

Concrète Collision: An Internet of Sound Remixing System and Live Instrument

Wade Marynowsky
Creativity and Cognition Studios
University of Technology Sydney
Sydney, Australia
wade.marynowsky@uts.edu.au

Sam Ferguson
Creativity and Cognition Studios
University of Technology Sydney
Sydney, Australia
samuel.ferguson@uts.edu.au

This paper presents *Concrète Collision*, an open-source, browser-native system and instrument for real-time generative composition using found sound retrieved dynamically from public audio repositories. The system embodies *Musique Concrète*'s core principle of deliberate sonic collision through a vibe-driven retrieval and mixing engine. The instrument operates without installation, synthesis, or pre-authored content, using the internet of sound as the source. The internet is treated not as a distribution medium but as a living sound library, drawing on the full spectrum of recorded sound to situate the work within the sound art tradition. The system retrieves audio from four heterogeneous public repositories, Freesound, the Internet Archive, the NASA Image and Video Library, and OpenVerse, assembling it into a four-track sound collage processed through a Web Audio API effects chain. The paper presents the system architecture, compositional framework, and theoretical lineage connecting Schaeffer's *typomorphologie*, Cage's indeterminacy, and contemporary Internet of Sounds research. The work demonstrates that browser-native, multi-source generative composition offers a compositional paradigm distinct from both traditional studio practice and existing single-source generative systems. *Concrète Collision* demonstrates that real-time, non-deterministic retrieval from public audio repositories constitutes a viable and theoretically grounded compositional method. The techniques described in this paper are designed for adoption and reuse. The complete source code is publicly available under an open-source license, and the system runs entirely in the browser.

Keywords: *musique concrète*, Web Audio API, Freesound, soundscape, found sound, ubiquitous music, browser-native audio

I. INTRODUCTION

The principle of *Musique Concrète*, [1] composition from recorded everyday sound, has endured as a paradigm for experimental music for nearly eight decades. Pierre Schaeffer's laboratory recordings of trains, spinning tops, and mechanical objects in 1948 established that the sonic world itself could serve as compositional material. His insight was simple but radical: any sound, captured and removed from its original context, becomes available for musical transformation through what he termed *réduction acousmatique*, the discipline of attending to sound purely as sound-object, free from the visual and contextual associations that normally govern listening. Schaeffer organised this material using *typomorphologie*, a taxonomy of sound-objects defined by properties including mass, grain, allure, and dynamic profile. For Schaeffer, the studio was a semantic laboratory in which the properties of sounds could be examined and recombined deliberately, with the composer's trained ear as the arbiter of all selection.

What Schaeffer could not have anticipated was the scale and accessibility of recorded sound seven decades later. The internet has become the largest sound archive in human history, continuously growing, globally distributed, and publicly accessible in real-time. Public audio repositories such as Freesound [2, 3], the Internet Archive [4], the NASA Image and Video Library [5], and OpenVerse [6] now aggregate millions of field recordings, archival material, broadcast audio, and environmental recordings from across the planet and beyond. These platforms have transformed the material conditions of *concrète* practice: where Schaeffer required a studio, reel-to-reel equipment, and physical access to recorded material, a contemporary practitioner can query millions of sounds from a browser within milliseconds.

The emergence of the Web Audio API [7, 8, 9] as a standardised browser audio processing framework has simultaneously made it possible to perform sophisticated real-time signal processing within that same browser environment, without plugins or native application installation. The combination of globally networked public sound archives and client-side real-time audio processing defines a genuinely new set of conditions for *concrète* practice, one that the traditional studio paradigm does not address.

The convergence of networked audio archives and browser-native processing situates the work within the Internet of Musical Things (IoMusT) paradigm, introduced by Turchet et al. [10] and its broader extension, the Internet of Sounds (IoS) [11], encompassing the online network of auditory devices, protocols, and applications that enable real-time interaction with sound-related information across the internet.

Concrète Collision [12] is an open-source, browser-native system that treats this networked archive as a living compositional resource. It retrieves found sound in real-time from four heterogeneous public repositories, assembles it into a four-track collage, and processes the result through a comprehensive Web Audio API effects chain. The system proposes non-deterministic archive retrieval as a compositional paradigm: the performer selects a semantic intent through what the system terms a *vibe*, the system resolves that intent into queries dispatched across multiple archives, and the encounter between the performer's intention and unpredictable retrieval constitutes the work. No two sessions produce identical material.

This paper addresses three research questions. First, how can a browser-native architecture support real-time generative composition from multiple heterogeneous networked sound

archives simultaneously? Second, can a semantic taxonomy of operators function as a viable generative mechanism for Internet of Sounds composition, producing aesthetically coherent yet non-deterministic results? Third, what compositional structures and affordances emerge when non-deterministic retrieval across institutionally and curatorially distinct sources, including sonified scientific data [13], is treated as a primary compositional method rather than a supplementary technique?

This paper describes the system architecture, compositional framework, and theoretical lineage connecting Schaeffer's typomorphologie [1], Cage's indeterminacy [14], Stockhausen's global archive thinking, and Schafer's soundscape theory [15]. It situates *Concrète Collision* within the Internet of Sounds paradigm, provides a detailed account of the Web Audio API processing chain, and compares the system with prior generative systems, examining what is theoretically and technically distinctive about the approach. The system is publicly accessible at <https://wademarynowsky.art/concrete> [12] and the complete source code is available under an open-source licence.

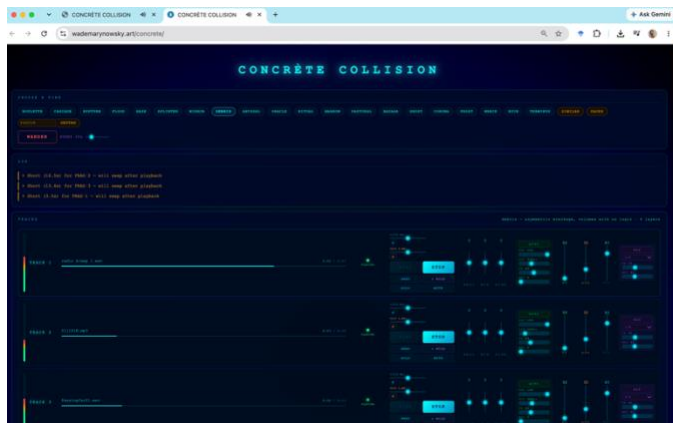


Fig. 1. *Concrète Collision*'s user interface, Chrome browser screenshot.

II. RELATED WORK

A. *Musique Concrète* and Recorded Sound Composition

Pierre Schaeffer's *Cinq Études de bruits* (1948) [1] and the subsequent work of the *Groupe de Recherche Musicale* (GRM) established *Musique Concrète* as a compositional practice in which recorded environmental sound is treated as raw musical material. Schaeffer's typomorphologie proposed a systematic taxonomy of *objets sonores* organised by their properties, enabling a compositional method grounded in the acoustic qualities of material rather than in conventional pitch relationships or rhythmic structure. This framework is foundational to *Concrète Collision*: its semantic 'vibe' taxonomy functions as a set of operators that specify acoustic domains to be retrieved and collided, adapting Schaeffer's analytical framework as a generative mechanism.

B. Cage: Radio, Tape, and Non-Intention

John Cage's *Imaginary Landscape No. 4* (1951) [16], scored for twelve radios and twenty-four performers, is the most direct historical precedent for *Concrète Collision*. The score governs only the manipulation of tuning and volume controls; sonic content is determined entirely by whatever is broadcasting at performance time, making each performance fundamentally unrepeatable. *Williams Mix* (1952) extended this into tape:

approximately 600 recordings assembled across six categories using I Ching [17] chance operations, with the final montage determined by aleatory procedure rather than compositional intention. Together these works establish the two poles that *Concrète Collision* occupies simultaneously: live uncontrolled retrieval and non-intentional assembly. Cage required physical tape, physical radios, physical performers, and a physical space. *Concrète Collision* requires only a browser and a network connection.

C. Stockhausen: Global Archive as Compositional Material

Karlheinz Stockhausen's *Telemusik* (1966) [18], composed at NHK's electronic music studio in Tokyo, draws on recordings of traditional music from Japan, Spain, Hungary, Vietnam, the Sahara, and the Amazon, electronically transforming them into a continuous unified texture. *Hymnen* (1967) [19] constructed a four-hour electronic work from national anthems intercepted from shortwave radio broadcasts, treating the world's political sound archive as compositional raw material. Stockhausen's intuition that the world's recorded sound constitutes a compositional resource is directly inherited by *Concrète Collision*. The critical distinction is between monument and stream: *Telemusik* is a fixed work built once in a controlled studio over many months, while *Concrète Collision* is a live channel through which the global archive flows continuously and differently with each session.

D. Soundscape and the World Soundscape Project

R. Murray Schafer's concept of the soundscape and his World Soundscape Project, begun at Simon Fraser University in the late 1960s and consolidated in *The Tuning of the World* [20], established a tradition of treating environmental sound as a culturally, ecologically, and politically significant object of study and composition. Schafer proposed a taxonomy of soundscape elements (keynotes, soundmarks, signals) and argued that listening to environmental sound constitutes an act of ecological attention. Public repositories such as Freesound realise at scale what the World Soundscape Project attempted with analogue means, creating a globally networked archive of environmental sound contributed by recordists worldwide. *Concrète Collision* treats these platforms not as libraries of samples but as a living globalised soundscape from which momentary compositions are instantiated in real-time, directly positioning the work within the Internet of Sounds paradigm.

E. Scanner and Networked Listening

Robin Rimbaud (Scanner) pioneered the interception of mobile phone transmissions as compositional material in the early 1990s [21]. Works such as *Mass Observation* (1994) assembled intercepted private conversations and urban sound into dense documentary collages, establishing a practice in which listening becomes a form of critical inquiry. Scanner's approach frames the network itself as a sound field to be sampled and reassembled, positioning the artist as a listener whose critical attention reveals sonic material ordinarily invisible. *Concrète Collision* inherits this critical framing while operating on publicly contributed Creative Commons licensed archives rather than privately intercepted transmissions, shifting the ethical position from interception to curation.

F. Internet of Sounds Systems

In related work, Buffa and Music [22] have published a guitar amplifier simulator built entirely with Web Audio API nodes, demonstrating that browser-native signal processing can achieve results comparable to dedicated audio software, while Buffa et al. [23] present a browser-based digital audio workstation supporting multi-track recording, editing, and real-time effects using Web Audio Modules, directly establishing the technical feasibility of the kind of browser-native compositional environment that *Concrète Collision* implements.

Recent work presented at the IEEE Internet of Sounds symposium has established a body of practice in networked and browser-native audio that provides direct context for *Concrète Collision*. Mikolajczyk et al. [24] describe an Internet of Sounds ecosystem for rendering object-based audio across wireless distributed devices, interfaced with gestural controllers to support embodied interaction with spatialised sound. The work demonstrates how IoS infrastructures can support live, performative engagement with distributed audio in ways that parallel *Concrète Collision*'s multi-source retrieval and four-track spatial assembly. Mazzocchio [25] presents a cooperative net art installation implemented in Web Csound, in which users connect via a browser and their connections affect real-time ambisonics playback of lockdown binaural recordings. The work establishes a precedent for browser-native audio installation in the IoS context, with user presence shaping compositional outcome in a manner analogous to *Concrète Collision*'s autonomous operation mode. Boem and Turchet [26] explore browser-based shared virtual sonic environments using WebXR and the Web Audio API, demonstrating that the browser has become a primary platform for networked musical experience. Together these works situate *Concrète Collision* within an active research trajectory in browser-native and networked audio composition.

G. Freesound-Based Generative Systems

Several prior systems have established Freesound as a viable resource for generative composition. Eigenfeldt and Pasquier's *Coming Together: Freesound* [27] deployed four autonomous software agents selecting sounds based on spectral content and metadata tags, with agents continuously analysing each other's audio output to minimise spectral masking between tracks. The system pursued acoustic coherence as its compositional goal, treating the simultaneous sounding of multiple retrieved materials as a problem to be resolved through spectral negotiation. Audio Metaphor [28] used natural language queries to retrieve and algorithmically compose soundscapes from Freesound's then database of over 390,000 recordings. Playsound.space [29] is a browser-based improvisation tool enabling real-time semantic retrieval from Freesound for free improvisation performance, providing spectrogram visualisation and Web Audio API playback. *Concrète Collision* builds on this lineage while extending it across multiple heterogeneous archives and supporting autonomous unattended operation.

III. SYSTEM ARCHITECTURE

A. Vibe-Driven Retrieval Engine

The central compositional interface is the *vibe*, a semantic keyword specifying a domain for each of the system's four audio tracks simultaneously. Each *vibe* maps to a set of four orthogonal acoustic domain terms dispatched as queries to different archive sources:

- TERMINUS specifies {train, station, crowd, signals};
- ABYSSAL specifies {hydrophone, whale, sonar};
- FROST specifies {glacier, aurora, melt, arctic wind};
- CORONA specifies {solar wind, plasma, aurora, infrasound}.

The named *vibes* span acoustic categories, each constructed to maximise contrast between the four domain terms assigned to its tracks.

Beyond named *vibes*, four discovery modes extend retrieval into the archive's relational structure.

- SIMILAR queries the Freesound API's similarity endpoint to find acoustically related sounds based on spectral distance from a seed recording.
- PACKS retrieves curated Freesound collections uploaded as grouped sets by individual contributors.
- GEOTAG queries sounds by geographic coordinate pair, enabling real-time soundscape retrieval from any point on the Earth's surface, directly realising Schafer's soundscape mapping as an interactive system.
- WANDER autonomously navigates between *vibes* at a configurable drift rate, enabling unattended long-form compositions in which the system evolves its own semantic trajectory over hours without performer input.

B. Audio Source Network

Concrète Collision queries four primary sources in parallel, each contributing a distinct curatorial identity to the retrieval pool:

- Freesound provides user-contributed recordings with Creative Commons licensing, rich community-assigned metadata tags, geolocation data, and automatically extracted spectral descriptors.
- The Internet Archive provides historical broadcast recordings, field recordings, oral history materials, and archival audio spanning over a century of recorded history.
- The NASA Image and Video Library contributes spacecraft telemetry, planetary probe data, radio telescope recordings, and radio emissions converted to audio files, providing acoustic domains unreachable by any human recordist.
- OpenVerse aggregates audio published under Creative Commons licenses (CC0, CC-BY, CC-BY-SA) from multiple public repositories including Freesound, Jamendo, and ccMixer, providing a unified search interface across over one million audio recordings.

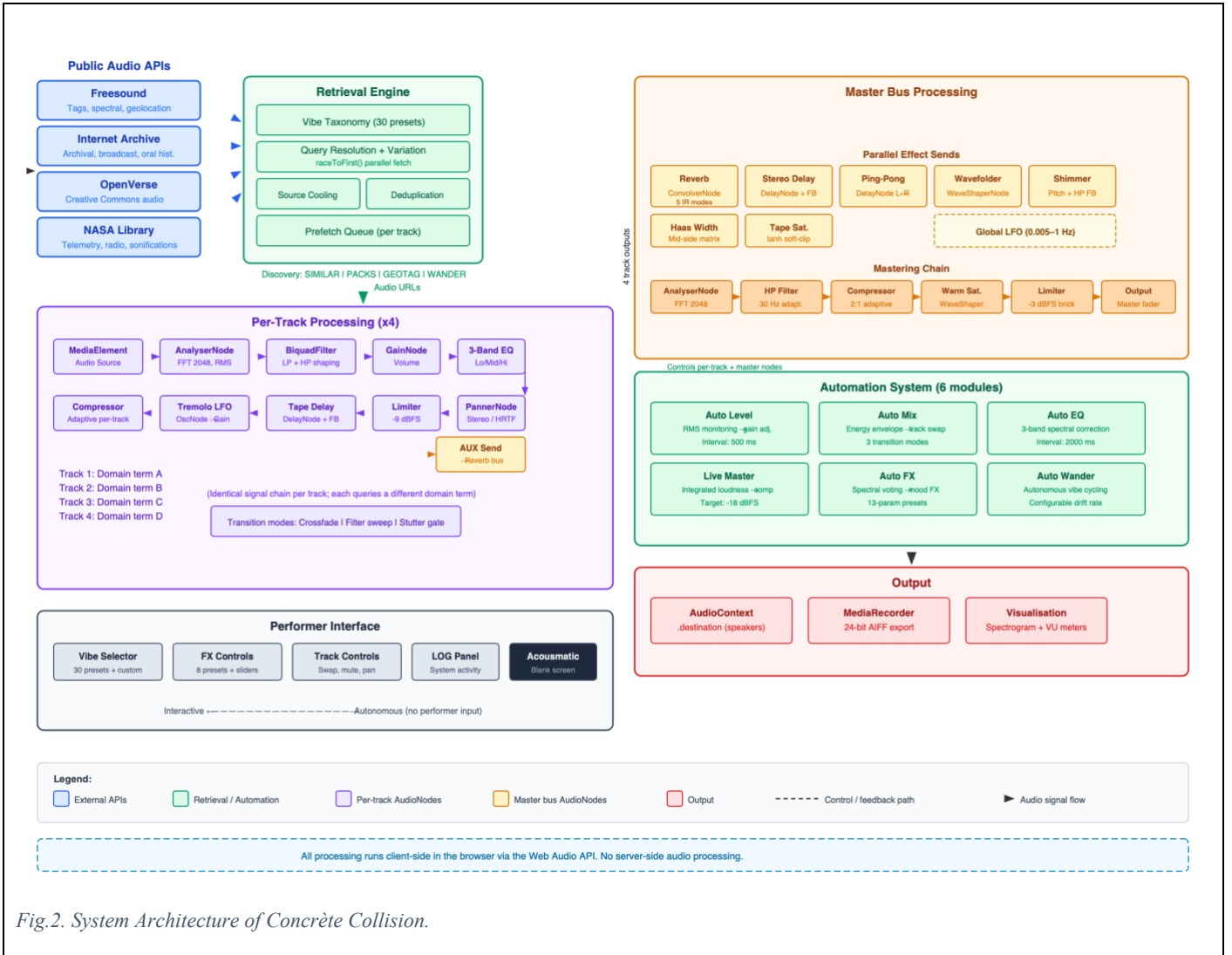


Fig.2. System Architecture of Concrète Collision.

Because all retrieved material is openly licensed, the system ensures that source audio is legally available for reuse in generative composition.

Source-cooling logic progressively deprioritises sources that have recently returned results, maintaining diversity across successive retrieval cycles. Session-level deduplication tracking prevents the repetition of previously heard material within a single session. Each track maintains an asynchronous prefetch queue, loading the next sound before the current one completes, ensuring continuous uninterrupted playback across substitutions.

Query construction follows a direct mapping strategy: each domain term from the active vibe is dispatched as a keyword query to the assigned source's search API. When a source returns no results, the system applies a fallback strategy using a synonym lookup table. Audio format differences are handled at the decoding stage via the Web Audio API's `decodeAudioData` method, which normalises all retrieved audio to the `AudioContext` sample rate regardless of source format. Retrieval failures due to network errors, rate limiting, or Cross-Origin Resource Sharing (CORS) restrictions are logged and the

affected track continues its current buffer while the system retries with an alternative source.

C. Web Audio API Processing Pipeline

The Web Audio API defines a directed graph architecture in which `AudioNode` objects are connected to form signal processing networks within an `AudioContext`. Concrète Collision implements a two-tier graph: a per-track processing tier and a shared master bus tier. Each track instantiates a `BiquadFilterNode` chain providing independent high-pass and low-pass spectral shaping. A `GainNode` controls track output level. A `StereoPannerNode` provides spatial placement in the stereo field. Per-track outputs feed into the master bus through individual send `GainNodes`.

The master bus implements a comprehensive effects chain using both native Web Audio API node types and custom `AudioWorklet` processors [23]. Reverberation is generated algorithmically using `ConvolverNodes` loaded with procedurally constructed impulse responses in five room character modes (synth, plate, cave, spring, room). Stereo delay and ping-pong delay employ `DelayNode` feedback networks. Wavefolder distortion uses `WaveShaperNode` with a custom

piecewise transfer function for harmonic enrichment. Tape saturation applies a tanh soft-clipping curve via WaveShaperNode. Shimmer uses a feedback path through pitch shift processing and a high-pass BiquadFilterNode to generate upward spectral regeneration. Tremolo and auto-pan use OscillatorNodes modulating GainNode amplitude and StereoPannerNode position. The stereo widener encodes a mid-side matrix using gain and polarity inversion nodes.

A global LFO sub-system provides continuous time-varying modulation using OscillatorNodes at rates between 0.005 and 1 Hz, routing through AudioParam connections to modulate filter cutoff, master gain, reverb send, and delay feedback. AnalyserNode instances provide real-time frequency data for spectrogram rendering at FFT size 2048, and four per-track AnalyserNodes supply RMS data for VU meter display.

D. Automation System

Concrète Collision implements a six-module automation stack enabling fully unattended operation. Auto Level monitors RMS energy across all four tracks and applies compensatory GainNode adjustments to maintain target loudness relationships. Auto Mix monitors per-track energy envelopes and triggers track substitution via three transition modes: crossfade, filter sweep, or stutter gating. Auto EQ applies per-track spectral correction through dynamic BiquadFilterNode adjustments derived from ongoing FFT analysis. Live Master implements adaptive lookahead compression and peak limiting on the master bus via custom AudioWorklet processors. Auto FX selects effect preset combinations based on vibe domain semantics. Auto Wander autonomously transitions between vibes at a configurable drift rate. A button for autonomous spatial mixing using HRTF-based binaural filtering is also provided.

IV. COMPOSITIONAL FRAMEWORK

A. Semantic Vibes as Operators

Vibes enact a theory of sonic material grounded directly in Schaeffer's typomorphologie. Each vibe specifies four acoustic domains that are intentionally non-adjacent and orthogonal, drawn from distinct registers of texture, pitch content, temporal behaviour, spectral density, and cultural association. The collision of glacier ice sounds with aurora recordings, melt water drips and Arctic wind in the FROST vibe produces emergent acoustic relationships unavailable to any pre-composed or synthesised approach, because the specific material characteristics of each retrieved sound cannot be known in advance. The system realises typomorphologie not as an analytical tool applied retrospectively but as a generative engine.

B. Indeterminacy as Compositional Structure

Retrieval results are non-deterministic by design at multiple levels. API response content varies with the time of day, recent repository contributions, rate-limiting conditions, network latency, and the stochastic outputs of the internal deduplication algorithm. No two instances of the same vibe produce identical material. The composition's identity lies not in its sonic content but in its process: the vibe taxonomy, the retrieval architecture, the four-track structure, and the automation rules constitute the composition. In Cagean terms, the system provides a

determinate structure within which the sound mix remains undetermined, directly mirroring Imaginary Landscape No. 4 where the score governs only the apparatus and never the content. Concrète Collision's indeterminacy is constrained by the vibe, giving the performer a lever on the acoustic character of retrieved material without granting control over specific sounds.

C. Ubiquitous Music and Autonomous Operation

Keller's ubiquitous music model [30] describes music-making that is continuous, location-independent, participatory, and woven into everyday context rather than confined to dedicated performance or studio situations. Concrète Collision realises this model through its automation stack, enabling the system to operate indefinitely as an installation piece, a broadcast compositional process, or a background generative environment without performer attention. In this configuration the system functions as a composition engine: the artist's role shifts from real-time performer to architect of a system whose ongoing behaviour constitutes the work, a role that Collins has associated with the tradition of algorithmic composition [31].

V. EXAMPLE SESSION

To illustrate the system's compositional behaviour in detail, consider a session initiated with the vibe ABYSSAL, specifying the domains {hydrophone, whale, sonar, pressure}. The retrieval engine dispatches parallel queries to Freesound, the Internet Archive, the NASA Library, and OpenVerse. Within several seconds four tracks begin loading independently. Track one carries underwater ice cracking from a Nordic oceanographic researcher uploaded to Freesound under Creative Commons attribution. Track two delivers a 1970s archival humpback whale song from an International Whaling Commission recording held by the Internet Archive. Track three loads a submarine sonar ping recorded at a naval research station. Track four plays plasma wave emissions from the Cassini spacecraft's approach to Saturn, sonified from raw electromagnetic sensor data by NASA's audio visualisation team. The session thus combines acoustic field recordings, documentary sound, and sonified scientific data, exemplifying the system's position at the intersection of Musique Concrète and data sonification.

None of these sounds were pre-selected, pre-listened, or anticipated. Their simultaneous sounding in the four-track mix is entirely the product of the retrieval event, constrained only by the four domain terms of the ABYSSAL vibe. The per-track BiquadFilterNode chains provide spectral shaping: the sonar ping's low-frequency energy is high-pass filtered to prevent it from masking the whale song's mid-range frequency content. The humpback recording loops; after four minutes Auto Mix detects an energy envelope drop in the ice cracking track and triggers a crossfade substitution, initiating a new Freesound query for hydrophone. A different researcher's contact microphone recording of underwater rock surfaces arrives within seconds and crossfades in.

The Cassini plasma emission, routed through the shimmer effect's upward pitch-shift feedback path, generates a continuous high-frequency spectral wash that transforms at each LFO modulation cycle. The sonar ping, delayed through the ping-pong DelayNode network at 440 milliseconds with 0.65

feedback, creates a decaying spatial call-and-response pattern whose timing is determined by the retrieved audio's own attack envelope interacting with the delay parameters. After 1 minute, Auto Wander initiates a drift from ABYSSAL toward CORONA, substituting domains progressively: whale and sonar transition toward solar wind and plasma. The session has become a continuous transformation from ocean depth to outer space, a trajectory that emerged entirely from the system's autonomous behaviour.

VI. PERFORMANCE MODES AND INTERFACE

The instrument supports a continuous spectrum from fully interactive performance to fully autonomous unattended operation. In interactive mode, the performer selects vibes, triggers individual track swaps, adjusts effect parameters via sliders mapped to Web Audio API AudioParam values, and monitors the system state through multiple real-time displays. A LOG panel provides transparency into all system decisions: retrieval queries, source selection, deduplication rejections, network errors, and Auto Mix events are displayed in chronological sequence, giving the performer a continuous read of the system's compositional activity as a form of score or transcript.

Eight effect presets (NATURE, OCEAN, DARK, SPACE, RITUAL, INDUSTRIAL, URBAN, VOICE) provide rapid transformations of the master processing chain by simultaneously adjusting multiple node parameters. SPACE routes material through maximum reverb decay, shimmer, and high-frequency emphasis. RITUAL applies slow LFO modulation, mid-frequency emphasis, and long ping-pong delay. INDUSTRIAL uses wavefolder distortion, short reverb, and high-pass filtering. Master output can be recorded as 24-bit AIFF using the MediaRecorder API, capturing the full processed mix without encoding loss.

Acousmatic mode suspends all visual feedback, presenting a blank black screen while audio continues uninterrupted. This directly implements Schaeffer's *réduction acousmatique* [32] as an interactive affordance: removing the visual interface compels focused listening to the sound itself as sound-object, free from the metadata and source information that would ordinarily contextualise listening retrieval.

VII. SYSTEM PERFORMANCE EVALUATION

Multiple system performance sessions were run from the Gandhi web host server on the University of Technology Sydney's internet connection. We tested Retrieval Performance, the audio engine and Material Non-Determinism.

A. Retrieval Performance.

The following performance data was conducted on the 7/8/26. The duration was 323.7 min (5.4 hours). Wander on, range 1 min, per 15 vibes. Auto mix on, auto swap off. Platform Chrome, 48.0 kHz. Host server Gandhi.net. Client UTS campus connection. Across multiple test sessions totalling 323.7 minutes, the system generated 8,924 retrieval events. Of these, 3,565 (40.0%) succeeded, yielding 2,222 unique sound URLs.

Table 1. Retrieval Latency by Source.

Source	Queries	Success Rate	Median Latency	P95 Latency	Share of Successes
Freesound	2,447	19-93.1%	399 ms	689 ms	14-55.4%
OpenVerse	2,159	72.7%	1,154 ms	1,722 ms	29.7%-55%
Internet Archive	2,159	34.3%	2,740 ms	7,194 ms	14.0%-26%
NASA	2,159	2.2%	2,913 ms	3,747 ms	0.9%-1.1%
All	8,924	32-54.9%	—	—	—

Freesound was the fastest source (399–425 ms median) but degraded from 93.1% to 19.5% success over extended use due to per-IP rate-limiting, with OpenVerse absorbing the load. OpenVerse maintained consistent reliability (72.7–75.3% success) across all sessions. Internet Archive contributed at a steady ~35% success rate with higher latency. NASA remained largely unavailable (~2% success), contributing to 4–6 vibes.

B. Audio Engine.

Table 2. Audio Engine Metrics.

Metric	Value
Audio glitches	0 (across 8,924 retrievals, 324 vibe transitions)
Active tracks	1–4 (mean 3.9)
JS heap memory	6–48 MB (no monotonic growth)
Base / output latency	21.3 ms / 56.0 ms (constant)
Auto mix active	99.8% of 1941 samples

Zero audio dropouts across 323.7 minutes of server-hosted operation. The server environment ran at 48.0 kHz with base latency of 21.3 ms and output latency of 56.0 ms, both constant throughout all sessions. Memory ranged 6–48 MB with no monotonic growth pattern. Auto mix was engaged for 99.8–99.9% of performance samples; FX processing for 98.6–99.1%.

C. Material Non-Determinism

100% URL uniqueness across all 15 vibes. 636 candidate duplicates were rejected by the deduplication filter. No two tracks within or across vibes played the same sound during either session. Throughput: 6.3–8.1 unique sounds/minute.

This data validates Concrète Collision's viability as a server-hosted deployment. Across 323.7 minutes of operation, the system retrieved 2,222 unique sounds across 15 vibes with zero audio glitches, zero URL overlap, and stable memory. Throughput ranged from 6.3 to 8.1 unique sounds/minute, with higher rates when Freesound was available. The audio engine operated at 48.0 kHz with constant latency and near-complete auto mix engagement. No user intervention was required. A metrics button is available in the UI of Concrète Collision for comparable testing.

VIII. DISCUSSION

Concrète Collision occupies a specific and definable position within the lineage traced above. Like Cage's radio works it surrenders material selection to an external uncontrolled source; like Stockhausen's *Telemusik* it treats the global archive as compositional resource; like Scanner it frames networked listening as ethically and aesthetically charged, though where Scanner intercepted private transmissions, Concrète Collision draws exclusively on Creative Commons licensed material, shifting the ethical ground from interception to curation. Prior Freesound-based systems illuminate what is theoretically distinctive about this positioning. In *Coming Together*: Freesound deploys autonomous spectral agents that pursue acoustic coherence, minimising masking between simultaneously sounding tracks. Concrète Collision explicitly inverts this logic: the vibe taxonomy deliberately maximises contrast between domains, treating spectral collision and acoustic conflict not as problems to be solved but as compositional properties to be inhabited. It should be acknowledged that the individual technical components of Concrète Collision are established technologies; the contribution lies in their specific combination under a theoretically grounded compositional framework, producing compositional outcomes unavailable to any of these components independently.

The system's LOG panel records every retrieval query, source URL, track assignment, effect parameter change, and automation event with millisecond timestamps. This transcript, combined with the captured audio output, provides a complete record of each session's compositional decisions. Future development will extend this to full session serialisation, enabling exact replay of a recorded session.

Playsound.space offers a performer-centred improvisation tool in which a human performer drives all retrieval decisions in real-time. Concrète Collision extends this from performer-driven or fully autonomous operation, making unattended long-form installation use a primary rather than incidental use case. Audio Metaphor uses natural language queries to assemble soundscapes from a single large database; Concrète Collision uses domain operators across heterogeneous archives with distinct institutional and curatorial identities, making source diversity itself a compositional variable.

The recent Internet of Sound works provide further points of comparison. Mazzocchio's browser-based net art installation shares Concrète Collision's commitment to the browser as a compositional medium, but differs in that user presence rather than semantic retrieval drives the sonic outcome. Boem and Turchet's browser-native shared sonic environments demonstrate the Web Audio API's capacity to support immersive networked experience, extending the browser from playback medium to performance environment. The Mikolajczyk et al. *IoS* ecosystem, while hardware-based rather than browser-native, shares the architectural concern with real-time spatialised multi-source audio assembly that defines Concrète Collision's four-track structure.

The Web Audio API's role in enabling browser-native instruments is worth examining directly. Prior to its standardisation, browser-based audio was limited to basic

HTML5 playback without real-time processing. The AudioContext graph model, its provision of filtering, gain, panning, convolution, analysis, and oscillator nodes, and the AudioWorklet interface for custom DSP in a dedicated real-time audio thread together enable a signal processing architecture that previously required a DAW or specialist audio programming environment [7, 33]. Every element of Concrète Collision's signal chain runs within the browser without any server-side audio processing, realising the entire compositional process in the same environment where the internet's sound archives are accessed.

The vibe taxonomy introduces a tension that predecessors do not address. Schaeffer's practice was fully intentional in material selection; Cage's was fully non-intentional. Concrète Collision occupies a deliberate middle position: the vibe specifies intent, while retrieval remains stochastic. The composer chooses what acoustic world to inhabit but not what is heard within it. This raises questions of semantic authority the system does not resolve. What cultural assumptions are embedded in the domain label *ABYSSAL*? Future work might enable user-defined vibes, community-contributed taxonomies, or machine-learning-driven retrieval that learns acoustic associations from listening behaviour. The system's dependency on public APIs also entangles its identity with the living ecology of these platforms, a dependency that is simultaneously a limitation and a constitutive feature of its network-native identity. It should also be acknowledged that the current vibe taxonomy is manually designed and has not been validated through formal listening experiments. Planned future work includes structured listening evaluations to assess the perceptual coherence and distinctiveness of vibe-driven retrieval outcomes.

The question of the work's status as composition or instrument is productively unresolved. As a system that can operate without performer input, it functions as a composition in the Cagean sense of instructions that determine a musical situation. When responding to real-time performer gestures, it functions as an instrument. The two modes coexist within the same architecture, and the performer's choice of engagement level is itself a compositional decision. Xenakis positioned the computer as a means of realising compositional structures too complex for manual execution [34]; Concrète Collision positions the network as the source of material too vast and unpredictable for any individual to curate, suggesting that the archive itself, in its living and changing totality, is the instrument's primary collaborator.

IX. CONCLUSION

Concrète Collision demonstrates that browser-native, network-reliant generative composition is a theoretically rich and technically viable practice. By treating the internet as a living sound library and non-deterministic archive retrieval as compositional structure, the system extends the lineage of Schaeffer, Cage, Stockhausen, Schafer, and Scanner into the specific material conditions of networked computing in the 2020s. It builds on prior Freesound-based generative systems and recent Internet of Sounds browser-native work while distinguishing itself through multi-source retrieval, vibe operators, full autonomous operation, and acousmatic interface design.

The system proposes that the unpredictable encounter between semantic intent and retrieved material constitutes a primary aesthetic resource for the Internet of Sounds era. Where Cage required twelve radios and twenty-four performers, and Stockhausen required months at NHK's Tokyo studio, Concrète Collision requires only a browser and a network connection. The democratisation of concrète practice this represents is not only a matter of access but also ontology: the living, continuously updated, globally contributed public sound archive constitutes a sonic world in constant change that no composition can fully anticipate or represent.

Future development directions include user-defined vibe creation, machine-learning-driven acoustic similarity retrieval, expanded archive integration, and multi-performer networked configurations enabling synchronised or interactive ensemble performance.

The techniques described in this paper are designed for adoption and reuse. The complete source code is publicly available under an open-source license, and the system runs entirely in the browser with no server-side dependencies, requiring only a free Freesound API key to begin retrieving sound. Researchers and practitioners can extend the vibe taxonomy with custom domain terms, integrate additional public audio APIs beyond the four described here, and modify the Web Audio API processing chain to suit their own compositional or installation contexts. The modular architecture separates retrieval, mixing, effects processing, and automation into independent components, enabling selective reuse of individual subsystems in other browser-native audio projects.

Example recordings of multiple live sessions from May to June 2026 are available at https://wademarynowsky.art/concrete_tracks [35]. The live sessions combine real-time effects and mixing as well as autonomous mode recordings. The recordings demonstrate the evolution of the mixing and mastering chain as well as experiments in autonomous spatial mixing using HRTF-based binaural filtering. The track names include the date of the remix and the first word of each of the first loaded tracks, creating a form of cut-up poetry in the track titles.

ACKNOWLEDGMENT

The authors acknowledge the Freesound project and its community of contributors, the Internet Archive and its community of audio archivists, the OpenVerse project, and the NASA Image and Video Library for their open audio repositories, without which this work would not be possible. The authors also acknowledge the W3C Web Audio API Working Group for standardising the browser audio infrastructure that makes this system viable.

REFERENCES

- [1] P. Schaeffer, *Traité des objets musicaux*. Paris: Éditions du Seuil, 1966.
- [2] Font, F., Roma, G., & Serra, X. (2013, October). Freesound technical demo. In *Proceedings of the 21st ACM International Conference on Multimedia* (pp. 411–412).
- [3] Fonseca, E., Pons, J., Favory, X., Font, F., Bogdanov, D., Ferraro, A., Oramas, S., Porter, A., & Serra, X. (2017). Freesound datasets: A platform for the creation of open audio datasets. *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, 486–493.
- [4] Internet Archive, "Internet Archive," 2001. [Online]. Available: <https://archive.org/>.
- [5] NASA, "NASA Image and Video Library," 2017. [Online]. Available: <https://images.nasa.gov/>.
- [6] Openverse, "Openverse: An Extensive Library of Free Stock Photos, Images, and Audio," WordPress, 2021. [Online]. Available: <https://openverse.org/>.
- [7] P. Adenot and C. Wilson, "Web Audio API," 2021/06// 2021. [Online]. Available: <https://www.w3.org/TR/webaudio/>.
- [8] P. Adenot and H. Choi, "Web Audio API," W3C Recommendation, 17 June 2021. [Online]. Available: <https://www.w3.org/TR/2021/REC-webaudio-20210617/>.
- [9] B. Matuszewski and O. Rottier, "The Web Audio API as a Standardized Interface Beyond Web Browsers," *Journal of the Audio Engineering Society*, vol. 71, no. 11, pp. 790–801, 2023.
- [10] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of Musical Things: Vision and Challenges," *IEEE Access*, vol. 6, pp. 61994–62017, 2018.
- [11] Turchet, L., Lagrange, M., Rottondi, C., Fazekas, G., Peters, N., Østergaard, J., ... & Fischione, C. (2023). The internet of sounds: Convergent trends, insights, and future directions. *IEEE Internet of Things Journal*, 10(13), 11264–11292.
- [12] W. Marynowsky, "Concrète Collision," 2026. [Online]. Available: <https://wademarynowsky.art/concrete/>.
- [13] Zanella, A., Harrison, C. M., Lenzi, S., Cooke, J., Damsma, P., & Fleming, S. W. (2022). Sonification and sound design for astronomy research, education and public engagement. *Nature Astronomy*, 6(11), 1241–1248.
- [14] J. Cage, *Silence: Lectures and Writings*. Middletown, CT: Wesleyan University Press, 1961.
- [15] R. M. Schafer, *The Tuning of the World*. New York: Knopf, 1977.
- [16] Valle, A., & Casella, A. (2016). Imaginary Landscape No. 4: study and annotation of the score. In *Extending interactivity-Atti del XXI CIM-Colloquio di Informatica Musicale* (pp. 69–76). DADI-Dip. Arti e Design Industriale. Università IUAV di Venezia.
- [17] Wilhelm, H. (1967). *The I Ching or book of changes*.
- [18] K. Stockhausen, *Telemusik*. Vienna: Universal Edition, 1969.
- [19] K. Stockhausen, "Nr. 22, Hymnen: elektronische und konkrete Musik.," in *Mittele-Partitur ("a score for reading")*. ed. Vienna: Universal Edition, 1967.
- [20] R. M. Schafer, *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster, 1993.
- [21] R. Rimbaud, "Concrete Resistance," *r/musicconcrete*, 2025/10// 2025. [Online]. Available: <https://scannerdot.com/concrete-resistance-with-robin-rimbaud/>.
- [22] M. Buffa and J. Music, "Real time guitar amplifier simulation and virtual pedal board using Web Audio," in *Proceedings of the 3rd Web Audio Conference (WAC)*, London, 2017.
- [23] M. Buffa, S. Ren, O. Campbell, T. Burns, S. Yi, J. Kleimola, and O. Larkin, "Web Audio Modules 2.0: An Open Web Audio Plugin Standard," in *Companion Proceedings of the Web Conference 2022 (WWW '22)*, Lyon, 2022, doi: 10.1145/3487553.3524225.
- [24] K. Mikolajczyk, S. Trolland, A. Ilisar, J. McCormack, S. Ferguson, and O. Bown, "Gestural interactions with object-based audio in an Internet of Sounds ecosystem," in *Proceedings of the 4th IEEE International Symposium on Internet of Sounds*, Pisa, 2023, doi: 10.1109/IEEECONF59510.2023.10335408.
- [25] M. Mazzocchio, "Design and implementation of an immersive, cooperative net art installation using Web Csound," in *Proceedings of the 4th IEEE International Symposium on Internet of Sounds*, Pisa, 2023, doi: 10.1109/IEEECONF59510.2023.10335251.
- [26] A. Boem and L. Turchet, "Musical metaverse playgrounds: Exploring the design of shared virtual sonic experiences on web browsers," in *Proceedings of the 4th IEEE International Symposium on Internet of Sounds*, Pisa, 2023, doi: 10.1109/IEEECONF59510.2023.10335297.
- [27] Eigenfeldt, A., & Pasquier, P. (2011, March). Negotiated Content: Generative Soundscape Composition by Autonomous Musical Agents in Coming Together: Freesound. In *International Conference on Computational Creativity* (pp. 27–32).

- [28] M. Thorogood and P. Pasquier, "Computationally generated soundscapes with Audio Metaphor," presented at the Proceedings of the International Computer Music Conference, Perth, 2013.
- [29] de Souza Stolfi, A., Ceriani, M., Turchet, L., & Barthelet, M. (2018). Playsound.space: Inclusive Free Music Improvisations Using Audio Commons. In Proceedings of the 18th International Conference on New Interfaces for Musical Expression, NIME 2018 (pp. 228-233).
- [30] D. Keller, V. Lazzarini, and M. S. Pimenta, *Ubiquitous Music*. Springer, 2014.
- [31] N. Collins, M. Schedel, and S. Wilson, *Electronic Music*. Cambridge, UK: Cambridge University Press, 2013.
- [32] P. Schaeffer, *À la recherche d'une musique concrète*. Paris: Éditions du Seuil, 1952.
- [33] S. Letz, Y. Orlarey, and D. Fober, "Compiling Faust audio DSP code to WebAssembly," in Proceedings of the 3rd Web Audio Conference (WAC '17), London, 2017.
- [34] I. Xenakis, *Formalized Music*. Bloomington, IN: Indiana University Press, 1971.
- [35] Concrète Collision, (26 tracks, 33minutes), 2026. https://wademarynowsky.art/concrete_tracks