

Supplementary Material

Part A: Sample Overview

Table A1. Participant List with Demographic Information

PersonID	Age group	Gender	Place of residence	Frequency of being in Würzburg	Household Size
p01	18-30	male	Erlangen	less than weekly	2
p02	18-30	female	Fürth	less than weekly	3
p03	18-30	male	Erlangen	less than weekly	2
p04	31-45	female	Nürnberg	never	N/A
p05	18-30	male	Erlangen	never	1
p06	31-45	female	Nürnberg	daily	1
p07	18-30	female	Nürnberg	less than weekly	1
p09	18-30	female	Erlangen	never	1
p11	18-30	female	Erlangen	never	2
p13	18-30	male	Nürnberg	less than weekly	2
p14	18-30	female	Nürnberg	never	1
p15	18-30	female	Erlangen	less than weekly	2
p16	18-30	male	Erlangen	never	1
p17	18-30	male	Erlangen	never	5
p18	18-30	male	Erlangen	never	1
p19	18-30	male	Buckenhof	never	3
l01	18-30	female	Würzburg	daily	1
l02	18-30	female	Frankfurt am Main	several times a week	2
l03	18-30	female	Würzburg	daily	1
sts01	18-30	female	Giebelstadt	several times a week	2
sts02	18-30	male	Würzburg	daily	2
sts03	18-30	female	Würzburg	daily	1
sts04	18-30	female	Würzburg	daily	6
sts05	18-30	female	Würzburg	daily	1
sts06	18-30	female	Würzburg	daily	1
sts07	18-30	female	Würzburg	several times a week	1
sts08	18-30	female	Würzburg	daily	3
sts09	18-30	female	Würzburg	daily	7
sts10	18-30	female	Würzburg	daily	2
sts11	31-45	female	Höchberg	daily	2

sts12	18-30	female	Würzburg	daily	4
sts13	18-30	female	Würzburg	daily	3
sts14	18-30	female	Würzburg	daily	1
sts15	18-30	male	Würzburg	daily	1
sts16	31-45	female	Nürnberg	less than weekly	1
sts17	18-30	male	Regensburg	never	3
sts18	31-45	male	Würzburg	daily	2
sts19	18-30	female	Würzburg	daily	3
sts20	18-30	female	Würzburg	daily	3
sts21	18-30	male	Würzburg	daily	3
sts22	31-45	female	Würzburg	daily	1

8 Part B: Environmental Indicators, Spatial Support, and Clustering Robustness

9 This Part Documents the methodological details underlying the environmental baseline analysis used in the
10 main paper. The purpose of the environment layer is to provide a robust, objective reference for comparison
11 with experiential data, not to introduce a novel clustering methodology. Consequently, technical details are
12 reported here to ensure transparency and reproducibility.

13 Built form and street geometry indicators were derived from OpenStreetMap (OSM) building footprints and
14 road centre lines, complemented where necessary by cadastral cross-sections. Three-dimensional building
15 height information was obtained from Level of Detail 2 (LOD2) building models. Vegetation indicators were
16 computed from atmospherically corrected Sentinel-2 imagery, and functional context was operationalised
17 using harmonised OSM points of interest (POIs).

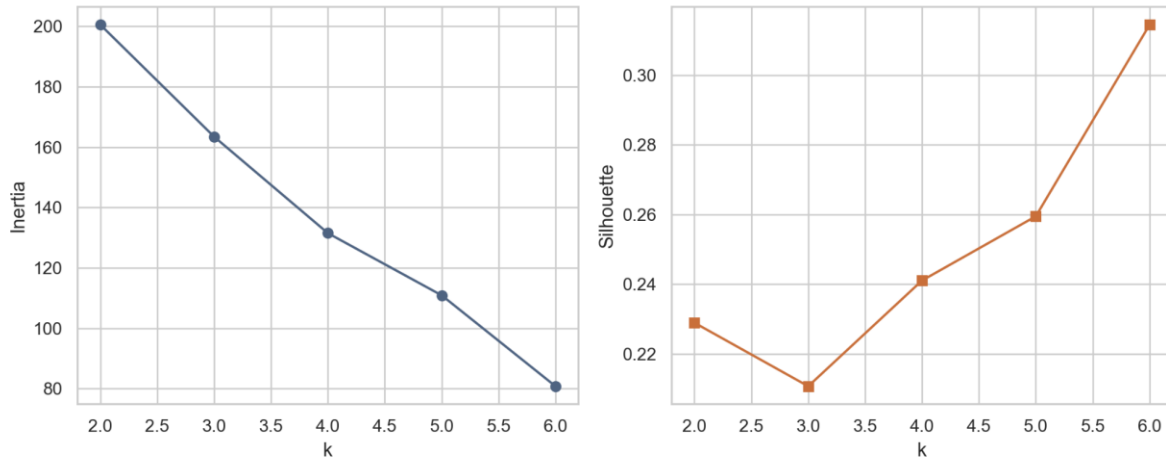
18 The environment layer is used exclusively as an objective structural baseline. It defines the environmental
19 envelope within which experiential data are interpreted in Levels 0–2 of the framework. Detailed
20 interpretation of lived experience, attention, and sense-making is intentionally reserved for subsequent levels
21 and is not derived directly from environmental indicators alone.

22 *B.1. Environmental Indicators and Spatial Support*

23 Environmental indicators characterising built form, street morphology, openness, vegetation, and functional
24 context were computed using street-centred circular buffers at multiple spatial scales (25 m, 50 m, and 90 m).
25 This multi-scale design accounts for differences between immediate street-level conditions and slightly
26 broader neighbourhood contexts. Variables included, among others, building density and height statistics, sky
27 view factor (SVF), vegetation indices (NDVI), openness measures, and point-of-interest (POI) density. All
28 indicators were z-standardised prior to analysis to ensure comparability across different units and magnitudes.

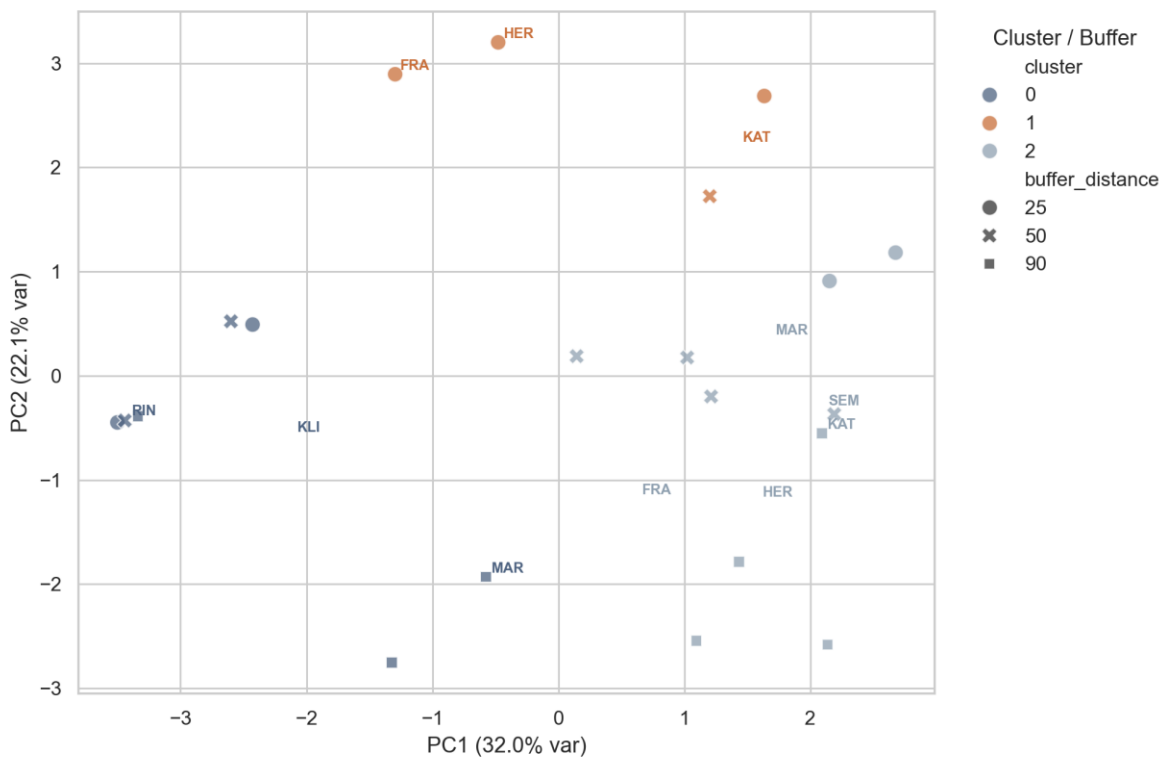
29 *B.2. Clustering Procedure and Model Selection*

30 Clustering was performed on the standardised environmental dataset using k-means. Model selection was
31 guided by the elbow criterion (inertia) and silhouette diagnostics across candidate solutions. A three-cluster
32 solution was retained as the canonical typology because it provided a parsimonious and interpretable
33 representation of structural variation while maintaining adequate separation quality. Diagnostic plots
34 supporting this choice are shown in Figures B1 and B2.



35

36 **Figure B1.** K-means model selection diagnostics for the environment-layer clustering. Notes: Left panel shows
 37 inertia (within-cluster sum of squares) by number of clusters k. Right panel shows mean silhouette score by k.
 38 The diagnostics support selection of k = 3 as the canonical solution used for the environment typology. Source:
 39 Own figure.



40

41 **Figure B2.** PCA biplot of environment-layer site profiles in the space of the first two principal components (PC1
 42 and PC2; variance explained shown on axes). Notes: Points are labelled by site and encoded by cluster
 43 assignment and buffer distance (25 m, 50 m, 90 m), illustrating separation between structural types and the
 44 effect of spatial support on site positioning. Source: Own figure.

45 To assess robustness, clustering was repeated across sensitivity scenarios that varied:

- 46 • spatial buffer size (25 m, 50 m, 90 m),
- 47 • feature subsets (excluding or including thematic groups of indicators),

- and standardisation and scaling choices.

Stability was further evaluated using pairwise co-assignment probabilities, capturing how consistently locations were assigned to the same cluster across scenarios. While some locations exhibit scale-dependent local heterogeneity, the results indicate that cluster membership patterns are not contingent on a single modelling configuration but are structurally stable.

While the overall three-cluster typology remains stable across scales, individual locations exhibit scale-dependent local heterogeneity, reflecting differences between immediate street segments and their surrounding context. These shifts are interpreted as indicators of spatial heterogeneity rather than instability of the clustering procedure itself. Scale-dependent site positioning and robustness of the environmental typology are illustrated in the PCA biplots shown in Figures A2 and A3.

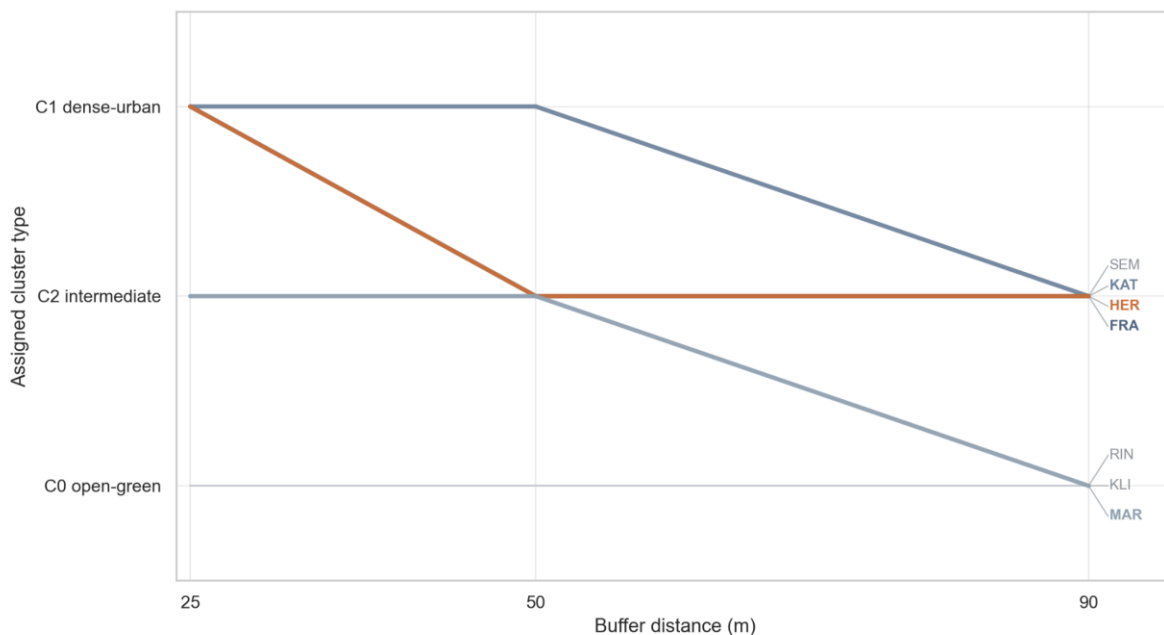


Figure B3. PCA biplot of environment-layer site profiles in the space of the first two principal components (PC1 and PC2; variance explained shown on axes). Notes: Points are labelled by site and encoded by cluster assignment and buffer distance (25 m, 50 m, 90 m), illustrating separation between structural types and the effect of spatial support on site positioning. Source: Own figure.

Part C: Level 0 – Aggregated Self-reports and Rating Analysis

This Part Documents the methodological details underlying Level 0 (L0) of the analytical framework. L0 constitutes the first experiential layer and serves as a diagnostic screening interface between the environment-derived structural baseline and subsequent attention- and interpretation-based analyses.

Level 0 is explicitly designed as a screening layer. It identifies locations where experiential similarity can be assumed and those where latent differences may exist despite structural comparability. However, L0 does not resolve which visual, acoustic, or situational cues shape these judgements, nor how different groups weight or interpret them.

Consequently, interpretation of attention, meaning, and group-specific relevance is intentionally deferred to Level 1 and Level 2, where documented cues and participant accounts are examined in greater detail. L0 thus defines where further analytical investment is warranted, rather than providing explanatory insight itself.

C.1. Self-reported Measures and Aggregation

At Level 0, participant evaluations were collected using standardised self-report items capturing overall experience at each survey location, which was assessed using 16 bipolar items rated on five-point Likert scales (Table B1). Sound-related items followed the bipolar descriptors established in soundscape research (Axelsson et al., 2010) which structures auditory experience along dimensions such as eventful–uneventful, exciting–monotonous, pleasant–unpleasant, and calm–chaotic. Wellbeing items captured momentary affective states based on the Warwick-Edinburgh Scale (Tennant et al., 2007), including anxious–safe, stressed–relaxed, depressed–cheerful, detached–attached, and exhausted–excited. Location-related items were drawn from constructs of environmental experience used in urban quality-of-place research (Baumann & Brooks-Cederqvist, 2023; Nasar, 2008), covering central–remote, accessible–inaccessible, clean–polluted, green–built-up, quiet–busy, relaxing–disturbing, and crowded–uncrowded. In addition to domain-based analyses, all 16 items were combined into an Integrated Place Appraisal.

Table C1. Questionnaire items, domain assignment, and pole coding (1 = left pole, 5 = right pole).

Item (variable)	Domain	1 (left pole)	5 (right pole)
sound_Eventful-Uneventful	Acoustic Appraisal	Atmosphere Eventful	Uneventful
sound_Exciting-Monotonous	Acoustic Appraisal	Atmosphere Exciting	Monotonous
sound_Pleasant-Unpleasant	Acoustic Appraisal	Atmosphere Pleasant	Unpleasant
sound_Calm-Chaotic	Acoustic Appraisal	Atmosphere Calm	Chaotic
wellbeing_Anxious-Safe	Personal Wellbeing State	Anxious	Safe
wellbeing_Stressed-Relaxed	Personal Wellbeing State	Stressed	Relaxed
wellbeing_Depressed-Cheerful	Personal Wellbeing State	Depressed	Cheerful
wellbeing_Detached-Attached	Personal Wellbeing State	Detached	Attached
wellbeing_Exhausted-Excited	Personal Wellbeing State	Exhausted	Excited
location_uncrowded-crowded	Perceived Spatial Context	uncrowded	crowded
location_accessible-unaccessible	Perceived Spatial Context	accessible	unaccessible
location_central-remote	Perceived Spatial Context	central	remote
location_green-built-up	Perceived Spatial Context	green	built-up
location_clean-polluted	Perceived Spatial Context	clean	polluted
location_quiet-busy	Perceived Spatial Context	quiet	busy
location_relaxing-disturbing	Perceived Spatial Context	relaxing	disturbing

C.2. Similarity-based Comparison

For each site and item, responses were summarised as interval-based profiles to represent sampling uncertainty and intra-site variability. Pairwise similarity between sites was quantified using an Intersection-over-Union (IoU) measure computed for the corresponding intervals and aggregated within domains and across all items. IoU values range from 0 (no overlap) to 1 (complete overlap), providing an indicator of the extent to which ratings converge or diverge. Summary statistics and aggregated rating profiles for all clusters

are reported in Table B2. The purpose of this aggregation was not to infer causal effects of individual environmental features, but to capture overall experiential appraisal under comparable structural conditions. This screening-based approach is used to guide subsequent, higher-resolution analysis and should not be interpreted as causal explanation of person-place relationships.

Table C2. Cluster-level medians of Likert ratings by domain and for the integrated appraisal (all items).

env_cluster	Acoustic Appraisal	Atmosphere Personal Wellbeing State	Perceived Context	Spatial Integrated Appraisal	Place
0	2.75	3.80	2.428571	2.8750	
1	2.25	3.70	2.928571	2.9375	
2	3.00	3.20	3.142857	3.1250	

Part D: Level 1 – Acoustic and Visual Attention as Perceptual Evidence

This appendix provides the conceptual and methodological background for Level 1 (L1) of the analytical framework. Level 1 operationalises perceptual attention as an intermediate layer between aggregated appraisal (Level 0) and interpretive sense-making (Level 2). The purpose of this appendix is to document how visual and acoustic attention traces were used to identify design- and regulation-relevant cues, without overloading the main text.

Level 1 builds on the assumption that differences in urban experience often originate not at the level of overall appraisal, but at the level of what is noticed in situ. Rather than treating perception as a passive response to environmental structure, L1 conceptualises attention as an active selection process through which certain acoustic and visual elements enter the experiential foreground. By documenting these selections, L1 provides an empirical basis for identifying actionable elements that may constitute potential levers for urban design and zoning.

Within the overall framework, Level 1 serves as the first empirical bridge between diagnostic screening (Level 0) and interpretive analysis (Level 2). By documenting which visual and acoustic cues are perceptually present under structurally comparable conditions, L1 reduces environmental complexity to a set of perceptually salient, potentially actionable elements. Importantly, Level 1 does not assess how these cues are weighted, valued, or negotiated by different groups. Questions of meaning, relevance, conflict, and planning assessment are therefore intentionally deferred to Level 2.

D.1. Acoustic Attention and Soundscape Perspective

Short audio recordings produced by participants were analysed as indicators of experienced acoustic environments. This approach follows the soundscape tradition, which understands urban sound not merely as physical noise exposure, but as a perceptual and contextual phenomenon shaped by listening, interpretation, and situational meaning (Axelsson et al., 2010). In line with this perspective, acoustic recordings in L1 were not treated as psychoacoustic measurements, but as attention traces capturing which sound elements participants chose to document when present on site.

Descriptive acoustic source profiles were derived by first resampling raw audio to 32 kHz and band-pass filtered (100 Hz to 20 kHz) to reduce handling noise and sensor artefacts. Sound event detection was performed using Pretrained Audio Neural Networks (PANNs), a convolutional model trained on AudioSet and suitable for general sound event classification (Kong et al., 2020). Model outputs were aggregated at two semantic levels: At the macro level, predicted events were categorised as anthropophony, biophony, geophony, or as an additional other category, following established soundscape-ecology framing (Pijanowski

et al., 2011). At the micro level, the most prevalent subclasses were examined to document the specific sources contributing to the macro composition and to avoid inference based solely on coarse grouping. Where required for pair-specific interpretation, psychoacoustic descriptors including roughness, sharpness, and tonality were computed as complementary summaries (Fastl & Zwicker, 2007).

The analytical focus therefore lies on experienced sound characteristics and their variation across sites, rather than on absolute sound levels or compliance with regulatory thresholds. This makes the acoustic component of L1 particularly relevant for planning questions related to spatial atmosphere, comfort, and the experiential quality of public space.

D.2. Image-based Analysis of Visual Attention

Visual attention was analysed using self-selected photographs captured by participants during their site visits. Images were pre-processed to standardise acquisition conditions and analysed using a combination of feature extraction and clustering to identify recurrent visual motifs at the image level (Figure D1). For each image, features were derived to quantify the proportions of green and grey pixels as proxies for vegetation and built surfaces, following established approaches to street-level greenery and scene-based urban image analysis (Li et al., 2015; Yang et al., 2015). Visual complexity was represented using entropy-based indicators, and spatial heterogeneity of green and grey structure was captured using variability metrics, including inequality-based measures, to distinguish more uniform from more patchy scene compositions. Object cues were additionally derived from detection counts of persons, faces, animals, cars, and other vehicles as descriptive indicators of social presence and mobility-related elements in the recorded scenes.

Image-level feature vectors were standardised and grouped using k-means clustering to derive a set of representative visual motifs. Site-level visual profiles were subsequently computed as composition vectors, defined by the proportions of images assigned to each motif per location. Similarity between site-level profiles was assessed using cosine similarity, allowing comparison of internal compositional structure independently of the number of images per site.

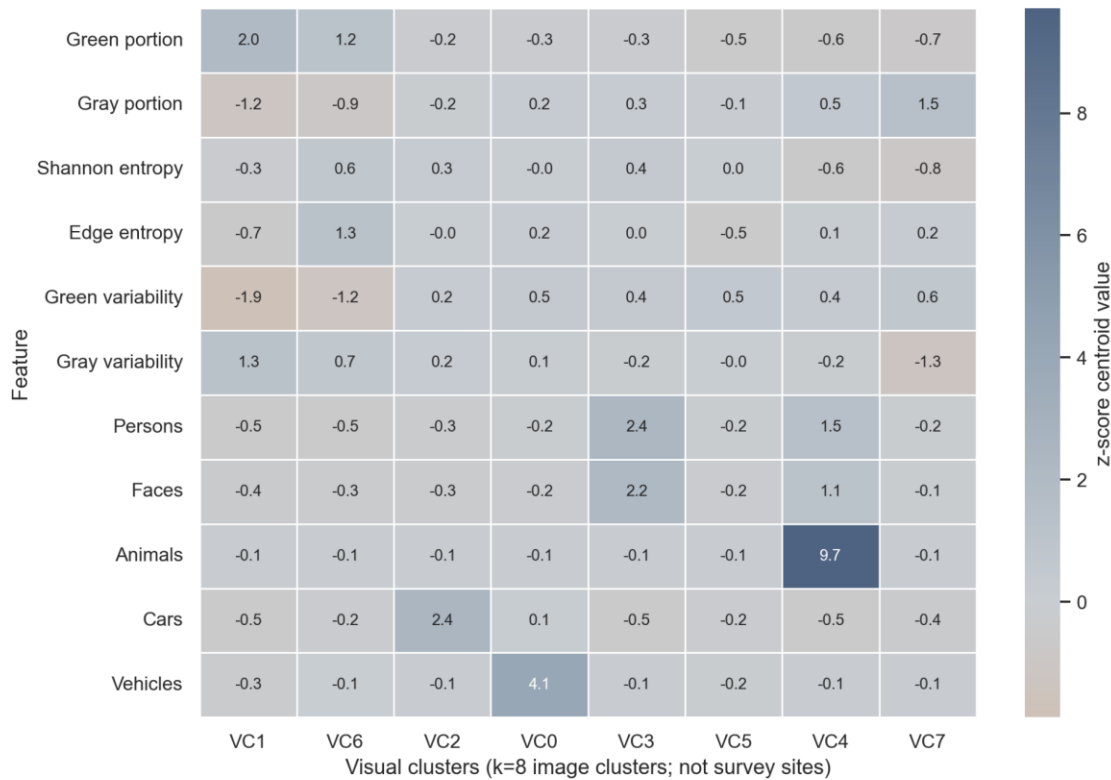


Figure D1. Standardised centroid values for the eight image-level visual clusters derived from participant photographs. Notes: Rows represent visual features and object cues (green and grey proportions, entropy metrics, variability metrics, and detection-based cues), and columns represent visual clusters (image motifs, not survey sites). Values are z-score centroids, indicating which features characterise each motif. Source: Own figure.

This analysis provides fine-grained insight into visual attention patterns and supports the interpretation of perceptual similarity and difference across locations. However, due to its methodological complexity, the image-based analysis is reported exclusively in this appendix. Its result largely align with the sound-based analysis indicating that composition-level similarity can coexist with measurable differences in specific visual properties, supporting the interpretation that visually similar central sites may still differ in localised cues relevant to design and regulation, such as the balance of green versus built surfaces, variability in built texture, and the presence of mobility-related objects (Figure D2). The image-based analysis is thus used in the main text as supporting evidence for attention-based differentiation at Level 1.

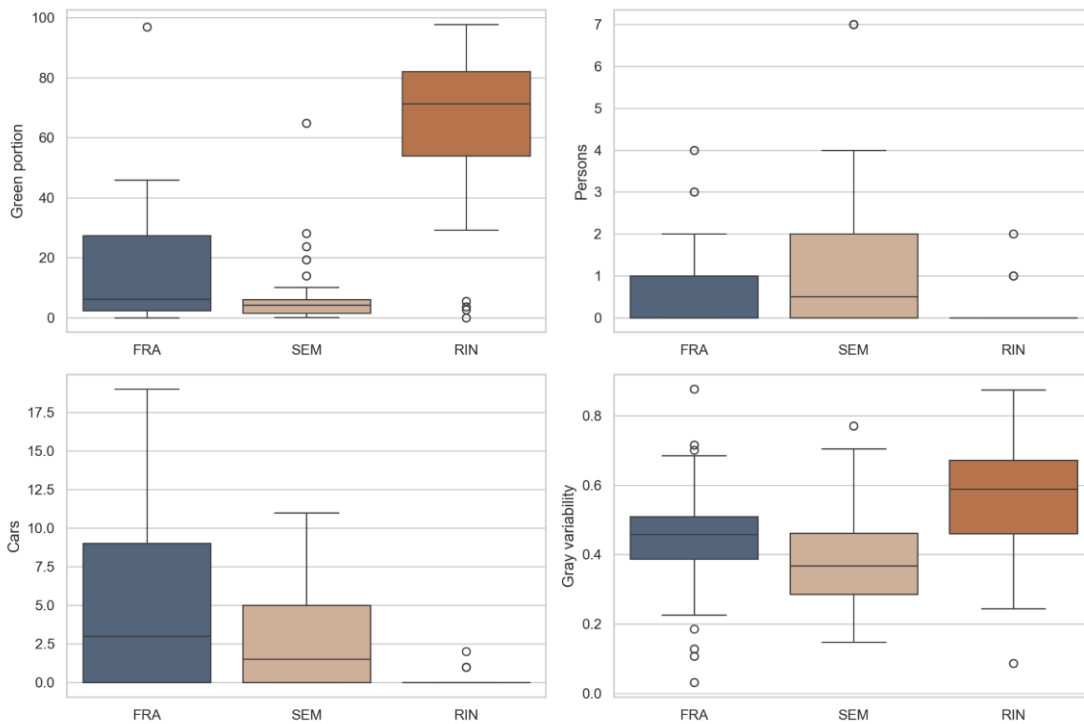


Figure D2. Distributions of selected visual features for FRA, SEM, and RIN, shown as boxplots across the images captured at each site. Notes: Features illustrate interpretable cue-level differences, including green portion, object presence proxies (persons, cars), and structural variability (grey variability), complementing composition-based similarity by highlighting feature-specific contrasts. Source: Own figure.

Part E: Level 2 – Geolinguistic Analysis

This Part Documents the analytical procedures underlying Level 2 (L2) of the framework. Level 2 focuses on sense-making and interpretive practices, examining how pedestrians articulate experience, relevance, and evaluation in their own words. The purpose of this appendix is to ensure transparency of the computational and interpretive steps used to analyse textual material, while keeping the main text focused on planning-relevant insights.

Within the overall framework, Level 2 provides the highest resolution insight into pedestrian experience by revealing how environmental and perceptual cues are transformed into arguments, priorities, and evaluative positions. The methods documented here support the construction of salience maps and compromise matrices discussed in the main text, but do not replace planning judgement. Instead, they prepare the transition to Level 3, where evidence is negotiated and translated into implementable measures, operational rules, and governance arrangements.

E.2. Transcription, Topic Modelling, and Sentiment Assessment

Participants' spoken descriptions of place and experience were recorded in situ and automatically transcribed using a multilingual speech recognition system (OpenAI, 2022). Transcriptions were manually reviewed to correct recognition errors and, where necessary, enriched to provide a more detailed representation of communicative devices, such as filled pauses or discourse makers. Each transcription was segmented into short, meaningful units, serving as the basis for subsequent topic and sentiment analysis.

To identify recurrent themes in participants' accounts, we applied the BERTopic framework (Grootendorst, 2022), a topic modelling workflow that combines neural language embeddings with density-based clustering.

The number and structure of topics were determined empirically using density-based clustering, followed by k-means aggregation. Resulting clusters were manually inspected, and thematic labels were assigned based on representative text units.

Topics associated with the same observation were subsequently concatenated to allow place-level comparison of recurring themes. This approach enables identification of salient experiential domains (e.g., built environment, traffic, tranquillity, points of interest) while preserving contextual variation in how these domains are articulated.

In addition to thematic content, textual units were analysed for evaluative polarity using a multilingual sentiment classification model (Yuan, 2023) that assigns probabilities for positive, neutral, and negative sentiment. For each unit, the polarity with the highest probability was retained. Frequencies of polarities were then compared across sites and topics.

This step was used to complement thematic interpretation by highlighting potential discrepancies between automated sentiment scores and discursively constructed meaning, rather than as a stand-alone measure of experience quality. To examine evaluative orientation in greater depth, a subset of observations was analysed using an appraisal-theoretical framework (Martin & White, 2005), distinguishing between:

- affect (expressions of feelings),
- appreciation (evaluations of places and phenomena),
- judgement (evaluations of behaviour and conduct).

This interpretive layer supports the identification of planning-relevant tensions, such as divergent constructions of tranquillity, noise, or activity, that are not consistently captured by sentiment polarity alone.

E.3. Cross-level Case Selection and Decision Matrix

To support the selection of site pairs for in-depth analysis at Level 2, we constructed a cross-level decision matrix that integrates evidence from the environment layer, Level 0, and Level 1. For each unique unordered site pair, the matrix combines structural similarity (environment cluster membership), robustness of co-assignment across sensitivity scenarios, overlap in integrated place appraisal at Level 0, and visual profile similarity at Level 1.

The resulting decision matrix is reported in Table D1. It provides a transparent, rule-based screening mechanism to identify site pairs that are structurally comparable, similarly appraised, and perceptually similar, yet still warrant interpretive analysis due to remaining differences. This procedure ensures that Level 2 analysis focuses on analytically informative cases rather than extreme contrasts.

The matrix does not constitute an optimisation or decision algorithm, but a planning-oriented screening tool that formalises the balancing logic across analytical levels. In this sense, it operationalises the transition from multi-level evidence to targeted sense-making analysis at Level 2.

Table E1. Cross-level decision matrix for selecting site pairs for Level 2 analysis. For each unique unordered site pair, the table reports whether both sites share the same environment cluster, the structure co-assignment probability across sensitivity scenarios, Level 0 integrated overlap (IoU), Level 1 visual profile cosine similarity, and the resulting Level 2 recommendation flag.

site_1	site_2	same_environment_ cluster	co_assignment_ probability	integrated_place_ appraisal_iou	visual_profile_ cosine	level_2_ recommended
FRA	SEM	Yes	0.583	0.641	0.885	True
KLI	SEM	No	0.083	0.714	0.426	False
KLI	FRA	No	0.125	0.573	0.513	False
KAT	HER	No	0.875	0.545	0.832	False
HER	SEM	Yes	0.792	0.518	0.664	False
HER	FRA	Yes	0.583	0.455	0.717	False
HER	MAR	Yes	0.917	0.391	0.589	False
HER	KLI	No	0.125	0.405	0.678	False
MAR	SEM	Yes	0.875	0.508	0.758	False
MAR	FRA	Yes	0.792	0.498	0.737	False
MAR	RIN	No	0.000	0.259	0.082	False
RIN	SEM	No	0.000	0.173	0.114	False
RIN	FRA	No	0.000	0.138	0.176	False
RIN	KLI	Yes	0.958	0.349	0.210	False
RIN	KAT	No	0.000	0.180	0.040	False
RIN	HER	No	0.000	0.233	0.063	False