Towards Sensory Attenuation as a Measure of Agency Provided by Digital Musical Instruments

Erik Peralta Løvaas
erikpl@protonmail.com
Max Planck Institute for Informatics
Saarland Informatics Campus
Saarbrücken, Germany

Courtney N. Reed
c.n.reed@lboro.ac.uk
Digital Technologies &
Creative Futures
London, United Kingdom

Paul Strohmeier
pastrohm@mpi-inf.mpg.de
Max Planck Institute for Informatics
Saarland Informatics Campus
Saarbrücken, Germany

Abstract

Questions of agency and authorship become central to the design of expressive digital musical instruments (DMIs). Agency is recognized as essential for meaningful musical performance but remains difficult to operationalize and measure. We here propose sensory attenuation as an implicit measure of self-attribution and a practical proxy for agency in DMI interaction. We implemented a pressure-based control system for a digital synthesizer, introducing nonlinear behaviors along pressure and time dimensions. Using a psychophysics experiment combined with qualitative interviews, we investigate how these nonlinearities affect performers' sense of agency. Results suggest that sensory attenuation provides a viable proxy for self-attribution, but only when output exceeds a volume threshold. Furthermore, we observe interaction effects between temporal and action nonlinearities, indicating that maintaining agency requires control across both dimensions. Our findings provide empirical grounding for using sensory attenuation for experimental agency evaluation methods.

Keywords

Agential experiences, self-attribution, linearity, parameter mapping, gesture, expression

ACM Reference Format:
Erik Peralta Løvaas, Courtney N. Reed, and Paul Strohmeier. 2026. Towards Sensory Attenuation as a Measure of Agency Provided by Digital Musical Instruments. In *Proceedings of International Conference on New Interfaces for Musical Expression (NIME '26).* ACM, New York, NY, USA, 13 pages.

1 Introduction

When playing a physical instrument, there is a complex, nonlinear relationship between the musician's actions and the resulting sound. In digital musical instruments (DMIs), this relationship becomes malleable, largely determined by the design of mappings between performer input and sound generation [33, 48]. Mapping strategies define how gestural variation translates into changes in timbre, dynamics, and other expressive-related cues [24, 25]. Simple one-to-one mappings offer transparency and ease of learning but may restrict expressive flexibility [11]; in contrast, more complex approaches—including one-to-many mappings, nonlinear response functions, and stateful control dynamics—have been explored to enable richer and more acoustic-like behavior [8, 34]. With the increasing proliferation of generative AI (GenAI) systems [4, 42, 53], mappings in creative tools can become more opaque and less directly traceable from action to sound. As AI systems take a larger role in music creation, the role and agency of the musician can become blurred [55]. While the work presented here is inspired by debates around the use of GenAI in art, the current work examines a more controllable and better understood manipulation of input mappings.

This trend inspires us to examine agency in interaction with DMIs, particularly in a specific aspect put into focus by the proliferation of AI systems: the *subjective experience of agency.* There is a fundamental tension in the design of input-to-sound mappings: while complex mappings can expand expressive possibilities, they may also weaken action to sound correspondence [24, 37], potentially making it harder for performers to anticipate how their actions shape sound. Prior work suggests that nonlinear or state-dependent mappings can increase learning demands and reduce predictability [21, 39]; however, they are also desirable as they provide expressivity and are often used as a stylistic element [39, 41], such as in feedback musicianship [31, 38, 40]. To ground this discussion in empirical observation, a method for measuring performer agency when using arbitrary instruments and mappings – with or without AI involvement – is necessary.

The *sense of agency* is often defined as the experience of controlling one's actions and their consequences [36]. A distinction is commonly made between the *judgment* of agency, which happens after the fact and is often influenced by context (e.g., beliefs about causality, social cues) and the *sense* of agency, which is the pre-reflective experience one has in the moment, before it is shaped by judgment [10?]. Agency judgments can be collected through self report. However, measuring the pre-reflective sense of agency those judgments are based on is more difficult. The go-to measure for assessing pre-reflective agency is *temporal binding* [2, 6, 10], but this method is not compatible with assessing continuous music performances. An alternative approach is inspired by *comparator approaches* to agency [16]. A core assumption in comparator models is that the sensory feedback we receive is compared to our mental model of the expected sensory feedback, and that, if these match, then stimuli are attenuated; that is, they are experienced less salient. While this does not cover the full scope of agency, it captures the sense of self-attribution as a relevant proxy measure in this context. While sensory attenuation is well understood and has featured in HCI design [50], it has so far not been directly used as an evaluation method.

Consequently, we investigate whether sensory attenuation can serve as an implicit measure of agency with DMIs. Specifically, we present an experiment in which we investigate (a) if magnitude estimation can capture attenuation effects, (b) how different action to sound mappings might effect such attenuation, and (c) how playback volume effects attenuation. We complement magnitude estimation data with qualitative interviews to contextualize the results.

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

2 Related Work

This work draws on agency research, implicit perceptual measures, and DMI interaction. We first review theoretical models of agency and approaches for measuring pre-reflective agency. We discuss limitations of commonly used measures such as temporal binding in continuous interaction contexts, motivating sensory attenuation as a methodological alternative. Finally, we position DMIs as environments where predictive mechanisms and mapping design shape agency-related perception.

2.1 Agency

The sense of agency refers to the experience of controlling one's actions and their effects in the world [20]. Beyond simple motor execution, agency encompasses attribution, responsibility, and the belief that one's intentions causally shape outcomes [6]. Psychology and HCI research show that agency is highly sensitive to interaction design, including factors such as feedback latency [23, 32], system autonomy [1, 37], and input modality [29, 58].

Agency is often modeled using predictive frameworks like the comparator model, which proposes that internally generated predictions of sensory consequences are compared with incoming feedback to determine self-attribution [16, 17]. When prediction and perception align, outcomes are experienced as self-caused; discrepancies may instead be attributed to external factors. Complementary accounts emphasize inferential processes, suggesting that agency judgments may arise from higher-level reconstruction of causal relationships between intentions, actions, and effects [61].

We adopt a two-level account distinguishing a *pre-reflective* sense of agency from *reflective* agency judgments [10]. The *pre-reflective* sense emerges from sensorimotor processing and predictive mechanisms, while *reflective* judgments incorporate contextual interpretation, beliefs, and goals. Because interface design primarily shapes sensorimotor contingencies, implicit measures of pre-reflective agency are particularly relevant for evaluating interactive systems.

2.2 Agency Measurement in Generative AI Interaction

The need for robust implicit measures of agency becomes particularly pressing in the context of GenAI systems [4, 10, 42]. Unlike traditional instruments or interfaces, GenAI tools often introduce partial autonomy, stochastic behavior, and opaque transformations between user input and system output. These characteristics can blur authorship and weaken users' ability to attribute outcomes to their own actions.

In creative domains such as image or music generation, agency relates not only to usability, but also to ownership, responsibility, and artistic identity. Explicit self-reports may be insufficient because users may rationalize outcomes after the fact; for example, when rationalizing if an award-winning image was created by GenAI or the person who selected the image from the GenAI output[49]. Implicit perceptual measures therefore provide a more robust way to examine how system behavior shapes underlying experiences of control, independent of conscious interpretation.

2.3 Measuring Pre-Reflective Agency

Two implicit perceptual phenomena are commonly used as proxies for pre-reflective agency: temporal binding and sensory attenuation.

2.3.1 Temporal Binding. Temporal binding refers to the compression of perceived time between an action and its outcome, where causally related events are experienced as closer together than they objectively are [19]. Measurement approaches include Libet clock paradigms and interval estimation tasks, both of which quantify perceived delays between discrete action–effect pairs. Temporal binding has been widely used in HCI research because it provides a quantifiable perceptual measure without relying solely on self-report [6].

While widely used as an implicit measure of agency, temporal binding experiments rely on discrete action–effect pairs, such as a button press followed by a delayed outcome. This assumption conflicts with many forms of musical interaction. Musical performance often involves continuous modulation with ongoing control loops linking gesture and sound [56, 57]. Complex mappings, layered synthesis processes, and dynamic feedback make it difficult to define clear causal intervals between action and outcome. As a result, temporal binding may fail to capture the experiential qualities most relevant to expressive instrument use. These limitations motivate alternative measures that operate within continuous interaction dynamics.

2.3.2 Sensory Attenuation as an Alternative Measure. Sensory attenuation refers to the reduced perceived intensity of self-generated sensory stimuli relative to externally-generated stimuli of comparable physical magnitude [26]. Predictive models explain attenuation through efference copy mechanisms that generate expectations about sensory consequences; when prediction matches feedback, sensory responses are attenuated [3, 64]. Classical examples include diminished perception of self-generated touch and the reduced ability to tickle oneself.

Because attenuation reflects pre-reflective self-attribution rather than explicit evaluation, it provides a promising measurement paradigm for continuous interaction scenarios. Unlike temporal binding, attenuation does not require discrete action–effect timing and can be assessed within ongoing sensorimotor loops.

2.4 Digital Music Performance Systems and Mapping Structure

DMIs are commonly conceptualized as closed-loop interactions where performer action and system response are continuously coupled through perception and feedback [59, 63]. Rather than discrete commands, musical interaction unfolds as an ongoing sensorimotor process, where performers form predictive models linking gesture to sound. Within this loop, the mapping layer plays a central role by transforming gestural input into control over synthesis parameters [22].

Mapping design has long been recognized as a key determinant of instrumentality and expressivity in New Interfaces for Musical Expression (NIMEs). While simple one-to-one mappings provide transparency and ease of learning, more complex strategies—including nonlinear response functions [39, 41], state-dependent behavior, and many-to-many parameter relationships—have been explored to expand expressive possibilities [8, 28]. However, increasing mapping complexity introduces trade-offs between expressive richness and intelligibility. Nonlinear or stateful mappings may produce qualitatively distinct behavioral regimes or history-dependent responses, similar to threshold phenomena observed in acoustic instruments [40] such as bowed strings or wind instruments [34, 52].

These mapping characteristics directly affect predictability and, consequently, sense of agency. Performers rely on stable predictive relationships to attribute sonic outcomes to their own actions [35, 37]; when mappings introduce hidden dynamics or nonlinear transformations, prediction may become more difficult, potentially altering perceived control. Prior NIME research suggests that successful instruments balance responsiveness with complexity, remaining *reliably unpredictable* such that exploration remains possible without undermining intelligibility [44].

Examining the feasibility of sensory attenuation as an evaluation method, we investigate this mediation in two distinct mapping types: response linearity and temporal state. Linear mappings preserve proportional relationships between gesture and outcome, supporting direct correspondence, whereas nonlinear mappings introduce threshold effects that may alter system predictability [39]. Similarly, time-dependent mappings introduce history into the interaction loop, affecting temporal predictability. By systematically comparing linear and nonlinear mappings across both gestural (pressure) and temporal domains, we investigate how these design choices influence pre-reflective self-attribution as measured through sensory attenuation.

3 Implementation

We implemented a configurable sound control system intended for expressive, real-time performance. The implementation targets five requirements: (1) low-latency sound generation and parameter modulation, (2) capture of expressive gestural input, (3) timbre control through multiple synthesis parameters, (4) support for both linear and nonlinear (including stateful) control dynamics, and (5) switching between simple and complex mappings. A full overview of the system can be found in Appendix B. The system is implemented as a sequencer and synthesizer in Max/MSP [7], where participants can modulate various parameters using an Expressive E Touché [13].

3.1 Platform for Flexible Sound Control

We used Max/MSP's visual dataflow environment to provide real-time scheduling, built-in synthesis and DSP objects, and a fast iteration workflow that is well-suited to prototyping mapping engines for DMIs [7, 12]. Text-based environments like SuperCollider [54] and other visual environments like Pure Data [46] would likewise be appropriate for these general requirements.

3.2 Controller for Expressive Gestural Input

For gestural input we used the Expressive E Touché [13], which affords performance gestures such as tapping, sliding, wiggling, and bending. The device reports pressure-derived values (top, bottom, left, right) as MIDI Continuous Controller (CC) messages. We implemented a single pressure stream as the primary control signal to keep the mapping space interpretable and isolate complexity in the mapping engine, rather than in multi-sensor coordination.

In Max/MSP, incoming CC data are received using `ctlin`, scaled into a normalized 0–1 control range, and then resampled and smoothed prior to use by the mapping engine. We sample the normalized control stream every 15.625 ms (64 Hz), which provides a stable update rate for parameter modulation while remaining responsive. Smoothing reduces audible zipper noise during rapid parameter changes and stabilizes the signal for stateful processing stages.

3.3 Timbre Manipulation Targets

Expressive control in DMIs performance often relies on timbre modulation [11, 45]. Combining common subtractive and additive synthesis approaches, we provide three complementary timbre control targets [51]: (1) low-pass filtering, (2) wave-folding, and (3) overdrive.

Pulse-wave duty cycle. The core sound source is a pulse oscillator with a continuously modulated duty cycle.

Low-pass filter cut-off. We apply a low-pass filter and modulate its cut-off frequency to control spectral brightness by opening the filter with increasing pressure. This provides a familiar performance metaphor of *opening* and *closing* the sound.

Wave folding. Finally, we apply a wavefolder, which creates additional harmonics by folding amplitude peaks into repeated segments once they exceed a threshold [5]. Greater volume produces stronger folding and a more aggressive distortion character.

3.4 Linear vs. Nonlinear Behavior

To model nonlinear and stateful interaction dynamics that are common in acoustic instruments [34, 39], we provide two behavior modes that operate *within* the mapping engine—we compare Linear to Non-Linear *Pressure Mappings*, and Linear to Non-Linear *Temporal Mappings*.

3.4.1 *Pressure Mapping.*

For the *Linear Pressure* condition, the pressure input was mapped to duty cycle (from 50% to 85%) over the entire input range.

We create a *Non-Linear Pressure* response via combined three piecewise-linear functions with different activation thresholds and slopes (Figure 1). Concretely, each timbral manipulation begins and plateaus at individual points of the available pressure range:

- Filter cut-off: active from 0% (100 Hz) and plateaus at 50% (6000 Hz),
- Duty cycle: starts at 33% (50% duty cycle) and plateaus at 66% (95% duty cycle),
- Wavefolder volume: activates from 50% (0.5×) and saturates at 100% (10×).

These ranges were selected to span clearly audible timbral changes while avoiding trivial saturation at low pressures. The combination of effects produces distinct *control zones* where different combinations of parameters are active, yielding qualitatively different timbral behaviors across the pressure range. In this sense, the system behaves like a mode-switching control surface without requiring explicit switching by the performer.

3.4.2 *Time Mapping.*

In the *Linear Time* condition, no further manipulation was applied. In the *Non-Linear Time* condition, we added temporal nonlinearity through hysteresis. The system maintains a latent state variable, *system energy*, computed as an exponential moving average (EMA) of recent input history [34, 39]:

$$SystemEnergy_t = (1 - \alpha)\, SystemEnergy_{t-1} + \alpha\, CurrentInput_t.$$

The effective control signal is then computed by combining current pressure with system energy. In our implementation, the two are weighted equally, yielding a control signal that is responsive to immediate gesture changes while still reflecting recent interaction history. This produces hysteresis-like behavior: the same instantaneous pressure can yield different outputs depending on prior pressure trajectories, enabling effects such as temporal

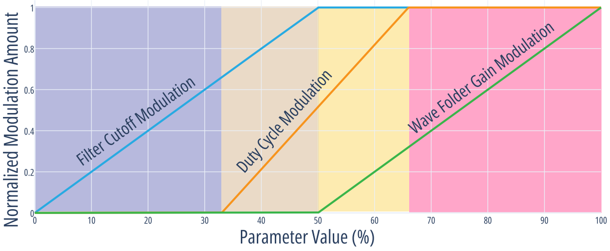


Figure 1: Nonlinear pressure response curves showing how filter cut-off (blue), duty cycle (orange), and wavefolding (green) vary with input pressure. Shaded regions indicate control zones where different parameter combinations are active, producing a globally nonlinear response from piecewise-linear mappings.

smoothing and history-dependent settling that are common in acoustic interaction dynamics (e.g., overblowing thresholds and return paths) [28, 39].

4 Study Overview

Musical expressivity is often framed as communication from performer to audience [18], mediated through the shaping of perceptible sound features [45]. For expressive DMIs, prior work argues that mappings should be sufficiently rich to support nuanced control [22], while still remaining understandable in live performance [62]. Motivated by expressive behaviors found in acoustic instruments, our system introduces nonlinear and stateful control dynamics (Section 3.4). The central question for the study is whether such added mapping complexity reduces performers’ ability to maintain gesture-effect correspondence during performance, and, consequently, a sense of agency.

We study this question through the lens of *self-attribution*, i.e., the sense that one caused a change in the environment [9]. Self-attribution is a key component of the sense of agency: in the two-level account, the pre-reflective feeling of agency provides an initial classification of outcomes as self-caused or not, which contributes to higher-level agency judgments (Section 2.1) [?]. In performance contexts, reduced mapping comprehension can plausibly weaken this self-caused classification by increasing uncertainty about whether an outcome matches the intended action effect.

To operationalize self-attribution without relying on explicit self-reports, we use *sensory attenuation* as an implicit measure [26]. We quantify the perceived intensity of stimuli using *magnitude estimation* [27], which is commonly used in psychophysics to obtain ratio-scale judgments and supports within-participant comparisons across conditions.

4.1 Participants

Twelve participants (9 male, 2 female, 1 non-binary) aged between 20–29 ($M = 24.42$, $SD = 2.60$) took part in the study ($N = 12$). Ten participants reported no hearing loss, while two reported slight, constant hearing loss.

All participants met the inclusion criterion of having a musical background and being comfortable playing at least one instrument that requires fine hand control. This criterion was selected because instrumental skills can transfer to unfamiliar musical controllers and interfaces [60]. We further expected that participants with performance experience would be better able to describe and contextualize their interaction experience during a follow-up interview.

4.2 Study Design

The study comprised two connected parts: (1) a within-subjects psychophysics experiment and (2) a short qualitative interview.

5 Psychophysics Experiment

5.1 Experiment Design

To examine whether mapping complexity and stateful control dynamics affect performers’ ability to maintain gesture-effect correspondence, we conducted a psychophysics experiment measuring *sensory attenuation*. Sensory attenuation was operationalized by comparing perceived loudness during active performance with perceived loudness during later playback of the same sound.

In each trial, participants produced a short performance while a fixed melody was supplied to the sound engine via MIDI playback. Participants were instructed to freely modulate the sound using the controller. Each performance was rated twice: once during recording (active control) and once during playback (passive listening). Differences between these two estimates served as an implicit measure of self-attribution.

We used a factorial design with the following independent variables *Pressure* (Linear/Non-Linear), *Time* (Linear/Non-Linear), and *Volume* (Quiet/Loud).

Pressure and Time conditions were blocked. **Volume Level** was randomized within blocks, to reduce repetitive responding. Each condition was repeated eight times per participant (four repetitions per volume level). Condition order was partially counterbalanced to mitigate order and learning effects.

To reduce the likelihood that participants would recognize their own recordings during playback and reproduce earlier estimates from memory, recording and playback trials were temporally decoupled. Trials were organized into alternating recording (REC) and playback (PB) blocks in the following fixed sequence:

$$REC_0 \longrightarrow REC_1 \longrightarrow PB_1 \longrightarrow REC_2 \longrightarrow PB_0 \longrightarrow REC_3 \longrightarrow PB_3 \longrightarrow PB_2$$

For each condition, we collected the difference between the estimate participants provided for the loudness of their live performance, and the estimate they provided for the loudness of the resulting recording. Differences in estimates were standardized per participant.

5.2 Procedure

Participants were first briefed on the task, provided informed consent, and completed a short demographic questionnaire. They were then instructed in the magnitude estimation procedure. Participants were asked to report relative loudness using any numerical scale, with larger numbers indicating louder sounds. They were encouraged to capture fine-grained differences between trials and were told they could assign an arbitrary reference value (e.g., 100) to the first trial. To reduce ambiguity, participants were instructed to estimate *peak loudness*.

Before the main experiment, participants completed a short familiarization session with the controller. During this phase, a random melody generator provided MIDI input, and both Pressure Linearity and Time Linearity were set to *nonlinear*. Participants were encouraged to explore a range of gestures (e.g., soft vs. forceful, slow vs. rapid movements) to become comfortable with the interaction.

5.3 Task

Across the experiment, each participant completed 32 recordings (4 Conditions × 2 Volume Levels × 4 repetitions). Each recording

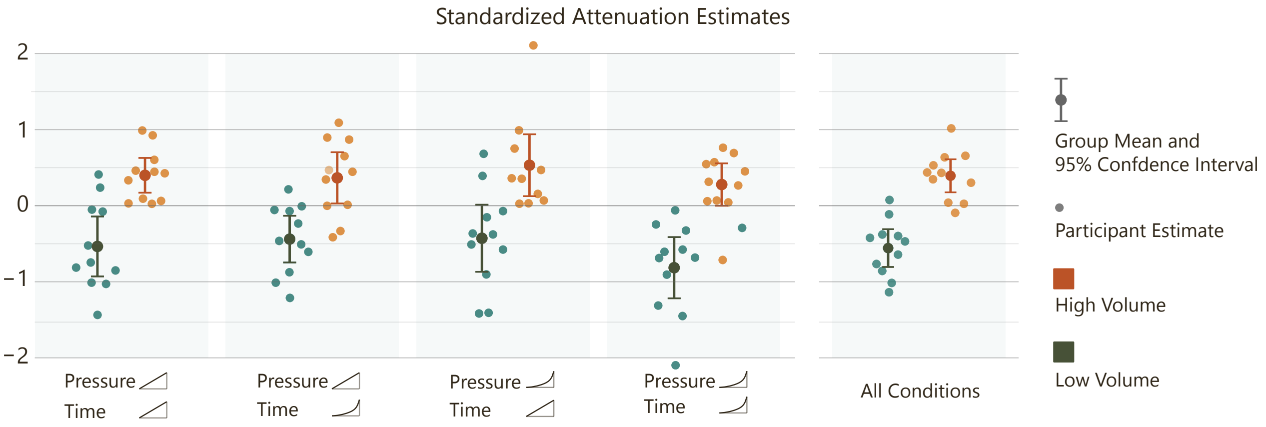


Figure 2: Differences in estimates of live audio and recorded audio. A positive difference indicates that the recording was rated louder. The Y axis represents distance from zero attenuation in Standard Deviations. Each point is the average of four estimates for the graphs broken down by pressure and time. For all conditions, each point is the average of 16 estimates. Linear and Non-Linear are indicated by Linear and Exponential Slopes respectively. *Non-Linear Pressure* combined with *Linear Time* resulted in highest attenuation, *Non-Linear Pressure* combined with *Non-Linear Time* resulted in lowest. The only statistically significant effect was observed for *Volume*.

was evaluated twice—once during recording and once during playback—yielding a total of 64 loudness estimation trials per participant. Responses were provided immediately after each trial, and the next trial began as soon as a response was recorded.

5.4 Setup

Mono audio output was delivered through a single FBT J 5A active loudspeaker [14], positioned approximately 80 cm in front of the participant. Gestural input was provided via the Touché with input sensitivity set to 40%.

5.5 Data Collection and Analysis

For each recording, participants provided estimates for **LoudnessRecording** and **LoudnessPlayback**, which were recorded manually into a spreadsheet. Because participants used self-chosen numerical scales, estimates were standardized within participants prior to analysis.

One participant was excluded due to repeatedly providing identical loudness estimates across trials, including trials with objectively different volume levels.

After exclusion, data from 11 participants were analyzed. Differences between playback and recording estimates (PB−REC) were computed for each sound, yielding 350 valid difference measurements (two missing values). Outliers were screened using the Median Absolute Deviation (MAD) method with a threshold of 4, revealing no problematic observations. Estimates were then standardized (by transforming to z-scores) per person, to ensure that all participants estimates were within a shared scale. The distribution of difference scores was approximately normal, with slight positive skew.

5.6 Analysis and Results

We conducted a three-way ANOVA with *Volume* (High, Low), *Pressure* (Linear, Non-Linear), and *Time* (Linear, Non-Linear) as fixed factors and Participant as a blocking factor. The dependent variable was Difference in Estimated Amplitude; for simplicity sake, we will refer to this as *Attenuation*.

Differences between Groups. The analysis revealed a significant main effect of *Volume*, $F(1, 70) = 93.09$, $p < .001$, indicating that *Attenuation* differed reliably between volume conditions. No significant main effects were observed for *Pressure*, $F(1, 70) = 0.32$, $p = .573$, or *Time*, $F(1, 70) = 2.16$, $p = .146$. The blocking factor Participant was significant, $F(10, 70) = 3.58$, $p < .001$, suggesting strong between-participant variability. There were no significant interactions effects.

Follow-up Check for Absolute Attenuation. To establish if the effect of volume observed in the ANOVA lead to attenuation in absolute terms (rather than just differences between groups), we conducted two-sided one-sample t-tests on participant means. Holm correction was applied to control for multiple comparisons across volume levels.

For Low Volume, the mean differed significantly from zero ($M = -0.55$, $SD = 0.37$, $n = 11$), $t(10) = -4.91$, $p_{\text{Holm}} = .001$, Cohen's $d_z = -1.48$, 95% confidence interval of $[-0.80, -0.30]$.

For High Volume, the mean was also significantly different from zero ($M = 0.40$, $SD = 0.32$, $n = 11$), $t(10) = 4.13$, $p_{\text{Holm}} = .002$, Cohen's $d_z = 1.24$, 95% confidence interval of $[0.19, 0.62]$.

This shows that, in both cases, there was a measurable attenuation effect; however, the direction changes. It appears that participants overestimated the live audio with respect to the recordings in the Low Volume condition and underestimated the live audio in the High Volume condition.

6 Semi-Structured Interview

6.1 Interview Protocol

A complementary semi-structured interview explored participants' explicit understanding of the system and judgments of agency. Broadening the psychophysics experiment's focus on sensory attenuation, the interview explored aspects of agency, including perceived control, perceived self-causation, and the ability to anticipate sonic outcomes.

First, participants were asked to describe what they believed the system did, to elicit their spontaneous mental models of the

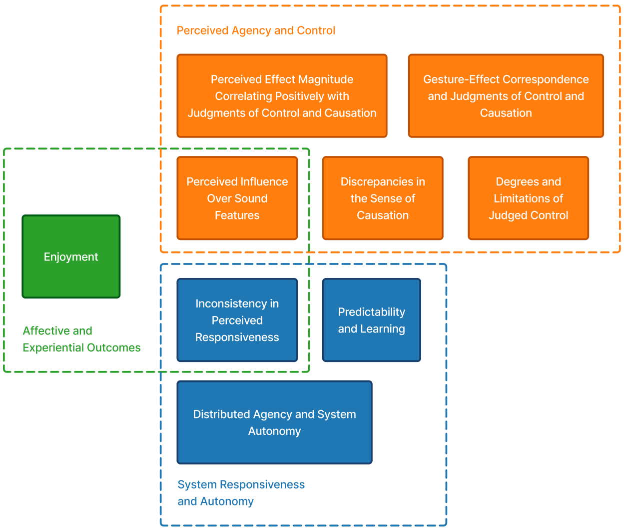


Figure 3: Overview of qualitative clusters grouped into three meta-clusters: (1) perceived agency and control, (2) system responsiveness and autonomy, and (3) affective and experiential outcomes.

mapping. Second, after a brief explanation of the intended control principles, participants were asked more directly about (i) self-causation, (ii) control, and (iii) predictability. Third, participants were asked about enjoyment, under the assumption that perceived agency and expressive potential are often intertwined. When applicable, participants were also asked whether effort was required to achieve their intentions; however, because the musical task was improvisational, explicit goals were not expected in all cases. A complete list of questions is provided in Appendix A.

6.2 Data Collection and Analysis

All 12 participants completed the interview. For consistency, the same participant excluded from the quantitative analysis (Section 5.5) was excluded here as well.

Interviews were audio recorded and transcribed using OpenAI Whisper [43]. Transcripts were manually checked for correctness. To identify recurring patterns, Løvaas coded the transcripts in Taguette [47] and used an visual affinity clustering [30] approach with Miro to organize themes informed both by the interview prompts (e.g., control, causation, prediction) and inductively from participants' accounts.

6.3 Results

The analysis yielded nine clusters across participants (see Appendix D) that describe how participants understood the system and judged their own influence over sonic outcomes, and how these judgments shaped gesture-effect correspondence and enjoyment.

6.3.1 Perceived influence over sound features. Most participants described an influence on timbre and loudness (8 participants each). Within timbre, wavefolding distortion was particularly salient (7 participants) and described as more *saturated* (P3), *wonkier* and harsher (P7), and in one case, *a bit more like a trombone* (P8). Filter effects were reported less frequently (P7, P8, P10), with participants describing the sound as becoming *broader* (P7) or comparing it to a wah-wah pedal (P8). Although explicit loudness control was not implemented, participants commonly inferred loudness changes from timbral changes (e.g., from filter opening and added distortion).

However, perceived loudness control was often inconsistent. P10 reported *sometimes, I'm just controlling the cut-off filter and sometimes I wasn't really controlling the volume.* P2 and P3 similarly noted that loudness changes were not reliable: *there were some times it didn't, sometimes it increased* (P2).

A small number of participants also reported perceived pitch changes (P5, P8), despite pitch being fixed by the MIDI input. P5 described the system as *randomly pitch[ing] up* with pressure, whereas P8 perceived the opposite trend (lower notes at higher pressure). This suggests that certain timbral transformations may have been interpreted as pitch variation.

6.3.2 Inconsistency in perceived responsiveness. Several participants reported that sometimes nothing seemed to happen in response to their gesture (P9) or wondered if they were experiencing a *placebo effect* when changes were not clearly perceivable.

Experiences of *Non-Linear Time* temporal smoothing were also inconsistent. Some experienced the system as smoothed, especially when attempting faster rhythmic gestures. P10 noted that their expectations were met primarily when they applied pressure slowly, even though they felt *able to do like some kind of rhythm.* Others reported the opposite, describing the sound as *wavy* (P11) and even suitable for tremolo-like effects (P12). Notably, participants disagreed most about the *Non-Linear Time* stateful conditions (*Linear Pressure & Non-Linear Time* and *Non-Linear Pressure & Non-Linear Time*). Some reported no meaningful difference (P2, P5), while others reported that these conditions produced the strongest perceived effect (e.g., P8 for *Linear Pressure & Non-Linear Time*).

6.3.3 Degrees and limitations of judged control. Several participants described only partial or intermittent control judgments (P9, P3). P10 described they could not always change a specific thing, but felt they could always change something. Others described stronger and more consistent control after forming a mental model of the mapping, with the system becoming *very easy to control once I got to know what a movement does* (P12).

A recurring theme was that perceived control depended on the system's available range. Participants described being *in control within the range* (P2, P7) but also criticized the system as overly constrained. P8 reported they could approach a desired sound but not fully reach it: *I could get like a bit towards the sound, but I couldn't like completely go there.* Others linked limited range directly to reduced control. P7 stated they did not feel in control because *the majority of the sound was predetermined.* P2 similarly felt they could *play around with it a bit* but lacked access to more extreme changes.

6.3.4 Distributed agency and system autonomy. Some participants described the interaction as collaborative, attributing a degree of autonomy, rather than controllability, to the system. The system was positioned as *reacting* (P11), requiring the user as an intermediary that *is going to translate those movements* (P10). This suggested mediated rather than direct control. P7 implied that the system sometimes acted independently, stating it *sometimes got louder anyway* without deliberate input.

Several participants needed to time their gestures to the system's dynamics, comparing it to a timing game where peaks needed anticipation and accurate reaction (P8) or like *like surfing a wave and I had to press it in the right moment* (P7).

6.3.5 Discrepancies in the sense of causation.

Some participants (P3, P7, P9, P10) reported inconsistency in causation without being able to specify when it occurred. P10 added that some changes felt like they did not originate from them. Others provided more specific contrasts: P6 reported strong self-causation during the training phase (*Non-Linear Pressure & Non-Linear Time*), whereas P5 reported reduced self-causation in the same configuration when they *pushed it and there was no tone.* P8 felt they *really changed* the sound in the stateful conditions and questioned whether the system was working in earlier blocks.

6.3.6 Gesture–effect correspondence shaping judgments of control and causation.

P3 described loudness as *random* when they could not detect correspondence and, therefore, less control. P12 similarly reported: *I was pressing it and it was kind of random. It didn't have the same effects uniformly. So it made me think maybe I was not controlling it.* P9 and P4 likewise reported reduced control when the system did not respond as expected.

Conversely, several participants described stronger causation when they perceived their gesture pattern reflected in the sound. P11 reported causation when the modulation matched their movement rhythm. P12 described making *a little wavy movement* and hearing *the sound being a little wavy.* P3 similarly stated: *I tried for example sometimes to do another rhythm. [...] I could hear the same phrase in a way, what I did.*

At the same time, perceived inconsistency was common. P12 reported being unable to form a *clear, direct correlation.* P9 emphasized inconsistency across repetitions: *you did the same action, but you didn't have the same results as before.* P10 reported forming a mental model early on (e.g., expecting volume control) that later broke down.

6.3.7 Effect salience correlating with judged control and causation.

Many participants reported stronger agency judgments when actions produced larger and more salient sonic changes, including both timbral intensity and overall loudness.

P4 stated that they felt more in control and more like they caused changes when *a lot happened.* P4 and P5 reported stronger self-causation at high pressure. P8 similarly stated that they only attributed changes to themselves when there was a *real difference.*

Several participants also linked agency to overall system volume. P11 reported stronger perceived impact when it was louder and noted that *when it was louder I kind of felt more in control of it.* P8 likewise reported noticing meaningful differences primarily when the output was *really loud.*

6.3.8 Predictability and learning.

Perceived predictability varied across individuals and conditions and often improved with exposure. P3 stated they could sometimes anticipate outcomes but did not know *if it was just coincidence.* P8, P9, and P10 likewise reported expectation mismatches. Others described building prediction through pattern recognition.

Several participants described learning effects. P5 reported being able to anticipate outcomes after two trials in a new configuration. P3 similarly reported understanding the system after a few tries. P2 and P11 reported being better able to anticipate toward the end of the experiment, attributing this to increased familiarity and, in P2's account, also to changes in system configuration. Participants tended to agree that *Non-Linear Pressure & Linear Time* was easier to anticipate than other configurations, but there was no consensus about the remaining conditions.

6.3.9 Enjoyment.

Enjoyment varied and was linked to effect perceivability, perceived control range, and personal preference for predictability versus exploration.

Several participants reported greater enjoyment when effects were clear and controllable. P8 reported finding the system more enjoyable when differences were perceivable. P4 stated: *it's more satisfactory to see that when you do something, there's a direct effect or at least in that same moment, you see a difference.* P7 similarly expressed that they would have enjoyed it more if they could *control the whole sound more.*

However, participants differed on whether strong control was necessary for enjoyment. For example, P10 reported enjoying the *mysterious* quality of the device. P12 similarly described enjoying exploratory sound *twisting* and surprising outcomes driven by curiosity.

7 Discussion

We assumed participants would underestimate the loudness of expressively controlled audio compared to recordings of their performances, and that this underestimation—sensory attenuation [26, 50]—could serve as an implicit measure of self-attribution. We quantified this using attenuation estimates, where positive values indicate playback judged louder than live performance and negative values indicate live performance judged louder than playback. We examined this under two output volume levels (Low/High), two pressure mappings (Linear/Non-Linear), and two temporal mappings (Linear/Non-Linear). At high volume, results followed the expected direction associated with sensory attenuation, whereas at low volume the direction reversed. This suggests that the measure may not behave monotonically across output levels and could reflect additional factors when stimuli approach perceptual thresholds. Overall, the results support the feasibility of using attenuation as a measure, but require further investigation.

7.1 Volume Effects

Output volume strongly modulated attenuation. At High Volume, participants rated actively controlled sounds as less intense than their later playback At Low Volume this pattern reversed, with participants tending to rate controlled sound as louder than the playback.

The High Volume condition aligns with established interpretations of attenuation as a predictive dampening of self-generated sensory events [26]. Rapid timbral changes can be perceptually salient or surprising when unexpected; when participants actively control the sound and anticipate these changes, their perceived intensity may be reduced, leading to attenuation. For quieter sounds, however, this mechanism may be less pronounced.

The observed inverse attenuation at Low Volume—participants rating live manipulation as louder than playback—is more difficult to explain. One possible account is attentional modulation: During active control, participants expected gesture-dependent changes and likely focused attention on subtle variations, increasing their perceived salience. During playback, without the action–perception coupling, these subtle timbral shifts may have received less attention, leading to comparatively lower perceived intensity and thus the observed inversion. This suggests that sensory attenuation is usable as an implicit proxy, but only above a perceptual salience threshold.

7.2 Mapping Effects

We expected linear mappings to produce the strongest sensory attenuation, based on the assumption that stable and proportional action–effect relationships support accurate forward models. Instead, the strongest attenuation was observed in the *Non-Linear Pressure & Linear Time* condition, while the weakest attenuation occurred in the *Non-Linear Pressure & Non-Linear Time* condition. Because we did not observe statistically significant mapping effects, these patterns should be interpreted cautiously. Nevertheless, they align with several themes emerging from the qualitative data.

Linear temporal mappings were generally described as more predictable, and conditions with linear time tended to show stronger attenuation overall. Temporal nonlinearity introduced history-dependent behavior that participants often found difficult to interpret consistently, likely reducing prediction accuracy and weakening attenuation.

For pressure mappings, the results were less straightforward. Although linear pressure mappings were expected to maximize attenuation, the strongest attenuation instead occurred with nonlinear pressure combined with linear time. Qualitative reports suggest a plausible explanation. Nonlinear pressure mappings produced more salient timbral events — particularly distortion and wavefolding — which were repeatedly noted by participants. These salient changes may have strengthened the sense that actions had meaningful and noticeable effects, thereby reinforcing self-attribution. At the same time, linear temporal behavior supported building strong action–effect links. Several participants described the *Non-Linear Pressure & Linear Time* condition as easiest to anticipate of all conditions.

The weakest attenuation occurred in the *Non-Linear Pressure & Non-Linear Time* condition. Here, salient timbral changes coincided with temporal unpredictability, creating situations in which effects were both noticeable and unexpected. Such combinations likely increase prediction error despite strong perceptual salience, reducing the extent to which outcomes are experienced as self-generated and therefore attenuated.

However, these interpretations are not built on a solid statistical basis. Future work might explore the link between mappings and self-attribution further with more carefully designed experiments focusing on high-volume, or with experiments using larger sample sizes.

7.3 The Role of (Generative) Algorithms in Music

This work was partially motivated by the increasing use of GenAI in music creation. In this context, we consider traditional nonlinear mappings as a relatively well-understood baseline. GenAI can then be conceptualized as a more extreme form of nonlinear input to output mapping. We expect that the trends in pre-reflective agency observed in the present experiment may become more pronounced in the context of music creation with GenAI systems.

This perspective introduces a potential paradox: some users claim authorship over music created with GenAI tools such as Suno [42, 53]. If the attenuation effects observed here generalize to GenAI-based music creation, self-attribution may be primarily driven by reflective judgment, while the pre-reflective experience of creating differs fundamentally. In other words, creating with GenAI may diverge from traditional instrumental performance at a basic experiential and sensorimotor level.

Understanding these sensorimotor foundations of creative practice is particularly important in a context where artistic production is increasingly framed in terms of efficiency and output rather than lived experience. The present study provides an initial step toward a more detailed investigation of this issue.

At the same time, generative processes have long been integral to music composition and performance. Contemporary classical practices, including twelve-tone composition, algorithmic composition, and electronic music, routinely involve rule-based or system-driven generation. Indeed, in our own experiment, melodic and rhythmic material was not produced by participants but predetermined by a sequencer. This highlights that attribution and algorithmic contribution do not form a simple zero-sum relationship. Rather, they interact in complex ways, both at the level of sensorimotor experience and within broader cultural and historical traditions of musical practice.

By introducing an experiential perspective to this debate, we aim to extend discussions of algorithms and GenAI in music beyond questions of output quality or authorship toward a deeper consideration of how different systems shape the lived experience of creating music.

8 Conclusion

In this work, we investigated how mapping structure and temporal dynamics influence performers' sense of agency during digital music interaction. By systematically manipulating pressure linearity and time linearity within a configurable sound-control system, and combining psychophysical measures with qualitative reports, we examined how variations in gesture–effect relationships shape self-attribution and perceived control. Our approach positions digital instruments as sensorimotor systems in which agency emerges from ongoing interaction.

Results suggest that sensory attenuation can serve as an implicit measure of self-attribution, but only when output volume exceeds a salience threshold. Qualitative findings further indicate that self-attribution was strongest when sonic changes were clearly perceivable, temporally aligned, and proportional to gestures. Conversely, when mappings were inconsistent, or outcomes violated expectations, performers reported reduced control and weaker self-attribution, at times interpreting the system more as a collaborator than as an instrument. Together, these findings highlight that agency depends less on mapping complexity itself and more on whether interaction supports stable prediction and perceivable action–effect correspondence.

From a theoretical perspective, the results align with accounts of agency that emphasize predictive processes and sensorimotor contingencies, suggesting that pre-reflective agency arises when performers can form reliable expectations about the consequences of their actions. Rather than treating agency as a binary property, our findings support a view of agency as graded and interaction-dependent, shaped by perceptual thresholds, temporal dynamics, and the learnability of mappings.

9 Ethical Standards

The authors have ensured that the study meets ethical guidance for research with human participants outlined by the Max Planck Institute for Informatics and aligns to the NIME Principles & Code of Practice on Ethical Research. The authors have no conflicts of interests to declare.

References

[1] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How does HCI Understand Human Agency and Autonomy?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 375, 18 pages. https://doi.org/10.1145/3544548.3580651

[2] Joanna Bergstrom-Lehtovirta, David Coyle, Jarrod Knibbe, and Kasper Hornbæk. 2018. I Really did That: Sense of Agency with Touchpad, Keyboard, and On-skin Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3173574.3173952

[3] Sarah-J. Blakemore, Daniel M. Wolpert, and Chris D. Frith. 1998. Central cancellation of self-produced tickle sensation. *Nature Neuroscience* 1, 7 (Nov 1998), 635–640. https://doi.org/10.1038/2870

[4] Gino Brunner, Andres Konrad, Yuyi Wang, and Roger Wattenhofer. 2018. MIDI-VAE: Modeling dynamics and instrumentation of music with applications to style transfer. In *19th International Society for Music Information Retrieval Conference (ISMIR)*. https://doi.org/10.48550/ARXIV.1809.07600

[5] Emmett Corman. 2015. *Simple Synthesis: Part 8, Wavefolding*. Keith McMillen Instruments. https://www.keithmcmillen.com/blog/simple-synthesis-part-8-wavefolding/

[6] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I Did That! Measuring Users' Experience of Agency in Their Own Actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2025–2034. https://doi.org/10.1145/2207676.2208350

[7] Cycling74. [n. d.]. What is Max?. Cycling '74. https://cycling74.com/products/max.

[8] Palle Dahlstedt. 2009. Dynamic Mapping Strategies for Expressive Synthesis Performance and Improvisation. In *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*. Springer Berlin Heidelberg, 227–242. https://doi.org/10.1007/978-3-642-02518-1_16

[9] Andrea Desantis, Carmen Weiss, Simone Schütz-Bosbach, and Florian Waszak. 2012. Believing and Perceiving: Authorship Belief Modulates Sensory Attenuation. *PLoS ONE* 7, 5 (May 2012), e37959. https://doi.org/10.1371/journal.pone.0037959

[10] Johanna K. Didion, Krzysztof Wolski, Dennis Wittchen, David Coyle, Thomas Leimkühler, and Paul Strohmeier. 2024. Who did it? How User Agency is influenced by Visual Properties of Generated Images. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 94, 17 pages. https://doi.org/10.1145/3654777.3676335

[11] Christopher Dobrian and Daniel Koppelman. 2006. The E in NIME: Musical Expression with New Computer Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Paris, France, 277–282. https://doi.org/10.5281/zenodo.1176893

[12] Jon Drummond. 2009. Understanding Interactive Systems. *Organised Sound* 14, 2 (Aug. 2009), 124–133. https://doi.org/10.1017/S1355771809000235

[13] Expressive E. 2025. Touché - Expressive Midi/CV/USB Controller. https://www.expressivee.com/1-touche

[14] FBT Elettronica. [n. d.]. Prodotto - FBT J Series. https://www.fbt.it/product/j-series/.

[15] Fabián Esqueda, Henri Pöntynen, Julian D. Parker, and Stefan Bilbao. 2017. Virtual Analog Model of the Lockhart Wavefolder. In *Proceedings of the 14th Sound and Music Computing Conference (SMC)*. http://research.spa.aalto.fi/publications/papers/smc17-wavefolder/

[16] Christopher D. Frith, Sarah-Jayne Blakemore, and Daniel M. Wolpert. 2000. Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 355, 1404 (Dec 2000), 1771–1788. https://doi.org/10.1098/rstb.2000.0734

[17] Chris D Frith, Sarah-Jayne Blakemore, and Daniel M Wolpert. 2000. Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews* 31, 2 (2000), 357–363. https://doi.org/10.1016/S0165-0173(99)00052-1

[18] Alf Gabrielsson and Patrik N. Juslin. 1996. Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychology of Music* 24, 1 (April 1996), 68–91. https://doi.org/10.1177/0305735696241007

[19] Patrick Haggard, Sam Clark, and Jeri Kalogeras. 2002. Voluntary action and conscious awareness. *Nature Neuroscience* 5, 4 (Mar 2002), 382–385. https://doi.org/10.1038/nn827

[20] Patrick Haggard and Manos Tsakiris. 2009. The Experience of Agency: Feelings, Judgments, and Responsibility. *Current Directions in Psychological Science* 18, 4 (Aug 2009), 242–246. https://doi.org/10.1111/j.1467-8721.2009.01644.x

[21] Andy Hunt and Ross Kirk. 2000. Mapping Strategies for Musical Performance. In *Trends in Gestural Control of Music*, Marcelo M. Wanderley and Marc Battier (Eds.). https://api.semanticscholar.org/CorpusID:54081997

[22] Andy Hunt and Marcelo M. Wanderley. 2002. Mapping performer parameters to synthesis engines. *Organised Sound* 7, 2 (Aug 2002), 97–108. https://doi.org/10.1017/s1355771802002030

[23] Robert H. Jack, Adib Mehrabi, Tony Stockman, and Andrew McPherson. 2018. Action-sound Latency and the Perceived Quality of Digital Musical Instruments. *Music Perception* 36, 1 (Sept. 2018), 109–128. https://doi.org/10.1525/mp.2018.36.1.109

[24] Robert H. Jack, Tony Stockman, and Andrew McPherson. 2017. Rich gesture, reduced control: the influence of constrained mappings on performance technique. In *Proceedings of the 4th International Conference on Movement Computing* (London, United Kingdom) *(MOCO '17)*. Association for Computing Machinery, New York, NY, USA, Article 15, 8 pages. https://doi.org/10.1145/3077981.3078039

[25] Alexander Refsum Jensenius. 2014. To gesture or Not? An Analysis of Terminology in NIME Proceedings 2001–2013. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Goldsmiths, University of London, London, United Kingdom, 217–220. https://doi.org/10.5281/zenodo.1178816

[26] Fabian Kiepe, Nils Kraus, and Guido Hesselmann. 2021. Sensory Attenuation in the Auditory Modality as a Window Into Predictive Processing. *Frontiers in Human Neuroscience* 15 (Nov 2021). https://doi.org/10.3389/fnhum.2021.704668

[27] Frederick A.A. Kingdom and Nicolaas Prins. 2016. Chapter 3 - Varieties of Psychophysical Procedures. In *Psychophysics (Second Edition)* (second edition ed.), Frederick A.A. Kingdom and Nicolaas Prins (Eds.). Academic Press, San Diego, 37–54. https://doi.org/10.1016/B978-0-12-407156-8.00003-7

[28] Tellef Kvifte. 2008. On the Description of Mapping Structures. *Journal of New Music Research* 37, 4 (Dec. 2008), 353–362. https://doi.org/10.1080/09298210902731394

[29] Hannah Limerick, James W. Moore, and David Coyle. 2015. Empirical Evidence for a Diminished Sense of Agency in Speech Interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3967–3970. https://doi.org/10.1145/2702123.2702379

[30] Andrés Lucero. 2015. *Using Affinity Diagrams to Evaluate Interactive Prototypes*. Springer International Publishing, 231–248. https://doi.org/10.1007/978-3-319-22668-2_19

[31] Thor Magnusson, Chris Kiefer, and Halldor Ulfarsson. 2022. Reflexions upon Feedback. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Andrew McPherson and Emma Frid (Eds.). Auckland, New Zealand, Article 19. https://doi.org/10.21428/92fbeb44.aa7de712

[32] Andrew McPherson, Robert Jack, and Giulio Moro. 2016. Action-Sound Latency: Are Our Tools Fast Enough?. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Queensland Conservatorium Griffith University, Brisbane, Australia, 20–25. https://doi.org/10.5281/zenodo.3964611

[33] Andrew McPherson, Landon Morrison, Matthew Davison, and Marcelo M. Wanderley. 2024. On mapping as a technoscientific practice in digital musical instruments. *Journal of New Music Research* 53, 1–2 (March 2024), 110–125. https://doi.org/10.1080/09298215.2024.2442356

[34] Dylan Menzies. 2002. Composing instrument control dynamics. *Organised Sound* 7, 3 (Dec 2002), 255–266. https://doi.org/10.1017/s1355771802003059

[35] Lia Mice and Andrew McPherson. 2020. From miming to NIMEing: the development of idiomatic gestural language on large scale DMIs. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Romain Michon and Franziska Schroeder (Eds.). Birmingham City University, Birmingham, UK, 570–575. https://doi.org/10.5281/zenodo.4813200

[36] James W. Moore. 2016. What Is the Sense of Agency and Why Does it Matter? *Frontiers in Psychology* 7 (Aug 2016). https://doi.org/10.3389/fpsyg.2016.01272

[37] Tom Mudd. 2023. Playing with Feedback: Unpredictability, Immediacy, and Entangled Agency in the No-input Mixing Desk. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 243, 11 pages. https://doi.org/10.1145/3544548.3580662

[38] Tom Mudd and Akira Brown. 2023. Musical pathways through the no-input mixer. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Miguel Ortiz and Adnan Marquez-Borbon (Eds.). Mexico City, Mexico, Article 56, 7 pages. https://doi.org/10.5281/zenodo.11189224

[39] Tom Mudd, Simon Holland, and Paul Mulholland. 2019. Nonlinear dynamical processes in musical interactions: Investigating the role of nonlinear dynamics in supporting surprise and exploration in interactions with digital musical instruments. *International Journal of Human-Computer Studies* 128 (Aug 2019), 27–40. https://doi.org/10.1016/j.ijhcs.2019.02.008

[40] Tom Mudd, Simon Holland, and Paul Mulholland. 2019. The Role of Nonlinear Dynamics in Musicians' Interactions with Digital and Acoustic Musical Instruments. *Computer Music Journal* 43, 4 (2019), 25–40. https://doi.org/10.1162/comj_a_00535

[41] Tom Mudd, Simon Holland, Paul Mulholland, and Nicholas Dalton. 2014. Nonlinear Dynamical Systems as Enablers of Exploratory Engagement with Musical Instruments. In *INTER-FACE: International Conference on Live Interfaces*. https://oro.open.ac.uk/43252/1/Mudd%20ICLI%202014.pdf

[42] Y. Yogi Tegar Nugroho and P. Paulus Metta Dwi Manggala. 2024. The Use of AI in Creating Music Compositions: A Case Study on Suno Application. In *Proceedings of the 7th Celt International Conference (CIC 2024)*. Atlantis Press SARL, 177–189. https://doi.org/10.2991/978-2-38476-348-1_13

[43] OpenAI. 2025. whisper: Robust Speech Recognition via Large-Scale Weak Supervision. https://github.com/openai/whisper. https://github.com/openai/

whisper

[44] Sile O'Modhrain. 2011. A Framework for the Evaluation of Digital Musical Instruments. *Computer Music Journal* 35, 1 (March 2011), 28–42. https://doi.org/10.1162/COMJ_a_00038

[45] Cornelius Poepel. 2005. On Interface Expressivity: A Player-Based Study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Vancouver, BC, Canada, 228–231. https://doi.org/10.5281/zenodo.1176802

[46] Miller Puckette and Pure Data Community. 2025. Pure Data. https://puredata.info/

[47] Rémi Rampin and Vicky Rampin. 2021. Taguette: open-source qualitative data analysis. *Journal of Open Source Software* 6, 68 (Dec 2021), 3522. https://doi.org/10.21105/joss.03522

[48] Courtney N. Reed, Adan L. Benito, Franco Caspe, and Andrew P. McPherson. 2024. Shifting Ambiguity, Collapsing Indeterminacy: Designing with Data as Baradian Apparatus. *ACM Trans. Comput.-Hum. Interact.* 31, 6, Article 73 (Dec. 2024), 41 pages. https://doi.org/10.1145/3689043

[49] Kevin Roose. 2022. An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy. *The New York Times* (2022). https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

[50] Nihar Sabnis, Maëlle Roche, Dennis Wittchen, Donald Degraen, and Paul Strohmeier. 2025. Motion-Coupled Asymmetric Vibration for Pseudo Force Rendering in Virtual Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1134, 22 pages. https://doi.org/10.1145/3706598.3713358

[51] Charalampos Saitis, Courtney N. Reed, Ashley Laurent Noel-Hirst, Giacomo Lepri, and Andrew McPherson. 2025. (De)Constructing Timbre at NIME: Reflecting on Technology and Aesthetic Entanglements in Instrument Design. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Doga Cavdir and Florent Berthaut (Eds.). Canberra, Australia, Article 29, 10 pages. https://doi.org/10.5281/zenodo.15698835

[52] John C. Schelleng. 1974. The Physics of the Bowed String. *Scientific American* 230, 1 (1974), 87–95. http://www.jstor.org/stable/24949986

[53] Suno. 2025. Suno — AI Music Generation Platform. https://suno.com/

[54] SuperCollider. [n. d.]. SuperCollider. https://supercollider.github.io/.

[55] Daisuke Tajima, Jun Nishida, Pedro Lopes, and Shunichi Kasahara. 2022. Whose Touch is This?: Understanding the Agency Trade-Off Between User-Driven Touch vs. Computer-Driven Touch. *ACM Trans. Comput.-Hum. Interact.* 29, 3, Article 24 (Jan. 2022), 27 pages. https://doi.org/10.1145/3489608

[56] Atau Tanaka, Balandino Di Donato, Michael Zbyszynski, and Geert Roks. 2019. Designing Gestures for Continuous Sonic Interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Marcelo Queiroz and Anna Xambó Sedó (Eds.). UFRGS, Porto Alegre, Brazil, 180–185. https://doi.org/10.5281/zenodo.3672916

[57] Bradley W. Vines, Marcelo M. Wanderley, Carol L. Krumhansl, Regina L. Nuzzo, and Daniel J. Levitin. 2004. *Performance Gestures of Musicians: What Structural and Emotional Information Do They Convey?* Springer Berlin Heidelberg, 468–478. https://doi.org/10.1007/978-3-540-24598-8_43

[58] Thomas Waltemate, Irene Senna, Felix Hülsmann, Marieke Rohde, Stefan Kopp, Marc Ernst, and Mario Botsch. 2016. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology (VRST '16)*. ACM, 27–35. https://doi.org/10.1145/2993369.2993381

[59] Marcelo M. Wanderley and Marc Battier. 2000. Trends in Gestural Control of Music. https://www-media.idmil.org/media/Trends_Ircam/DOS/RESSOURCES.pdf.

[60] Marcelo Mortensen Wanderley and Nicola Orio. 2002. Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Computer Music Journal* 26, 3 (Sept. 2002), 62–76. https://doi.org/10.1162/014892602320582981

[61] Daniel M Wegner. 2003. The mind's best trick: how we experience conscious will. *Trends in Cognitive Sciences* 7, 2 (Feb 2003), 65–69. https://doi.org/10.1016/s1364-6613(03)00002-0

[62] Travis West, Baptiste Caramiaux, Stéphane Huot, and Marcelo M. Wanderley. 2021. Making Mappings: Design Criteria for Live Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 31. https://doi.org/10.21428/92fbeb44.04f0fc35

[63] Todd Winkler. 2001. *Composing Interactive Music: Techniques and Ideas Using Max*. MIT Press, Cambridge, MA London.

[64] Daniel M. Wolpert, Zoubin Ghahramani, and Michael I. Jordan. 1995. An Internal Model for Sensorimotor Integration. *Science* 269, 5232 (Sept. 1995), 1880–1882. https://doi.org/10.1126/science.7569931

A Semi-Structured Interview Questions

(1) What do you think the system allowed you to do?

[After hearing the participant's response:] This system lets you control sound parameters of a synthesis engine using the pressure input. In some configurations, you only controlled the cutoff frequency of a filter, and others, you also controlled the audio waveform and a distortion effect. In some configurations, there was a direct and linear connection between your input and the parameter values, whereas in others it was nonlinear. In some cases, the parameter changes also depended on an internal system state.

(2) Did you feel like you caused the change of the sounds which came out of the system?
  - Why/why not?
  - Was this more the case with certain system settings than with others?

(3) Did you feel like you were in control of the sound by using the system?
  - Why/why not?
  - Was this more the case with certain system settings than with others?

(4) Did you feel like you could predict how the sound would end up being based on how you used the input controller?
  - Why/why not?
  - Was this more the case with certain system settings than with others?

(5) Did you enjoy the effect your actions had on the sound?
  - Why/why not?
  - Was this more the case with certain system settings than with others?

(6) Did you have a specific sonic outcome or goal that you wanted to achieve? (If yes:) Did you feel like effort was required to achieve it?
  - Why/why not?
  - Was this more the case with certain system settings than with others?

Note: as this was a semi-structured interview, the follow-up questions were asked when deemed appropriate, and additional questions might have been asked in between depending on the participant's answers.

B System Overview

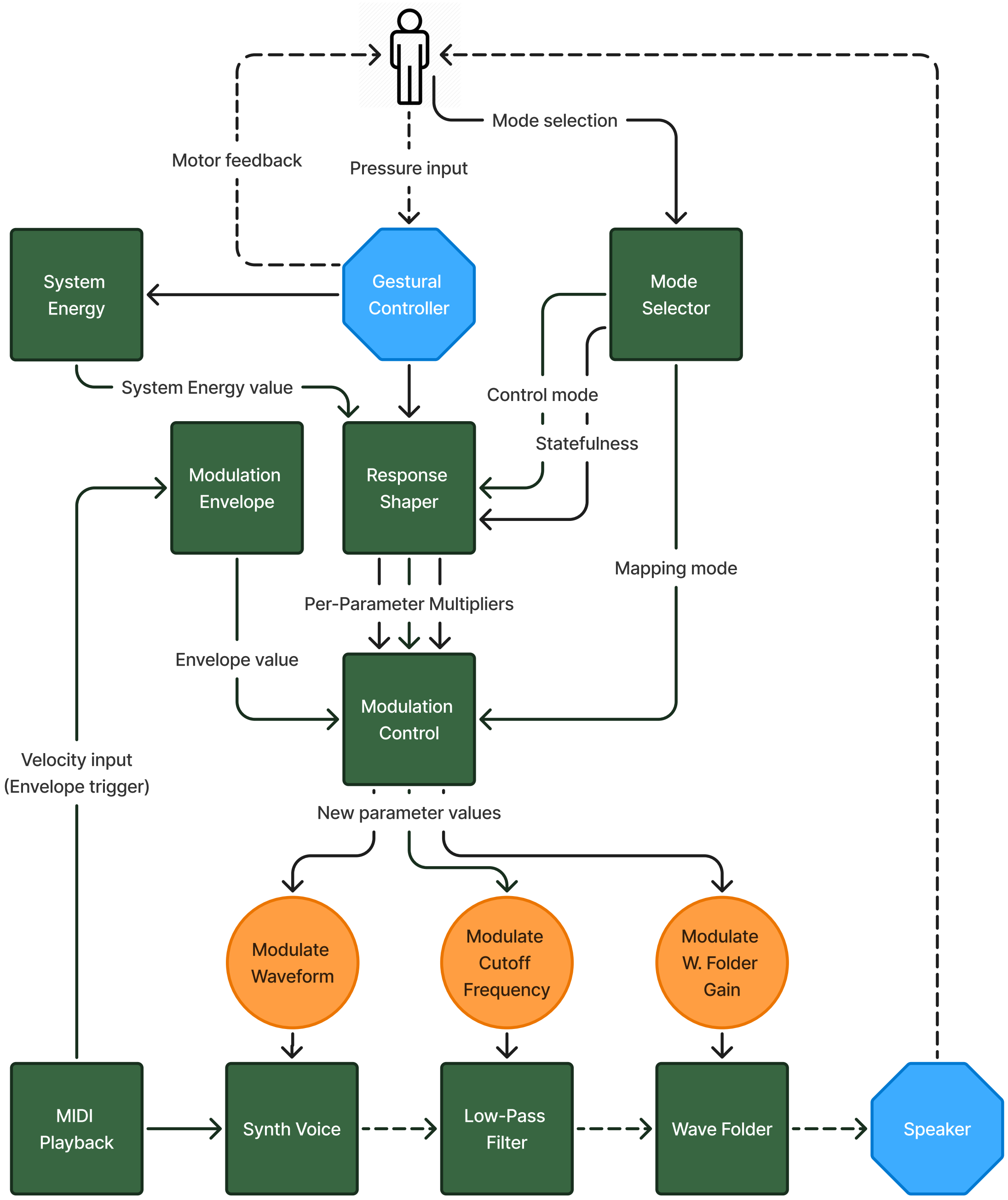


Figure 4: System architecture showing how pressure input and mode selection modulate sound parameters. The Response Shaper computes parameter multipliers based on control mode and stateful control dynamics. The Modulation Control then scales these multipliers based on mapping mode. The resulting multipliers modulate duty cycle, filter cut-off frequency, and wavefolder volume. See Appendix C for implementation details.

C Max/MSP Patcher & Signal Flow

Sections of the Max patcher for handling input, timbral manipulation, and response shaping. Please consult Section 3 for implementation details.

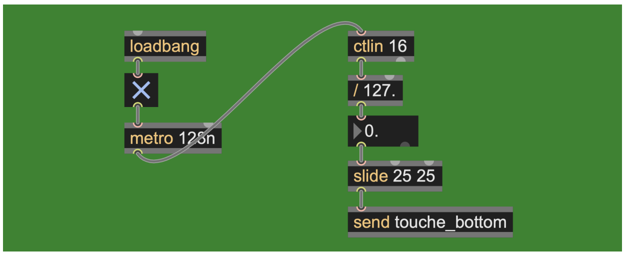


Figure 5: ***Input handling module.*** **Pressure input value is received as a MIDI CC message and sampled every 15.625 ms. It is then normalized and smoothed before the next stage in the signal flow.**

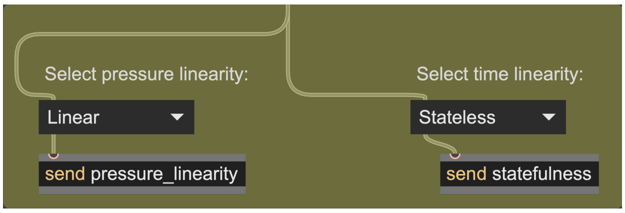


Figure 6: ***Mode selector module.*** **Responsible for configuring Pressure Linearity and Time Linearity system-wide. See Section 5.1 for a brief overview of settings.**

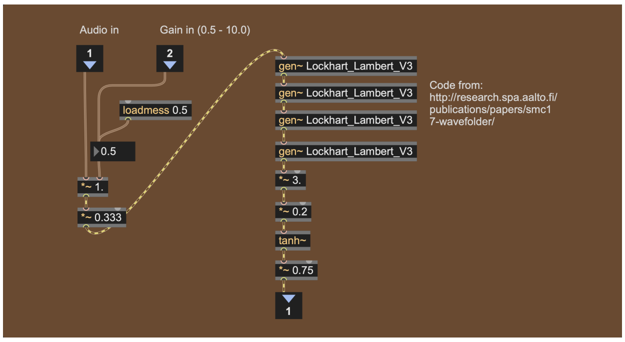


Figure 7: ***Wave Folding module.*** **Receives the audio signal as well as a gain multiplier as an input. When the gain increases, the output generally gets louder and more distorted. This specific implementation has been adapted from Esqueda et al. [15]**

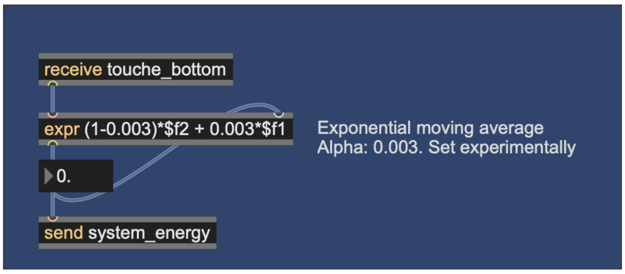


Figure 8: ***System energy module.*** **Maintains a system state that gradually builds and decays based on pressure input. Used for** ***nonlinear*** **Time Linearity. Calculated using an exponential moving average, with an α value of 0.003.**

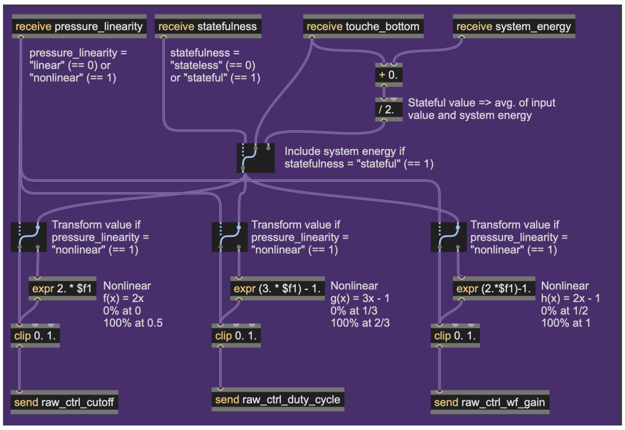


Figure 9: ***Response shaper module.*** **Responsible for incorporating the system energy (see Figure 8), if applicable. The module can also be configured to modulate all parameters at the same rate, or use independent mapping functions per parameter. See Section 3.4 further.**

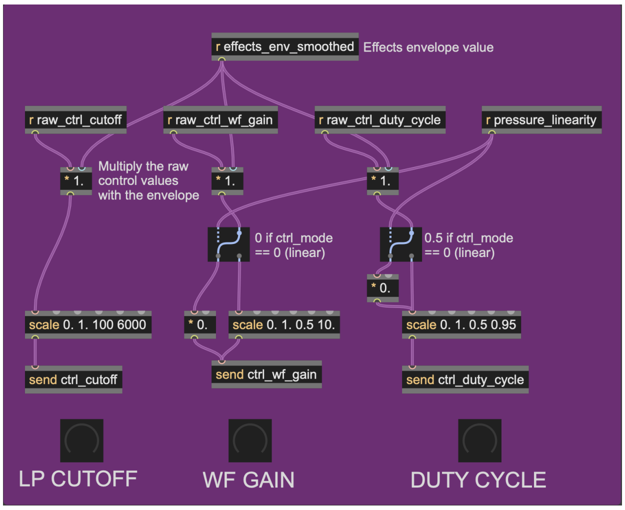
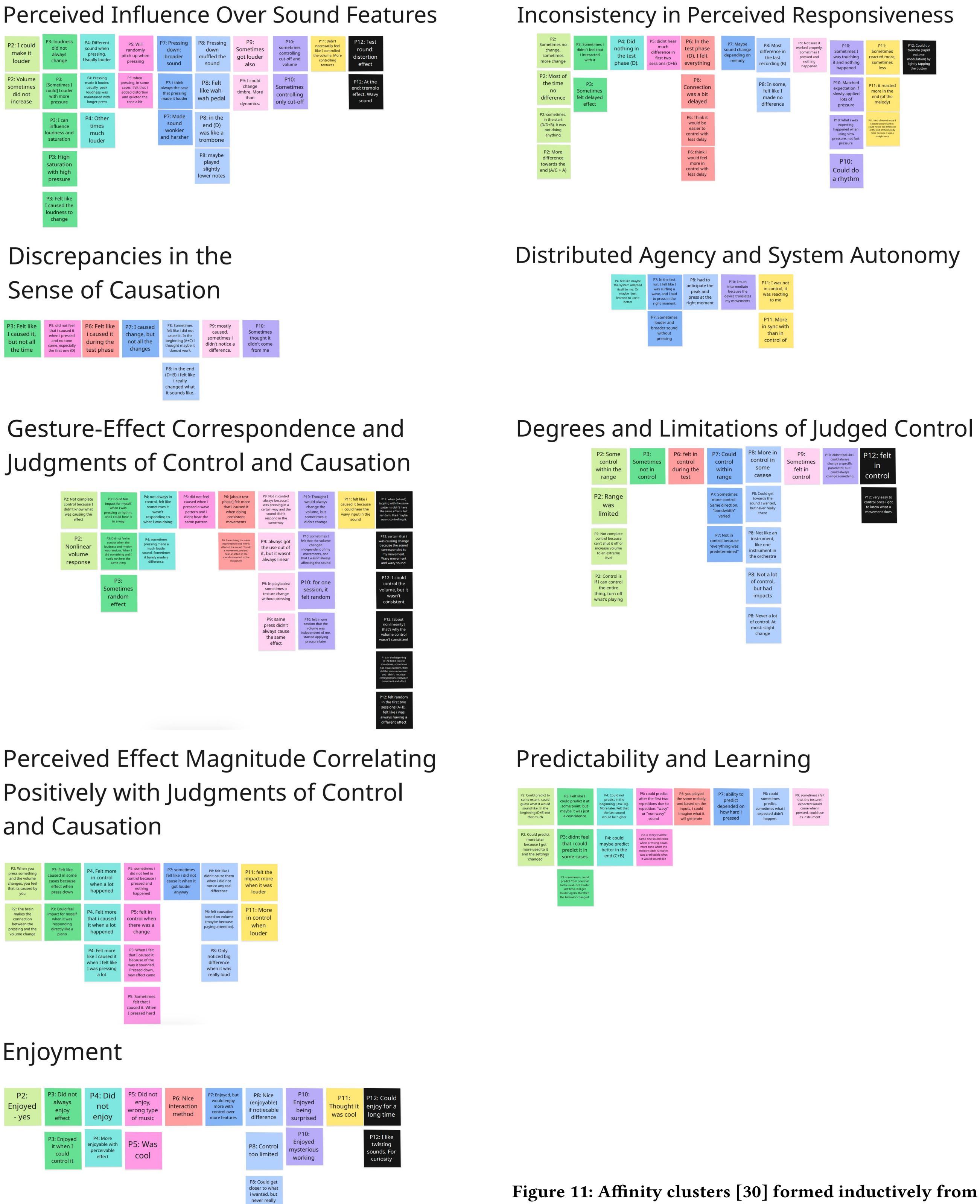


Figure 10: ***Modulation Control module.*** **Raw parameter values are shaped using the effects envelope before the normalized parameter values are scaled to their desired output ranges. If Pressure Linearity is** ***linear,*** **the default wave folder gain and waveform duty cycle values are used.**

D Affinity Clusters



Perceived Influence Over Sound Features

- P2: I could make it louder
- P2: Volume sometimes did not increase
- P3: loudness did not always change
- P3: [Sometimes I could] Louder with more pressure
- P3: I can influence loudness and saturation
- P3: High saturation with high pressure
- P3: Felt like I caused the loudness to change
- P4: Different sound when pressing. Usually louder
- P4: Pressing made it louder. usually peak loudness was maintained with longer press
- P4: Other times much louder
- P5: Will randomly pitch up when pressing
- P5: when pressing, in some cases i felt that i added distortion and quieted the tone a bit
- P7: Pressing down: broader sound
- P7: i think always the case that pressing made it louder
- P7: Made sound wonkier and harsher
- P8: Pressing down muffled the sound
- P8: Felt like wah-wah pedal
- P8: in the end (D) was like a trombone
- P8: maybe played slightly lower notes
- P9: Sometimes got louder also
- P9: I could change timbre. More than dynamics.
- P10: sometimes controlling cut-off and volume
- P10: Sometimes controlling only cut-off
- P11: Didn't necessarily feel like I controlled the volume. More controlling textures
- P12: Test round: distortion effect
- P12: At the end: tremolo effect. Wavy sound

Inconsistency in Perceived Responsiveness

- P2: Sometimes no change, sometimes more change
- P2: Most of the time no difference
- P2: sometimes, in the start (D/D+B), it was not doing anything
- P2: More difference towards the end (A/C + A)
- P3: Sometimes i didn't feel that i interacted with it
- P3: Sometimes felt delayed effect
- P4: Did nothing in the test phase (D).
- P5: didnt hear much difference in first two sessions (D+B)
- P6: In the test phase (D), I felt everything
- P6: Connection was a bit delayed
- P6: Think it would be easier to control with less delay
- P6: think i would feel more in control with less delay
- P7: Maybe sound change depending on melody
- P8: Most difference in the last recording (B)
- P8: In some, felt like I made no difference
- P9: Not sure it worked properly. Sometimes I pressed and nothing happened
- P10: Sometimes I was touching it and nothing happened
- P10: Matched expectation if slowly applied lots of pressure
- P10: what i was expecting happened when using slow pressure, not fast pressure
- P10: Could do a rhythm
- P11: Sometimes reacted more, sometimes less
- P11: it reacted more in the end (of the melody)
- P11: kind of waved more if i played around with it. could notice the difference at the end of the melody more because it was a straight note
- P12: Could do tremolo (rapid volume modulation) by lightly tapping the button

Discrepancies in the Sense of Causation

- P3: Felt like I caused it, but not all the time
- P5: did not feel that i caused it when i pressed and no tone came. especially the first one (D)
- P6: Felt like i caused it during the test phase
- P7: I caused change, but not all the changes
- P8: Sometimes felt like i did not cause it. In the beginning (A+C) i thought maybe it doesnt work
- P8: in the end (D+B) i felt like i really changed what it sounds like.
- P9: mostly caused. sometimes i didn't notice a difference.
- P10: Sometimes thought it didn't come from me

Distributed Agency and System Autonomy

- P4: felt like maybe the system adapted itself to me. Or maybe i just learned to use it better
- P7: In the test run, I felt like I was surfing a wave, and I had to press in the right moment
- P7: Sometimes louder and broader sound without pressing
- P8: had to anticipate the peak and press at the right moment
- P10: I'm an intermediate because the device translates my movements
- P11: I was not in control, it was reacting to me
- P11: More in sync with than in control of

Gesture-Effect Correspondence and Judgments of Control and Causation

- P2: Not complete control because I didn't know what was causing the effect
- P2: Nonlinear volume response
- P3: Could feel impact for myself when i was pressing a rhythm, and i could hear it in a way
- P3: Did not feel in control when the loudness and rhythm was random. When I did something and I could not hear the same thing
- P3: Sometimes random effect
- P4: not always in control, felt like sometimes it wasn't responding to what I was doing
- P4: sometimes pressing made a much louder sound. Sometimes it barely made a difference.
- P5: did not feel caused when i pressed a wave pattern and i didnt hear the same pattern
- P6: [about test phase] felt more that i caused it when doing consistent movements
- P6: I was doing the same movement to see how it affected the sound. You do a movement, and you hear an effect in the sound connected to the movement
- P9: Not in control always because I was pressing in a certain way and the sound didn't respond in the same way
- P9: always got the use out of it, but it wasnt always linear
- P9: In playbacks: sometimes a texture change without pressing
- P9: same press didn't always cause the same effect
- P10: Thought I would always change the volume, but sometimes it didn't change
- P10: sometimes I felt that the volume changed independent of my movements, and that I wasn't always affecting the sound
- P10: for one session, it felt random
- P10: felt in one session that the volume was independent of me. started applying pressure later
- P11: felt like i caused it because i could hear the wavy input in the sound
- P12: when [when?] tapping with the same patterns didn't have the same effects. felt random, like i maybe wasnt controlling it.
- P12: certain that i was causing change because the sound corresponded to my movement. Wavy movement and wavy sound.
- P12: I could control the volume, but it wasn't consistent
- P12: [about nonlinearity] that's why the volume control wasn't consistent
- P12: in the beginning (B+A): felt it control sometimes, sometimes not. it was random. then did the same movement and i didn't. not clear correspondance between movement and effect
- P12: felt random in the first two sessions (A+B). felt like i was always having a different effect

Degrees and Limitations of Judged Control

- P2: Some control within the range
- P2: Range was limited
- P2: Not complete control because can't shut it off or increase volume to an extreme level
- P2: Control is if i can control the entire thing, turn off what's playing
- P3: Sometimes not in control
- P6: felt in control during the test
- P7: Could control within range
- P7: Sometimes more control. Same direction, "bandwidth" varied
- P7: Not in control because "everything was predetermined"
- P8: More in control in some casese
- P8: Could get towards the sound I wanted, but never really there
- P8: Not like an instrument, like one instrument in the orchestra
- P8: Not a lot of control, but had impacts
- P8: Never a lot of control. At most: slight change
- P9: Sometimes felt in control
- P10: didn't feel like I could always change a specific parameter, but I could always change something
- P12: felt in control
- P12: very easy to control once i got to know what a movement does

Perceived Effect Magnitude Correlating Positively with Judgments of Control and Causation

- P2: When you press something and the volume changes, you feel that its caused by you
- P2: The brain makes the connection between the pressing and the volume change
- P3: Felt like caused in some cases because effect when press down
- P3: Could feel impact for myself when it was responding directly like a piano
- P4. Felt more in control when a lot happened
- P4. Felt more that i caused it when a lot happened
- P4: Felt more like I caused it when I felt like I was pressing a lot
- P5: sometimes i did not feel in control because i pressed and nothing happened
- P5: felt in control when there was a change
- P5: When I felt that I caused it: because of the way it sounded. Pressed down, new effect came
- P5: Sometimes felt that i caused it. When I pressed hard
- P7: sometimes felt like i did not cause it when it got louder anyway
- P8: felt like i didn't cause them when i did not notice any real difference
- P8: felt causation based on volume (maybe because paying attention).
- P8: Only noticed big difference when it was really loud
- P11: felt the impact more when it was louder
- P11: More in control when louder

Predictability and Learning

- P2: Could predict to some extent, could guess what it would sound like. In the beginning (D+B) not that much
- P2: Could predict more later because I got more used to it and the settings changed
- P3: Felt like I could predict it at some point, but maybe it was just a coincidence
- P3: didnt feel that i could predict it in some cases
- P3: sometimes i could predict from one trial to the next. Got louder last time, will get louder again. But then the behavior changed.
- P4: Could not predict in the beginning (D/A+D)). More later. Felt that the last sound would be higher
- P4: could maybe predict better in the end (C+B)
- P5: could predict after the first two repetitions due to repetition. "wavy" or "non-wavy" sound
- P5: in every trial the same one sound came when pressing down. more tone when the melody pitch is higher. was predictable what it would sound like
- P6: you played the same melody, and based on the inputs, i could imagine what it will generate
- P7: ability to predict depended on how hard i pressed
- P8: could sometimes predict. sometimes what i expected didn't happen.
- P9: sometimes i felt that the texture i expected would come when i pressed. could use as instrument

Enjoyment

- P2: Enjoyed - yes
- P3: Did not always enjoy effect
- P3: Enjoyed it when I could control it
- P4: Did not enjoy
- P4: More enjoyable with perceivable effect
- P5: Did not enjoy, wrong type of music
- P5: Was cool
- P6: Nice interaction method
- P7: Enjoyed, but would enjoy more with control over more features
- P8: Nice (enjoyable) if noticeable difference
- P8: Control too limited
- P8: Could get closer to what i wanted, but never really what i wanted
- P10: Enjoyed being surprised
- P10: Enjoyed mysterious working
- P11: Thought it was cool
- P12: Could enjoy for a long time
- P12: I like twisting sounds. For curiosity

Figure 11: Affinity clusters [30] formed inductively from participant interviews, as coded and organised in Miro.