The Readymade Synth: Prepared Synthesis with Everyday Objects

Antoni Rayzhekov
antoni.rayzhekov@plus.ac.at
University of Salzburg
Salzburg, Austria

Martin Murer
martin.murer@plus.ac.at
University of Salzburg
Salzburg, Austria

Figure 1: The Readymade Synth: a polyphonic hybrid instrument that transforms everyday objects into sound through *Prepared Synthesis*, using computer vision and spectral shape analysis.

Abstract

John Cage's prepared piano reimagined instruments as spaces where everyday objects reshape timbre and articulation. We extend this ethos digitally through introducing Prepared Synthesis, a method that treats the geometry of arbitrary, everyday objects as direct input for spectral sound generation. We present The Readymade Synth, an integrated instrument implementing this approach: A keyboard controls pitch and timing while a computer-vision-based system transforms objects placed on a sensing surface into live preparations. The objects' contours are decomposed into elliptic Fourier descriptors that define additive oscillator spectra. The objects' areas define the respective amplitude envelopes and object position controls stereo panning and mix balance. Any object becomes timbral material, and multiple objects blend spectral contributions. We formalise these as perceptually motivated feature-to-sound mappings and investigate the approach through a performance-oriented implementation. The Readymade Synth and a series of compositional explorations with a professional musician demonstrate how unmodified objects can function as repeatable yet flexible timbral material, producing coherent musical outcomes.

Keywords

hybrid musical instrument, everyday objects, tangible interaction, spectral sound synthesis, experimental composition

NIME '26, June 23–26, 2026, London, UK

1 Introduction

Inspired by the tradition of prepared instruments initiated by John Cage, in which objects radically alter timbre and become compositional acts, we aim to transpose that ethos into digital synthesis. We introduce *Prepared Synthesis*, a method that treats the geometry of everyday objects as input for spectral sound generation. We developed this approach through designing and implementing *The Readymade Synth*, a hybrid tangible-digital instrument named after Marcel Duchamp's readymades. Resonating with Duchamp's transformation of ordinary objects into art through recontextualization, our instrument transforms objects into sound through spectral mediation: any object on the sensing surface becomes a form of tangible timbre.

Tangible music interfaces have explored the use of objects, yet most rely on fiducial markers or predefined tokens [13, 21]. Similarly, graphical-sound lineages from Murzin's ANS to contemporary wavetable drawing (e.g., [19]) treat authored marks as sonic data, not found geometry. What remains unexplored is rendering *any* object's shape into spectral content, envelope, and spatial mix.

Our method addresses this gap by introducing *Shape-to-timbre* mapping synthesis, using elliptic Fourier descriptors [16] to decompose contours into frequency components that further drive additive oscillators, thereby transforming shape into sound. In addition, we apply *Area-to-envelope* mapping, translating object size into amplitude dynamics using an embodied metaphor: larger objects yield sustained attacks; small objects sound percussive, while their position controls stereo panning and mix balance. The *Readymade Synth* pairs computer vision with a MIDI keyboard. Through a compositional study with a professional musician and composer, we show how unmodified objects function as flexible yet reproducible sonic resources, making material choice a form of artistic control.

2 Background and Related Work

Early graphic sound systems, from Soviet optical experiments and Murzin's ANS [15] to Xenakis's UPIC [18] and MetaSynth [26], established visual-to-auditory translation using drawn marks. Contemporary successors include Peschke's geometrical oscillator [22]. Such systems operate on authored representations.

In our earlier work [24], we presented an approach that unfolds object contours directly into wavetable oscillators. By cutting the shapes horizontally into two halves and concatenating them, they mapped the horizontal position of contour points to time and the vertical displacement to amplitude. While intuitive, this literal translation can produce harsh, noisy output, disconnected from the shape's visual qualities, due to stretched concave regions or sharp corners in the contour, which can create discontinuities that generate broadband noise rather than a pitched timbre. Scanned synthesis [25] similarly reads a spatial pattern as a waveform, scanning the evolving state of a dynamic system rather than a captured contour directly.

Sonification literature offers guiding principles that inform, though do not fully specify, shape-to-sound mappings. Barrass [1, 2] argues that comprehensible auditory displays align data with perceptual structure: continuous dimensions suit continuous data, categorical distinctions better suit discrete categories. Dubus and Bresin's review [6] confirms that pitch and timbre are common targets for geometric data, yet fewer than seven percent of mappings have been empirically assessed. Tangible systems typically implement shape-to-sound categorically: PerForm [27] classifies shapes to trigger pre-assigned sounds. Shape selects timbre but does not constitute it.

Our approach in this paper differs: geometry constitutes sound rather than selecting it. Contour determines timbre, area shapes the envelope, and position controls spatialization. This positions our approach in the middle ground in Barrass's [2] taxonomy: not audification (contour is spectrally analysed, not played as waveform), yet more direct than parametric sonification because the shape's spectral structure transfers to the sound's spectral structure.

Previous research on digital instruments has examined how mapping structures shape instrumental character. Hunt et al. [12] showed cross-coupled mappings feel "more like an instrument" than one-to-one mappings, yet complexity risks opacity. Fels et al. [7] argue that mappings remain comprehensible when grounded in legible metaphors, what they term *mapping transparency*. Magnusson [17] and Gurevich [10] further argue that deliberate constraints help audiences perceive performer intent. All three positions concern what listeners can read off the instrument's mapping. *The Readymade Synth* makes this audience-side legibility a property of the design itself, since preparations sit on a visible surface rather than inside the instrument body. Our mappings (*shape-to-timbre*, *area-to-envelope*, *position-to-mix*) are structurally one-to-one yet perceptually rich, for example: with larger objects producing sustained envelopes and smaller objects sounding percussive.

Dahlstedt's Foldings [5] presents a related hybrid approach: keyboard playing controls electronic processing through dynamic vector mapping, preserving pitch and timing while opening timbre to expressive control.

Tangible music interfaces must navigate a persistent tension between openness and precision. Token-based systems such as the reacTable achieve precision through fiducial tracking but might limit spontaneity to pre-prepared objects [13]. Kiefer's deformable foam achieved gestural expressivity but sacrificed repeatability [14]. Holzmann and Hader's pressure-imaging exposed fundamental sensing limits [11]. Algorithmic corrections like Jam-O-Drum's quantisation were found to constrain [3]. We resolve this through separation of concerns: a MIDI keyboard handles pitch and timing, while objects handle timbre and envelope. Any outline becomes a unique timbral source without tagging, while the keyboard ensures a reliable musical structure.

3 The Prepared Synthesis and the Sonic Space of Tangible Timbres

In this section, we introduce our concept of *Prepared Synthesis* as an approach for translating any object's silhouette into sound. Our approach aims to create timbres that resonate with our perception of the objects' shapes. Such *tangible timbres* should prioritise musical coherence and the preservation of geometric structures. While we developed this approach together with the instrument presented in Section 4, which uses computer-vision-based contour extraction of object silhouettes, it can be applied to other methods of transforming shapes into timbres. We provide a generic implementation using SuperCollider that demonstrates the *Prepared Synthesis* approach using synthetic shapes.[1] In the following, we will introduce the core aspects of our synthesis approach: using (complex) Fourier descriptors to decompose contours, and translating them into coherent synthesis parameters via tailored *shape-to-timbre*, *area-to-envelope*, and *position-to-mix* mappings.

3.1 Contours as Complex Signals

A closed 2D contour can be parameterised periodically (e.g., by normalised arc length), resulting in a consistent ordered point sequence $x[n], y[n]$. While it is tempting to treat this sequence directly as an audio waveform, sample-by-sample playback is brittle: local corners and parameterisation artefacts can dominate the sound, and concavities may be over-represented depending on how the curve is traversed. Instead, we follow the approach of elliptic Fourier descriptors [9] – describing the contour as a complex signal $z[n] = x[n] + i\,y[n]$ – and derive synthesis parameters from its spectral decomposition.

By treating X and Y coordinates as real and imaginary components of a complex signal, we bypass the need to 'unfold' contours into 1D waveforms, a process that distorts concave regions and creates discontinuities at corners. As a result, we gain explicit control over which frequency components are fed into the sound synthesis process. Rather than using the contour as a waveform, we decompose its geometry into a spectral fingerprint. The contour is resampled (to either a fixed FFT size, e.g., $N = 2048$, or adaptive based on contour length) and a single FFT yields coefficients at positive and negative frequencies, following the complex contour parameterisation introduced by Granlund [9]. The resulting coefficients form an elliptic Fourier descriptor representation [16], in which each harmonic represents a counter-rotating phasor pair tracing an elliptic locus. We extract harmonic ratios, amplitudes, and phase relationships that drive a bank of sine oscillators performing additive sound synthesis to produce interesting, playable timbres.

Before analysis, we preprocess each contour by (1) centering it at its centroid, (2) resampling to a FFT size N (determined adaptively by object size to avoid oversampling small objects),

[1] The Readymade Synth GitHub repository: https://github.com/the-beyond-repair/the-readymade-synth

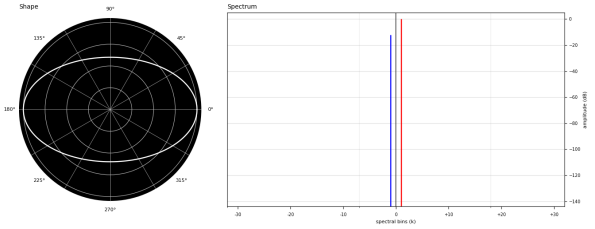

Figure 2: Spectral decomposition of an ellipse. Left: input shape. Right: FFT output showing positive (red) and negative (blue) frequency components.

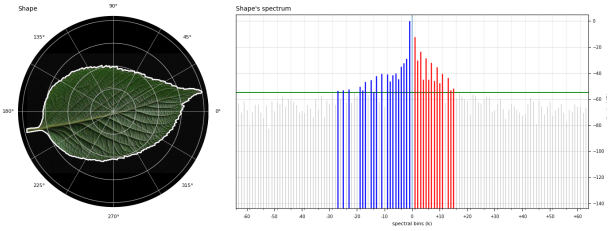

Figure 3: Spectral decomposition of a leaf. The leaf's complex contour produces a richer spectrum with more prominent partials than the ellipse (see Figure 2).

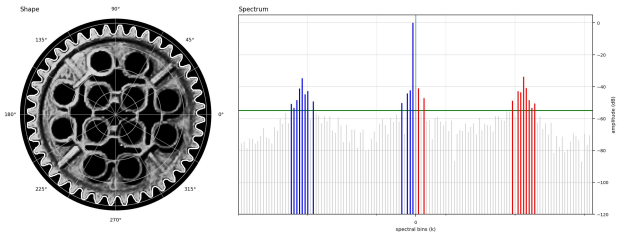

Figure 4: Spectral decomposition of a plastic gear. The gear's distinct angular contour produces a spectrum with more prominent partials in two groups: the gear overall round shape and its 40 teeth

and (3) enforcing consistent counter-clockwise winding so that raw input point order does not alter the descriptor.

3.2 Spectral Decomposition

When designing our *shape-to-timbre* mapping, we aimed for relationships to feel natural: a jagged gear should sound different from a smooth pebble, and that difference should be intuitive. Rather than arbitrary aesthetic choices, the shape's own structure should drive the timbre.

The contour's X and Y coordinates form the complex signal described in Section 3.1. Because the contour is encoded as a complex signal $z[n] = x[n] + i\,y[n]$, its Fourier transform decomposes the shape into rotating complex exponentials $e^{i2\pi kn/N}$. In this representation, "rotation" is literal: each coefficient C_k describes a vector in the complex plane that rotates around the origin as the contour parameter advances. Positive indices $+k$ correspond to counter-clockwise (CCW) rotations, while negative indices $-k$ correspond to clockwise (CW) rotations. These rotations are not metaphorical but geometric building blocks: each one traces a circular motion, and their superposition reconstructs the contour. For each spatial scale $|k|$, the pair $+k$ and $-k$ represents opposite directions of rotation at the same frequency. The first non-zero pair, $k = +1$ and $k = -1$, captures the dominant large-scale motion of the contour, which is its primary elliptical mode (see Figure 2). For richer contours such as a leaf, additional pairs at higher $|k|$ become prominent (see Figure 3), reflecting finer geometric detail. Their magnitudes $|C_{+1}|$ and $|C_{-1}|$ indicate how strongly this global rotation is expressed, and their phases ϕ_{+1} and ϕ_{-1} determine how these counter-rotating components align, describing the overall orientation and balance of the form.

Before any normalisation, we measure the global maximum amplitude across both spectra and discard bins that fall below a relative threshold. In our implementation, we used -55 dB as a threshold, empirically determined to separate perceptually meaningful spectral content from camera-induced noise artefacts. The surviving bins are then normalised across the full spectrum. Joint normalisation preserves the natural amplitude ratio between the positive and negative components and yields cleaner, more stable timbres while retaining the contour's directional asymmetry as a timbral feature. Independent normalisation of the two halves is a viable alternative when parity between CW and CCW components is musically motivated. From the normalised spectrum, we retain only the strongest N (configurable; we use $N = 64$ by default) bins by peak picking.

3.3 From Spectrum to Synthesis Parameters

To utilise this Fourier transformation for creating timbres, one particular challenge arises: the bidirectional rotational structure must be translated into the one-dimensional ordering of audible frequencies. At the same time, we aimed to maintain the harmonic relationships of the Fourier transform in the synthesis.

Using only one half of the spectrum would eliminate one direction of rotation, thereby collapsing the contour into a single-handed description and discarding a significant portion of its spatial information. We therefore explored different approaches to retain both directional halves and interleave them so that $+k$ and $-k$ (i.e., opposite directions of rotation at the same frequency) remain perceptually adjacent. Interleaving is implemented using an offset so that negative-half partials fall between positive-half partials, producing a denser partial series. We explored three offsets: (1) fixed offset of 0.5, (2) a phase-derived offset computed from the phase difference between the first positive and first negative bins, and (3) a magnitude-only symmetry-guided offset using cosine similarity between the positive and negative magnitude spectra. We adopted (2): the phase-derived offset produced the most perceptually expressive timbres in our explorations.

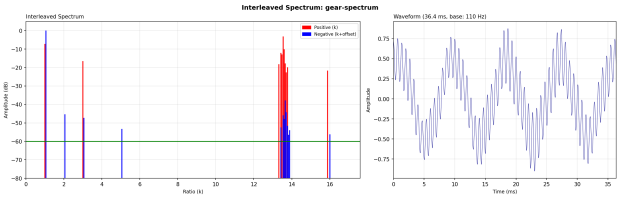

Figure 5: From spectrum to waveform. Left: interleaved positive/negative partials from the gear in Figure 4, with higher frequencies compressed to the audible range. Right: resulting waveform after additive synthesis with 64 sine oscillators.

As each partial ratio is multiplied by the played note's base frequency, the resulting frequencies might become inaudible or

exceed the Nyquist frequency of the audio rendering. To preserve geometric detail at higher pitches, our approach employs additional compression: We apply a simple Nyquist-derived threshold that compresses (0.25 power law) rather than discards. Its amplitude attenuates by the fourth power of the compression ratio, so heavily compressed partials fade rather than dominate (see Figure 5 for an example of the interleaved and compressed partials derived from the gear shown in Figure 4). We additionally apply perceptual loudness compensation (SuperCollider's UGen AmpCompA[2]) to compensate for perceived level across frequency.

3.4 Area-to-Envelope Mapping

While shape determines spectral content, the *area-to-envelope* mapping translates object size into amplitude dynamics. The envelope is driven by area alone, independent of contour structure, so two differently shaped small objects produce similar envelopes but distinct timbres. We use a two-stage attack-release shape rather than a full ADSR, with attack and release times scaling continuously with normalised area. This mapping follows natural intuition, motivated by ecological acoustics [8], where larger physical objects produce sounds with longer attacks and decays, whereas smaller objects sound percussive. For example, a small seed (area ∼0.5 %) triggers a 10 ms attack and 500 ms release; a large leaf (area ∼15 %) results in 500 ms attack and 5 s release. Between these extremes, the envelope morphs continuously, and a medium-sized object produces balanced, versatile articulation. This single parameter, extracted automatically from the contour's enclosed area, provides immediate expressive control without requiring explicit interface elements.

3.5 Position-to-Mix Mapping

When multiple objects occupy the sensing surface, each contributes independently to the output based on its position in polar coordinates. Distance from the center of the sensing surface controls amplitude, where central objects sound at full level, attenuating linearly toward the edge (center = 1.0, edge = 0.1). Angular position controls stereo panning: 0° (top) pans center, 90° pans hard right, 270° hard left, with 180° wrapping back to center. Together, these mappings enable spatial composition through physical arrangement, where clustering objects near the center creates dense, focused texture, while spreading them outward yields diffuse, attenuated contributions with stereo width.

Geometry and panning are largely decoupled in practice: the same object tends to yield similar spectral descriptors across angular positions, while its stereo placement changes. However, full independence is not guaranteed because edge placement simultaneously attenuates amplitude and introduces lens-dependent deformation, so level and spectral deformation remain partially coupled. When multiple objects are present, their spectral contributions blend according to the *position-to-mix mapping*; envelope contrast between small and large objects becomes a deliberate compositional technique, discussed in Section 5.3.

3.6 Spectral Output Summary

The complete analysis pipeline outputs a spectral descriptor set for each detected object: *frequency ratios* (harmonic positions), *amplitudes*, and *phases* for up to N partials.Combined with *area* (for envelope) and *polar position* (for mix and panning), these parameters fully specify a sound source derived from arbitrary geometry.

[2]SuperCollider AmpCompA UGen: https://doc.sccode.org/Classes/AmpCompA.html

Figure 6: *The Readymade Synth:* A 49-key MIDI controller provides pitch and expression. The circular sensing surface (40 cm diameter, matte black fabric) is captured by an overhead IR camera (3840×2160 px, 850 nm). An Arduino UNO handles auxiliary controls: capture button and a dedicated pedal enabling foot-triggered capture. Dimensions: 84×98×72 cm (W×H×D).

While our primary implementation uses additive synthesis, where partials directly drive sine oscillators, the spectral descriptors can inform other types of synthesis paradigms. Together, the mappings (*shape-to-timbre*, *area-to-envelope*, and *position-to-mix*) constitute the *Prepared Synthesis* method. Any object placed on the sensing surface becomes a complete sound source, where its silhouette determines timbre, its size determines articulation, and its position defines spatial presence. In this sense, object selection becomes timbral composition, extending the prepared piano's technique into the digital-tangible domain. Whereas Cage's prepared piano concealed its preparations inside the instrument body, our approach foregrounds them as part of the performance, shifting the locus of preparation from a private compositional act to a visible one.

4 The Readymade Synth

To investigate the idea of translating the tradition of the prepared instrument to the hybrid domain and explore its sonic and compositional potential, we developed a performance-oriented hybrid instrument integrating a MIDI keyboard with a computer-vision system (see Figure 6) that draws on the aesthetics of early 1970s synths. This section describes our design goals, the collaborative process that shaped the final implementation and the system architecture.

4.1 Design goals and process

We named our instrument *The Readymade Synth* after Duchamp's readymades. Where Duchamp's selection turned ordinary objects into art, our instrument turns them into timbre through spectral

analysis. This focus informed every design decision we made, from materials and signal routing to the choices of *shape-to-timbre* mappings: the instrument is built for stage performance, where reliability and reproducibility serve the performer's compositional intentions.

Our goal was not to digitise the prepared piano but to extract its core idea – that material selection is itself composition – and transpose it into a domain where computation mediates between object and sound. This required an architecture that preserves what makes tangible interaction powerful: direct manipulation of familiar everyday objects, embodied exploration, and material significance, while providing what professional musical performance and digital audio synthesis demand: precise pitch control, repeatable timing and dynamic expression. We therefore split control by design: a MIDI keyboard provides pitch, timing, and velocity, while unmodified everyday objects on a circular tabletop shape timbre, envelope, stereo field and spectral character.

An overhead infrared camera captures objects placed in the sensing surface, enabling the performer to "construct" sound by arranging and moving objects rather than editing parameters on screens or manipulating traditional knobs, sliders, and patch cables. Geometrical features, such as: shape contour, centroid, area, distance to the centre of the circular sensing surface, are mapped to sound synthesis parameters, where the objects' outlines are transformed into spectral analytical data and envelopes for additive synthesis. Our implementation extends previous system design [24], omitting the overhead projector to focus attention mainly on *shape-to-timbre* transformation and tangible manipulation alongside the keyboard playing. We use a high-resolution compact IR camera and custom computer-vision software developed in vvvv[3] with the popular computer-vision library OpenCV [23], where extracted features are sent via Open Sound Control (OSC)[28] to SuperCollider[4] for analysis and sound synthesis.

Barrel distortion from the wide-angle lens deforms objects, which we treat as a timbral modifier rather than correcting it computationally (see Section 3.5). Multiple-object handling likewise exploits the circular geometry: distance from the centre sets amplitude in the overall mix, while angle (in polar coordinates) determines stereo panning, wrapping at 180°.

Our design aimed for a performance-oriented artefact in which oscillators, mixers, and envelope generators are replaced with unmodified tangibles, examining how such an approach can combine precision, variation, and artistic expressivity, interconnected with the cultural context offered by the objects' materiality.

4.2 System Architecture

The instrument integrates three elements into a single performable unit: a MIDI keyboard for pitch and timing, a sensing surface for object placement, equipped with an overhead IR camera for computer vision, and system-specific dedicated controls. The instrument's three integrated components communicate through the well-established protocols OSC and MIDI (see Figure 7).

4.2.1 The Input layer. A 49-key MIDI controller provides velocity-sensitive note input, sustain pedal, pitch bend, volume slider, and modulation wheel, all familiar keyboard controls. Three additional controls supplement the keyboard: a capture button that triggers vision analysis, a three-way switch (selecting the spectral interleaving mode: based on a fixed offset, phase, or symmetry), a slider controlling the spectral morphing time between captures,

[3]VVVV Gamma: https://visualprogramming.net
[4]SuperCollider: https://supercollider.github.io

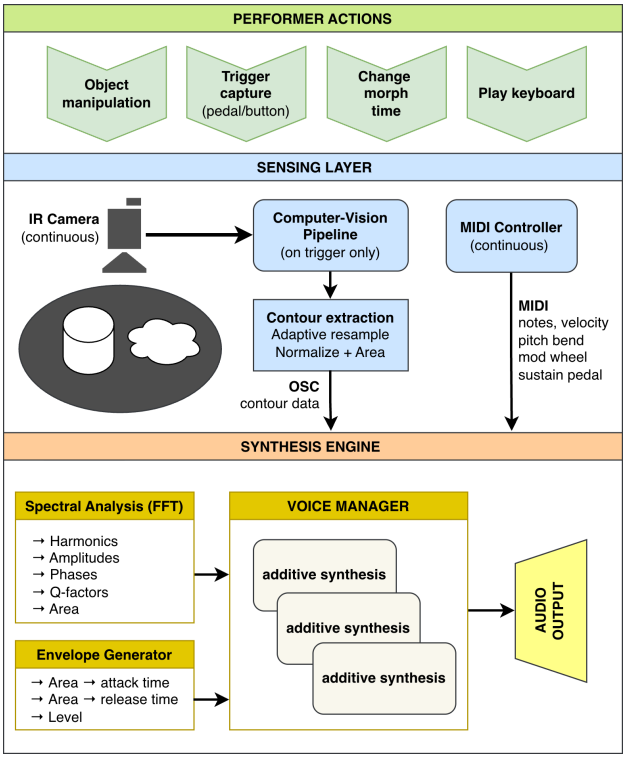

Figure 7: System architecture: Performer actions feed two pathways: continuous MIDI input and triggered vision capture. The computer-vision pipeline extracts contours and area; SuperCollider performs spectral analysis and renders audio. Object polar position controls amplitude (distance) and panning (angle).

and a system reset button. A second pedal duplicates capture functionality, freeing hands during complex passages.

4.2.2 The vision layer. The computer-vision component uses an overhead IR camera that captures the circular sensing surface (40 cm in diameter) at 3840×2160 px resolution, cropped and masked to a 2160×2160 px circle. An 850 nm infrared ring ensures robust capture by providing consistent lighting across performance environments. The matte black cotton fabric surface maximises contrast, delivering anything brighter to register as foreground. As described in Section 4.1, barrel distortion and polar coordinate mapping provide additional timbral and spatial control. We adaptively resample contours, scaling point count with object size to avoid oversampling small objects while preserving detail in larger ones. For example, a small object (a seed) may yield 32 points and a large leaf 512. This "adaptive resolution" ensures consistent spectral density across the object size range, preventing small objects from producing artificially complex spectra due to excessive sampling, while retaining geometric fidelity for larger, more detailed outlines. Additionally, we limit the smallest detectable object size to 5 mm × 5 mm, since resolution below that size does not allow us to extract meaningful contour data.

The system assumes connected, high-contrast silhouettes. Objects with internal voids (rings, damaged leaves) register as multiple independent contours. This is a deterministic property of the contour-extraction logic rather than a stochastic failure, so

performers can predict it from the object and exploit it compositionally. Fragmented contours create unexpected polyphony, and partial detection produces timbres distinct from a more "clean" form.

4.2.3 The Sound engine. The sound engine is implemented in SuperCollider and maintains a polyphonic voice bank, using SuperCollider's DynKlang UGen[5], driven by MIDI notes; each active note renders an additive spectrum parameterised by the currently captured object set. In the current instrument, we support up to 32 detected objects and up to 10 concurrent MIDI notes, with spectral updates applied to sustaining notes without re-triggering, enabling continuous timbral morphing over a controllable transition time (0.01–10 s).

4.2.4 Dataflow. The data flow is straightforward: three camera frames are averaged to reduce noise, then enter the computer-vision pipeline, which thresholds the frame and extracts and normalises object contours. The pipeline calculates geometric data (normalised outline, area, bounding box, and centroid), and transmits them as OSC messages. SuperCollider receives these messages, performs spectral analysis, updates synthesis parameters, and renders audio in response to MIDI input. Shape data is transmitted when the performer explicitly triggers capture via a pedal or button, treating the capture gesture as a compositional decision rather than an automatic process. Object movement during held notes does not affect timbre until the next capture is manually triggered, so performers can reposition objects "silently", then commit (capture) at musically appropriate moments.

4.2.5 Latency considerations. Latency accumulates across the signal chain from camera capture to computer-vision processing, OSC data encoding and transmission, spectral analysis, and audio rendering. We measured latency on our development workstation (24-core Windows 11, dedicated GPU, vvvv Gamma 7.0, SuperCollider 3.14 (Supernova) with 256-sample buffer at 48 kHz) with up to 64 partials per voice.

The musical response latency (*NoteOn-to-sound*) is ∼20 ms and thus within perceptual limits for live performance of a keyboard instrument (as confirmed by the participating external musician in the project), ensuring that pitch and timing remain tight. The total latency for spectral timbre updates (*capture-to-data-ready* plus *NoteOn-to-sound*) of ∼94 ms, possibly is comparable to turning several knobs simultaneously rather than playing a note, and also acceptable given the computational load of up to 64 partials per voice. The system scales efficiently: adding objects incurs only ∼0.5 ms per object, and spawning additional voices adds ∼0.2 ms per voice.

5 Compositional Exploration

The design and development of the *Readymade Synth* and the *Prepared Synthesis* followed an iterative art-led process intertwining phases of conceptualising, building, playing and refinement. Both authors are practising artists, so technical decisions and artistic intuitions developed together rather than in sequence. However, such processes can introduce a methodological tension, outlined by McPherson and Kim [20] as the *problem of the second performer*, noting that NIME instruments often remain confined to the intentions and views of their creators, limiting claims about learnability and expressive range. We addressed this by inviting an external collaborator, the professional musician and composer Matthias Leboucher[6], to explore the compositional potential of the instrument. Deliberately positioned at a phase in the process when the instrument was already fully functional and playable, yet still open to refinements, we worked together across iterative sessions. The collaborating musician has over 15 years of experience in experimental and electroacoustic music, including prior work with prepared piano, electronic instruments, and sound-art installations. This background provided an informed critique, while unfamiliarity with our instrument allowed us to explore its learnability and expressive potential.

During five sessions spanning four weeks (approximately 20 hours in total), conducted in a rehearsal-studio setup, the external musician took the part of the leading performer. The compositional exploration took place in close collaboration with the first author, who also observed, took field notes, and implemented modifications – sometimes during sessions for minor adjustments, sometimes between sessions for larger architectural changes.

Rather than formal coding of transcripts, our analysis followed a reflective practice-based approach [4], where, after each session, we discussed what worked, what felt awkward, and what the system afforded or constrained. These reflections directly informed both instrument refinements and compositional input. The resulting composition emerged from this dialogue between artistic intention and professional experience, the researcher's input, and the instrument's capabilities. The collaborative exploration with the invited collaborator finally culminated in an open rehearsal of the composition (see supplementary video for a video documentation excerpt). For this open rehearsal, we hosted 15 people from diverse backgrounds, including several professional musicians, producers, visual artists and experts in HCI.

5.1 In Advance of the Broken Sine

The composition *In Advance of the Broken Sine for Readymade Synth, three plastic gears, four geometric shapes, pavement grit, a white thread, and a handful of leaves*, spans approximately 20 minutes of structured improvisation in three movements corresponding to material classes: industrial (gears, plastic forms), transitional (white thread), and organic (pavement grit, leaves). The keyboard provides melodic and harmonic continuity while objects supply timbral evolution. The piece begins with an empty sensing surface, and thus a pure sine wave, then builds toward complex compound shapes through gradual object introduction, then objects spread, and it ends again with an empty surface.

Objects were curated for both geometric and semantic properties. Manufactured artifacts (gears, folded paper geometric forms) offered repeatable contours; pliable materials, such as the white thread, allowed fine real-time adjustments to their shape; natural objects (leaves, pavement grit) introduced organic variation, as no two leaves produce identical spectra. The performer assembled an "object palette" analogous to instrument patches, but one carrying material associations: a metal gear evokes different meanings than a handful of pavement grit. These interwoven qualities – geometric (what it sounds like) and semantic (what it suggests) – informed both rehearsal and the compositional arc described above.

[5]SuperCollider DynKlang UGen: https://doc.sccode.org/Classes/DynKlang.html

[6]Matthias Leboucher homepage: https://www.matthiaslebocher.com

5.2 Instrument Refinements

Several refinements of the instrument emerged directly from the collaboration and were integrated into the concept and implementation. During the collaborative sessions, explicit capture became evident as a crucial compositional gesture, a deliberate "commit" to the new timbral state – rather than a background process triggered by inactivity on the sensing surface, as in our earlier system [24]. In addition to a dedicated button, we implemented a pedal, freeing the hands for playing the keyboard and manipulating objects.

The morphing time between two timbral states emerged as a very expressive parameter. The addition of an according slider allows control of the transition duration between subsequent capture states, enabling both abrupt timbral shifts and gradual crossfades. In the instrument, we repurposed the keyboard level slider and mapped it to morph time (0.01–10 s range).

In addition, a resonant low-pass filter (RLPF) was mapped to the mod-wheel as part of the original system design. For the particular compositional exploration and open rehearsal, we adjusted the pitch-bend control to be continuous (±1 octave, exponential) rather than discrete steps, and added additional output processing: subtle reverb and a limiter.

5.3 Emergent Performance Techniques

Coordinating three simultaneous activities, object manipulation, capture timing, and keyboard playing, required a three-limb independence similar to organ technique. Rapid object-capture-play sequences required extended practice. Learning which objects produce coherently pitched timbres versus noisy or thin results also demanded experimentation. After the first session (approximately four hours), the performer had internalised when to use the capture button versus the pedal and developed intuition about how shapes might sound.

Rehearsal proceeded in three phases. First, exploring individual objects, learning their timbral characters and envelope behaviours. Second, choreographing object sequences: which materials enter when, how they accumulate, when they exit, and which playing technique fits the timbre (e.g., staccato for small objects, sustained chords while rearranging objects). Third, composing keyboard parts in accordance with this object dramaturgy. Object pairing emerged as a deliberate timbral sculpting technique. A small, edgy object contributes sharp attack and bright spectral content, while a large, rounded object contributes sustained release and darker, broader spectra. Placing both at the same time on the sensing surface layers their contributions: the small object "strikes," the large object "resonates." This affords composite timbres through material selection rather than parameter adjustment.

The capture gesture itself emerged as a rhythmic element: the performer developed timing intuitions, treating the pedal as an expressive control. An unexpected technique, "micro-capture sequencing," arose when the performer attached an object to a thread, made small positional adjustments, and rapidly triggered captures, creating sequences of related but varied timbres from a single object (see Figure 8).

By the third session, the performer began arranging objects not only for their sonic properties but also visually, "drawing" abstract faces on the sensing surface. The spatial layout became a score-like composition visible to the audience, adding a performative dimension beyond sound, situating the sensing surface as both instrument and visual canvas.

Figure 8: An unexpected technique emerged during the early exploratory composition sessions: The performer attached an object to a thread and made small positional adjustments. Rapidly triggering captures created sequences of related, but varied timbres from a single object.

6 Discussion

Our work introduces a novel synthesis method – termed *Prepared Synthesis* – that combines mappings designed to be perceptually intuitive with a performance-oriented instrument. These are interconnected: the instrument embodies the method, the mappings make it playable, and the compositional exploration demonstrates how the whole coheres as artistic practice.

The core of *Prepared Synthesis* is treating object contours as complex signals for spectral decomposition. This reframes *shape-to-sound* not as recognition (identify the object, trigger a sound) but as a form of audification (the shape is the sound). Where prior tangible systems mapped detected shapes to pre-assigned timbres (e.g., [14]), and our prior work [24] unfolded contours directly into wavetables, the FFT-based approach presented here extracts a spectral fingerprint, such as harmonic ratios, amplitudes, and phases, which drive additive synthesis. The result is a stable, pitched timbre that aims at preserving geometric structure: a gear sounds jagged; an ellipse sounds smooth; a leaf sounds complex and rich.

Complementing this spectral mapping, we ground additional parameters in embodied metaphors and spatial intuition. *Area-to-envelope* maps object size to temporal behaviour: small objects are percussive; large objects sustain. *Position-to-mix* maps object placement in polar coordinates: distance from center controls amplitude balance (center is loudest, edges fade), while angle determines stereo panning (wrapping at 180°). These mappings require no explanation, so performers and audiences can perceive the relationships intuitively, achieving what Fels et al. term "mapping transparency" [7]. Together, *contour-to-spectrum*, *area-to-envelope*, and *position-to-mix* constitute a minimal but complete transformation from physical form to sonic behaviour: shape determines timbre, size determines articulation, and placement determines spatial presence.

The system captures 2D silhouettes only. Displacing the object from the camera's optical axis introduces barrel distortion; hence, the same 3D object at different positions and orientations produces different timbres. Utilising this as a feature seems meaningful. However, our current design couples it with the

amplitude controlled by the polar distance to the center of the sensing surface (used in the position-to-mix).

Our practice-based approach documents one expert musician's adoption but also hints at the potential for further playing techniques: For instance, micro-capture sequencing (Section 5.3, Figure 8), showed that capture timing itself affords rhythmic variation independent of keyboard input. Visual arrangement also emerged as more central to performance practice than anticipated, with the instrument's surface serving as a visible score.

The open rehearsal (see the supplementary video for an excerpt of the documentation) motivated the audience to freely explore the instrument for approximately two hours and led to a multitude of curiosity-driven experiments with whatever objects were at hand. The anecdotal feedback collected from those explorations suggested a variety of future potentials of *Prepared Synthesis*, ranging from sound design for performance and music production to applying our approach at different stages of modular synthesis.

The hybrid control model between arbitrary tangibles and precision input may transfer beyond music to design tools or educational interfaces. More fundamentally, the instrument foregrounds the materiality of everyday objects: a plastic gear, a small stone, a fallen leaf carry different associations that *Prepared Synthesis* renders audible.

7 Conclusion and Future Work

The Readymade Synth transforms unmodified arbitrary objects into sound sources through *Prepared Synthesis* informed by spectral shape analysis, which can be situated as an extension of the conceptual art tradition of prepared instruments. By separating precise control (keyboard) from exploratory control (objects), the instrument balances repeatability with discovery. Our compositional exploration with a professional musician demonstrated that object choice becomes a compositional act, capture timing emerges as a performative gesture, and material dramaturgy structures musical form. In addition, our approach offers a transferable design pattern for negotiating precision and openness in tangible musical interfaces.

Beyond the technical transformation lies a deeper contribution: the performer's choice of objects, whether a plastic gear, a thread, or a fallen leaf, carries material associations into the sonic domain. Where Cage's prepared piano concealed its preparations inside the instrument body, ours sit visibly on the sensing surface, so the audience can read the gesture-to-sound link alongside the performer. In this sense, *Prepared Synthesis* extends the prepared piano's central insight: that what we place inside an instrument defines not only its sound but also suggests its meaning.

In future work, we plan to produce a fully self-contained unit around an embedded computer to simplify stage setup. In parallel, we plan to expose the OSC (analytical spectral data) and MIDI streams for connection to external systems (e.g., DAWs, modular synths). We also plan to further explore additional synthesis paradigms, for instance, subtractive and FM synthesis driven by spectral ratios, and physical modelling through banks of resonators.

8 Ethical Standards

Informed consent was obtained from the collaborating musician for participation in the compositional exploration and for publication of photographic and video documentation. The collaborator, a professional musician and composer with expertise in experimental music and sound art, engaged voluntarily and retained artistic agency throughout the process, including co-determination of the final composition. They also received remuneration in line with the national guidelines for fees in the cultural sector. The research presented in this paper, in conformity with the ethical standards of the hosting university, did not require formal ethics board approval. A generic implementation of our presented approach is publicly available[7] under the GPL 3.0 license.

Acknowledgments

The research presented in this publication has been supported by the III.Lab (Intelligent Interfaces Innovation) funded by AIT - Austrian Institute of Technology, University of Salzburg and Land Salzburg under grant number 20102/F2300416-KZP.

We thank Matthias Leboucher for his essential contribution as collaborating performer and composer, and for their patience with an evolving instrument.

References

[1] Stephen Barrass. 1996. Sculpting a sound space with information properties. *Organised Sound* 1, 2 (1996), 125–136. https://doi.org/10.1017/S1355771896000040

[2] Stephen Barrass and Paul Vickers. 2011. Sonification Design and Aesthetics. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 7, 145–171. http://sonification.de/handbook/chapters/chapter7/

[3] Tina Blaine and Tim Perkis. 2000. The Jam-O-Drum interactive music system: a study in interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques (DIS00)*. ACM. https://doi.org/10.1145/347642.347705

[4] Benjamin Carey and Andrew Johnston. 2016. Reflection On Action in NIME Research: Two Complementary Perspectives. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Queensland Conservatorium Griffith University, Brisbane, Australia, 377–382. https://doi.org/10.5281/zenodo.1176006

[5] Palle Dahlstedt. 2015. Mapping Strategies and Sound Engine Design for an Augmented Hybrid Piano. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (Baton Rouge, Louisiana, USA) *(NIME 2015)*. The School of Music and the Center for Computation and Technology (CCT), Louisiana State University, Baton Rouge, Louisiana, USA, 271–276.

[6] Gaël Dubus and Roberto Bresin. 2013. A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLOS ONE* 8, 12 (12 2013), 1–28. https://doi.org/10.1371/journal.pone.0082491

[7] Sidney Fels, Ashley Gadd, and Axel Mulder. 2002. Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound* 7, 2 (Aug. 2002), 109–126. https://doi.org/10.1017/s1355771802002042

[8] William W. Gaver. 1993. What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception. *Ecological Psychology* 5, 1 (1993), 1–29. https://doi.org/10.1207/s15326969eco0501_1

[9] G. H. Granlund. 1972. Fourier Preprocessing for Hand Print Character Recognition. *IEEE Trans. Comput.* C-21, 2 (1972), 195–201. https://doi.org/10.1109/TC.1972.5008926

[10] Michael Gurevich, Paul Stapleton, and Adnan Marquez-Borbon. 2010. Style and Constraint in Electronic Musical Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Sydney, Australia, 106–111. https://doi.org/10.5281/zenodo.1177785

[11] Clemens Holzmann and Andreas Hader. 2010. Towards tabletop interaction with everyday artifacts via pressure imaging. In *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction* (Cambridge, Massachusetts, USA) *(TEI '10)*. Association for Computing Machinery, New York, NY, USA, 77–84. https://doi.org/10.1145/1709886.1709901

[12] Andy Hunt, Marcelo M. Wanderley, and Matthew Paradis and. 2003. The Importance of Parameter Mapping in Electronic Instrument Design. *Journal of New Music Research* 32, 4 (2003), 429–440. https://doi.org/10.1076/jnmr.32.4.429.18853

[13] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction (TEI07, 8)*. ACM, New York, NY, USA, 139–146. https://doi.org/10.1145/1226969.1226998

[14] Chris Kiefer. 2010. A malleable interface for sonic exploration. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 291–296.

[7] The Readymade Synth GitHub repository: https://github.com/the-beyond-repair/the-readymade-synth

[15] Stanislav Kreichi. 1995. The ANS Synthesizer: Composing on a Photoelectronic Instrument. *Leonardo* 28, 1 (1995), 59–62. http://www.jstor.org/stable/1576159

[16] Frank P Kuhl and Charles R Giardina. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* 18, 3 (1982), 236–258. https://doi.org/10.1016/0146-664X(82)90034-X

[17] Thor Magnusson. 2010. An Epistemic Dimension Space for Musical Devices. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Sydney, Australia, 43–46. https://doi.org/10.5281/zenodo.1177837

[18] Gérard Marino, Marie-Hélène Serra, and Jean-Michel Raczinski. 1993. The UPIC System: Origins and Innovations. *Perspectives of New Music* 31, 1 (1993), 258–269. http://www.jstor.org/stable/833053

[19] Kyosuke Nakanishi, Paul Haimes, Tetsuaki Baba, and Kumiko Kushiyama. 2016. NAKANISYNTH: An Intuitive Freehand Drawing Waveform Synthesiser Application for iOS Devices. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Zenodo, 143–145. https://doi.org/10.5281/zenodo.1176086

[20] Andrew P. McPherson and Youngmoo E. Kim. 2012. The Problem of the Second Performer: Building a Community Around an Augmented Piano. *Computer Music Journal* 36, 4 (12 2012), 10–27. https://doi.org/10.1162/COMJ_a_00149

[21] James Patten, Ben Recht, and Hiroshi Ishii. 2006. Interaction techniques for musical performance with tabletop tangible interfaces. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Hollywood, California, USA) *(ACE '06)*. Association for Computing Machinery, New York, NY, USA, 27–es. https://doi.org/10.1145/1178823.1178856

[22] Joshua Peschke and Axel Berndt. 2017. The Geometric Oscillator: Sound Synthesis with Cyclic Shapes. In *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences* (London, United Kingdom) *(AM '17)*. Association for Computing Machinery, New York, NY, USA, Article 15, 6 pages. https://doi.org/10.1145/3123514.3123522

[23] Kari Pulli, Anatoly Baksheev, Kirill Kornyakov, and Victor Eruhimov. 2012. Real-time computer vision with OpenCV. *Commun. ACM* 55, 6 (June 2012), 61–69. https://doi.org/10.1145/2184319.2184337

[24] Antoni Rayzhekov and Martin Murer. 2025. The Well-Tempered Objects: Using Everyday Things for Tangible Composition. In *Proceedings of the Nineteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '25)*. Association for Computing Machinery, New York, NY, USA, Article 80, 7 pages. https://doi.org/10.1145/3689050.3705994

[25] Bill Verplank, Max Mathews, and Rob Shaw. 2001. Scanned synthesis. *The Journal of the Acoustical Society of America* 109, 5 Supplement (05 2001), 2400–2400.

[26] Eric Wenger. 1997. MetaSynth [Computer software]. U&I Software and Arboretum Systems, San Francisco, CA. https://uisoftware.com/metasynth/ Sound design and composition environment using images to generate/process sound (Macintosh)..

[27] Irmandy Wicaksono, Caroline Rozendo, Runzhou Ye, Jaleesa Trapp, V. Michael Bove Jr., Canan Dagdeviren, and Hiroshi Ishii. 2018. PerForm: Deformable Interface for Exploring Sound through Shapes. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188478

[28] Matthew Wright and Adrian Freed. 1997. Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. In *International Computer Music Conference (ICMC)*. International Computer Music Association (ICMA), International Computer Music Association (ICMA), Thessaloniki, Hellas, 101–104. http://cnmat.berkeley.edu/http://hdl.handle.net/2027/spo.bbp2372.1997.033/open_sound_control_new_protocol_communicating_sound_synthesizers