The Sublime Ordinary: A Tool for Analysing Temporal City Soundscapes

Elina Oikonomaki
Independent Researcher
elinaoik@alum.mit.edu

Lukas Lesina Debiasi
Independent Researcher
lukas.debiasi@gmail.com

Abstract

As cities hurtle toward ever more data-driven futures, The Sublime Ordinary offers an alternative perspective on how we record, understand, and ultimately design urban environments by examining the temporal and sensory dimensions of city life from a first-person perspective. Through a multimodal dataset of synchronized audio, video, and GPS recordings collected in Harvard Square, Cambridge, Massachusetts, the project analyzes how the rhythms of everyday urban life shift over time and in response to policy and environmental changes. By combining semantic segmentation (SegFormer), object detection (YOLO), and sound classification (YAMNet), our system generates linked spatial–temporal–acoustic representations that form the data foundation of a new notational language informed by graphic notation to visualize the interplay of sound, activity, and place. Presented as an interactive web interface, the work enables users to explore recurring sound profiles and similarities between locations—asking: which city block sounds most like another, and how does its acoustic identity change over time? Engaging the NeurIPS Creative AI theme of Humanity, it examines how human and machine perception complement one another in a shared authorship that enables a more sensory, human-centered understanding of urban environments.

1 Description of the Work and Roles of AI/ML

The Sublime Ordinary functions as both artistic exploration and analytical instrument, transforming personal urban perceptions into tools for understanding and reimagining city life. It attempts to return focus to embodied, street-level experience—countering the prevalent way we record city change through static, top-down maps that fail to capture temporal dynamics and the human perspective.

The system operates through a web interface that captures synchronized video, audio, and GPS data, collected via phone during city strolls. Once uploaded, it enables analysis and visualization, letting users explore individual walks, build personal archives, and compare recordings over time to track acoustic and visual changes. We demonstrate this through a case study around Harvard Square during COVID-19 over 60 days, revealing how policy and behavior reshaped urban rhythms.

Recordings are segmented into equal time units for comparison. Three machine learning models form its core capability: SegFormer segments street scenes, YOLO identifies fine-grained objects, and YAMNet classifies sound events. These models convert raw sensory data into structured spatial-temporal-acoustic information, presented in an interactive interface combining geospatial maps with time-based notation inspired by graphic music scores.

Users navigate linked map and timeline views to explore patterns, compare walks over time, and investigate specific locations. For example, they might track how construction or seasonal changes alter a sidewalk's sensory profile across visits. This approach generates parallel interpretations of the

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Creative AI Track.

Figure 1: Visualization of the models for processing recorded video and audio data.

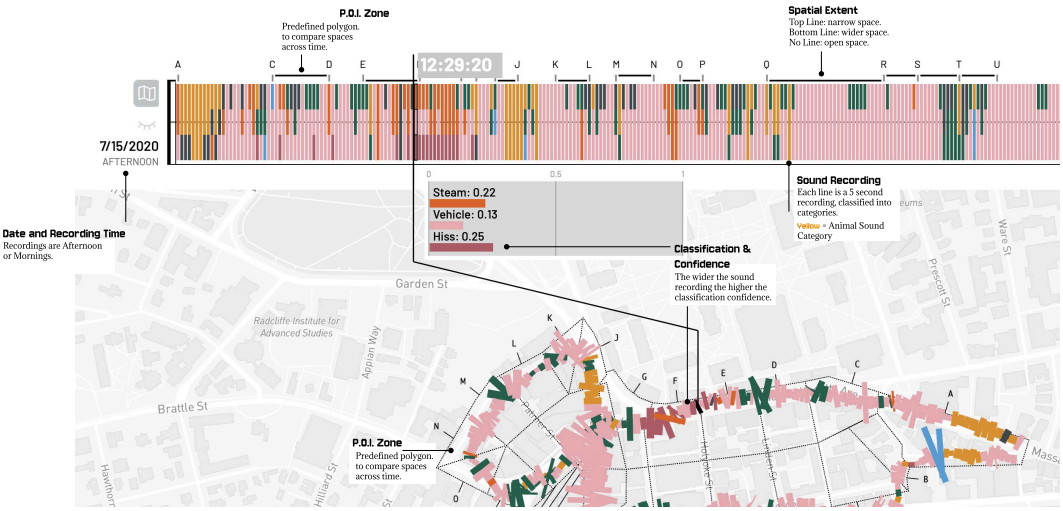

Figure 2: Web interface overview.

city, helping users examine subtle details of their lived experience and create richer, more complete memories through computational augmentation of perception.

The interface displays each recording as a musical staff, with colored bars indicating sound classes over time. Users can hover over segments for confidence scores, isolate specific sounds, group data by sidewalk, and listen to full walks or selected segments. A comparison mode allows overlaying multiple recordings to observe acoustic evolution over time.

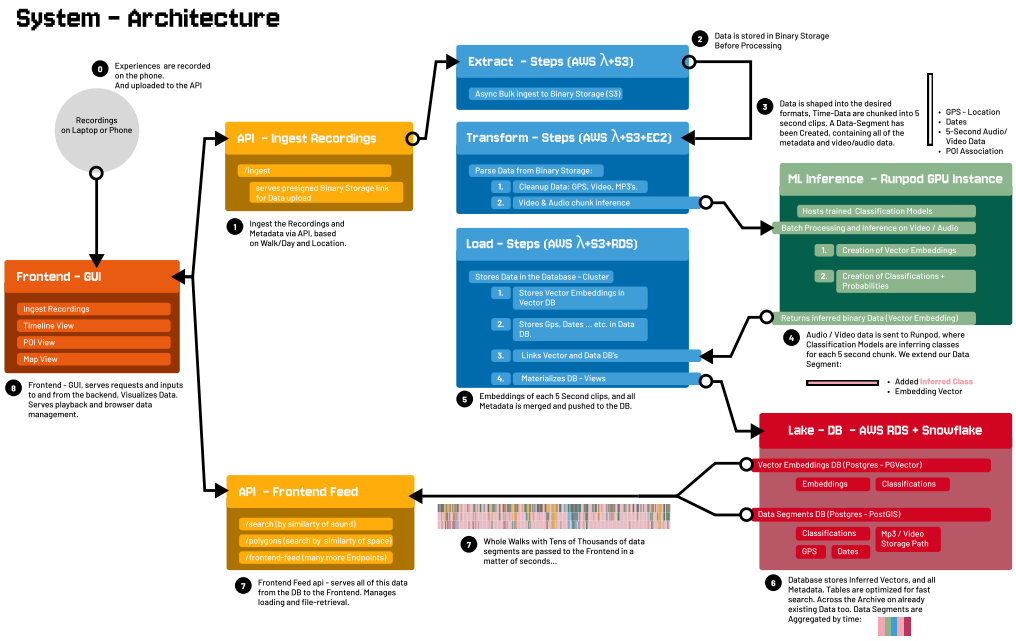

Figure 3: System architecture overview. Front-end and Data-Processing Setup.

2 Addressing the Theme of *Humanity*

The work engages with the *Humanity* theme by examining how human perception and machine perception can be combined to document and interpret public space. While the models identify and categorize sounds and visuals, the human role shapes the framing, walking routes, and interpretation of results. This shared authorship invites reflection on what aspects of the sensory city we choose to preserve, and how those choices may shift as we rely more on algorithmic perception. By aligning technological capabilities with human curiosity, the project captures the lived, ephemeral qualities of urban environments—elements that are felt as much as measured—and asks how we might value and remember these experiences in an age where machines see and hear alongside us.

3 Short Biographies of Authors

Elina Oikonomaki is a data visualization designer, engineer, and researcher working at the intersection of data visualization, urban informatics, and HCI. She has led visualization RD at Mobi.AI and taught Advanced Data Visualization Techniques at Northeastern University. She holds dual Master of Science degrees from the Massachusetts Institute of Technology (MIT) in Computer Science, and in Design and Computation. Before MIT, she earned her Diploma in Architectural Engineering from the National Technical University of Athens and was a visiting scholar at Istanbul Technical University. Website: `https://elinaoikonomaki.com`.

Lukas Lesina Debiasi is a software engineer based in New York, and is interested in data science and backend software development. Currently he is working as a founding engineer for a fintech startup, previously he worked at Samsung (SRA) and at Fathom Information Design as a data visualization engineer, developing epidemiological tools for COVID-19 research in collaboration with the Broad Institute at MIT and Harvard. He holds degrees from MIT and Princeton in both computer science and architecture. Website: `https://www.lukas-lesina-debiasi.info`.