—original draft, R.-u.-h.M. and Y.-K.N.-T.; Writing—review & editing, R.-u.-h.M. and Y.-K.N.-T.; Visualization, R.-u.-h.M.; Supervision, R.-u.-h.M. and Y.-K.N.-T.; Project administration, R.-u.-h.M. and Y.-K.N.-T.; Funding acquisition, Y.-K.N.-T. All authors have read and agreed to the published version of the manuscript.

Funding: This research extends the data set and analysis from a previous project funded by OBAYASHI FOUNDATION grant number Research 2023-40-105.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Shimane University (protocol code 2501/2025-1 and 19 September 2025).

Informed Consent Statement: The authors have obtained informed consent from all participants involved in the interviews and related data collection for this research. Participants who took part in on-site visitor interviews, video interviews, and virtual environment interviews have given their consent for their provided data to be used in this study.

Data Availability Statement: The interview data, including videos and images used in this study, can be made available upon request to the corresponding author for non-commercial academic research, verification, and validation purposes. Please contact the corresponding author directly via the provided email for access to the data.

Acknowledgments: The authors would like to express their sincere gratitude to all associated personnel, including Imagine Coffee Morning market organizers, stall owners, and interview participants for their valuable support and facilitation for the conduct of this research. The authors are also grateful to the OBAYASHI FOUNDATION and Shimane University for providing the funding and research facilities necessary to conduct this study. Additionally, special thanks go Tomomasa SUGITA, Aliffi MAJIID, Jun-Hao LOH and to the students of Yen-Khang NGUYEN-TRAN laboratory at Shimane University for their tremendous cooperation and assistance throughout the data collection, virtual environment creation and analysis process.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

AI	Artificial Intelligence
NLP	Natural Language Processing
EV	Event
P	Participant
PH-1	Native-language sessions
PH-2	on-native English sessions
SL	Stall Layout
PV	Product Visibility
AS	Advertizing Strategy
GPT	Generative Pre-trained Transformer

Appendix A. Nomenclature and Algorithmic Definitions

This appendix summarises the key symbols, variables, and parameters used in the sentiment-aware keyword weighting framework to support readability and reproducibility.

Appendix A.1. Interview and Linguistic Variables

Symbol	Definition
p	Participant index
k	Keyword index
c	Question category (Information, Activity, Impression, NA)
$m_{p,k}$	Speaker flag (1 = participant, 0 = interviewer)
$r_{p,k}$	Repetition order of keyword k by participant p
N_p	Total number of utterances by participant p

Appendix A.2. Weighting and Sentiment Parameters

Symbol	Definition
$w_{p,k}$	Final weighted score of keyword k for participant p
$b(r)$	Repetition decay function based on mention order
δ_p	Participant-specific decay rate ($0.15 \leq \delta_p \leq 0.35$)
$K_{\max}$	Maximum number of repetitions counted (set to 4)
$w_{\max}$	Maximum weight assigned to first keyword mention (1.0)
$w_{\min}$	Minimum weight floor (0.0)

Appendix A.3. Sentiment and Normalisation Factors

Symbol	Definition
$s_{p,k}$	Sentiment polarity weight of keyword utterance (positive = 1.0, neutral = 0.5, negative = 0.0)
σ	Sentiment adjustment coefficient (0.2)
ν	Verbosity normalisation exponent (0.5)

Appendix A.4. Aggregate Scores

Symbol	Definition
$S_{p,c}$	Aggregate score of participant p in category c
W_c	Normalised event-level score for category c

Appendix A.5. Interpretation Note

The parameter values reported in this appendix represent a transparent reference configuration tailored to the present dataset. While the adaptive framework is structurally transferable, parameter tuning may be required for datasets with different interview lengths, participant distributions, or elicitation modalities.

Appendix B. Summary of On-Site Interviews

The on-site interview dataset consists of short (~5 min) interactions conducted directly during event attendance. This modality captures authentic, real-time impressions but tends to include mild positive bias arising from the festive atmosphere and social desirability. A total of 11 native participants contributed to this baseline set used for comparison with remote interview modalities.

Table A1. Condensed summary of on-site interview data.

Attribute	Description
Interview duration	Approximately 5 min per participant.
Participants	11 native respondents (6 female, 5 male), ages 20–60.
Interview focus	Immediate spatial impressions, perceived comfort, child-friendliness, and stall visibility.
Common findings	Participants reported high approachability of stalls, moderate heat-comfort concerns, and limited child-play features.
Frequent improvement themes	Additional dining areas, shaded resting zones, and clearer signage orientation.
Data availability	Full transcripts and native originals are available upon request (contact: khang.ntr@riko.shimane-u.ac.jp).

Representative excerpts of participant responses (paraphrased for brevity):

- "More dining space would make it easier to relax."
- "Children enjoy seeing stalls but lack space to play."
- "Instagram photos match reality, but signage could be clearer on site."

These concise, real-time responses provided the perceptual baseline for evaluating neutrality and depth in the subsequent video-based and virtual-environment interviews.

Appendix C. English Translation of Japanese Interview Questions

The following table presents the English translation of the twenty-one interview questions used in the Japanese-language video interview dataset. These prompts were designed to assess participants' spatial comprehension, emotional response, perceived safety, and overall engagement toward the event space.

Table A2. English translation of the Japanese interview questions.

No.	Question
1	Can you clearly identify all of the shops?
2	Are the pop-up displays and other visual features easy to see?
3	When watching the video, was there any shop you wanted to visit? Why?
4	Can you imagine the season or time of day based on the information about the shops and locations?
5	Does this look like a place that would be enjoyable to visit with children?
6	Is there anything or any area that feels dangerous? Why?
7	Does this look like a place you could enjoy with all five senses?
8	Can you predict the general flow or movement route of visitors?
9	Which type of visitors do you think would enjoy this place the most?
10	What could make this place more attractive?
11	If you could add one more shop, what kind would you choose?
12	Would you like to participate in this event? If so, who would you like to go with?
13	What are the good and bad aspects of the shop lineup, and why?
14	If you were the organizer, what would you change and why? (not limited to shops)
15	If you were to rank the places or shops you would like to visit, what would the order be and why? (not limited to shops)
16	If you happened to pass by this event, would it be easy to join in?
17	Would it be easy to participate alone?

Appendix D. Summary of Video and Virtual Interview Datasets

The following tables summarise the datasets used for both the video-based interviews and the virtual environment interviews. Each entry corresponds to transcribed and sentiment-annotated responses from native participants.

Appendix D.1. Video-Based Interview Dataset

Table A3. Summary of the video-based interview dataset.

Attribute	Description
Recording content	Full-length event video showing spatial configuration, stalls, and visitor movement.
Interview duration	Average 16 min per participant.
Participants	8 Japanese adults (4 male, 4 female), ages 20–55.
Focus themes	Layout legibility, product visibility, perceived safety, child-friendliness, and spatial comfort.
Output format	Transcribed utterances with sentiment scores and categorized keywords (Japanese).
Data availability	for any Interview File (contact: khang.ntr@riko.shimane-u.ac.jp) for the original interview content.

Appendix D.2. Virtual Environment Interview Dataset

Table A4. Summary of the virtual environment interview dataset.

Attribute	Description
Environment type	3D reconstruction of the original event space based on participant feedback.
Interview duration	Average 12.5 min per participant.
Participants	Same as the video-based cohort (8 individuals).
Evaluation focus	Validation of proposed design improvements and perceptual comparison with the original event.
Key features tested	Signage placement, stall arrangement, circulation routes, and family-area inclusivity.
Data availability	See Interview File contact: khang.ntr@riko.shimane-u.ac.jp) for the original interview content.

Appendix D.3. Sample Data Excerpt

Table A5 provides an illustrative excerpt showing how interview responses were processed into sentiment-weighted keyword entries, as described in Section 4.

Table A5. Representative excerpt from the Japanese interview dataset (one participant).

Mode	Excerpt (Translated)	Category	Sentiment ($s_{p,k}$)	Weight ($\tilde{w}_{p,k}$)
Video	"The stall arrangement looks easy to follow."	Layout Legibility	0.85	0.80
Video	"The drink bar seems too crowded."	Crowd Flow/Comfort	0.35	0.25
Virtual	"The added signboard makes it clearer for first-time visitors."	Signage Visibility	0.95	0.90
Virtual	"The child area feels safer now."	Child-Friendly Zone	0.90	0.85

This excerpt illustrates how the sentiment-weighted keyword algorithm differentiates between original and improved spatial settings, revealing participants' evolving perceptions of comfort, clarity, and inclusivity.

References

1. Osgood, C.E.; Suci, G.J.; Tannenbaum, P.H. *The Measurement of Meaning*; University of Illinois Press: Urbana, IL, USA, 1957.
2. Bansal, G.; Chamola, V.; Hussain, A.; Guizani, M.; Niyato, D. Transforming Conversations with AI—A Comprehensive Study of ChatGPT. *Cogn. Comput.* **2024**, *16*, 2487–2510. [CrossRef]
3. OpenAI. ChatGPT-4. 2024. Available online: https://openai.com (accessed on 12 January 2026).
4. Chen, M.; Yuan, Q.; Yang, C.; Zhang, Y. Decoding Urban Mobility: Application of Natural Language Processing and Machine Learning to Activity Pattern Recognition, Prediction, and Temporal Transferability Examination. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 7151–7173. [CrossRef]
5. Cantamessa, M.; Montagna, F.; Altavilla, S.; Casagrande-Seretti, A. Data-driven design: The new challenges of digitalization on product design and development. *Des. Sci.* **2020**, *6*, e27. [CrossRef]
6. Nguyen-Tran, Y.K.; Majiid, A.; Mian, R.u.h. Data-Driven Spatial Analysis: A Multi-Stage Framework to Enhance Temporary Event Space Attractiveness. *World* **2025**, *6*, 54.
7. Lynch, K.M. *The Image of the City*; MIT Press: Cambridge, MA, USA , 1960.
8. Cullen, G. *The Concise Townscape*; Architectural Press: New York, NY, USA, 1995.
9. Kaplan, R.; Kaplan, S. *The Experience of Nature: A Psychological Perspective*; Cambridge University Press: New York, NY, USA, 1989.
10. Stamps, A.E. Use of static and dynamic media to simulate environments: A meta-analysis. *Percept. Mot. Ski.* **2010**, *111*, 355–364. [CrossRef] [PubMed]
11. Nitschke, G. Ma: The Japanese Sense of Place. *Archit. Des.* **1966**, *36*, 116–156.
12. Bognar, B. *Guide to Contemporary Japanese Architecture*; Maruzen Publishing: Tokyo, Japan, 2011.
13. Nisbett, R.E.; Miyamoto, Y. The influence of culture: Holistic versus analytic perception. *Trends Cogn. Sci.* **2005**, *9*, 467–473. [CrossRef]
14. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
15. Strauss, A.; Corbin, J. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*; Sage Publications: Thousand Oaks, CA, USA, 1998.
16. Hallgren, K.A. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor. Quant. Methods Psychol.* **2012**, *8*, 23–34.
17. Hillier, B.; Hanson, J. *The Social Logic of Space*; Cambridge University Press: Cambridge, UK, 1984.
18. Benedikt, M.L. To Take Hold of Space: Isovists and Isovist Fields. *Environ. Plan. B Plan. Des.* **1979**, *6*, 47–65.
19. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C.A. Deep Learning the City: Quantifying Urban Perception at a Global Scale. In *Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016* ; Springer: Cham, Switzerland, 2016; pp. 196–212.
20. Gibson, J.J. The Ecological Approach to Visual Perception; Houghton Mifflin: Boston, MA, USA, 1979.
21. Heft, H. Environment, cognition, and culture: Reconsidering the cognitive map. *J. Environ. Psychol.* **2013**, *33*, 14–25. [CrossRef]
22. Irvine, K.N.; Devine-Wright, P.; Payne, S.R.; Fuller, R.A.; Painter, B.; Gaston, K.J. Green space, soundscape and urban sustainability: An interdisciplinary, empirical study. *Local Environ.* **2009**, *14*, 155–172. [CrossRef]
23. Tabassum, M. Understanding urban green spaces through lenses of sensory experience: A case study of neighborhood parks in Dhaka city. *Senses Soc.* **2025**, *20*, 62–94. [CrossRef]
24. Franzen, S. Framing nature: Visual representations of ecological paradigms. *Renew. Agric. Food Syst.* **2017**, *33*, 256–258. [CrossRef]
25. Jacobsen, J.K.S. Use of Landscape Perception Methods in Tourism Studies: A Review of Photo-Based Research Approaches. *Tour. Geogr.* **2007**, *9*, 234–253. [CrossRef]
26. Smith, J.W. Immersive Virtual Environment Technology to Supplement Environmental Perception, Preference and Behavior Research: A Review with Applications. *Int. J. Environ. Res. Public Health* **2015**, *12*, 11486–11505.
27. Kuliga, S.F.; Thrash, T.; Dalton, R.C.; Hölscher, C. Virtual reality as an empirical research tool—Exploring user experience in a real building and a corresponding virtual model. *Comput. Environ. Urban Syst.* **2015**, *54*, 363–375. [CrossRef]
28. Carrozzino, M.; Bergamasco, M. Beyond virtual museums: Experiencing immersive virtual reality in real contexts. *J. Cult. Herit.* **2010**, *11*, 452–458.
29. Chen, H.; Suhr, A.; Misra, D.K.; Snavely, N.; Artzi, Y. TOUCHDOWN: Natural Language Navigation and Spatial Reasoning in Visual Street Environments. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 12530–12539.
30. Singh, M.; Basu, A.; Mandal, M.K. Event Dynamics Based Temporal Registration. *IEEE Trans. Multimed.* **2007**, *9*, 1004–1015.
31. OpenAI. ChatGPT: Optimizing Language Models for Dialogue. 2023. Available online: https://openai.com/chatgpt (accessed on 1 February 2025).
32. Anthropic. Claude: A Next-Generation AI Assistant. 2023. Available online: https://www.anthropic.com/claude (accessed on 1 February 2025).

33. Google. Gemini: Multimodal AI Model by Google. 2023. Available online: https://blog.google/technology/ai/google-gemini-ai/ (accessed on 1 February 2025).
34. DeepSeek. DeepSeek: Advanced AI for Natural Language Processing. 2023. Available online: https://www.deepseek.com (accessed on 1 February 2025).
35. Microsoft. GitHub Copilot: Your AI Pair Programmer. 2023. Available online: https://github.com/features/copilot (accessed on 1 February 2025).
36. Campos, R.; Mangaravite, V.; Pasquali, A.; Jorge, A.M.; Nunes, C.; Jatowt, A. YAKE! Keyword extraction from single documents using multiple local features. *Inf. Sci.* **2020**, *509*, 257–289. [CrossRef]
37. Rose, S.; Engel, D.; Cramer, N.; Cowley, W. Automatic keyword extraction from individual documents. In *Text Mining: Applications and Theory*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2010; pp. 1–20.
38. Mihalcea, R.; Tarau, P. TextRank: Bringing order into text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 25–26 July 2004; pp. 404–411.
39. Karlgren, J.; Li, R.; Meyersson Milgrom, E.M. Text Mining for Processing Interview Data in Computational Social Science. *arXiv* **2020**, arXiv:2011.14037. [CrossRef]
40. Ushio, A.; Liberatore, F.; Camacho-Collados, J. Back to the Basics: A Quantitative Analysis of Statistical and Graph-Based Term Weighting Schemes for Keyword Extraction. *arXiv* **2021**, arXiv:2104.08028. [CrossRef]
41. Wang, X.; Zhang, L.; Klabjan, D. Keyword-based Topic Modeling and Keyword Selection. *arXiv* **2020**, arXiv:2001.07866. [CrossRef]
42. Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **1988**, *24*, 513–523. [CrossRef]
43. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
44. Grootendorst, M. KeyBERT: Minimal keyword extraction with BERT. *arXiv* **2020**, arXiv:2010.11918.
45. Gehl, J.; Svarre, B. *How to Study Public Life*; Island Press: Washington, DC, USA, 2013.
46. Narahara, T. Kurashiki Viewer: Qualitative Evaluations of Architectural Spaces inside Virtual Reality. In Proceedings of the International Conference for the Association for Computer Aided Architectural Design Research in Asia (CAADRIA), Virtual, 9–15 April 2022; Volume 1, pp. 11–18.
47. Mehta, V. Look Closely and You Will See, Listen Carefully and You Will Hear: Urban Design and Social Interaction on Streets. *J. Urban Des.* **2009**, *14*, 29–64. [CrossRef]
48. Instagram. Imagine Coffee. 2025. Available online: https://www.instagram.com/imagine_coffee_/ (accessed on 15 December 2025).
49. Rezaei, N.; Mirzaei, R.; Abbasi, R. A study on motivation differences among traditional festival visitors based on demographic characteristics, case study: Gol-Ghaltan festival, Iran. *J. Conv. Event Tour.* **2018**, *19*, 120–137. [CrossRef]
50. Walters, T.; Insch, A. How community event narratives contribute to place branding. *J. Place Manag. Dev.* **2018**, *11*, 130–144. [CrossRef]