SOUND AS SCORE IN ARCHITECTURAL PLANS: A PATH-BASED NOTATION WORKFLOW USING AMBISONICS + 360° VIDEO

DI Nicolaj Kirisits

University of applied arts Vienna
Institute fine arts & media art
Department digital arts
nicolaj.kirisits@uni-ak.ac.at

ABSTRACT

Architecture is experienced in physical space but is represented as a non-temporal object in 2D plans, whereas sound is framed as a temporal phenomenon in the time–frequency domain. Bridging this divide requires conceiving architecture as event and sound as a spatial, architectural element. We propose a plan-centric notation that “folds” listening events along a walked path into the plan for re-enactment, comparison, and design discussion. Building on the *Soundspace-Sequenz*—an improvisational listening walk documented for later analysis—our method combines first-order Ambisonics (FOA) audio with 360° video. Python-based processing extracts descriptors (amplitude envelope, spectral complexity, tonal segments, onsets) and maps time to walked distance using video-assisted position checks. The resulting path-line is registered to the route and embedded in the plan; interpretive icon annotations (atmospheres, sound effects, soundmarks) are added manually under an explicit rule set. A case study in an urban public space shows how the plan-embedded score functions as a memory-stable record of listening and an actionable artifact for communicating sonic design concerns. Unlike static sound maps that merely locate sound, the notation documents listener–space interaction as constitutive of sonic reception.

INTRODUCTION

Architectural design workflows largely operate through 2D representations of physical space (plans and sections), whereas sound is commonly treated as an engineering constraint or reduced to aggregated soundscape indicators. While such indicators effectively describe larger urban spaces (macro: 10–100 m; macro+: >100 m), they often miss meso- and micro-scale phenomena crucial to architectural perception: thresholds, courtyards, passageways, edges, and the rhythmic coupling between locomotion and shifting audible situations (meso: 0.1–10 m; micro: <0.1 m).

This paper introduces a workflow that makes such situations legible in architectural plans—not by rendering sound as a static architectural element, as in many soundscape representations [1] and early sound-art visualizations [2] [3], but as plan-embedded scores of situated listening. The contribution is a plan-native, derived score that couples perceptual events to route geometry through time-to-distance transfer, enabling repeatable reading and comparison within architectural drawing conventions.

The plan thus becomes a medium in which time is folded into space: a listening performance is recorded and converted into a path-based representation, where descriptor layers and annotated events are arranged along the traversal. The method is motivated by three observations: (1) architectural listening is enacted, not merely observed; (2) a traversal is both time and distance and can therefore be scaled and drawn; and (3) if a plan shows when and where sonic features occur along a path, the drawing can operate like an audio score for re-listening, comparison, and design discussion.

Beyond documenting “where sound is,” the goal is to represent how sound is co-produced through movement, orientation, and architectural affordances such as edges, openings, reflective surfaces, and thresholds. The plan is therefore treated as a coordinate system for a listening performance: changes in pace, stopping, and turning are part of the sonic situation and should remain readable in the notation.

We ask: (i) How can a spectral-temporal description of sound be translated into a plan-native representation without losing sequential and embodied aspects of listening? (ii) What minimal descriptor set remains legible for designers while enabling comparison across traversals? (iii) How can the output function both as documentation and as a re-performable audio score?

BACKGROUND AND RELATED WORK

From soundscape maps to scored traversals

Soundscape mapping and standardized vocabularies offer valuable frameworks for describing acoustic environments [4], but typical outputs (contours, indices, averages) can downplay sequential, embodied listening. For architectural purposes the challenge is to represent not only where certain levels occur but how sonic situations unfold while listeners interact with space by moving, stopping, accelerating, or turning.

Soundwalk and listening-walk practices address this gap if they treat listening as situated action rather than as passive exposure. However, their outputs often remain narrative (field notes) [5] or media-based (recordings) [6] and are difficult to integrate into plan-centered design workflows. A plan-embedded, path-based score keeps route choice, speed changes, and thresholds visible, supporting side-by-side comparison between repeated traversals and making the interaction between listener and built form part of the representation.

Notation, plans, and folding

The technique of the *revolved section* (German: *Klappschnitt*) inserts sectional information into plans to clarify spatial change. Here a similar logic is applied: not a building section, but a perceptual event sequence is folded into the plan, preserving the plan as a shared design medium while adding a score-like time/distance dimension.

Folding is not merely a graphic trick but a way to establish commensurability between domains. Here, plan coordinates (meters) are connected to the spectral-temporal domain (seconds and descriptor values) via the walked path. This preserves the plan as a shared medium while avoiding the illusion that sound can be captured as a static layer independent of traversal and attention.

Soundspace-Sequenz: listening as architectural method

The workflow is grounded in the concept of the *Soundspace-Sequenz*: an improvisational traversal guided by *architectural listening*—an embodied, situated mode of attention that renders sound-based spatial elements legible as part of architectural form—and understood as a compositional act within a site. A single sequence is a singular event; a space can be considered a topology of many possible sequences. From such performances, a derived score can be generated and returned to representation as a re-performable artifact.

Audio scores, listening guides, and derived notation

Within expanded notions of scores—including audio scores, listening guides, and representational formats that organize attention—a score can mediate a performance even when the primary activity is listening. From this perspective, the *Soundspace-Sequenz* aligns with practices that treat listening as performative and generative: it produces a structured sequence of perceptual events that can be made amenable to notation. Related work also points toward derived forms of scoring, where notations are generated from recorded traversals rather than authored in advance, yielding re-enactable artifacts. Situated within architectural representation, such derived scores connect performative, time-based experience to plan-based conventions, supporting critique and iteration while keeping rhythmic and atmospheric change legible along movement through space.

From visual path notation to audio-based scoring

Historical precedents for path-based notation can be found in architectural and urban design traditions that describe perception in sequence, such as serial-vision and route-based analyses (e.g., townscape approaches and "view-from-the-road" notations) [7], as well as practices that frame movement through space as a guided promenade. [8] Related lineages also emerge outside architecture: dance choreographers and notation pioneers developed systems to score movement and timing—codifying trajectories, pauses, and rhythmic transitions of bodies in space. [9] Across these precedents, the path becomes a representational device by registering changes in views, landmarks, edges, thresholds, and experiential rhythm as one moves. Yet what was typically "scored" remained overwhelmingly visual (or purely kinematic): they documented what can be seen, and sometimes how bodies move, while sonic phenomena—atmospheres, sound effects, and audio-based events co-produced through listener–space interaction—were largely absent or relegated to anecdotal description. The present work extends path notation beyond visual sequencing and movement coding by making sound-based spatial elements part of the ontology of architecture and notable within plan conventions, so that interaction with the built environment is documented not only as changing sightlines and motion, but also as changing listening conditions.

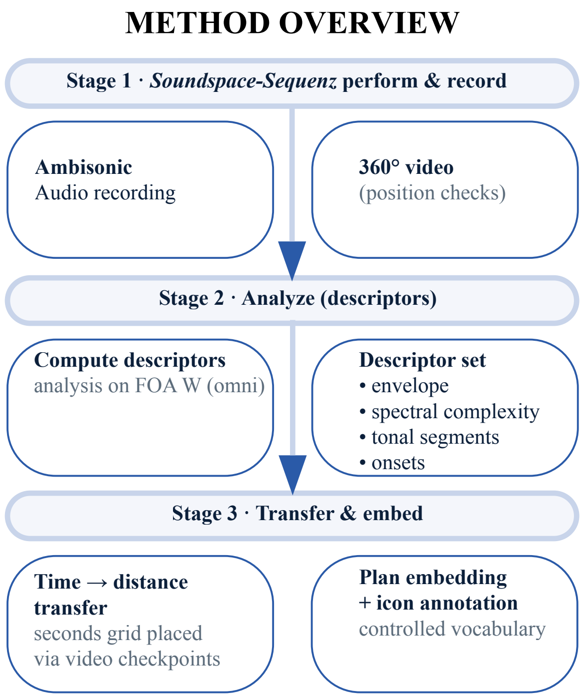

Figure 1. Workflow from traversal to plan-embedded score: FOA + 360° recording, descriptor extraction (FOA W), time-to-distance transfer via video checkpoints, and plan embedding with controlled icon annotation

The workflow has three stages: (1) perform and record, (2) analyze the recorded signal, and (3) transfer time to distance and embed computed descriptor layers in the plan, complemented by rule-based interpretive icon annotations.

.Stage 1: Perform and record (FOA + 360° video)

A *Soundspace-Sequenz* is conducted as an attentive, improvisational walk through the site. Because the practice demands sustained concentration, instrumentation and annotation are kept minimal during the traversal; measurement is deferred to post-processing. The walk is documented using a 360° camera and a first-order Ambisonics (FOA) microphone, enabling subsequent re-listening with spatial and visual context.
Recordings can be replayed either binaurally over headphones or in a multichannel loudspeaker environment. The 360° video provides a positional reference that supports reconstructing turning points, stops, threshold crossings, and other micro-situations, and enables periodic position checks for later alignment.
To ensure comparability across sessions, a basic metadata set is logged (time of day, weather, approximate activity level, and gain settings) and the route is noted. The video functions as a positional scaffold for linking micro-actions (e.g., stopping, turning, crossing thresholds) to audible transitions. This coupling is critical, as the same acoustic event may carry different architectural implications depending on whether it occurs while approaching, passing through, or lingering within a spatial condition.

Figure 2. Synchronized recording setup for a Soundspace-Sequenz: first-order Ambisonics (FOA) microphone and 360° video camera used for re-listening and positional verification.

Stage 2: Audio analysis (interpretable descriptor set)

For plan-readable descriptor layers within the score, we emphasize descriptors that remain interpretable without specialist signal-processing knowledge and that relate to perceived change: envelope (dynamic contour), spectral complexity (density of spectral peaks), tonal segments (voiced or musical tendency), and onsets (impulsive change points).
In the demonstrated implementation, analysis is run on the FOA W channel (omnidirectional component) to reduce directionality bias in descriptor estimation.
Descriptor settings are tuned for plan readability. We compute features on short analysis windows but aggregate them to a coarser time step (≈0.5–1 s) so that changes align with the pace of architectural actions such as walking, stopping, and crossing thresholds. Detection thresholds are set conservatively so the score highlights robust, repeatable features rather than incidental fluctuations of the audio signal.

Stage 3: Fold time into distance, then into plan

The analyzed descriptors are initially time-indexed. To embed them in the plan, time is mapped to walked distance by placing a seconds grid along the path using position checks from the 360° video. This makes accelerations and slowdowns visible through unequal spacing. The scaled path-line is then geometrically aligned with the route and inserted into the plan. In addition to the computed descriptor layers, interpretive icon annotations (atmospheres [10], sound effects [11], soundmarks [12]) are added using a controlled vocabulary.
When embedding the score into a plan, readability is treated as a design constraint. In practice, a contextual plan view is paired with one or more insets where the path-line is shown at notation scale so that descriptor bands and icon annotations can be read without excessive compression. This mirrors musical practice, where an overall form and detailed passages are both necessary for interpretation, and helps maintain the plan's role as a shared medium rather than turning the score into a separate analytic diagram.

IMPLEMENTATION DETAILS

We distinguish between computed descriptor layers (signal-derived and reproducible) and interpretive icon annotations (rule-based and manually assigned), which together form the plan-embedded score.

Signal format and analysis channel choice

A FOA (1st Order) recording contains four channels (X, Y, Z directional components plus W omnidirectional). For descriptor stability across head turns and source directions, the W channel is used as the primary analysis input. The library Essentia was used for the analysis. [13]

Descriptor extraction

Envelope.

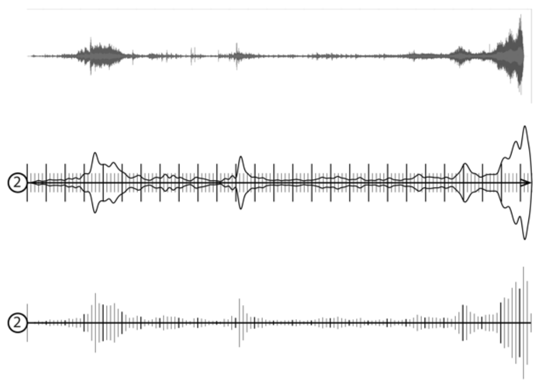

Figure 3. Descriptor example: amplitude envelope (dynamic contour) extracted from FOA W and rendered as a plan-readable layer along the seconds grid.

The envelope represents the loudness trajectory; a seconds grid can be visually weighted by the envelope so that time ticks become salient where the signal is strong.

Spectral complexity.

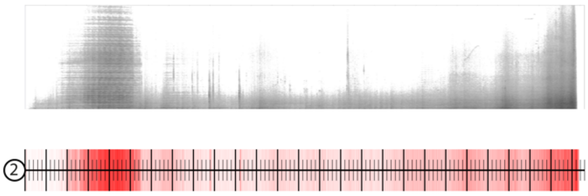

Figure 4. Descriptor example: spectral complexity (density change) rendered as an intensity layer along the seconds grid.

Spectral complexity indicates whether the spectrum is narrow/simple or broad/dense. Results can be drawn as a one-dimensional intensity layer along the path.

Tonal segments.

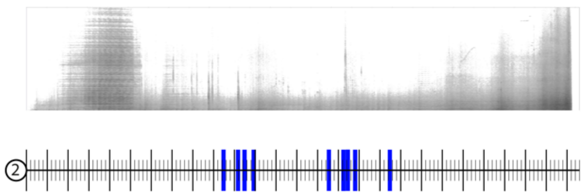

Figure 5. Descriptor example: tonal segments (voiced/tonal material) estimated from spectral flatness and pitch confidence after thresholding.

Tonal segments are estimated from spectral flatness and pitch confidence; thresholding isolates segments likely dominated by voiced speech or musical/tonal material.

Onsets.

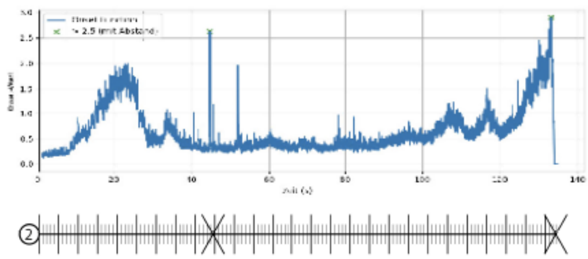

Figure 6. Descriptor example: strong onsets after thresholding, rendered as discrete marks indicating impulsive change points.

Onsets are extracted as peaks indicating sudden change. After thresholding, strong onsets appear as discrete marks along the path, suggesting rhythmic articulation.

Synchronization and temporal alignment

Since the notation links descriptor time stamps to positions in the plan, alignment between audio and video is treated as a dedicated step. The workflow uses the shared recording start as reference and verifies alignment during re-listening using visible/aural checkpoints such as turns, door crossings, or short impulse sounds. If drift is observed over longer takes, alignment is corrected at a small number of checkpoints so that time-to-distance transfer remains consistent.

Mapping time to path distance

A sequence is simultaneously a duration and a walked length. Mapping time to distance is not assumed linear; instead, the seconds grid is positioned via video observation. This exposes speed changes and supports interpretation: Tick spacing exposes speed variation; onset and tonal layers can then be interpreted against this movement context.

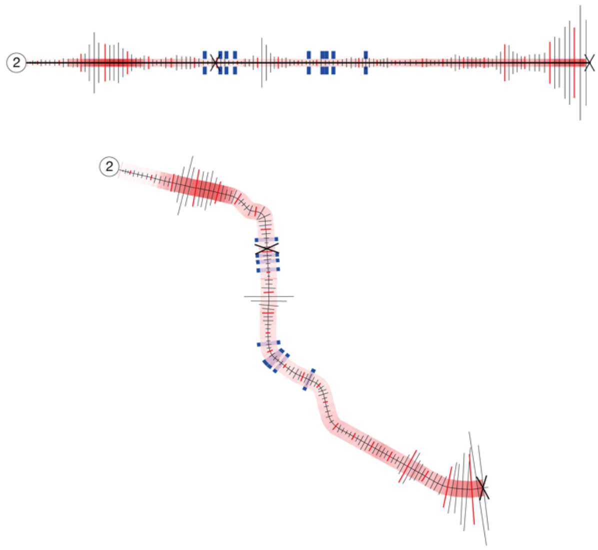

Figure 7. Seconds grid aligned to observed positions (video checkpoints). Tick spacing encodes walking speed; in this example, near-uniform spacing indicates largely steady pace.

Icon layers: atmospheres, effects, soundmarks

Atmospheres are treated as perceivable segments with transitions along the path and annotated using a controlled vocabulary (Figs. 8–10). Each atmosphere segment is labeled with exactly three icons—one for functional setting (e.g., shopping street/park/residential), one for social use (e.g., intimate/communicative/representational), and one for an acoustic descriptor (e.g., open / warm / reverberant / damped)—to maintain legibility and comparability. Sound effects are marked at perceived positions as localized events or short spans using the sound-effect categories listed in [11] (e.g., filtering/occlusion, resonance, reverberant bloom). Soundmarks are placed manually and categorized in line with ISO 12913 categories [1] when they function as identifiable, recurring anchors. Additional effect and soundmark icons may be added beyond the atmosphere label, but only when events remain salient and repeatable during re-listening, keeping the score readable in architectural drawing contexts.

Figure 8. Atmosphere icon set (examples) used to annotate perceptual background segments along the traversal.

Figure 9. Sound-effect icon set (examples) used to annotate localized perceptual transformations (e.g., filtering, resonance, reverberant bloom)

Figure 10. Soundmark icon set (examples) used to annotate identifiable, recurring elements that support orientation.

CASE STUDY: URBAN PUBLIC SPACE TRAVERSAL

Site and recordings

To demonstrate the workflow, multiple *Soundspace-Sequenzen* were recorded in an urban public space. The recording session took place in Spring 2025 in stable weather conditions. Several sequences were captured to show variation across routes and micro-situations.

The site was selected because it contains multiple acoustically distinct micro-zones within short walking distance: a relatively open area, edges exposed to traffic, semi-enclosed corners, and transitional passages where occlusion and reflection change quickly. Such conditions make threshold effects audible and therefore suitable for demonstrating how a scored traversal can capture alternating phases of exposure and shelter that would be flattened by static averaging.

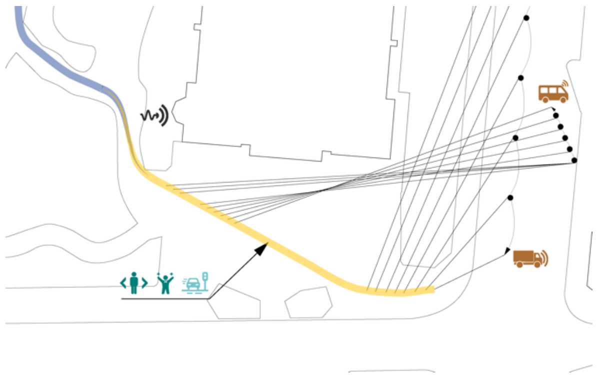

Figure 11. *Soundspace-Sequenz 2 (detail): plan-embedded score with interpretive layers (atmospheres, effects, soundmarks).*

Worked example: Sequence 2

Sequence 2 is used to illustrate the analysis and folding steps. Computed descriptor layers provide the plan-readable analytic backbone of the score: envelope as dynamic contour; spectral complexity as density transition; tonal segments as voiced or musical moments; and strong onsets as discrete rhythmic articulations. Because the seconds grid is positioned by video-based checks rather than constant-speed assumptions, the drawing can register accelerations, slowdowns, and pauses as changes in tick spacing—information that supports interpretation and comparison across repeated traversals. In the worked example, walking speed was largely steady; accordingly, dense clusters of onsets are read primarily as brief impulse-rich passages rather than as dwell-time effects, while tonal-segment density indicates where voiced/tonal material becomes sustained or recurrent as auditory conditions change along the route (e.g., masking or enclosure), with dense bands suggesting increased presence of voiced sources or acoustically supportive micro-situations and sparse segments indicating predominantly broadband or diffuse backgrounds. The folded representation makes these distinctions visible and supports more precise discussion of where attention was drawn and why.

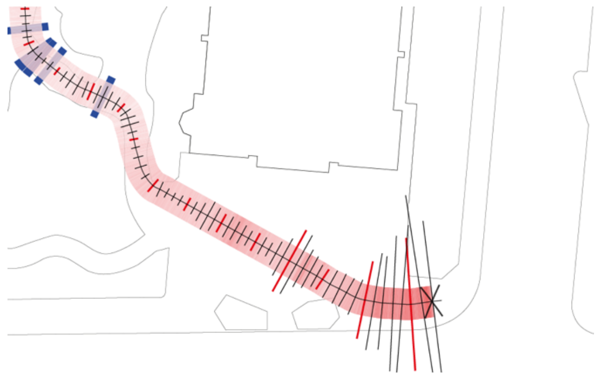

Figure 12. *Soundspace-Sequenz 2 (detail): computed descriptor layers (envelope, spectral complexity, tonal segments, onsets) folded into the plan.*

Plan integration and annotation

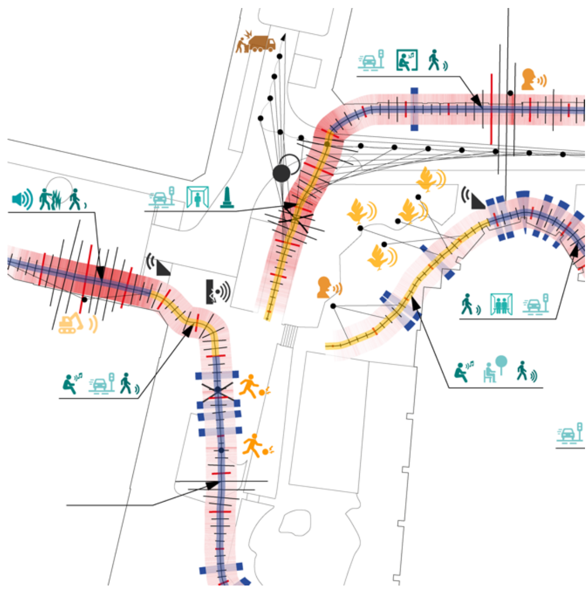

Figure 13. Soundspace-Sequenzen 1–3 (detail): comparative plan-embedded scores enabling side-by-side reading across traversals.

After folding, the path is aligned to match its geometry and inserted into the plan. Atmosphere segments are marked, sound effects are plotted, and soundmarks are positioned either on or near the path. The resulting plan-embedded score drawing functions as an audio score in two senses: as a listening guide for re-walking and as an analytic artifact for comparison across sequences.

DISCUSSION

Why this counts as sound as score

Unlike conventional scores that prescribe sound production, the proposed score is derived from situated listening and returned to representation as a guide for re-enactment and discussion. Computed descriptor layers are reproducible from the recordings, while icon annotations capture perceptual categories through rule-based interpretation. The site becomes the instrument, the moving listener the performer, and the notation encodes rhythm and sonic transitions along the path.

The workflow is not intended as a predictive acoustic model or a replacement for compliance metrics; instead, it provides a plan-native representation that supports architectural critique and iteration. Its validation criterion is therefore representational: whether repeated traversals yield scores that remain interpretable, comparable, and useful as guides for re-listening and re-walking within drawing conventions.

Because the score preserves context dependence—movement, orientation, and occupancy—differences between traversals become legible as design-relevant variation rather than being averaged out. At the same time, the method trades spatial exhaustiveness for legibility: a minimal descriptor set and print-oriented smoothing stabilize the score but omit finer-grained directionality and source separation, leaving room for optional extensions when higher analytic resolution is required.

Strengths

• Plan-native communication: inserting the score into plan drawings lowers the barrier to discussing sonic qualities in architectural workflows.
• Interpretability: the descriptor set is intentionally small and legible, supporting comparative reading without black-box metrics.
• Rhythm as design-relevant: the score foregrounds the coupling of movement decisions and sonic transitions as an architectural quality.

In design settings, this plan-native format supports concrete use cases: documenting a baseline listening route during site analysis; comparing repeated traversals across times of day or after an intervention; and supporting review discussions where circulation and material choices can be linked to audible transitions without collapsing the conversation into purely quantitative compliance metrics.

Practical use in public space also raises ethical considerations: recordings may capture speech or identifiable persons in the 360° video. In many cases, sharing the plan-based score without distributing raw recordings is sufficient for communication, reducing exposure while preserving the representational benefits of the method.

Limitations and future work

• Manual annotation load: atmosphere and soundmark annotation remains labor-intensive and depends on expertise.
• Single-channel analysis simplification: using the W channel discards spatial directionality that could enrich the score; future work could add optional spatial layers while keeping readability.
• Scale: the workflow is most effective for micro and meso sites; macro-scale mapping should combine it with broader soundscape methods.

A productive direction is to keep the current layer set as a readable core and add optional modules when needed, such as directional energy measures derived from FOA components or confidence indicators for event localization. This would preserve accessibility for designers while enabling deeper analytic detail in research-oriented applications.

CONCLUSION

We presented a workflow that turns a recorded listening traversal into a path-based audio score embedded in architectural plans. Using FOA + 360° video documentation, a small set of interpretable audio descriptors, and a time-to-distance folding procedure, the method produces drawings that document experience and guide re-enacted listening. The approach extends sound as score from musical production toward architectural perception by scoring a way of hearing a site rather than merely mapping sound.

REFERENCES

[1] ISO / TS 12913-2:2019(E), *Acoustics - Soundscape - Part 2: Data collection and reporting requirements,* Geneva, 2019.

[2] B. Leitner, TON : RAUM, Köln: DuMont Buchverlag, 1978.

[3] M. Neuhaus, Drawings, Sound Works, Volume 2, Ostfildern: Cantz Verlag, 1994.

[4] J. Kang, F. Aletta, T. Oberman, M. Erfanian, M. Kachlicka, M. Lionello and A. Mitchell, "Towards soundscape indices," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen , Germany, 2019.

[5] M. R. Schafer, Ed., European Sound Diary, Burnaby, CANADA: A.R:C: Publications, Vancouver, 1977.

[6] S. Tanzer, "Soundcities," 1997. [Online]. Available: https://soundcities.com/info.php. [Accessed 07 01 2026].

[7] D. Appleyard, K. Lynch and J. R. Myer, The view from the road, Cambridge, Mass.: MIT Pr., 1966.

[8] G. Debord, "Theory of the Dérive," 1956. [Online]. Available: https://www.cddc.vt.edu/sionline/si/theory.html. [Accessed 14 03 2023].

[9] P. Talwar, *Notation systems in architecture,* Massachusetts Institute of Technology, 1972.

[10] G. Böhme, Architektur und Atmosphäre, München: Wilhelm Fink Verlag, 2006.

[11] J.-F. Augoyard and H. Torgue, Eds., Sonic experience: A guide to Everyday Sounds, Montreal & Kingston, London, Chicago: McGill-Queen's University Press, 2019.

[12] M. R. Schafer, The Soundscape: Our Sonic Environment and the Tuning of the World, Rochester, Vermont: Destiny Books, 1977.

[13] "Essentia," [Online]. Available: https://essentia.upf.edu/. [Accessed 07 01 2026].