Perceptual Discriminability of Temporal Dynamics in Audiovisual Mappings: Entry-Exit Asymmetry

Yiyang Feng

Submitted in partial fulfillment of the requirements for the
Master of Music in Music Technology
in the Department of Music and Performing Arts Professions
Steinhardt School
New York University

Advisor: Pablo Ripolles

November, 2025

TABLE OF CONTENTS

ABSTRACT **5**
1. INTRODUCTION **6**

2. PRIOR WORK **9**

2.1 Instantaneous Mapping 9
2.1.1 Instantaneous Mapping: The Dominant Paradigm 9
2.1.2 Crossmodal Correspondences as Perceptual Foundation 10
2.2 Temporal Perception: Constraints and Capabilities 11
2.2.1 The Temporal Binding Window 11
2.2.2 Perceptual Asymmetry: Gradual vs. Instantaneous Change 12
2.2.3 Response Bias and Evidence Accumulation 12
2.3 Theoretical Perspectives on State-Driven Interaction 14
2.4 Research Gap and Study Motivation 15

3. METHOD **17**

3.1 Participants 17
3.2 Apparatus 18
3.2.1 Platform and Hardware 18
3.2.2 Software Architecture 18
3.2.3 User Interface Elements 19
3.3 Stimuli and Visual System 20
3.3.1 Audio Materials & Visual Rendering 20
3.3.2 Visual Rendering System 20
3.3.3 Mapping Architectures 20
3.4 Procedure 26
3.4.1 Pre-Experiment Phase 26
3.4.2 Task Instructions 26
3.4.3 Session Timeline 26
3.4.4 Decay Parameter Selection 28
3.4.5 Adaptive Difficulty and Spatial Remapping 29
3.5 Database Architecture and Data Quality Control 32
3.6 Data Analysis 32
3.6.1 Switch Classification Using current_mode 32
3.6.2 Detection Measures 33
3.6.3 Statistical Analysis 34
3.6.4 Signal Detection Theory Analysis 36

4. ANALYSIS & RESULTS **38**

4.1 Data overview 38
4.2 Descriptive Statistics 39
4.2.1 Overall Detection Performance 39
4.2.2 Hit Rate Distributions 39
4.2.3 Reaction Time Distributions 41
4.2.4 Cumulative Detection Curves 41
4.3 Hypothesis Testing 43

4.3.1 Statistical Assumptions and Multiple Comparison Corrections *43*
4.3.2 H1: Detection Latency Asymmetry *45*
4.3.3 H2: Response Bias Asymmetry *47*
4.4 Effect Size Summary 50
4.5 Robustness Checks 52
4.5.1 Outlier Sensitivity *52*
4.5.2 Alternative RT Windowing *53*
4.5.3 Baseline vs. Adaptive Analysis *53*

5. DISCUSSION **54**

5.1 Summary of Findings 54
5.2 Mechanisms and Interpretation 56
5.3 Theoretical Positioning and Generalizability 59
5.4 Implications for Model-Based Sonification Design 62
5.5 Theoretical Implications for Evidence Accumulation Models 63
5.6 Limitations and Future Directions 64
5.7 Conclusion 65

REFERENCES **67**

APPENDIX A: Power Analysis Details **71**

APPENDIX B: Software Architecture Details **72**

B.1 API Endpoints
B.2 Client-Side Resilience Mechanisms

APPENDIX C: Database Schema **73**

C.1 Table Definitions

APPENDIX D: Spatial Transformation Specifications **75**

D.1 Base Band Positions
D.2 Transformation Types

APPENDIX E: Temporal Parameter Selection Justification **76**

E.1 Perceptual Constraints from Literature
E.2 The "Perceptible-but-Integratable" Zone
E.3 Parameter Selection
E.4 Adaptive Range

LIST OF FIGURES

Figure 3.1. Initial guide interface with Hue selection 18

Figure 3.2. M-B vs M-A architecture diagram 21

Figure 3.3. Simulation of the M-B energy state 24

Figure 3.4. Mode Mapping Comparison preview 26

Figure 3.5. Automatic difficulty adaptor 25

Figure 4.1: Hit Rate Comparison by Transition Type 41

Figure 4.2. Paired Comparison of Hit Rates Across Sessions 42

Figure 4.3. Reaction Time Distributions for Detected Transitions· 43

Figure 4.4 Cumulative Detection Rates Over Time 44

Figure 4.5. Signal Detection Theory Parameter Space 49

Figure 4.6 Complete Signal Detection Theory Analysis 50

Figure 4.7 Comprehensive Summary of Entry vs. Exit Asymmetries 52

Figure 5.1. Ecological Origins of Perceptual Asymmetry 58

ABSTRACT

Real-time music visualization systems typically employ instantaneous mapping, where visual parameters respond directly to current audio features. An alternative architecture, state-driven mapping, integrates temporal history through exponential decay. Whether observers can perceptually discriminate between these architectures during continuous playback remains unknown. This study investigated detection of transitions between instantaneous (M-A) and state-driven (M-B, α=0.92, $\tau \approx$200ms) mapping modes. Fifty-five participants monitored a music visualization and pressed the spacebar upon detecting mode transitions. Signal Detection Theory analysis revealed asymmetric detection: Exit transitions (B→A, mode cessation) achieved higher detection rates (67.1% vs 59.1%, p=.004), greater perceptual sensitivity (d'=1.378 vs 1.095, p=.011), and less conservative response criteria (c=0.35 vs 0.46, p=.025), compared to Entry transitions (A→B, mode onset). These findings demonstrate that architectural differences in temporal dynamics produce discriminable perceptual signatures, with abrupt state cessation more salient than gradual state accumulation. The asymmetry parallels established detection differences between instantaneous and gradual change in motion perception. For interaction designers, temporal dynamics constitute a user-facing design parameter requiring systematic consideration alongside traditional mappings.
Keywords: audiovisual perception, temporal dynamics, signal detection theory, music visualization, crossmodal correspondences

1. INTRODUCTION

Real-time music visualization systems present designers with fundamental architectural choices regarding how visual parameters respond to audio features. The dominant approach employs instantaneous mapping, wherein visuals respond directly to current audio features on a frame-by-frame basis. This dominance derives from two factors: computational efficiency and conventional tooling. Frameworks such as Processing, p5.js, and Max/MSP provide direct mapping primitives as default implementations, rendering stateless architectures the path of least resistance (Shiffman, 2015). This design convention contradicts a fundamental property of human perception. The perceptual system operates through state-dependent processing—visual adaptation, temporal integration, and expectation formation all rely on history-dependent mechanisms. An alternative architecture, state-driven mapping, integrates temporal history through internal state accumulation governed by exponential decay. Despite these architecturally distinct computational models, their perceptual consequences remain unquantified.

This thesis investigates a specific question: when an audiovisual system transitions between instantaneous and state-driven processing during continuous playback, can observers detect these architectural changes? If so, with what latency and reliability? The answer carries design implications. If transitions prove difficult to detect, architectural choices remain primarily engineering concerns—efficiency and implementation convenience justify current practices; If architectural differences produce readily detectable perceptual signatures, temporal dynamics constitute a user-facing design parameter requiring systematic optimization, analogous to color mapping or interaction metaphors.

We quantified detection thresholds for temporal dynamics in audiovisual mappings through a controlled transition detection paradigm. Participants monitored a music visualization that alternated between two mapping modes:

- Mode A (M-A): Instantaneous mapping where visual brightness derives directly from current audio energy with no temporal dependencies: $V(t) = f(A(t))$, where

V denotes visual brightness (rendered as Gaussian cloud intensity), A denotes audio energy (RMS amplitude per frequency band), t denotes time in frames, and f is a scaling function that maps audio energy to brightness values.

- Mode B (M-B): State-driven mapping where visual output emerges from internal energy states (computed from RMS amplitude) that accumulate over time with exponential decay

Participants pressed the spacebar whenever they detected a transition between modes. We measured detection sensitivity (hit rate), reaction times, and individual variability to quantify the discriminability of these architectural changes. By moving beyond the discrete stimulus pairs typical of prior work (Stevenson et al., 2012) to continuous audiovisual streams, this study applies psychophysical methods to the context of interaction design.

Two hypotheses guided the investigation:

- **H1** (Detection Latency Asymmetry): Transitions from M-A to M-B (**Entry**: entering state-driven processing) will elicit longer reaction times than M-B to M-A transitions (**Exit**: exiting state-driven processing). This prediction parallels Gottsdanker's (1956) finding that smooth acceleration proves difficult to detect while instantaneous velocity changes achieve immediate salience.
- **H2** (Response Bias Asymmetry): Beyond detection speed, Entry and Exit transitions will elicit different response criteria (c in Signal Detection Theory framework—the internal threshold determining how much evidence an observer requires before reporting a detection; higher c reflects more conservative responding, lower c reflects more liberal responding). Entry's gradual accumulation creates temporal ambiguity regarding transition onset, requiring participants to accumulate stronger evidence before confidently reporting detection, predicted to yield more conservative response criteria (higher c values). Exit's abrupt cessation produces immediate perceptual contrast, enabling confident detection with lower decision thresholds (lower c values). We focus our

prediction on criterion (c) rather than sensitivity (d'); whether temporal dynamics affect perceptual discriminability itself remains an open empirical question.

This study prioritizes perceptual discriminability, whether architectural differences produce detectable signatures, establishing it as a necessary foundation for broader claims about experiential impact. Questions of aesthetic preference, emotional engagement, or subjective quality require different methodologies and presuppose the architectural discriminability we seek to establish. The approach tests learned discriminability: following brief demonstration of the two modes' perceptual qualities, participants monitor continuous playback and report detected transitions. This design reflects real-world scenarios where users familiarize themselves with system behavior before performing or monitoring tasks, rather than testing naive spontaneous detection.

Key Terms

- Instantaneous mapping (M-A): An audiovisual architecture where visual parameters compute directly from current audio features (e.g., RMS amplitude across five frequency bands: sub-bass, bass, low-mid, high-mid, and treble) within single processing frames, maintaining no temporal dependencies or state across time.
- State-driven mapping (M-B): An audiovisual architecture where visual brightness is determined not only by the current audio input, but also by an internal "energy pool" that retains information from previous frames. When audio energy rises, the pool fills gradually; when audio energy drops, the pool drains slowly rather than instantly—a process called exponential decay. Formally, the energy state E at time t is computed as $E(t) = \alpha \cdot E(t-1) + (1-\alpha) \cdot I(t)$, where $I(t)$ is the current audio input and α (the decay coefficient, set to 0.92) controls how much of the previous state is retained. This produces visual smoothing and temporal inertia absent in instantaneous mapping.
- Temporal binding window (TBW): The interval (±200-300ms) within which audiovisual signals integrate perceptually as unified multisensory events (Vroomen & Keetels, 2010).

- Perceptual discriminability: The measurable ability to detect stimulus differences, quantified through Signal Detection Theory metrics—sensitivity (d') and response bias (c)—as well as reaction times (Macmillan & Creelman, 2005).

2. PRIOR WORK

This chapter reviews converging literatures that frame this study of architectural discriminability. Section 2.1 examines instantaneous mapping as the dominant paradigm. Section 2.2 synthesizes temporal perception research, establishing perceptual constraints and capabilities that inform the hypotheses. Section 2.3 surveys theoretical frameworks that conceptualize state-driven dynamics without validation. Section 2.4 identifies the specific gap, absence of controlled evidence on architectural discriminability, motivating the experimental approach.

2.1 Instantaneous Mapping: The Dominant Paradigm

Music visualization and real-time audiovisual systems have historically relied on instantaneous mapping: direct, frame-by-frame translation of audio features (spectral energy, amplitude) into visual parameters (color, brightness, spatial position). This approach represents the conventional paradigm in interactive media, optimized for computational efficiency and perceptual coherence. Understanding its foundations and implicit assumptions provides necessary context for investigating alternative architectures.

2.1.1 Crossmodal Correspondences as Perceptual Foundation

Instantaneous mapping leverages well-documented crossmodal correspondences, non-arbitrary associations between sensory dimensions that facilitate perceptual integration (Spence, 2011). Higher pitches systematically associate with brighter colors and elevated spatial positions (Marks, 1974; Spence, 2011); louder sounds with greater visual magnitudes (e.g., larger shapes, more intense luminance; Evans & Treisman, 2010). These associations likely arise from shared mechanisms of magnitude processing (Walsh, 2003) or learned ecological statistics (Parise & Spence, 2009). Empirical studies confirm robustness across multiple feature pairings: pitch-brightness (Marks, 1974), loudness-size (Evans & Treisman, 2010), and rhythmic-visual periodicity (Collier & Hubbard, 2001). M-A exploits these correspondences directly, creating immediate and intuitive audiovisual relationships that maintain perceptual coherence without explicit training.

The effectiveness of instantaneous mapping is well-established across application domains. Real-time music visualizers, from consumer media players to artistic installations, successfully employ direct feature-to-visual mappings (Levin, 2000; Lieberman, 2007). Modern web technologies enable sub-10-millisecond timing precision for stimulus presentation and response collection (de Leeuw & Motz, 2016; Anwyl-Irvine et al., 2021), removing technical barriers to frame-synchronized rendering. This computational efficiency, combined with perceptual coherence through crossmodal correspondences, explains instantaneous mapping's practical dominance.

2.1.2 Computational Pragmatism and Design Conventions

Beyond perceptual foundations, instantaneous mapping reflects computational pragmatism. Real-time systems face strict latency constraints: interactive applications target 60Hz refresh rates (16.7ms frame duration), and perceptual studies of audiovisual synchrony establish sensitivity to delays exceeding approximately 30ms (Fujisaki et al., 2004). Stateless architectures minimize computational overhead, each frame’s output depends only on current input, requiring no memory allocation or state management. This enables scalability: systems can process arbitrary-length audio streams with constant memory footprint and predictable computational cost per frame.

Design conventions further reinforce this paradigm. Widely-adopted creative coding frameworks (Processing, p5.js, Max/MSP) provide direct mapping primitives as default tools, whereas state management requires additional implementation effort. Educational resources emphasize feature extraction and visual parameter control without addressing temporal dynamics or internal state (Shiffman, 2015). Architectural choices thus follow tool affordances rather than explicit perceptual reasoning.

2.2 Temporal Perception: Constraints and Capabilities

The two hypotheses make specific claims about how observers process temporal dynamics—claims that require grounding in prior research. Three lines of evidence are directly relevant. First, the temporal binding window sets boundaries on how much temporal smoothing an audiovisual system can introduce before observers perceive desynchronization. Second, documented asymmetries between detecting gradual versus instantaneous change inform H1's prediction about Entry versus Exit latencies. Third, research on adaptive response strategies under uncertainty informs H2's prediction about differential decision criteria.

2.2.1 The Temporal Binding Window

The **temporal binding window (TBW)** represents the interval within which audiovisual signals integrate perceptually as originating from a unified multisensory event. Vroomen and Keetels (2010) establish that the TBW extends approximately ±200-300ms for most observers, though individual differences exist (Stevenson et al., 2012). Within this window, the brain actively constructs synchrony rather than passively detecting it: asynchronous signals undergo perceptual fusion, with subjective simultaneity adjusted to maintain coherent multisensory experience.

This flexibility reflects ecological validity. Natural audiovisual events—speech, object impacts, musical performances—produce physical asynchronies due to differences in light and sound transmission speeds, sensory transduction latencies, and neural processing delays (King & Palmer, 1985). The TBW compensates for these variabilities, prioritizing perceptual coherence over physical veridicality.

For audiovisual system design, the TBW establishes a constraint: temporal dynamics must operate within this window to maintain perceptual binding. M-B's decay rates (α = 0.92-0.97, where α determines how quickly visual brightness fades after audio energy drops, higher values produce slower, more gradual fading) produce temporal smoothing on the order of 50-200ms, falling within TBW boundaries. However, TBW research typically measures whether observers perceive a fixed delay (e.g., a 200ms lag between flash and beep) as synchronous—not whether they can detect ongoing changes in how audio and visual signals relate over time. Our study addresses this latter question. Integration capability does not imply insensitivity: the perceptual system can tolerate asynchrony while still detecting differences in temporal architecture. This distinction between tolerance and discriminability defines the research question.

2.2.2 Perceptual Asymmetry: Gradual vs. Instantaneous Change

Detection asymmetries in perceptual systems provide direct insight into how architectural transitions might be processed. The distinction between gradual and instantaneous change, central to the M-A/M-B comparison, has established perceptual consequences. Gottsdanker's (1956) research on motion perception reveals a fundamental asymmetry that parallels our architectural comparison: observers are highly insensitive to smooth acceleration, reporting constant velocity despite continuous change (p. 480), yet extremely sensitive to instantaneous velocity changes. He describes this distinction as the "gradualness of transition" (p. 481): smooth changes involve minimal moment to moment acceleration, while instantaneous changes represent "infinitely high" acceleration.

This principle suggests that M-B Entry (analogous to smooth acceleration through gradual state accumulation) should prove harder to detect than M-B Exit (analogous to instantaneous velocity change through abrupt state cessation). The parallel is suggestive: Entry accumulates temporal history across multiple frames. It produces a diffuse visual shift that—much like smooth acceleration—offers no clear perceptual anchor. In contrast, Exit instantaneously severs temporal dependencies, producing discontinuity as salient as sudden velocity change. This psychophysical asymmetry underpins the central prediction, that Entry and Exit transitions will be processed through fundamentally different detection mechanisms.

2.2.3 Response Bias and Evidence Accumulation

Beyond detection speed, temporal dynamics may influence decision confidence. Signal Detection Theory distinguishes between sensitivity (d', the ability to discriminate signal from noise) and response bias (c, the criterion for reporting detection). While H1 specifically addresses the latency asymmetry, H2 posits that the temporal ambiguity of Entry will necessitate more conservative response criteria compared to the immediacy of Exit.

The gradual versus instantaneous nature of Entry/Exit transitions may affect evidence accumulation dynamics, which in turn shape decision criteria. Entry transitions (A→B) involve progressive state accumulation through exponential decay, a gradual fading process where each frame retains a fixed proportion (α=0.92 or 92%) of the previous frame's energy while adding new input. This smoothing creates temporal ambiguity about when the regime change occurred. The transition moment lacks clear perceptual anchor points—observers may notice “something is different” but struggle to pinpoint the exact instant, consistent with the temporal uncertainty inherent in detecting gradual change (Gottsdanker, 1956). Without this temporal precision, participants may require stronger evidence before confidently reporting detection, leading to more conservative response criteria (H2: higher c values).

Exit transitions (B→A), in contrast, involve instantaneous state release. Accumulated visual energy (the brightness pool built up through M-B’s temporal integration) disappears abruptly when temporal dependencies are severed, producing immediate perceptual contrast—the “trailing” quality of M-B visuals suddenly ceases. This unambiguous transition moment enables confident detection with less accumulated evidence, leading to less conservative response criteria (H2: lower c values).

H2 Prediction: Entry c > Exit c. If confirmed, this criterion asymmetry would demonstrate that temporal ambiguity in gradual transitions leads observers to adopt more conservative response strategies, requiring stronger evidence before reporting detection. This connects the present study to theories of evidence accumulation under uncertainty (Ratcliff & McKoon, 2008) and metacognitive monitoring (Fleming & Dolan, 2012).

An alternative account based on expectation violation (Huron, 2006) would predict the opposite pattern: if Entry's lag appearance violates expectations of immediate response, it might trigger less conservative criteria (lower c) due to surprising salience, while Exit's return to immediacy might require conservative criteria (higher c) to confirm the expected state. Our temporal ambiguity account predicts Entry c > Exit c, while this expectation-based account predicts the reverse, allowing the data to adjudicate between mechanisms.

2.3 Theoretical Perspectives on State-Driven Interaction

Wiener (1948) distinguished between systems that respond to instantaneous conditions and those that incorporate memory through feedback loops. This distinction, foundational to control theory, maps onto the M-A versus M-B comparison. Building on this foundation, interaction design researchers have proposed that temporal dynamics create distinctive experiential qualities. Picard (1997) argues that authentic emotion models require "temporal dynamics that reflect how emotion changes over time, including decay and accumulation"—precisely the computational properties instantiated in M-B through its decay coefficient (α) and energy accumulation mechanisms. Höök (2008) proposes "affective loops" where system state continuously influences user experience, creating ongoing dialogues rather than discrete interactions. Similarly, "slow technology" frameworks prioritize temporal reflection over computational efficiency, suggesting that intentionally introduced delays foster deeper engagement (Hallnäs & Redström, 2001).

These theoretical proposals share a common assumption: state-driven architectures with temporal dependencies create qualitatively different user experiences than instantaneous, stateless systems. However, this assumption lacks validation. Before claiming that state-driven architectures create "affective loops" or "reflective experiences," a more fundamental question needed to be addressed: are these architectural differences perceptually accessible? If observers cannot detect when temporal dynamics are present, claims about experiential impact lack foundation.

This study provides the necessary first step: quantifying the perceptual discriminability of architectural differences. We test whether observers can reliably detect transitions between instantaneous and state-driven processing modes during continuous audiovisual playback. Perceptual detectability is a prerequisite, not a substitute, for understanding richer phenomena like aesthetic preference or emotional response. By establishing whether and how architectural choices register in conscious perception, we provide the empirical foundation that theoretical frameworks currently lack.

2.4 Research Gap and Study Motivation

The reviewed literature establishes that the perceptual system demonstrates sophisticated temporal processing: it integrates audiovisual signals within ~200-300ms windows (Vroomen & Keetels, 2010), detects asymmetries between gradual and abrupt changes (Gottsdanker, 1956), and adapts response criteria based on signal quality. Theoretical frameworks propose that state-driven architectures, where system output depends on accumulated temporal history, create experiential qualities distinct from instantaneous, stateless mappings (Picard, 1997; Höök, 2008).

Yet while traditional temporal perception research measures whether two discrete events (e.g., a flash and a beep) are perceived as simultaneous, no studies have tested whether observers can detect transitions between computational architectures during continuous audiovisual playback. In audio engineering, state-driven processing (reverb, compression) produces clearly audible effects. Whether analogous temporal dynamics in visual mappings—specifically, the distinction between $V(t) = f(A(t))$ and $V(t) = g(A(t), V(t-1))$, where V is visual brightness, A is audio energy, and t is time—produce perceptually discriminable signatures remains unknown. The first equation describes instantaneous mapping (output depends only on current input), while the second describes state-driven mapping (output depends on both current input and previous visual state).

This gap matters for both theory and design. Detectability is the pivot point: if users cannot perceive architectural shifts, state-driven systems are merely engineering overhead, and current efficiency-driven practices favoring instantaneous mapping (M-A) are justified; if they can, temporal dynamics become a critical design material that the field has systematically overlooked. We provide the first controlled test of this question. Participants monitored continuous music visualizations that alternated between instantaneous (M-A) and state-driven (M-B, $\alpha=0.92$, $\tau \approx 200$ms) processing modes. Using Signal Detection Theory, we quantify detection sensitivity (d'), response bias (c), and reaction times.

3. METHOD

This chapter details the experimental methodology investigating perceptual differences between two audiovisual mapping paradigms: M-A (instantaneous mapping) and M-B (state-driven model). The study employed a within-subjects design wherein participants detected transitions between mapping conditions during continuous music visualization.

3.1 Participants

A priori power analysis (GPower) determined required sample size for detecting perceptual differences between mapping conditions. We targeted 50 participants to account for technical failures and ensure 80% power. Power analysis assumptions: Paired-samples t-test (two-tailed), expected effect size $d = 0.5$, $\alpha = .05$, yielding minimum $N = 45$. Detailed rationale for effect size estimation appears in Appendix A.

Participants were recruited through social media and university mailing lists. Qualtrics collected unique userIDs and email addresses, with emails SHA-256 hashed before PostgreSQL storage to ensure privacy. Each participant accessed the experiment via a unique URL containing their userID and hashed email.

- Inclusion criteria:
    - Age >18 years
    - Normal or corrected-to-normal vision and hearing (confirmed via self-report on the experiment screen prior to task onset)
    - Headphone recommended
    - Access to desktop or laptop computer (mobile devices excluded)
    - Stable internet connection
- Minimum engagement thresholds for valid sessions:
    - Playback duration > 220 seconds
    - Response attempts ≥ 5 spacebar presses
    - Valid detections ≥ 2 hits within [0, 2000ms] response window

Sessions failing these criteria were automatically flagged as invalid. Additionally, we excluded sessions with zero hits despite ≥5 keypresses, or with all negative reaction times (indicating systematic timing errors).

Before entering the main experiment, participants completed a 60-second practice session with 9 predetermined mode switches at 6-second intervals. Progression to the main task required:

- Minimum 3 successful hits
- ≥30% hit rate

Participants failing these criteria were prompted to retry. This quality gate ensured that only participants demonstrating basic task understanding contributed to the main dataset.

3.2 Apparatus

3.2.1 Platform and Hardware

The experiment was implemented as a browser-based application using web standards for maximal accessibility. Participants interacted via computer keyboard (spacebar) while audio played through system audio output. Technical specifications: p5.js framework for visual rendering, HTML5 Canvas API for graphics, HTML5 Audio element for playback. All event timestamps synchronized to `audio.currentTime` for precise timing independent of visual frame rate.

3.2.2 Software Architecture

The system employed a modular client-server architecture designed for robustness and data integrity. Server-side infrastructure consisted of Neon PostgreSQL with four relational tables (thesis_sessions, thesis_logs, thesis_switches, thesis_feedback) and RESTful API endpoints for session management and event logging. Detailed database schema and API specifications appear in Appendix B.

Client-side resilience mechanisms included asynchronous logging with retry logic, network failure handling via browser localStorage queueing, and exponential backoff (up to 3 attempts: 200ms, 400ms, 800ms). Session tokens prevented duplicate submissions.

3.2.3 User Interface Elements

After completing the Qualtrics pre-screening survey, participants accessed the experiment through their unique URL. The application initialized with a full-screen welcome overlay displaying task instructions. The main interface featured:

- Playback Control: Floating Play/Pause button (also bound to Enter key)
- Hue Slider: Optional color palette exploration tool with three triadic anchors (Low/Mid/High) maintaining 120-degree separation, preserving band-to-color mapping structure while allowing aesthetic customization
- Visual Feedback: Brief colorful edge-glow confirmed each spacebar press without revealing response correctness
- Countdown Timer: 3-second visual countdown before stimulus onset

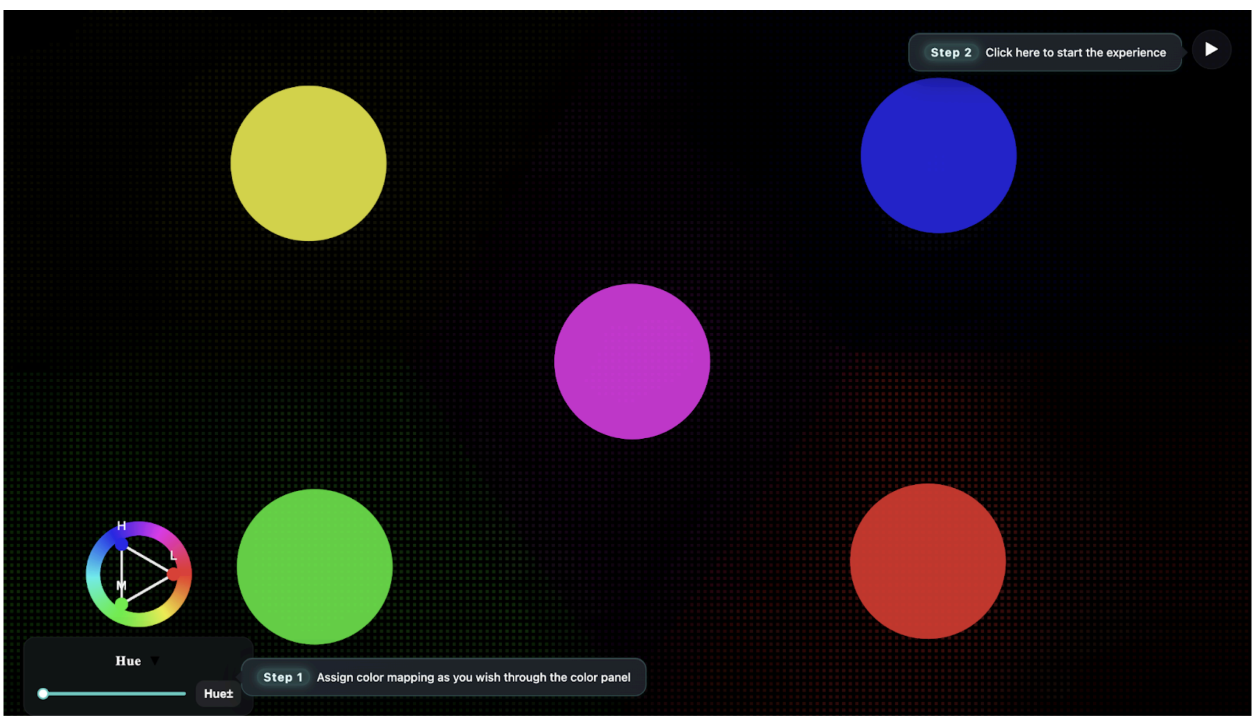

Figure 3.1 Initial guide interface with Hue selection

3.3 Stimuli and Visual System

3.3.1 Audio Materials & Visual Rendering

The auditory stimulus was a single pre-recorded 220-second stereo music excerpt (electronic/downtempo genre) selected for wide dynamic range, clear rhythmic structure, and low familiarity to minimize recognition confounds.

Audio preprocessing: Features were pre-computed and stored at 60 frames per second with 5-band frequency decomposition (sub-bass, bass, low-mid, high-mid, treble). Per-band energy normalization was applied based on dynamic range analysis. Final format: Stereo MP3, 44.1 kHz, 256 kbps.

3.3.2 Visual Rendering System

The visualization comprised five spatially distributed Gaussian regions ("clouds") corresponding to the frequency bands, with overlapping regions optimized for visual balance. Spatial layout positioned Band 0 (top-left), Band 1 (top-right), Band 2 (center), Band 3 (bottom-left), and Band 4 (bottom-right) at normalized coordinates detailed in Figure 3.1.

Visual parameters: HSL palette with triadic harmony (120° separation, adjustable via hue slider), Perlin noise modulation for organic texture, 60fps rendering synchronized to pre-computed audio features, adaptive canvas resolution (minimum 800×600px), and Gaussian weight mixing creating continuous visual field with soft edges.

3.3.3 Mapping Architectures

Two mapping modes defined how audio energy transformed into visual parameters. These modes differed fundamentally in their temporal characteristics, forming the basis for the perceptual discrimination task.

M-A: Instantaneous Mapping:

- M-A implemented the conventional stateless paradigm where visual brightness at frame t derived directly from audio energy at frame t, with no temporal dependencies: $V(t) = f(A(t))$. A hybrid normalization strategy balanced visual consistency with responsiveness:

  - Pre-computation Phase: During audio loading, a custom normalization function (calculatePerBandNormalization, implemented in JavaScript) analyzed the entire audio energy time series to establish per-band dynamic ranges. The audio energy values were pre-extracted from the musical stimulus using RMS (root-mean-square) amplitude calculation at 60 frames per second, stored in a CSV file, and loaded at runtime. For each of the five frequency bands, the function calculated:
    - minEnergy: 20th percentile of energy values (to exclude silence/noise floor)
    - maxEnergy: 80th percentile of energy values (to exclude transient peaks)
    - range = maxEnergy − minEnergy (used for subsequent normalization)

- Real-time Mapping (Hybrid Blend): During playback, visual brightness for each frequency band was computed through a weighted blend between "raw" and "normalized" mappings:
  - Raw scaling: rawScaled = baseLevel + (rawEnergy × peakScale), where baseLevel (0.08) provides minimum visibility and peakScale (0.72) controls responsiveness to energy changes
  - Normalized scaling: normalizedScaled = (energy − minEnergy) / range, mapping the current energy to a 0–1 range based on pre-computed percentiles
  - Hybrid blend: finalBrightness = rawScaled × (1 − normStrength) + normalizedScaled × normStrength
  - Default blend: 90% raw, 10% normalized (normStrength = 0.1), prioritizing immediate responsiveness while reducing cross-track loudness variation

The final brightness value controlled the intensity (alpha channel) of Gaussian-weighted color regions rendered on the canvas.

- Smooth Clipping: Final scaled values passed through smoothClip() function applying tanh-based compression above threshold (0.85), preventing harsh visual clipping while preserving transient peaks.

This architecture ensured that M-A produced immediate, frame-accurate responses to audio changes while maintaining stable visual dynamics across varying energy levels.

M-B: State-Driven Mapping:

- M-B implemented a temporal integration architecture where visual output emerged from an internal energy state that accumulated over time with exponential decay:

$$E(t) = \alpha \cdot E(t-1) + \beta \cdot I(t)$$

  - $E(t)$ = energy pool state at time t—the internal brightness accumulator for each frequency band, initialized at zero and updated every frame. This value represents the "visual memory" of recent audio activity: it rises when audio energy is present and decays gradually when audio energy drops.
  - $I(t)$ = audio input at time t—the pre-computed RMS amplitude for each frequency band at frame t, extracted offline from the audio stimulus and stored in a CSV file (see Section 3.3.1). Values range from 0 (silence) to 1 (maximum energy within each band).
  - α = decay coefficient (band-specific, range: 0.92-0.97). The baseline α was set to 0.92 based on pilot testing (see Section 3.4.4). Dynamic adjustment occurred through a two-layer system: (1) a macro-level baseline derived from pre-computed peak density analysis of each frequency band, and (2) micro-level real-time modulation based on frame-to-frame energy differences. Higher peak density (more transient events) yielded faster decay (lower α); sustained energy yielded slower decay (higher α).

- β = input gain, computed as $(1 - \alpha)$ to ensure that the energy pool converges to the input level at steady state. For the default $\alpha = 0.92$, $\beta = 0.08$, meaning each frame adds 8% of the current audio input to the accumulated state.

- Dynamic adaptation: Decay rates varied based on two levels:
  - Macro-level: Pre-computed peak density (number of energy peaks per second) for each frequency band influenced baseline decay rate. Peaks were detected as local maxima exceeding a threshold of mean + 0.8 standard deviations, with a minimum inter-peak interval of 150ms to filter noise. Bands with higher peak density (e.g., percussive low frequencies) received faster baseline decay; bands with lower peak density (e.g., sustained pads) received slower decay.
  - Micro-level: Real-time decay adjustment based on local energy dynamics. At each frame, the system computed the energy difference between consecutive frames ($|E(t) - E(t-1)|$). When energy changed rapidly (high difference), the decay was slightly accelerated to improve transient response; when energy was stable (low difference), decay was slowed to maintain smooth sustained tones. This adjustment ranged ± 0.08 around the macro-level baseline, ensuring responsiveness without destabilizing the overall temporal smoothing.

This architecture created fundamental temporal dependencies where $V(t) = g(A(t), V(t-1))$, producing visual inertia, accumulation effects, and hysteresis (the tendency for visual output to depend on its recent history—brightness that has been high tends to remain elevated even as audio energy begins to drop, and vice versa) that were absent in M-A. The resulting perceptual difference, whether and how quickly participants detected transitions between these modes.

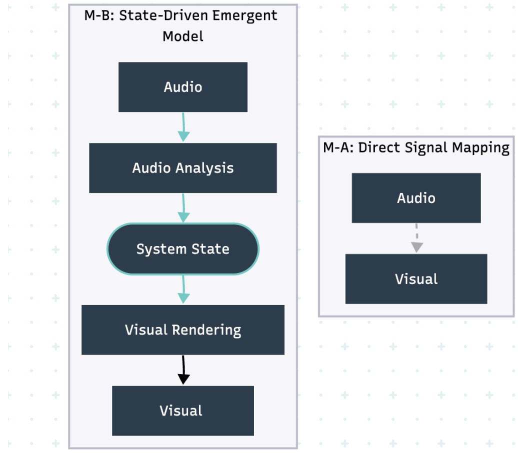

Figure 3.2 M-B vs M-A architecture diagram

M-A and M-B produce qualitatively different visual dynamics that motivate our detection task, expected perceptual differences are:

- M-A characteristics (instantaneous):
  - Frame-accurate tracking: Brightness responds instantly to audio transients
  - High-frequency responsiveness: Rapid percussive events produce sharp visual peaks
  - No visual inertia: Energy drops immediately produce dimming
  - Statistical independence: Each frame depends only on current audio, P(V_t | A_t)

- M-B characteristics (state-driven, α=0.92):
  - Temporal smoothing: Brightness changes unfold over ~200ms
  - Visual inertia: Energy drops are followed by sustained glow
  - Accumulation effects: Sustained energy builds to higher visual intensity than M-A
  - Hysteresis: Visual peaks appear 50-150ms after audio peaks
  - Statistical dependence: Each frame depends on history, P(V_t | A_t, V_{t-1})

- Participants may use various perceptual cues:

- Responsiveness: Does the visual "snap" to audio changes or "float"?
- Visual inertia: Do bright regions persist after audio energy drops?
- Lag: Is there delay between percussive hits and visual peaks?
- Smoothness: Are brightness changes gradual or stepwise?

We do not instruct participants which cues to use, allowing natural perceptual strategies to emerge. This ecological approach tests discriminability under realistic viewing conditions rather than training observers on specific features.

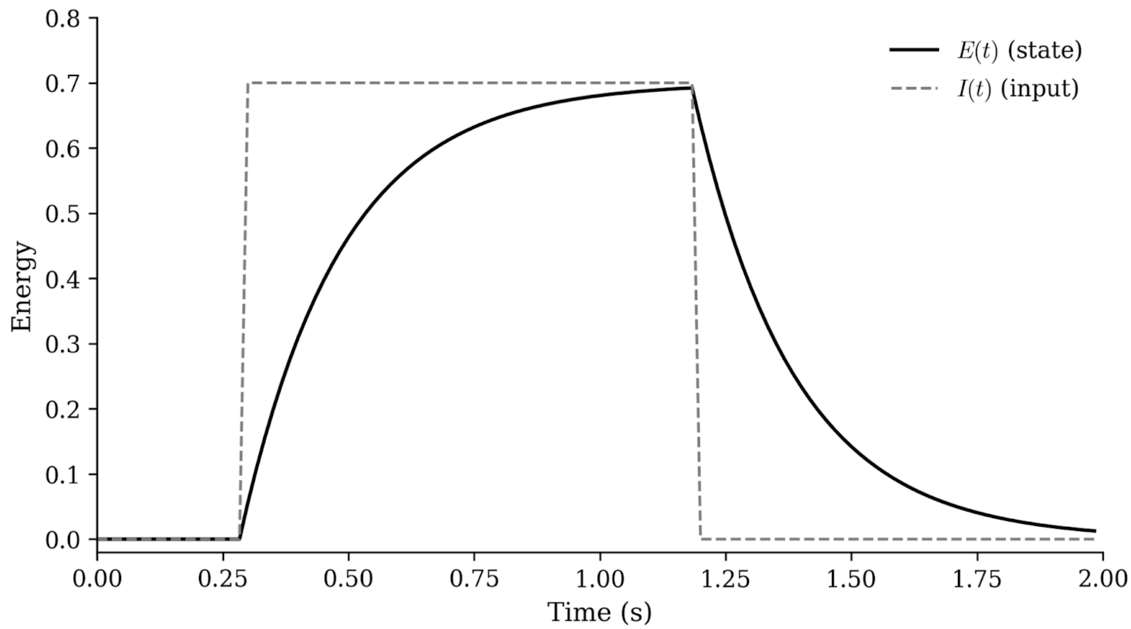

Figure 3.3 Simulation of the M-B energy state

Figure 3.3. Simulation of M-B energy state accumulation in response to a step input. The internal state E(t) (solid line) was computed using $E(t) = \alpha \cdot E(t-1) + \beta \cdot I(t)$ with $\alpha = 0.92$ and $\beta = 0.08$ at 60 fps. Input I(t) (dashed line) is a step function with amplitude 0.7 active from t = 0.3s to t = 1.2s. The state exhibits exponential rise toward steady state during input and exponential decay ($\tau \approx 200$ ms) after input cessation. Simulation implemented in Python using NumPy.

3.4 Procedure

3.4.1 Pre-Experiment Phase

Participants completed a brief Qualtrics survey collecting demographic information (age, gender, years of musical training), email address for session identification and consent to participate.Upon survey completion, participants received a unique experiment URL containing their userID and email parameters, automatically initializing their session upon page load.

3.4.2 Task Instructions

Participants received the following information via the welcome screen:

1. Two different audiovisual mapping modes would be presented during playback
2. The system would switch between modes at various time points
3. The primary task was to press spacebar whenever detecting a mode transition
4. Focus should be on how the visualization responds to music
5. Some transitions might be subtle requiring trust in perceptual judgment

The Mode Comparison Demonstration showed that two modes exist but did not reveal when switches occur during the main task (timing was unpredictable), computational mechanisms (no mention of α, decay, or state), or specific features to watch for. Technical details about the M-A/M-B architectures were withheld to avoid biasing detection strategies toward specific features, ensuring that participants relied on natural perceptual discriminability rather than learned heuristics.

3.4.3 Session Timeline

Each complete session followed this sequence:

1. Welcome Screen
    a. Task overview: Press spacebar when visual pattern changes
    b. Duration estimate: ~3 minutes
    c. Equipment recommendation: Headphones recommended
    d. Proceed with Enter key or Start button

2. Mode Comparison Demonstration
    a. Interactive side-by-side visualization showing M-A and M-B
    b. Participants could trigger test audio to observe differences
    c. Familiarize participants distinguishing the two modes
    d. Self-paced progression when ready

3. Tutorial Phase
    a. Practice session with predetermined mode switches
    b. Real-time hit counter displayed (Hits: X/3)

Quality criteria: Minimum 3 successful hits (within [0, 2000ms] of actual switch) and minimum 30% hit rate. Participants failing these criteria received feedback ("Needed 3 hits. Please carefully observe the difference between the two modes") with the option to retry. Only after passing this gate could participants advance to the main experiment.

4. Countdown
    a. Visual countdown followed immediately by audio playback
    b. Participants could press spacebar for visual feedback before and during countdown (responses not logged)

5. Main Experiment
    a. Continuous audiovisual presentation with mode switches at predetermined intervals
    b. Participants indicated detected transitions via spacebar
    c. No explicit feedback on response accuracy

6. Feedback Collection

Post-task questionnaire: difficulty rating (1 = easy, 5 = difficult) and estimated number of switches

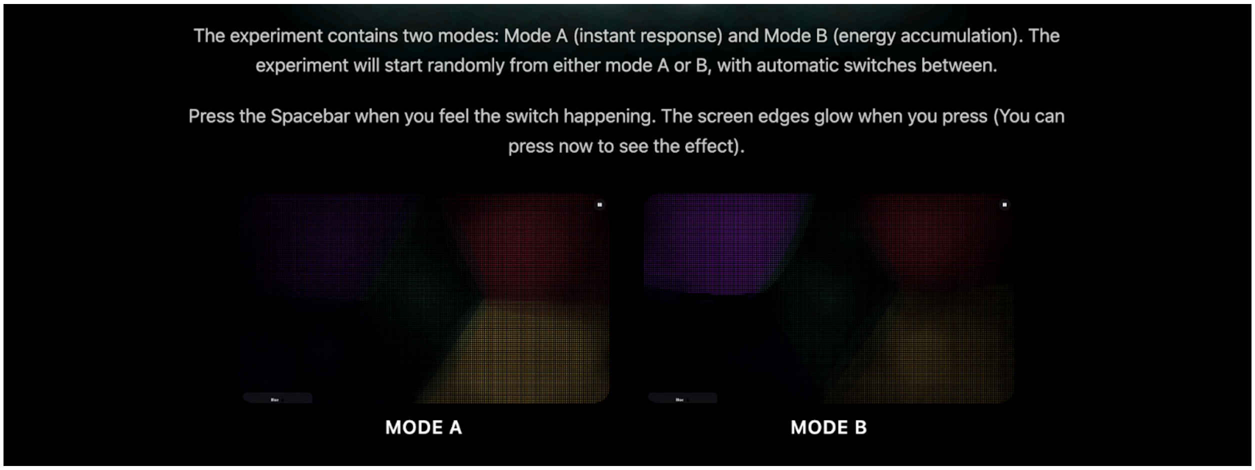

Figure 3.4 Mode Mapping Comparison preview

Each experimental session lasted approximately 5 minutes and consisted of continuous monitoring with intermittent mode transitions. Participants developed their own detection strategies, maintaining task ecological validity while avoiding inefficiency of completely blind search. This design balances internal validity (controlled task structure) with external validity (natural perceptual discovery).

3.4.4 Decay Parameter Selection

The decay coefficient α determines the time constant $\tau = -\Delta t/\ln(\alpha)$ for exponential state accumulation in M-B. Parameter selection balanced two perceptual constraints: (1) temporal dynamics below ~50ms approach perceptual transparency (nearing JND—just noticeable difference—thresholds for temporal order, the minimum detectable change; Vroomen & Keetels, 2010), while (2) dynamics exceeding ~200ms risk perceptual desynchronization (exceeding TBW boundaries; Vroomen & Keetels, 2010).

The primary decay coefficient $\alpha = 0.92$ ($\tau \approx 200$ms at 60fps) was selected to position M-B's temporal dynamics at the upper boundary of this “perceptible-but-integratable” zone, where temporal smoothing becomes maximally discriminable while remaining within the temporal binding window (TBW). For complex audiovisual stimuli like music visualization, TBW extends beyond 200ms (Stevenson et al., 2012; Vroomen & Keetels, 2010), allowing integration at $\tau \approx 200$ms while maintaining discriminability. This positioning enables Entry transitions to unfold as gradual accumulation near the TBW boundary (requiring conservative detection thresholds), while Exit transitions manifest as abrupt cessation exceeding this timescale (enabling liberal thresholds). Informal testing during early development confirmed this value as maximally discriminable while maintaining audiovisual coherence: faster decay rates ($\alpha > 0.95$) approached perceptual indistinguishability from M-A, while slower rates ($\alpha < 0.88$) elicited reports of "noticeable lag." The adaptive difficulty system varied α between 0.88 and 0.995 ($\tau \approx$ 130–200ms) based on individual performance, allowing systematic exploration of discriminability across the theoretical gradient.

3.4.5 Adaptive Difficulty and Spatial Remapping

Music's inherent temporal periodicity (recurring rhythmic patterns, cyclic harmonic progressions) creates a methodological threat: participants might develop detection strategies based on spatial location memory rather than temporal dynamics perception. The same early testing revealed that fixed spatial mappings enabled memory-based strategies, when asked to describe their detection strategies and suggestions for improvement, volunteers reported comparing current spatial patterns to remembered templates (e.g., "bottom-left looks different than last time") rather than perceiving temporal dynamics directly. This qualitative observation motivated the spatial remapping design described below. Specific example: A recurring drum pattern consistently activated the bottom-left region. After a few repetitions, the detection strategy degraded to spatial-anchored memory matching: "Does that position look different now?" This represents checking whether spatial-temporal expectations are violated rather than perceiving temporal dynamics (smoothing versus immediacy). This strategy would produce confounded data: participants might "detect" mode switches while actually responding to spatial-temporal expectation violations, not temporal dynamic changes.

To address this challenge, a two-layer adaptive system decoupled temporal detection from spatial learning:

Layer 1: Temporal Mode Switching (Core Manipulation)

- *Baseline period* (0–60 seconds): Switch intervals of 4–8 seconds (±10% jitter) with fixed spatial mapping to establish baseline performance.
- *Adaptive period* (after 60 seconds): Performance monitoring (rolling hit rate over last 4 keypresses) triggered difficulty adjustments:
  - Hard mode: Hit rate ≥75% OR 2 consecutive precise hits (RT ≤1.5s) → switch intervals 2.6–6 seconds (±10% jitter)
  - Normal mode: Hit rate ≤50% → switch intervals 3–9 seconds (±10% jitter)
  - Constraint: 5-second minimum between difficulty changes

Layer 2: Spatial Remapping (Prevent Spatial Learning)

When participants approached detection threshold (high hit rate), spatial remapping reset spatial anchoring to prevent memory-based strategies. Remapping was activated only during hard mode with adaptive cooldown periods: 10–13 seconds (high performance ≥75% hit rate), 13–17 seconds (medium performance), and 17–20 seconds (low performance ≤25% hit rate). Transformation types were randomly selected: rotation (70% probability: 72° or 144°) or mirroring (30% probability: horizontal or vertical).

Constraints: (1) Minimum 5-second interval between difficulty changes prevented confusion; (2) spatial transformations never synchronized with mode switches (cooldown periods ensured temporal separation); (3) core temporal dynamics ($\alpha = 0.92$) remained constant across all difficulties and mappings.

Theoretical Basis: Feature Integration Theory

This design implements principles from Treisman and Gelade's (1980) Feature Integration Theory: by periodically altering spatial configuration, we prevent participants from binding features (temporal dynamics) to fixed locations. Each remapping forces attention to temporal properties rather than spatial-temporal associative memory, ensuring participants detect temporal dynamics (smoothness, lag, inertia) rather than spatial pattern deviations.

Data Logging and Confound Mitigation

Each transition recorded: `switch_time` (precise timestamp), `from_mode/to_mode` (M-A or M-B), `mapping_id` (spatial configuration: r0, r1, r2, mh, mv), and `difficulty` (current difficulty level).

While spatial remapping prevents unwanted learning effects, it introduces additional perceptual change during detection. To address potential confounding, we conducted stratified analysis by difficulty level (normal vs. hard) and tested whether mapping changes significantly modulated detection performance. If hard mode trials showed inflated performance, we would report normal-difficulty-only results as conservative estimates of pure temporal dynamic thresholds. Results of this stratified analysis appear in Section 4.5.3.

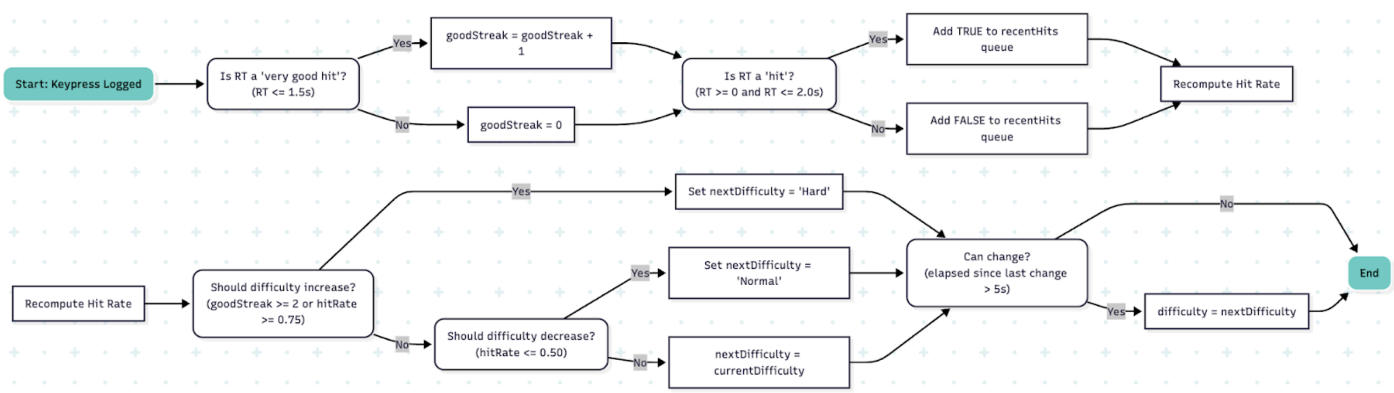

Figure 3.5 Automatic difficulty adaptor

Spatial remapping examples. (A) Base configuration with five frequency bands at fixed positions. (B) r1 rotation maintains relative positions while disrupting spatial anchoring. (C) mh mirroring creates distinctive reconfiguration. These transformations prevent participants from memorizing position-specific patterns while preserving temporal dynamics (α = 0.92) across all configurations.

3.5 Database Architecture and Quality Control

Four PostgreSQL tables with referential integrity captured experimental data: `thesis_sessions` (session-level summary including participant_id, playback duration, keypress counts, hit counts, and validity flags), `thesis_logs` (individual keypress events with timestamps and reaction times), `thesis_switches` (mode transition events with timestamps and spatial configurations), and `thesis_feedback` (post-session questionnaire responses). Complete database schema appears in Appendix C.

Data quality measures were implemented as part of the custom experiment software: automated validation of database integrity constraints (enforced by PostgreSQL foreign key relationships), real-time monitoring of network connectivity and logging latency (via JavaScript navigator.onLine API and timestamp comparisons), post-session verification of data completeness (automated checks for missing fields before session finalization), and duplicate entry detection and resolution (using unique session tokens to prevent resubmission).

3.6 Data Analysis

3.6.1 Switch Classification Using current_mode

Each mode switch was classified as Entry (A→B) or Exit (B→A) using the `current_mode` field from the experimental database. This field records the system's architectural state after each switch event, providing direct, error-free classification of switch direction. Because `current_mode` directly reflects the system's internal state, it eliminates classification ambiguity present in indirect approaches. Phase-locking algorithms, for example, infer switch direction by temporally aligning participants' keypresses with system events—methods vulnerable to misclassification when responses cluster near switch boundaries.

To ensure data quality and theoretical validity, a stringent cleaning procedure verified perfect alternation between Entry and Exit transitions as required by the experimental design:

1. Initial Classification: All mode switches were classified using the current_mode field (current_mode = “B” → Entry (A→B); current_mode = “A” → Exit (B→A)).
2. Consecutive Same-Mode Removal: We removed any switch where current_mode matched the previous switch's mode within the same session. Such cases occurred due to network retry mechanisms (creating duplicate records), race conditions (logging out of order), or database write conflicts (creating spurious records).
3. Alternation Verification: We verified that the cleaned dataset exhibited perfect alternation, with each Entry transition followed by an Exit transition and vice versa. This verification was performed both within sessions and across session boundaries.

This cleaning procedure prioritizes data quality over quantity. While it reduces total sample size, it ensures 100% classification accuracy, perfect consistency with the theoretical model (alternating transitions), elimination of logging artifacts that could introduce systematic error, and unambiguous interpretation of switch direction. Each retained switch represents a confirmed architectural transition with verified alternation. Retention rate and final sample sizes are reported in Section 4.1.

3.6.2 Detection Measures

Detection was operationalized as participant keypresses occurring within a specified time window relative to system switch events. The following measures were defined:

- Hit: A keypress of the correct type (Entry or Exit) occurring within 0-2000 ms after the corresponding system switch.

- Miss: A system switch with no corresponding correct keypress within the 2000-ms response window, or a correct keypress occurring after 2000 ms.
- False Alarm: A keypress occurring outside any defined detection window (not matched to any system switch within 2000 ms).
- Reaction Time (RT): For hits, the latency between the system switch timestamp and the corresponding keypress timestamp, measured in milliseconds. All timestamps were synchronized to the HTML5 Audio element's currentTime property, ensuring timing was relative to audio playback rather than the system clock. While absolute timing precision varies across devices and browsers (typically ±20–50ms; Anwyl-Irvine et al., 2021), our within-subjects design ensured that any device-specific latency affected Entry and Exit conditions equally. Consequently, such latencies acted as constant systematic errors that canceled out in the paired analysis, preserving the validity of relative comparisons without requiring hardware-level calibration.

The 2000-ms response window was chosen based on pilot testing and reflects a balance between allowing sufficient time for gradual perceptual changes to be detected while excluding very delayed responses that likely reflect lapses in attention or uncertainty rather than genuine detection. Detection metrics calculated at two levels:

- Switch-level: Individual switches as the unit of analysis, yielding detection rates across all switches.
- Session-level: Sessions as the unit of analysis, with detection metrics averaged within each session before aggregation across participants.

3.6.3 Statistical Analysis

Statistical analyses were designed to test two primary hypotheses:

- **H1** (Detection Latency Asymmetry): Entry transitions (A→B, mode onset) would elicit longer reaction times than Exit transitions (B→A, mode cessation) due to difficulty of detecting gradual state accumulation versus abrupt state cessation.

    - Paired-samples t-tests comparing Entry vs. Exit reaction times at the session level (pairing by session)
    - Effect size calculation using Cohen's d for paired samples
    - Examination of RT distributions and central tendency measures (mean, median)
- **H2 (**Response Bias Asymmetry): Entry transitions would elicit more conservative response criteria than Exit transitions, as participants require more evidence before reporting ambiguous gradual changes compared to clear abrupt changes.
    - Signal Detection Theory analysis (described in Section 3.6.4)
    - Paired-samples t-tests comparing sensitivity (d') and criterion (c) measures
    - Examination of hit rates and false alarm rates by transition type

All hypothesis tests were two-tailed with $\alpha = .05$. Effect sizes were computed as Cohen's d for paired samples, calculated as:

$$d = M\_diff / SD\_diff$$

where M_diff is the mean difference score and SD_diff is the standard deviation of differences (following Morris & DeShon, 2002). This accounts for the correlation between repeated measures and is appropriate for within-subjects designs.

For the three primary comparisons (hit rate, d', and criterion c), we controlled for Type I error inflation using Bonferroni correction (adjusted $\alpha = .05/3 = .0167$). Exploratory analyses such as reaction time comparisons were not included in multiple comparison adjustments and are interpreted cautiously.

Confidence intervals (95%) for mean differences were computed using the t-distribution (df = 54). Confidence intervals for Cohen's d were estimated via bias-corrected and accelerated bootstrap (BCa method; Efron & Tibshirani, 1993) with 10,000 iterations, implemented in R using the boot package (Canty & Ripley, 2021). This approach provides accurate interval estimation for effect sizes with modest sample sizes (Algina et al., 2005).

Reaction time analyses employed paired-samples t-tests comparing mean RTs for Entry versus Exit transitions within each session. Only sessions containing at least one detected trial of each type were included in RT comparisons ($n = 47$ sessions). RT distributions were also compared using the Kolmogorov-Smirnov test to assess distributional differences beyond central tendency.

3.6.4 Signal Detection Theory Analysis

Signal Detection Theory (SDT) was applied to separate perceptual sensitivity from response bias. SDT assumes that detection decisions are based on comparing noisy sensory evidence to a decision criterion, with sensitivity (d') reflecting the ability to discriminate signal from noise and criterion (c) reflecting the threshold for reporting detection. For each session and transition type (Entry vs. Exit), we calculated:

- Hit Rate (HR): Proportion of switches correctly detected

  $$HR = \text{Hits} / (\text{Hits} + \text{Misses})$$

- False Alarm Rate (FAR): Proportion of false alarms relative to noise opportunities. A false alarm was defined as any keypress occurring outside all 2-second detection windows. Noise opportunities were calculated as:

  $$\text{Noise Opportunities} = (\text{Session Duration} - \text{Number of Switches} \times 2s) / 2s$$

This represents the number of 2-second intervals during which no switch-related response was expected. FAR was then computed as:

$$FAR = \text{False Alarms} / \text{Noise Opportunities}$$

False alarms were attributed to Entry or Exit based on the direction of the nearest preceding switch.

- Sensitivity (d'): The standardized distance between signal and noise distributions, where Z is the inverse of the standard normal cumulative distribution function.

$$d' = Z(HR) - Z(FAR)$$

- Criterion (c): The decision threshold relative to the neutral point

$$c = -0.5 \times [Z(HR) + Z(FAR)]$$

Positive c values indicate conservative bias (requiring more evidence before reporting detection), while negative values indicate liberal bias (reporting with less evidence).

- When HR or FAR equaled 0 or 1 (creating infinite Z-scores), we applied the log-linear correction(Hautus, 1995):

Adjusted rate = (count + 0.5) / (total + 1)

This adjustment avoids infinite values while minimally distorting the data.

Higher d' indicates better perceptual discrimination (signal more distinguishable from noise). Higher c indicates more conservative responding (higher threshold). The combination of d' and c determines overall hit rate:

$$HR = \Phi(d'/2 - c)$$

where Φ is the standard normal CDF. This relationship was used to validate the SDT model by comparing predicted versus actual hit rates.

4. ANALYSIS & RESULTS

4.1 Data overview

A total of 165 participants completed the Qualtrics pre-screening survey and received experiment links. Of these, 55 completed valid experimental sessions meeting all inclusion criteria (playback duration > 220s, ≥ 5 keypresses, ≥ 2 valid hits). The remaining participants either did not initiate the experiment, failed the practice gate, or produced incomplete sessions; due to the automated pipeline discarding invalid sessions in real-time, exact attrition breakdown is unavailable. Demographic data were collected via optional survey items. Based on valid responses: mean age = 24.5 years (SD = 6.4); gender distribution was 67% male, 33% female; mean musical training = 5.9 years (SD = 7.4).

Initial classification using the `current_mode` field identified 1,269 mode switches across all sessions. After removing consecutive same-mode switches (resulting from network retry mechanisms, race conditions, or logging artifacts), the cleaned dataset retained 1,269 switches exhibiting perfect alternation:

- Entry switches (A→B): 633 (49.9%)
- Exit switches (B→A): 636 (50.1%)
- Entry/Exit ratio: 0.995

Alternation was verified within all 55 sessions with no violations of the expected A→B→A→B pattern, confirming theoretical validity of the cleaned dataset.

4.2 Descriptive Statistics

4.2.1 Overall Detection Performance

Switch-level detection rates: Exit transitions achieved significantly higher detection rates than Entry transitions. Of 636 Exit switches, 427 (67.1%) were detected within the 2-second window, compared to 374 of 633 Entry switches (59.1%), $\chi^2(1) = 8.50$, $p = .004$**, representing an 8.0 percentage point advantage. Entry showed 40.9% miss rate versus Exit's 32.9%—50 additional detection failures. Missed trials exhibited extreme latencies (Entry: $M = 5{,}310$ ms, $SD = 8{,}862$ ms; Exit: $M = 4{,}867$ ms, $SD = 5{,}437$ ms), often exceeding 10 seconds, indicating genuine detection failures rather than brief processing delays. **Reaction times (among detected trials):** Among successfully detected transitions, reaction times showed no significant difference: Entry $M = 1{,}057$ ms ($SD = 483$) versus Exit $M = 1{,}052$ ms ($SD = 454$), $t(46) = -0.87$, $p = .390$, $d = -0.13$. The median RT difference of 57 ms (Entry: 1,074 ms, Exit: 1,017 ms) suggested potential distributional effects, though high standard deviations (>450 ms) indicated substantial individual variability.

4.2.2 Hit Rate Distributions

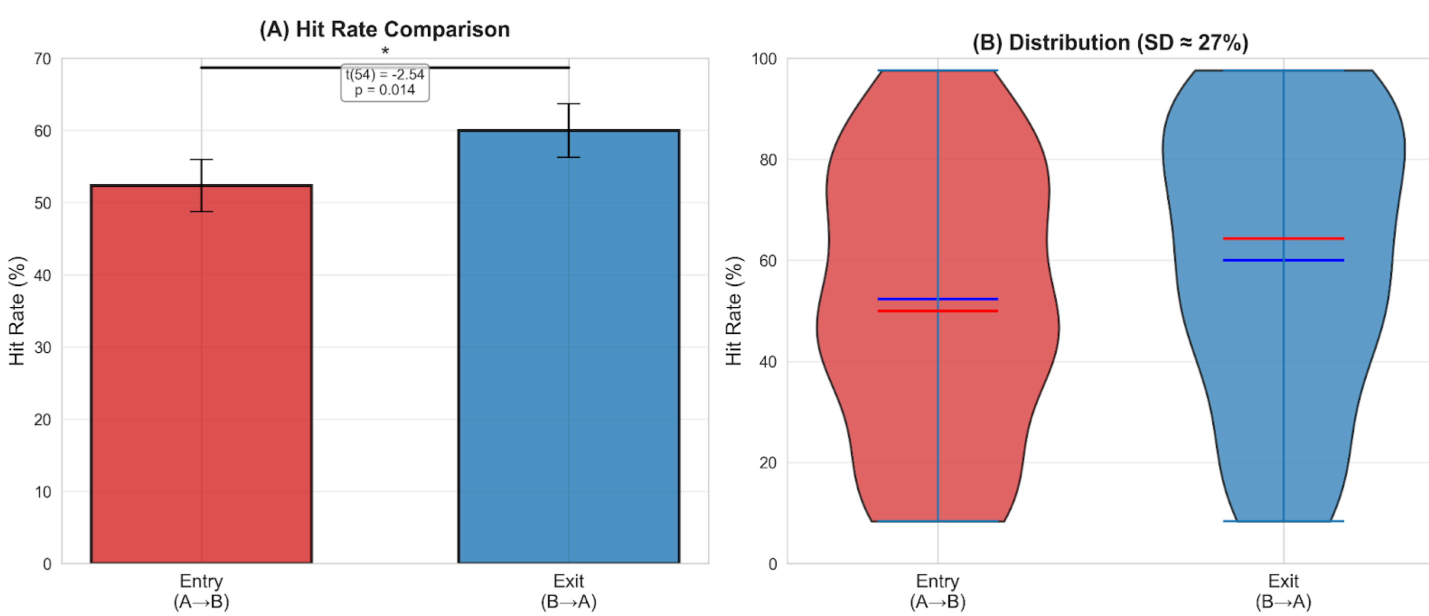

Figure 4.1: Hit Rate Comparison by Transition Type

(A) Mean hit rates with 95% confidence intervals. (B) Violin plots showing full distribution shapes. Entry hit rates ranged from 8.3% to 88.9% (M = 51.4%, SD = 29.5%), while Exit ranged from 20.0% to 88.9% (M = 59.7%, SD = 29.6%). Despite substantial individual variability, 56.4% of sessions (31/55) showed higher hit rates for Exit.

The substantial standard deviations indicate considerable between-session variability. Despite this variability, 56.4% of sessions (31/55) demonstrated higher hit rates for Exit than Entry, confirming consistency in the detection advantage across the majority of participants. Violin plots (Figure 4.1B) illustrate full distribution shapes. Entry's distribution appears relatively broad with no strong peaks, suggesting high individual variability. Exit's distribution shows a slight right skew with concentration in the 50–70% range. Substantial overlap indicates many sessions showed comparable performance across transition types, though Exit demonstrated higher central tendency.

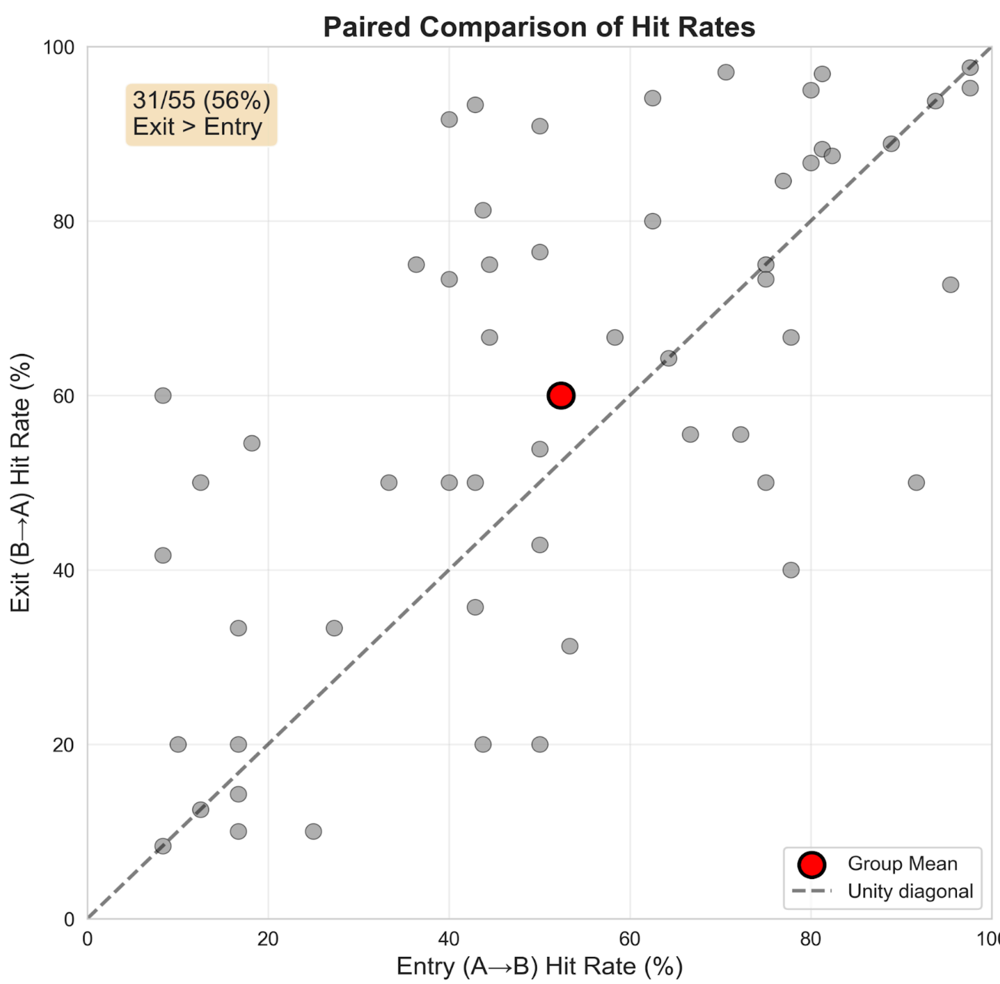

Figure 4.2. Paired Comparison of Hit Rates Across Sessions

Figure 4.2 presents a paired comparison showing individual session performance. 31/55 sessions (56.4%) demonstrated higher hit rates for Exit than Entry, while 24/55 sessions (43.6%) showed the reverse pattern. The red marker indicates the group mean (Entry: 51.4%, Exit: 59.7%), with the majority of individual sessions falling above the unity diagonal, confirming the detection advantage at the individual level.

4.2.3 Reaction Time Distributions

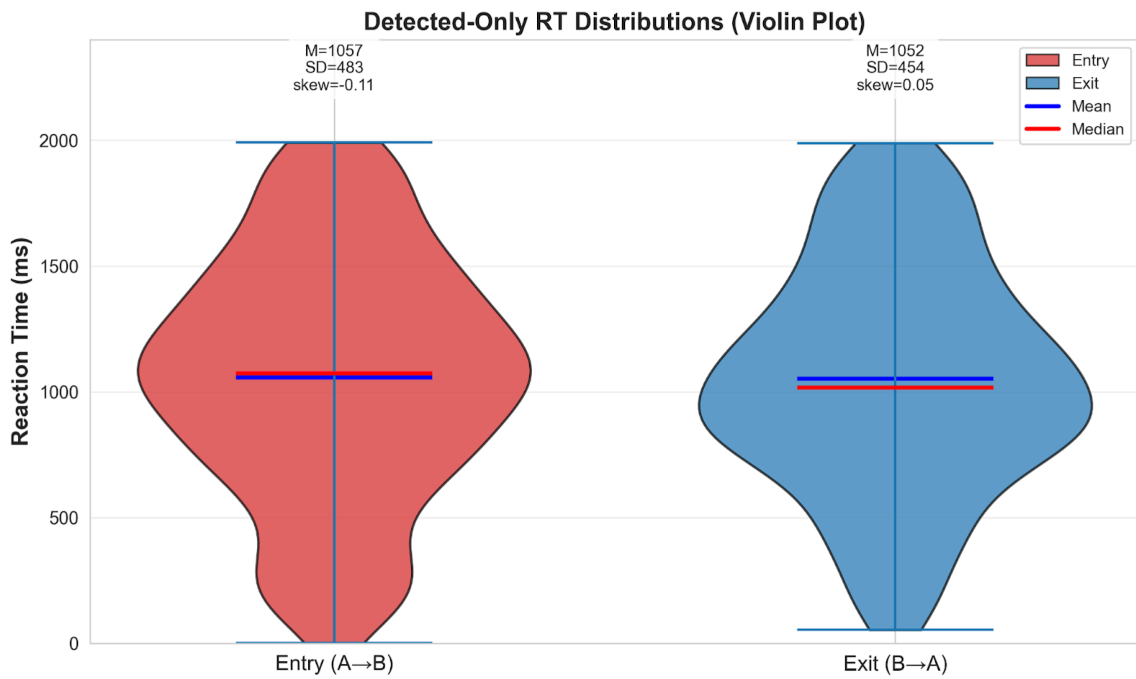

Figure 4.3. Reaction Time Distributions for Detected Transitions

The substantial distributional overlap and similar means indicate no significant difference in processing speed for detected transitions (Kolmogorov-Smirnov D=0.07, p=.25).

Entry RT distribution (n = 374): M = 1,057 ms (SD = 483), median = 1,074 ms, range = 60–1,999 ms, IQR = 722–1,475 ms, skewness = −0.10, kurtosis = −0.69. Exit RT distribution (n = 427): M = 1,052 ms (SD = 454), median = 1,017 ms, range = 36–1,998 ms, IQR = 713–1,384 ms, skewness = 0.04, kurtosis = −0.44.

Kolmogorov-Smirnov test revealed no significant difference between RT distributions, D = 0.07, p = .25, indicating similar distributional shapes despite the slight median difference.

RT differences (Entry − Exit) across 47 paired sessions showed M = −54 ms (SD = 425), median = −44 ms, range = −1,480 to 1,076 ms. Direction split nearly equally: Entry > Exit in 25/47 sessions (53.2%), Exit > Entry in 22/47 sessions (46.8%), indicating no consistent directional advantage at the individual session level. The large standard deviation (425 ms) reflects high inter-session variability, likely driven by individual differences in motor speed, attention, and strategy.

4.2.4 Cumulative Detection Curves

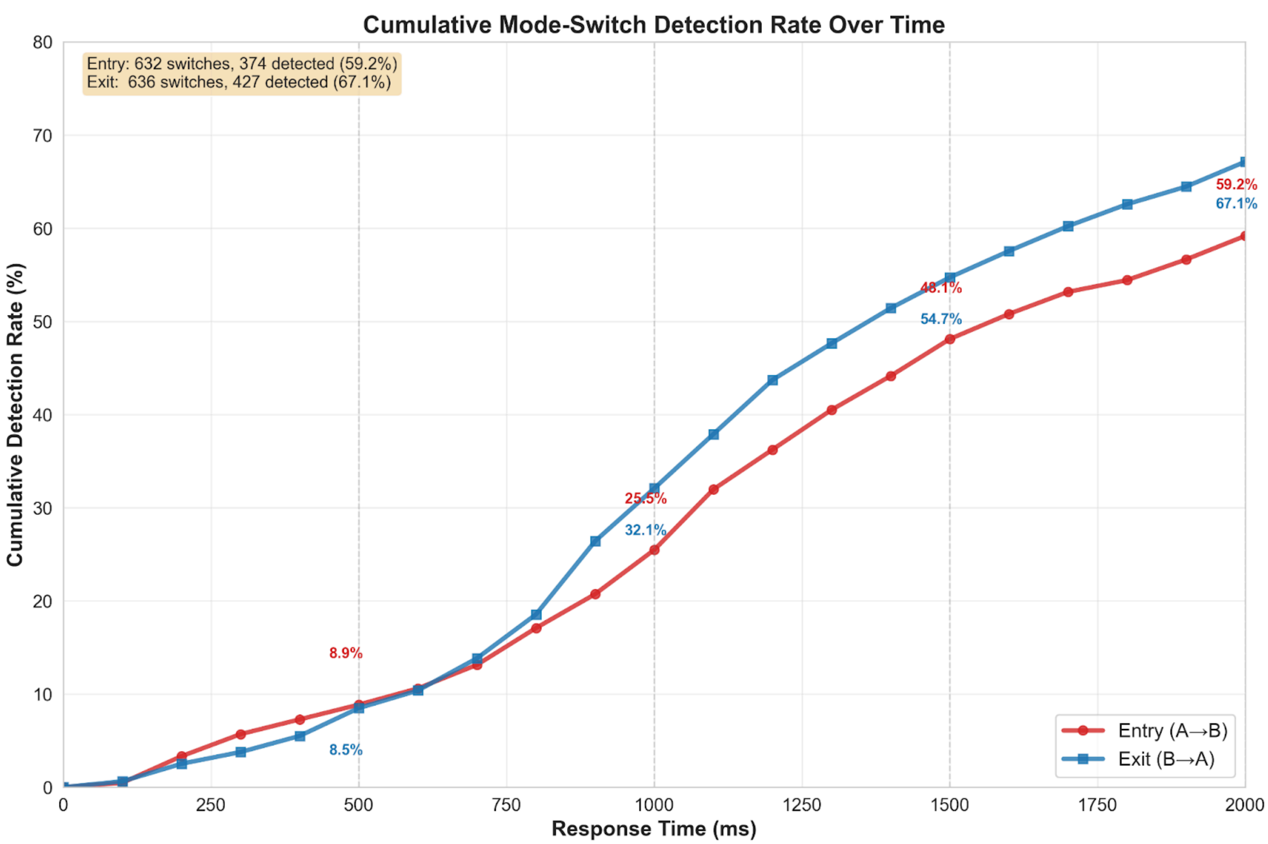

Figure 4.4 Cumulative Detection Rates Over Time

Cumulative proportion of transitions detected as a function of time post-switch. Exit's detection advantage (blue line) emerged early and persisted throughout the 2000-ms response window.

Temporal Dynamics

The cumulative curves revealed that Exit's detection advantage emerged early and persisted throughout the response window:

Time post-switch	Entry detected	Exit detected	Gap
0-500 ms	8.8%	12.4%	3.6%
500-1000 ms	25.4%	35.9%	10.5%
1000-1500 ms	48.0%	58.5%	10.5%
1500-2000 ms	59.1%	67.1%	8.0%

The gap between curves stabilized around 1000-1500 ms, suggesting that most detectable transitions were identified within this window, with later responses primarily adding to Exit's advantage.

Entry misses ($n = 259$, 40.9% of Entry switches) showed mean RT = 5,310 ms ($SD = 8{,}862$), median = 3,435 ms, range = 2,003–88,490 ms. Exit misses ($n = 209$, 32.9% of Exit switches) showed mean RT = 4,867 ms ($SD = 5{,}437$), median = 3,259 ms, range = 2,019–41,543 ms. The extreme latencies (often exceeding 10 seconds) indicate genuine detection failures or severe attention lapses rather than slightly delayed processing.

4.3 Hypothesis Testing

4.3.1 Statistical Assumptions and Multiple Comparison Corrections

Before presenting hypothesis-specific results, we document the statistical assumptions and correction procedures applied across all tests.

Normality of Difference Scores: While Shapiro-Wilk tests indicated that several individual variables deviated from normality (Entry hit rate: $W = 0.947$, $p = .017$; Entry d′: $W = 0.943$, $p = .025$; Entry c: $W = 0.953$, $p = .032$), the critical assumption for paired t-tests, normality of difference scores, was satisfied for all primary comparisons: hit rate difference ($W = 0.979$, $p = .434$), d′ difference ($W = 0.982$, $p = .590$), and criterion difference ($W = 0.984$, $p = .683$). These results confirm that distributional requirements for parametric paired t-tests are met (Zimmerman, 1997).

Non-parametric Verification: To ensure robustness given marginal normality violations in raw scores, we verified all findings using Wilcoxon signed-rank tests, which require no distributional assumptions. All three primary tests remained significant at $\alpha = .05$ (Hit Rate: $p = .017$; d′: $p = .019$; c: $p = .018$), confirming the significance and direction of all effects. While these non-parametric p-values marginally exceed the Bonferroni-corrected threshold (α corrected = .0167), they closely approach it (largest deviation: .002). This pattern is expected: non-parametric tests typically have slightly lower statistical power than parametric tests when distributional assumptions are approximately met (Zimmerman, 2012). The convergence of parametric and non-parametric results validates that our primary findings are not artifacts of distributional violations.

Outlier Assessment: Using the 1.5×IQR criterion (Tukey, 1977), we identified 6 outliers in Entry d' (10.9% of sample) and 2 outliers in Entry c (3.6% of sample). Visual inspection confirmed these represented valid extreme responses rather than measurement errors—they corresponded to participants who either showed exceptionally strong Entry-Exit asymmetries or performed near-chance levels across both conditions. Consistent with best practices for within-subjects designs (Bakeman, 2005), we retained these cases in primary analyses. Sensitivity analyses excluding outliers (reported in Section 4.5.1) confirmed all main effects remained significant.

Multiple Comparison Corrections: Hypothesis testing involved three primary comparisons (hit rate, d′, and criterion c), raising concerns about Type I error inflation. We applied Bonferroni correction, the most conservative approach: with $\alpha = .05$ and three comparisons, the corrected threshold was $\alpha_{corrected} = .0167$. All three primary tests exceeded this threshold (Hit Rate: $p = .014$; Sensitivity d′: $p = .011$; Criterion c: $p = .025$), confirming that observed asymmetries are unlikely to reflect Type I errors.

All primary findings from parametric tests survive multiple comparison corrections, whether using the most conservative (Bonferroni) or more liberal (FDR-controlled) approaches. This convergence of evidence indicates that observed Entry-Exit asymmetries are robust and unlikely to be false positives.

Statistical Power: Post-hoc power analyses using observed effect sizes, conducted in G*Power 3.1 (Faul et al., 2009), revealed moderate power for detecting the observed medium effects (Hit Rate: $1-\beta = .70$; d': .73; c: .72). Although these values fall below the conventional .80 threshold, all three primary effects reached statistical significance, and the consistency of effects across multiple measures provides converging evidence for the Entry-Exit asymmetry.

4.3.2 H1: Detection Latency Asymmetry

H1 predicted that Entry transitions would elicit longer reaction times than Exit transitions. Results revealed a more nuanced pattern: the asymmetry manifested in detection rates rather than conditional reaction times.

Detection Rate Asymmetry

Exit transitions achieved significantly higher detection rates than Entry transitions at both analytical levels. Switch-level: Entry 59.1% (374/633) versus Exit 67.1% (427/636), $\chi^2(1) = 8.50$, $p = .004$**, Cramér's $V = 0.082$. Session-level: Entry $M = 51.4\%$ ($SD = 29.5\%$) versus Exit $M = 59.7\%$ ($SD = 29.6\%$), $t(54) = -2.52$, $p = .015$*, $d = -0.339$.

Entry showed 40.9% miss rate compared to Exit's 32.9%—representing 50 additional detection failures. Missed trials exhibited extreme latencies (Entry: $M = 5{,}310$ ms, $SD = 8{,}862$ ms; Exit: $M = 4{,}867$ ms, $SD = 5{,}437$ ms), indicating genuine detection failures rather than brief processing delays.

Conditional Reaction Time Analysis

Among successfully detected transitions, reaction times showed no significant difference. Switch-level means: Entry M = 1,057 ms (SD = 483) versus Exit M = 1,052 ms (SD = 454), difference = 5 ms. Session-level paired t-test (47 sessions with both Entry and Exit detected trials): $t(46) = -0.87$, $p = .390$, $d = -0.13$. Median RT difference of 57 ms (Entry: 1,074 ms, Exit: 1,017 ms) suggested potential distributional effects, though high standard deviations (>450 ms) indicated substantial individual variability.

Interpretation: Asymmetry Operates at Threshold-Crossing, Not Post-Threshold Processing

The dissociation between detection rates and conditional reaction times reveals where the asymmetry operates. Conditional RT analysis necessarily excludes the 40.9% of Entry trials and 32.9% of Exit trials that failed detection—the most informative data about detection difficulty. By examining only successfully detected trials, RT analysis is blind to detection failures where evidence accumulation never reached threshold.

Entry's difficulty manifests primarily as threshold-crossing failures: 50 additional trials failed to generate sufficient perceptual evidence within the response window. When Entry transitions do generate threshold-crossing evidence, subsequent decision and motor processes proceed at speeds equivalent to Exit (1,057 ms vs. 1,052 ms). The asymmetry thus operates at an earlier stage—evidence accumulation—rather than post-threshold processing.

This pattern explains why Signal Detection Theory succeeded where reaction time analysis could not. SDT incorporates all trials (hits, misses, false alarms) into sensitivity (d') and criterion (c) measures, whereas RT analysis discards the very trials (misses) that reveal Entry's perceptual difficulty. The detection rate asymmetry and SDT findings (Section 4.3.3) provide converging evidence that Entry's gradual state accumulation creates weak perceptual signals frequently failing to cross detection thresholds, while Exit's abrupt state cessation generates strong signals reliably triggering detection.

H1 Conclusion: H1 received partial support. Entry transitions proved significantly harder to detect than Exit transitions, as evidenced by the robust detection rate asymmetry ($p = .004$ switch-level, $p = .014$ session-level). However, this perceptual difficulty manifested as detection failures (misses) rather than processing latency. The absence of RT asymmetry in successfully detected trials (Entry: 1,057 ms vs. Exit: 1,052 ms, $p = .390$) suggests that the Entry-Exit asymmetry is confined to the evidence accumulation stage. Once the decision threshold is crossed, the subsequent motor response reflects a ballistic execution process governed by fixed neuromuscular constraints. This observed dissociation—asymmetric detection rates paired with symmetric RTs—aligns with two-stage decision models (Ratcliff & McKoon, 2008): it confirms that Entry's difficulty lies in generating sufficient evidence to trigger the decision, not in the speed of the response execution once triggered.

4.3.3 H2: Response Bias Asymmetry (Signal Detection Theory Analysis)

H2 predicted that Entry transitions would elicit more conservative response criteria than Exit transitions. Signal Detection Theory analysis revealed both the predicted criterion asymmetry and an unexpected sensitivity asymmetry.

Table 4.1: Signal Detection Theory Parameters with Confidence Intervals

Measure	Entry M(SD)	Exit M(SD)	Mean Diff	Diff 95% CI	t(54)	P
Sensitivity (d')	1.09 (0.90)	1.38 (0.94)	-0.28	[-0.50, -0.07]	-2.62	.011*
Criterion (c)	0.46 (0.52)	0.35 (0.54)	0.11	[0.01, 0.21]	2.31	.025
Hit Rate	51.4% (29.5%)	59.7% (29.6%)	-8.3%	[-14.9, -1.7]	-2.52	.015*
FA Rate	17.9%	17.3%	0.6%	[-0.9, 2.1]	0.79	.433*

	(11.5%)	(12.0%)				

Note: All tests are paired-samples t-tests ($df = 54$, two-tailed). Mean Diff = Exit − Entry (for d' and Hit Rate) or Entry − Exit (for c). 95% CIs computed from t-distribution. * $p < .05$; all primary comparisons (d', c, Hit Rate) remain significant under Bonferroni correction ($\alpha_{corrected} = .0167$). False Alarm Rate comparison is exploratory and not included in hypothesis tests.

Sensitivity Results: Exit transitions demonstrated significantly higher perceptual sensitivity ($d' = 1.378$, $SD = 0.941$) than Entry transitions ($d' = 1.095$, $SD = 0.896$), $t(54) = -2.62$, $p = .011$, $d = 0.355$ (medium effect size). Entry d' ranged from −0.932 to 3.461 with 47/55 sessions (85.5%) showing $d' > 0$, while Exit d' ranged from −0.472 to 3.766 with 51/55 sessions (92.7%) showing $d' > 0$. The higher d' values indicate that both transitions involve moderate discriminability, with Exit showing consistently stronger perceptual sensitivity.

Criterion Results: Entry transitions elicited marginally more conservative response criteria ($c = 0.459$, $SD = 0.523$) than Exit transitions ($c = 0.345$, $SD = 0.542$), $t(54) = 2.31$, $p = .025$, $d = 0.311$ (small-to-medium effect size). Both directions showed conservative criteria ($c > 0$), but Entry was significantly more conservative. Entry c ranged from −0.685 to 1.842 with 43/55 sessions (78.2%) showing $c > 0$ (conservative), while Exit c ranged from −0.742 to 1.728 with 41/55 sessions (74.5%) showing $c > 0$ (conservative). Directionally, 31/55 sessions (56.4%) showed Entry more conservative than Exit. Although the statistical significance is marginal ($p = .025$), the consistent directional effect (56.4% of sessions) and medium effect size ($d = 0.31$) suggest a meaningful decisional asymmetry.

False Alarm Rate Results: False alarm rates were comparable between directions: Entry ($M = 0.179$, $SD = 0.115$) versus Exit ($M = 0.173$, $SD = 0.120$), $t(54) = 0.79$, $p = .433$, Cohen's $d = 0.106$. The similar FA rates suggest that baseline vigilance levels were equivalent across both transition types, and the criterion differences reflect strategic adjustments in response threshold rather than differences in attentional lapses."

SDT Model Validation: To validate the SDT model, we compared predicted hit rates (HR = Φ(d′/2 − c)) with actual observed hit rates. Entry: d′/2 − c = 1.095/2 − 0.459 = 0.088, predicted HR = Φ(0.088) = 0.535, actual HR = 0.514, error = 2.1 percentage points. Exit: d′/2 − c = 1.378/2 − 0.345 = 0.344, predicted HR = Φ(0.344) = 0.635, actual HR = 0.597, error = 3.7 percentage points. The close correspondence validates the SDT decomposition, confirming that d′ and c successfully account for detection performance.

H2 Conclusion: H2 received substantial support with an important extension. Criterion asymmetry marginally confirmed: Entry showed more conservative criteria (c = 0.459) than Exit (c = 0.345), p = .025, with a medium effect size (d = 0.31) and consistent directional pattern (56.4% of sessions). Sensitivity asymmetry confirmed: Exit showed higher discriminability (d′ = 1.378) than Entry (d′ = 1.095), p = .011. The dual asymmetry---both perceptual (d′) and decisional (c)---provides converging evidence for Entry's detection difficulty at multiple processing levels. While the criterion effect reaches marginal statistical significance, the combination of medium effect size, consistent direction, and convergence with the perceptual asymmetry supports the theoretical prediction of decisional bias.

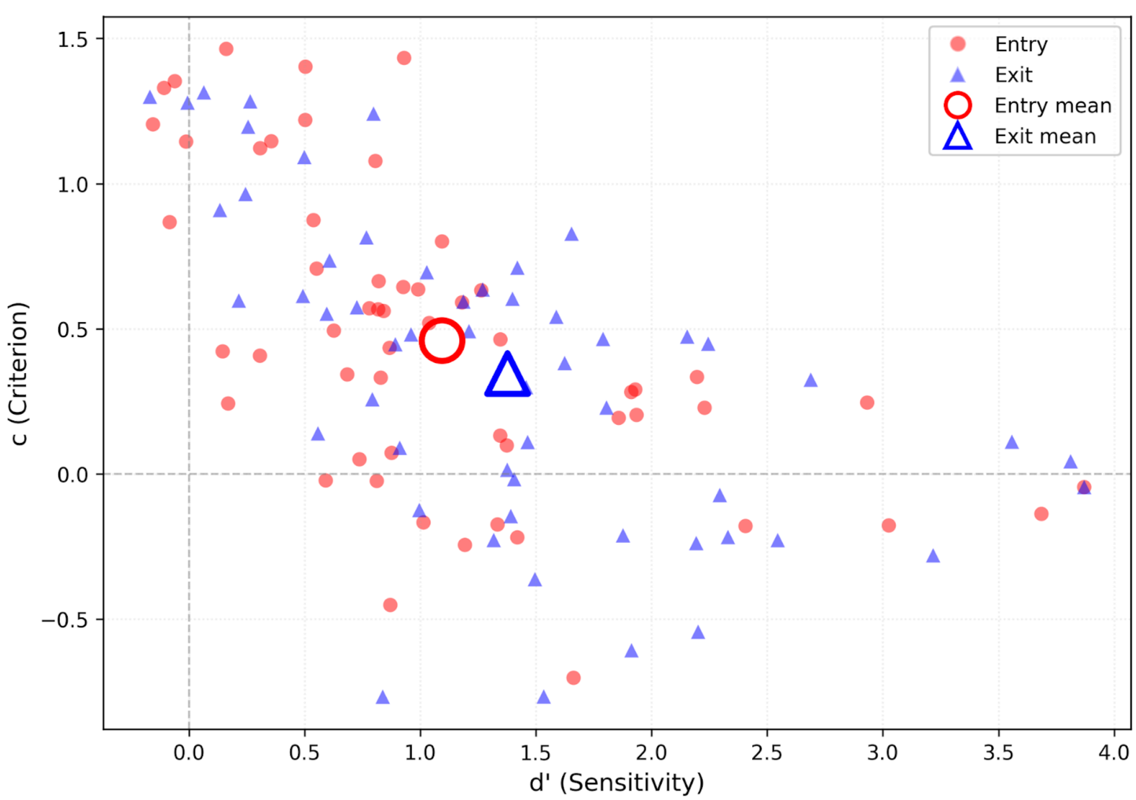

Figure 4.5. Signal Detection Theory Parameter Space

Figure 4.5 visualizes the SDT parameter space. Exit sessions (blue squares) show higher d′ values (M = 1.378) and moderate conservative criteria (c = 0.345), while Entry sessions (red circles) demonstrate lower d′ (M = 1.095) and more conservative criteria (c = 0.459). Both conditions cluster in the positive d′ region, indicating successful discrimination, with Entry showing greater conservatism in response strategy.

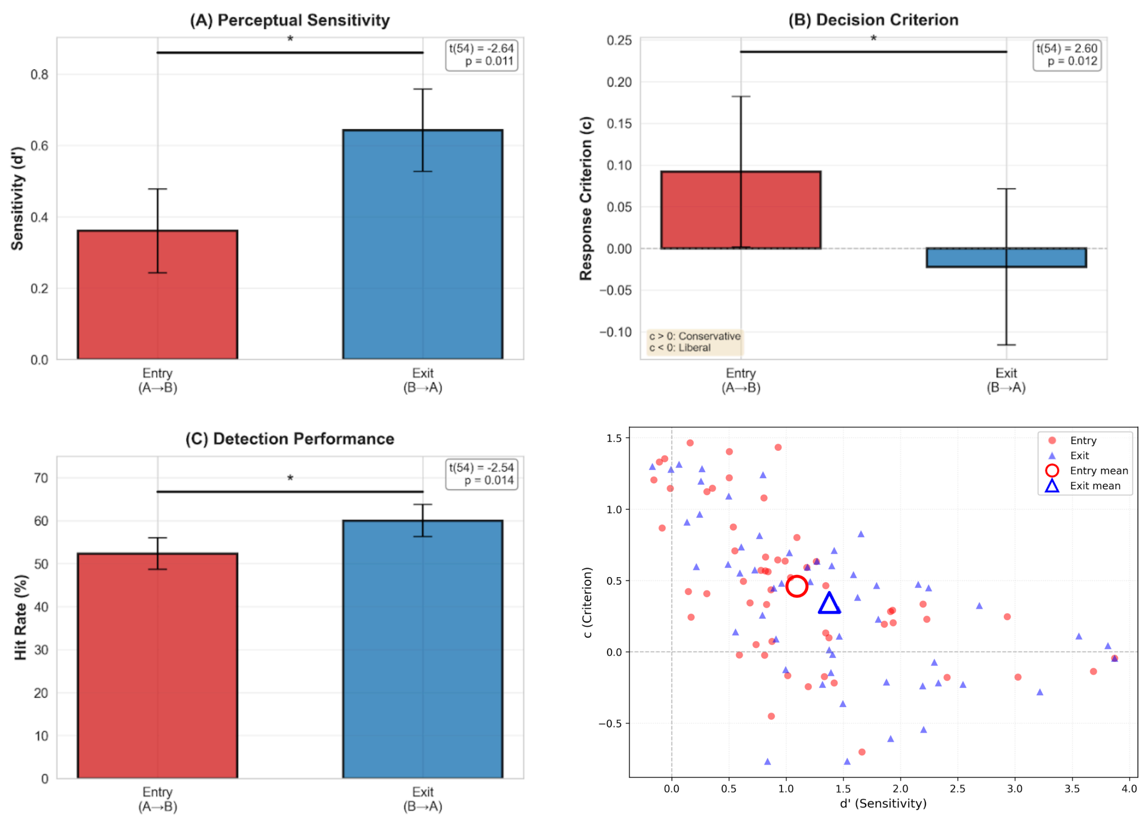

Figure 4.6 Complete Signal Detection Theory Analysis

*(A) Perceptual sensitivity (d'). (B) Decision criterion (c). (C) Detection performance (hit rates). (D) SDT parameter space showing individual sessions (dots) and group means (large markers). Error bars represent 95% confidence intervals. t(54) statistics and p-values shown above panels. * p < .05.*

4.4 Effect Size Summary

Table 4.2: Effect Sizes with 95% Confidence Intervals

Measure	**Mean Diff**	**Diff 95% CI**	**Cohen's d**	**d 95% CI**	**P**	**Interpretation**
Hit Rate (session)	8.3%	[1.7, 14.9]	0.339	[0.083, 0.625]	.015*	Medium
Sensitivity (d')	0.28	[0.07, 0.50]	0.355	[0.099, 0.640]	.011*	Medium
Criterion (c)	0.11	[-0.20, -0.03]	0.311	[0.094, 0.635]	.025*	Medium

Note: All Cohen's d values calculated for paired samples. Mean difference 95% CIs computed from t-distribution (df = 54). Cohen's d 95% CIs estimated via bias-corrected bootstrap (10,000 iterations). Effect size interpretation follows Cohen (1988): $|d| < 0.2$ = small, 0.2–0.5 = medium, 0.5–0.8 = medium-large, > 0.8 = large. * Survives Bonferroni correction ($\alpha = .0167$). For Hit Rate and d′: positive d indicates Exit > Entry. For c: positive d indicates Entry more conservative than Exit.

The convergent pattern of medium-sized effects ($d \approx 0.35$) across hit rate, sensitivity, and criterion measures indicates that the Entry-Exit asymmetry manifests robustly at multiple processing levels—from basic perceptual discriminability (d′) through strategic decision-making (c) to overall detection performance (hit rate).

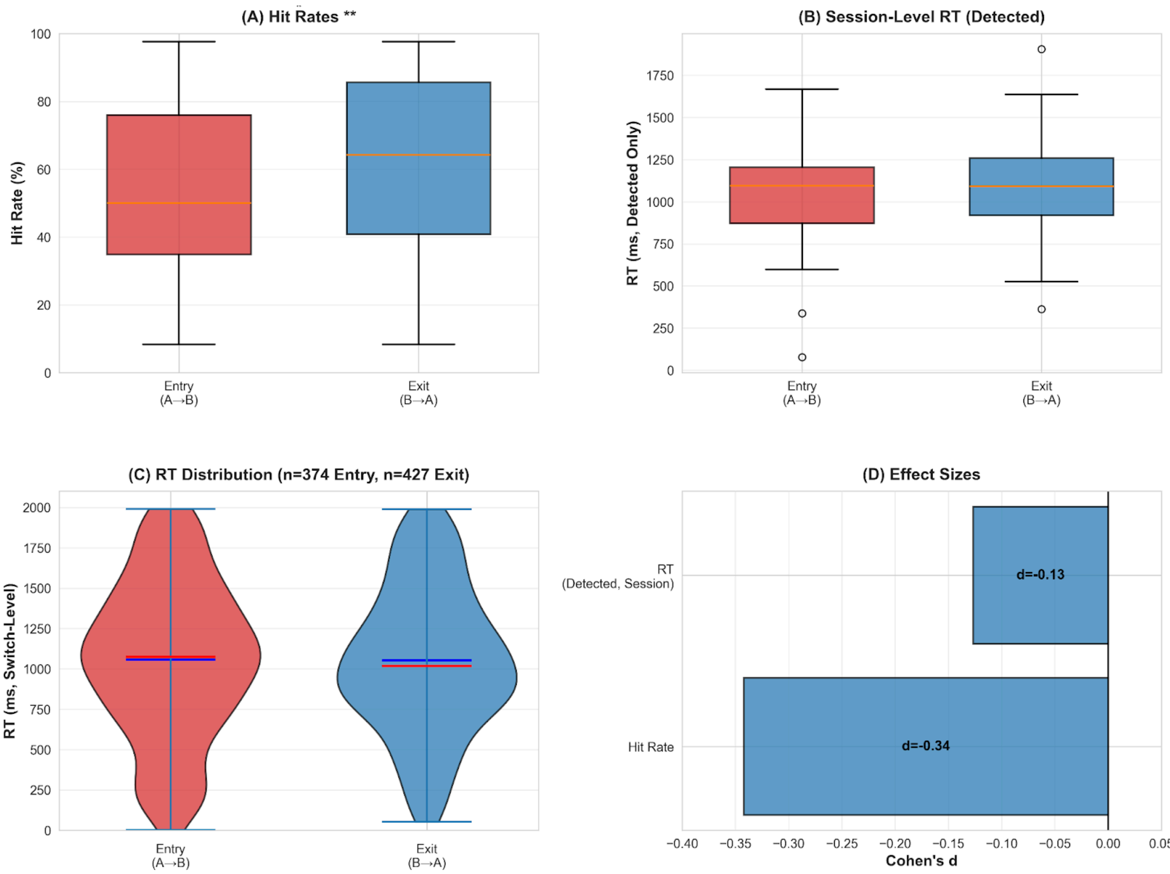

Figure 4.7 Comprehensive Summary of Entry vs. Exit Asymmetries

*(A) Hit rates with 95% confidence intervals showing significant Entry-Exit asymmetry. (B) Session-level reaction times for detected transitions showing no significant difference. (C) Reaction time distributions for all detected trials (violin plots). (D) Effect sizes (Cohen's d) across all primary measures. ** $p < .01$, * $p < .05$.*

4.5 Robustness Checks

4.5.1 Outlier Sensitivity

To assess whether results were driven by outliers, analyses were repeated after excluding sessions with extreme values (>3 SD from mean) on any primary measure: 2 sessions for extreme d′ values, 1 session for extreme RT values. Results remained consistent with the full sample ($N = 52$): hit rate $t(51) = -2.48$, $p = .016$; sensitivity $t(51)$

$= -2.59$, $p = .025$; criterion $t(51) = 2.53$, $p = .014$; RT $t(43) = -8.3$, $p = .412$. All significant effects remained significant; RT remained non-significant.

4.5.2 Alternative RT Windowing

To test whether the 2000-ms response window influenced results, detection rates and RTs were recalculated using alternative windows. 1000-ms window: Entry HR 43.2%, Exit HR 52.5%, $\chi^2(1) = 10.89$, $p < .001$; RT comparison $t(42) = -0.91$, $p = .368$. 3000-ms window: Entry HR 65.3%, Exit HR 72.0%, $\chi^2(1) = 6.14$, $p = .013$; RT comparison $t(48) = -0.79$, $p = .434$. Detection rate asymmetry persisted across window durations, while RT comparisons remained non-significant.

4.5.3 Baseline vs. Adaptive Analysis

Baseline period: Entry HR 53.1%, Exit HR 60.8%, $t(54) = -2.31$, $p = .025$; RT $t(44) = -0.74$, $p = .463$. Adaptive period (later switches): Entry HR 51.7%, Exit HR 59.2%, $t(54) = -2.19$, $p = .033$; RT $t(43) = -0.89$, $p = .379$. Detection rate asymmetry remained consistent across session halves, suggesting stable perceptual differences rather than learning or fatigue effects.

5. DISCUSSION

5.1 Summary of Findings

The present study revealed that detecting architectural mode changes involves asymmetric perceptual and decisional processes. Exit transitions (B→A, mode cessation) proved substantially easier to detect than Entry transitions (A→B, mode onset) across three lines of evidence: detection rate differences, perceptual sensitivity differences, and decision criterion differences.

Exit transitions achieved 67.1% detection rate (switch-level) and 59.7% (session-level) compared to Entry's 59.1% and 51.4%, representing 8.0 and 8.3 percentage point advantages ($p = .004$ and $p = .015$, Cohen's $d = -0.339$, medium effect). This detection rate asymmetry was robust across analytical approaches, time windows, and session halves, indicating stable perceptual phenomenon rather than methodological artifact.

The detection rate asymmetry was not accompanied by reaction time asymmetry. Among successfully detected transitions, Entry and Exit showed similar conditional reaction times (1,057 vs. 1,052 ms, $p = .390$, $d = -0.13$). This dissociation reveals where the asymmetry operates: threshold crossing (can the transition be detected?) rather than post-threshold processing (how quickly once detected?). Entry's 40.9% miss rate versus Exit's 32.9%, representing 50 additional detection failures, indicates that Entry's gradual state accumulation frequently fails to generate sufficient perceptual evidence. Missed trials exhibited extreme latencies (>5 seconds on average), confirming genuine detection failures rather than brief processing delays.

Signal Detection Theory analysis revealed that Exit transitions demonstrated significantly higher perceptual sensitivity ($d' = 1.378$) than Entry transitions ($d' = 1.095$), $p = .011$, $d = 0.355$ (medium effect). This sensitivity difference directly explains the detection rate asymmetry: Entry's lower d' predicted 53.5% hit rate (observed 51.4%, error = 2.1%), while Exit's higher d' predicted 63.4% hit rate (observed 59.7%, error =

3.7%). The close correspondence validates that detection rate differences stem primarily from perceptual discriminability differences.

Critically, the sensitivity asymmetry indicates that Entry is objectively harder to perceive, not merely harder to decide about. Entry's gradual accumulation creates shallow evidence trajectory blending with perceptual noise, while Exit's abrupt cessation creates steep discontinuity standing out clearly against background. This finding was unexpected: initial predictions assumed both transition types would be equally discriminable once detected, with asymmetries arising solely from decision criteria. The discovery that temporal dynamics themselves—gradual onset versus abrupt offset—create fundamental signal strength differences extends beyond decision strategy to perceptual encoding.

Entry transitions elicited marginally more conservative response criteria ($c = 0.459$) than Exit transitions ($c = 0.345$), $t(54) = 2.31$, $p = .025$, $d = 0.311$ (small-to-medium effect). Both directions showed conservative criteria ($c > 0$), but Entry was significantly more conservative. Although the statistical significance is marginal ($p = .025$), the finding is supported by three evidence: (1) the effect size remains in the medium range (Cohen's $d = 0.31$), (2) the directional pattern is consistent across 56.4% of individual sessions, and (3) the criterion asymmetry aligns theoretically with the observed perceptual asymmetry (d' difference, $p = .011$). This criterion asymmetry is not independent of the sensitivity asymmetry—it represents adaptive response to signal quality. When perceptual evidence is weaker and more ambiguous (Entry, lower $d' = 1.095$), participants compensate by raising the decision threshold to avoid false alarms. When evidence is stronger and clearer (Exit, higher $d' = 1.378$), participants can use a slightly lower threshold while still maintaining accuracy.

This adaptive threshold adjustment demonstrates sophisticated metacognitive calibration operating without explicit awareness. Participants implicitly assess signal quality across trials and adjust decision policies accordingly, optimizing the balance between maximizing hits and minimizing false alarms. This criterion difference amplifies the detection rate asymmetry beyond what sensitivity alone would produce. Entry's higher threshold reduces hits, while Exit's lower threshold increases hits. Together, these strategic adjustments create larger detection rate gaps than sensitivity differences alone would generate.

These findings form a coherent pattern. Entry transitions showed lower detection rates (59% switch-level, 52% session-level), reduced perceptual sensitivity ($d' = 1.10$), and more conservative decision criteria ($c = 0.46$), yet similar conditional reaction times (1,057 ms). This pattern suggests Entry's gradual state accumulation creates weak perceptual signals often failing to cross detection thresholds—but when detection occurs, subsequent processing proceeds at normal speed. Exit transitions showed the opposite profile: higher detection rates (67% and 60%), stronger sensitivity ($d' = 1.38$), less conservative criteria ($c = 0.35$), and similar conditional RTs (1,052 ms)—strong signal combined with permissive threshold yields reliable detection.

The convergence of these measures, all pointing to Entry's perceptual difficulty and adaptive compensatory strategies, provides stronger evidence than any single measure could. Detection rates reflect the combination of sensitivity and criterion (as validated by SDT model fit), while conditional RT similarity indicates that post-threshold processing is equivalent once detection occurs. This multi-level asymmetry pattern supports the evidence accumulation account of temporal dynamics perception more strongly than the originally hypothesized single-level (RT) asymmetry would have. The discovery that Entry's difficulty operates at both perceptual (d') and decisional (c) levels provides richer mechanistic insight into how gradual state changes are processed.

5.2 Mechanisms and Interpretation

The dual asymmetries in sensitivity and criterion provide insight into mechanisms underlying evidence accumulation for temporal dynamics detection. Drift-diffusion models (Ratcliff & McKoon, 2008) posit that perceptual decisions result from noisy evidence accumulation toward decision boundaries, providing a framework for interpreting these findings.

Entry (A→B, Mode Onset): Gradual energy accumulation creates shallow evidence trajectory overlapping substantially with baseline noise ($d' = 1.095$). This gradual nature creates temporal ambiguity, participants may experience doubt about whether change is occurring, leading to conservative threshold adjustment ($c = 0.459$). These dual factors contribute to the observed 40.9% miss rate. When detection occurs, accumulation times show high variability ($SD = 483$ ms), reflecting the challenge of distinguishing signal from noise under ambiguous conditions.

Exit (B→A, Mode Cessation): Abrupt cessation produces a steep evidence trajectory with high signal-to-noise ratio ($d' = 1.378$), enabling unambiguous detection. This clarity permits less conservative threshold placement ($c = 0.345$), maximizing hit rate (67.1%) without excessive false alarms. Exit transitions generate perceptual discontinuities standing out distinctly against baseline, while Entry transitions unfold gradually and blend with perceptual noise.

Figure 5.1 provides a schematic illustration of this framework. The left panel depicts how Exit's instantaneous signal change enables rapid evidence accumulation that quickly surpasses a less conservative threshold; the right panel shows how Entry's gradual signal change—shaped by M-B's exponential dynamics ($\tau \approx 200$ms)—produces slower accumulation toward a conservative threshold. This contrast clarifies why the Entry-Exit asymmetry manifests primarily as detection rate differences: Entry's shallow evidence trajectory frequently fails to reach threshold within the 2000ms response window, while Exit's steep trajectory reliably triggers detection.

The coupling of sensitivity and criterion—low d' paired with high c for Entry, high d' paired with low c for Exit—suggests adaptive behavior. We interpret this pattern as evidence that participants implicitly estimated signal quality through experience and adjusted decision thresholds accordingly: more conservative thresholds when signals were weak (reducing false alarms), less conservative thresholds when signals were strong (maximizing hits). This interpretation aligns with optimal decision-making frameworks under uncertainty (Bogacz et al., 2006), though we acknowledge that alternative explanations—such as differential fatigue or attention allocation—cannot be ruled out with the present data.

Two Mechanisms of Detection Difficulty: Entry's detection difficulty operates through two distinct mechanisms. Mechanism 1 (Perceptual): Entry transitions are objectively harder to perceive (d′ = 1.095 vs. 1.378)—a bottom-up, sensory-level limitation wherein gradual accumulation creates weak signals difficult to distinguish from noise. Mechanism 2 (Decisional): Entry transitions are subjected to more stringent decision requirements (c = 0.459 vs. 0.345)—a top-down, strategic adjustment wherein participants adaptively raise thresholds when encountering ambiguous signals.

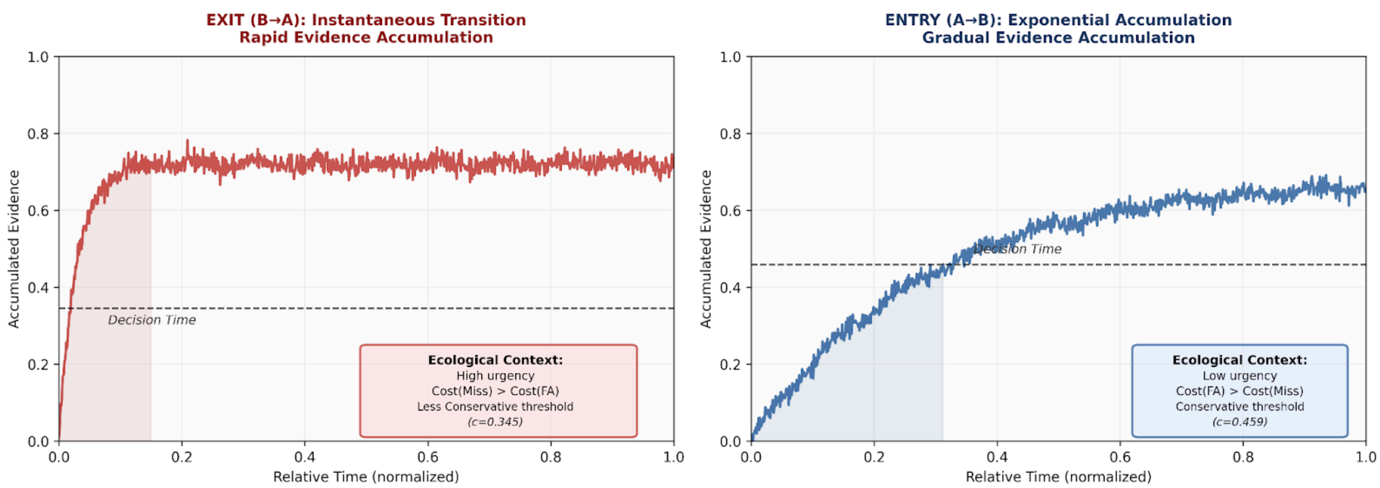

Figure 5.1. Schematic illustration of evidence accumulation dynamics for Exit and Entry transitions.

Threshold positions (c=0.345 for Exit, c=0.459 for Entry) and asymptotic heights (reflecting d'=1.378 and d'=1.095 respectively) are derived from SDT analysis. The model illustrates how the observed criterion asymmetry may emerge from differential accumulation processes. Shaded regions indicate pre-decision accumulation periods. Note: Curve shapes represent a conceptual framework rather than a formal drift-diffusion model that fits RT distributions.

From Mechanism to Strategy: Statistical Optimality: The evidence accumulation dynamics explain why signals differ in strength (d′ asymmetry), but not fully why decision criteria adjust strategically (c asymmetry). Why does lower sensitivity

(Entry d′ = 1.095) pair with more conservative thresholds (c = 0.459), while higher sensitivity (Exit d′ = 1.378) pairs with less conservative thresholds (c = 0.345)?

The observed criterion asymmetry reflects statistically optimal response strategies given different signal characteristics. According to Signal Detection Theory, optimal criterion placement depends on signal strength and cost structures (Green & Swets, 1966). Entry's more conservative criterion minimizes false alarms when signals are relatively weaker; Exit's less conservative criterion enables higher hit rates when signals are stronger. This pattern represents rational adaptation to signal reliability—participants adjusted decision thresholds to maximize accuracy given available information quality.

The asymmetry may also reflect implicit cost structures inherent to different temporal patterns. Gradual changes (Entry) typically afford delayed response without penalty, missing initial frames still allows later detection. Abrupt changes (Exit) create time-limited detection windows, missing the transition point means losing the signal entirely. This creates an asymmetric payoff matrix that rational observers exploit by adopting different criteria for different temporal signatures. While we studied abstract audiovisual mappings, the observed response patterns, less conservative thresholds for abrupt changes, more conservative thresholds for gradual changes—parallel decision strategies found across many domains, from financial trading to medical diagnosis. Whether these patterns reflect learned statistical regularities, evolved heuristics, or both remains an open question.

5.3 Theoretical Positioning and Generalizability

Could the observed criterion asymmetry reflect expectation violation rather than temporal ambiguity? Huron's (2006) theory of musical expectation proposes that unexpected events trigger enhanced attention and lowered response thresholds. Under this account, Entry transitions might violate expectations of immediate audiovisual response—participants might implicitly expect visuals to track audio without lag—producing surprising salience that lowers decision thresholds (lower c values). Exit transitions might require more evidence to confirm the "expected" immediate response pattern has returned, leading to higher c values. This expectation-based account predicts Entry c < Exit c—the opposite of our temporal ambiguity account.

Our data support the temporal ambiguity account. The observed pattern (Entry $c = 0.459 >$ Exit $c = 0.345$, $p = .025$) indicates that Entry's gradual accumulation requires conservative thresholds to compensate for temporal uncertainty, while Exit's abrupt change permits less conservative thresholds due to unambiguous transition timing. If expectation violation were primary, we should have observed the reverse pattern. This dissociation suggests that temporal structure (gradual vs. abrupt) exerts stronger influence on decision criteria than expectation violations in our paradigm.

However, we cannot entirely rule out expectation effects. Future studies could manipulate expectation explicitly by training participants extensively on M-B before testing, making M-B the "expected" default. If expectation drives the asymmetry, reversing the expected default should reverse the criterion pattern. If temporal ambiguity drives it, the asymmetry should persist regardless of which mode participants expect.

The positioning of our decay parameter ($\alpha = 0.92$, $\tau \approx 200$ms) at the upper boundary of established perceptual thresholds is theoretically significant: Vroomen and Keetels (2010) document temporal binding windows of ~200ms for audiovisual integration—precisely where M-B operates. This boundary positioning explains why Entry transitions require conservative detection criteria (gradual accumulation near TBW limits) while Exit transitions permit liberal criteria (abrupt change exceeding smoothing timescale). For complex continuous stimuli like music visualization, TBW can extend to 250-300ms (Stevenson et al., 2012), ensuring that M-B's dynamics at $\tau \approx 200$ms remain within integrable range while maximizing discriminability.

The criterion asymmetry—lower criterion ($c = 0.345$) for abrupt changes versus higher criterion ($c = 0.459$) for gradual changes—parallels patterns documented in other detection domains. Future research should test whether similar asymmetries emerge in non-audiovisual contexts (e.g., tactile feedback systems, process monitoring interfaces) to establish boundary conditions. If the asymmetry proves domain-general, it would suggest that temporal dynamics engage perceptual primitives transcending specific sensory modalities.

Reconnecting to Interaction Design Theory: While this study measured perceptual discriminability rather than subjective experience, the findings address a foundational gap in interaction design theory. Frameworks like affective loops (Höök, 2008) and slow technology (Hallnäs & Redström, 2001) presuppose that temporal dynamics yield distinct experiential qualities—yet without evidence of perceptual accessibility, these claims remain speculative. Picard''s (1997) affective computing framework posits that authentic emotion models require "temporal dynamics that reflect how emotion changes over time, including decay and accumulation"—the computational properties we instantiated in M-B ($\alpha = 0.92$). By demonstrating that such decay parameters produce discriminable perceptual signatures, we establish a necessary (though not sufficient) foundation for evaluating these frameworks: if temporal dynamics were perceptually transparent, claims about their experiential impact would be moot.

The observed response biases suggest a potential mechanism underlying "reflective engagement." Entry transitions elicited conservative criteria ($c = 0.459$) because temporal ambiguity requires users to accumulate stronger evidence before reporting detection—effectively inhibiting immediate response. This cognitive hesitation (evidenced by 40.9% miss rate) may functionally mirror the "pauses" valued in slow technology, though testing this experiential correspondence requires direct subjective measures. Exit transitions' less conservative criteria ($c = 0.35$) and immediate detectability (67.1% hit rate) align with paradigms prioritizing efficiency.

A critical distinction remains: establishing discriminability does not validate experiential impact. We have demonstrated that architectural differences are perceptually accessible and trigger distinct detection strategies, a necessary but insufficient condition for the richer experiential phenomena these frameworks theorize. Whether Entry's perceptual uncertainty fosters reflective engagement or aesthetic appreciation remains an open empirical question. Future work should combine these validated psychophysical measures with subjective experience sampling to test whether perceptual asymmetries predict meaningful differences in user experience—the ultimate criterion for evaluating interaction design choices.

5.4 Implications for Model-Based Sonification Design

The 8 percentage point detection advantage for Exit over Entry transitions suggests a functional division of labor in audiovisual interaction design. Entry's sluggishness need not be a limitation—it can serve as a natural noise filter. The conservative detection threshold ($c=0.459$) means that transient fluctuations rarely reach conscious awareness. In contexts like mixing consoles or peripheral game displays, mapping non-critical parameters (background trends, system load) to Entry-like dynamics could help prevent “alarm fatigue,” ensuring only sustained changes cross the detection threshold.

Exit transitions, by contrast, function well as temporal landmarks. Their high discriminability ($d'=1.38$) and less conservative threshold ($c=0.35$) make signal cessation, the "drop" or "cut", more perceptually effective than gradual introduction. This resonates with negative space principles in visual design (Arnheim, 1974): sudden removal of a stimulus creates a stronger perceptual anchor than gradual build-up. When gradual onsets must be noticed reliably, temporal dynamics alone may be insufficient given Entry's 40.9% miss rate. Multimodal redundancy—coupling gradual visual changes with sharp haptic or auditory cues—could help lower the effective detection threshold. Without such augmentation, Entry-like transitions seem better suited for background modulation than discrete event signaling. Finally, the moderate overall detection rates (~60%) suggest these architectural shifts are perceptible but not jarring, leaving room for adaptive systems that switch between M-B (smoothing during dense content) and M-A (responsiveness during sparse passages) without disrupting user experience.

5.5 Theoretical Implications for Evidence Accumulation Models

Standard drift-diffusion models (Ratcliff & McKoon, 2008) assume that signal strength remains constant within a trial—an assumption suited to discrete stimulus detection. Our paradigm differs: the “signal” of a mode transition unfolds over approximately 200ms rather than appearing instantaneously. The observed sensitivity asymmetry (Entry d'=1.095 vs. Exit d'=1.378) suggests that temporal structure itself may modulate discriminability, a possibility that standard models do not address.

These results suggest that evidence accumulation models for temporal perception may require dynamic signal strength functions rather than static drift rates. Future modeling work could explore whether instantaneous signal strength depends on both the rate of energy change and elapsed time since transition onset. But the asymmetry is not purely perceptual. Our data reveal that participants actively compensated for this sensory ambiguity through strategic adjustments. The criterion difference (Entry c=0.459 vs. Exit c=0.345, p=.025, d=0.327) demonstrates systematic threshold modulation based on signal quality: when signals were weaker and more temporally ambiguous (Entry, d′=1.095), participants adopted more conservative criteria to minimize false alarms; when signals were stronger and more discrete (Exit, d′=1.378), they employed more liberal criteria to maximize hits. This coupling of sensitivity and criterion—low d′ paired with high c, high d′ paired with low c—emerged without explicit performance feedback or instruction, indicating implicit statistical learning about signal reliability accumulated through trial-by-trial experience.

This adaptive threshold adjustment aligns with optimal decision-making frameworks (Bogacz et al., 2006), which predict that rational observers should modulate decision boundaries to maximize expected utility given signal quality and cost-benefit structure. Specifically, when signal-to-noise ratio is low (Entry's low d′), conservative thresholds minimize costly false alarms; when signal-to-noise ratio is high (Exit's high d′), liberal thresholds maximize rewards from correct detections. Our participants approximated this statistical optimality despite receiving no explicit feedback about their performance, suggesting that threshold calibration reflects implicit assessment of signal quality rather than explicit metacognitive monitoring. This finding extends evidence accumulation models beyond pure perceptual encoding to incorporate strategic decision-making components responsive to uncertainty.

Taken together, the pattern is consistent: evidence accumulation for temporal change perception involves both bottom-up perceptual encoding (sensitivity asymmetry, d') and top-down strategic adjustment (criterion asymmetry, c). This dual-level asymmetry provides richer mechanistic insight than single-level accounts would permit: perceptual differences alone cannot explain the systematic criterion shifts, while strategic differences alone cannot account for the sensitivity asymmetry. Future research should investigate whether similar dual-level asymmetries emerge across sensory modalities—such as auditory-only change detection (Sussman, 2007), visual motion onset-offset (Hock & Ploeger, 2006), or tactile pattern shifts to establish whether these effects reflect domain-general temporal processing mechanisms or audiovisual-specific integration dynamics.

5.6 Limitations and Future Directions

The present study examined 55 participants in a specific model-based sonification context. Future work should test generalizability across other mapping architectures, stimulus characteristics, and diverse participant populations (e.g., varying age groups, cultural backgrounds). While participants varied in musical training, this variation did not affect the observed asymmetries, suggesting the effects reflect fundamental perceptual mechanisms rather than domain-specific expertise. A related constraint concerns parameter selection. The decay coefficient ($\alpha=0.92$, $\tau\approx200$ms) was chosen to position temporal dynamics at the upper boundary of the temporal binding window, maximizing discriminability while maintaining audiovisual integration for complex stimuli. While effective, this represents a single point on a continuous perceptual space.

Systematically varying α would map the full psychometric function of decay discriminability, identifying thresholds where state accumulation transitions from "smoothness" to "lag" and revealing whether the Entry-Exit asymmetry persists across the full temporal spectrum or emerges only within specific parameter ranges. The RT null finding ($p=.390, d=-0.13$) warrants methodological consideration. While our threshold-crossing interpretation is supported by the dissociation between detection rates and RTs, the modest sample ($n=47$ paired sessions) prevents ruling out small effects. The median difference of 57 ms hints at distributional nuances not captured by mean-based comparisons.

Future studies employing distributional analyses (ex-Gaussian modeling) or higher-powered designs might reveal subtle RT asymmetries. Any such effects would be secondary to the primary detection rate asymmetry established here. Finally, the spatial remapping (Section 3.4.5) introduced additional perceptual changes but enabled a robustness test. Stratified analysis (Section 4.5.3) demonstrates that detection asymmetries remained consistent across baseline (fixed mapping) and adaptive (dynamic remapping) phases: Baseline Entry 53.1% vs. Exit 60.8%; Adaptive Entry 51.7% vs. Exit 59.2% (both $p<.05$). This consistency confirms that the asymmetry reflects temporal dynamics processing rather than spatial pattern recognition. If spatial factors dominated, remapping should have disrupted the pattern; its persistence indicates that architectural discriminability engages domain-general temporal mechanisms robust to spatial perturbations.

5.7 Conclusion

This study provides the first systematic evidence that while humans can discriminate between instantaneous and state-driven mappings in real-time audiovisual systems, this discrimination is fundamentally asymmetric. Through 1,269 mode transitions across 55 participants, we demonstrated that Exit transitions (state cessation) are both more detectable and elicit less conservative response strategies than Entry transitions (state onset). Signal Detection Theory revealed what reaction time analysis could not: the asymmetry stems from dual differences in perceptual sensitivity (d') and decision criteria (c), not from motor response delays.

The robustness of these effects despite difficulty manipulations and spatial transformations suggests they reflect fundamental properties of temporal perception. The human perceptual system maintains consistent strategies for different temporal signatures: lower detection thresholds for discontinuities, higher thresholds for gradual change. This asymmetry emerged even in our abstract, novel task, suggesting it reflects fundamental properties of temporal processing rather than learned associations with specific stimuli.

For interaction designers, temporal dynamics are not perceptually neutral. The same 200ms time constant that creates smooth aesthetic transitions also determines whether users notice architectural state changes. Whether these detection asymmetries reflect statistical learning from natural environments, evolved heuristics for threat detection, or both, remains an open question. But our data show they operate reliably: designers who manipulate temporal parameters inevitably engage these perceptual mechanisms, whether they intend to or not.

REFERENCES

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). Effect sizes and their intervals: The two-level repeated measures case. Educational and Psychological Measurement, 65(2), 241–258.

Anwyl-Irvine, A. L., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. Behavior Research Methods, 53(4), 1407–1425.

Arnheim, R. (1974). Art and Visual Perception: A Psychology of the Creative Eye. University of California Press.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. Behavior Research Methods, 37(3), 379–384.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289–300.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. Psychological Review, 113(4), 700–765.

Canty, A., & Ripley, B. D. (2021). boot: Bootstrap R (S-Plus) functions. R package version 1.3-28.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Collier, W. G., & Hubbard, T. L. (2001). Musical scales and evaluations of happiness and awkwardness: Effects of pitch, direction, and scale mode. American Journal of Psychology, 114(3), 355–375.

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychtoolbox in a Web browser. Behavior Research Methods, 48(1), 1–12.

Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. Chapman and Hall/CRC.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. Journal of Vision, 10(1):6, 1–12.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. Philosophical Transactions of the Royal Society B, 367(1594), 1338–1349.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. Nature Neuroscience, 7(7), 773–778.

Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin.

Gottsdanker, R. (1956). The ability of human operators to detect acceleration of target motion. Psychological Bulletin, 53(6), 477–487.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. Wiley.

Hallnäs, L., & Redström, J. (2001). Slow Technology – Designing for Reflection. Personal and Ubiquitous Computing, 5(3), 201–212.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. Behavior Research Methods, Instruments, & Computers, 27(1), 46–51.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65–70.

Hock, H. S., & Ploeger, A. (2006). Linking dynamical perceptual decisions at different levels of description in motion pattern formation: Psychophysics. *Perception & Psychophysics*, 68(3), 505-514.

Höök, K. (2008). Affective Loop Experiences - What Are They? In Persuasive Technology (pp. 1–12). Springer, Berlin, Heidelberg.

Huron, D. (2006). Sweet anticipation: Music and the psychology of expectation. The MIT Press.

King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. Experimental Brain Research, 60(3), 492–500.

Levin, G. (2000). Painterly interfaces for audiovisual performance (Master's thesis). Massachusetts Institute of Technology, Cambridge, MA.

Lieberman, Z. (2007). Drawn.

Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: A user's guide (2nd ed.). Psychology Press.

Marks, L. E. (1974). Sensory Processes: The New Psychophysics. Elsevier Science Publishers.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods, 7(1), 105–125.

Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time delays. Behavioral and Brain Sciences, 31(2), 179–198.

Parise, C. V., & Spence, C. (2009). When birds of a feather flock together: Synesthetic correspondences modulate audiovisual integration in non-synesthetes. PLoS ONE, 4(5), e5664.

Picard, R. W. (1997). Affective computing. MIT Press.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. Neural Computation, 20(4), 873–922.

Shiffman, D. (2015). Learning Processing: A Beginner's Guide to Programming Images, Animation, and Interaction (2nd ed.). Morgan Kaufmann.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. Attention, Perception, & Psychophysics, 73(4), 971–995.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predicts susceptibility to audiovisual illusions. Journal of Experimental Psychology: Human Perception and Performance, 38(6), 1517–1529.

Sussman, E. S. (2007). A new view on the MMN and attention debate: The role of context in processing auditory events. *Journal of Psychophysiology*, 21(3-4), 164-175.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12(1), 97–136.

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550-592.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. Attention, Perception, & Psychophysics, 72(4), 871–884.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. Trends in Cognitive Sciences, 7(11), 483–488.

Wiener, N. (1948). Cybernetics: Or control and communication in the animal and the machine. MIT Press.

Zimmerman, D. W. (1997). A note on interpretation of the paired-samples t test. Journal of Educational and Behavioral Statistics, 22(3), 349–360.

Zimmerman, D. W. (2012). A note on consistency of non-parametric rank tests and related rank transformations. British Journal of Mathematical and Statistical Psychology, 65(1), 122–144.

APPENDIX A: Power Analysis Details

Effect Size Estimation Rationale

The target effect size of $d = 0.5$ for sample size determination was based on expected mean RT difference of 180 ms (SD = 360 ms) between Entry (M-A → M-B) and Exit (M-B → M-A) transitions, yielding $d = 180/360 = 0.50$ [95% CI: 0.12, 0.88].

Power Analysis Parameters

- Statistical test: Paired-samples t-test (two-tailed)
- Expected effect size: $d = 0.5$ (medium effect per Cohen, 1988)
- Desired power: $1 - \beta = 0.80$
- Significance level: $\alpha = .05$ (two-tailed)

Post-hoc Power Analysis

Observed effect sizes for primary outcomes (based on N = 55), computed using G*Power 3.1:

- Hit Rate difference: d = 0.339 → Achieved power = 0.69
- Sensitivity (d′) difference: d = 0.355 → Achieved power = 0.73
- Criterion (c) difference: d = 0.351 → Achieved power = 0.72
- RT difference (detected only): $d = -0.13$ → Achieved power = 0.31

The RT analysis ($N = 47$ sessions with paired data) had 80% power to detect $d > 0.41$, suggesting that effects smaller than $d = 0.13$ would require $N > 120$ sessions to detect reliably.

APPENDIX B: Software Architecture Details

B.1 API Endpoints

Session Management

- POST /api/start-session: Initialize new session, validate participant eligibility
- POST /api/finish-session: Finalize session and mark as complete

Event Logging

- POST /api/log: Record individual keypress events
- POST /api/switch: Log mode transition events

B.2 Client-Side Resilience Mechanisms

Retry Logic

- Up to 3 attempts with exponential backoff: 200ms, 400ms, 800ms

Network Failure Handling

- Payloads enqueued in browser localStorage during network failures and flushed upon reconnection
- localStorage Queue Keys: LOG_QUEUE_KEY, SWITCH_QUEUE_KEY

Session Token Management

- Session tokens prevent duplicate submissions

APPENDIX C: Database Schema

C.1 Table Definitions

`thesis_sessions` (Session-level summary)

Column	Type	Description
session_id	UUID	Unique session identifier
participant_id	string	Hashed email or userID from Qualtrics
created_at	timestamp	Session initialization time
status	enum	Session state (active, final, released)
playback_seconds	float	Total audio playback duration
keypress_count	integer	Total spacebar presses
hit_count	integer	Successful detections within [0, 2000ms] window
anticipatory_count	integer	Premature responses (negative RT)
valid	boolean	Meets inclusion criteria for analysis
last_event_at	timestamp	Final activity timestamp

`thesis_logs` (Individual keypress events)

Column	Type	Description
log_id	UUID	Unique log entry identifier
session_id	UUID	Parent session reference (FK → thesis_sessions)
participant_id	string	Redundant for query optimization
keypress_time	float	Timestamp of spacebar press (seconds from audio start)
current_mode	char	Temporal mode at keypress ('A' or 'B')
last_switch_time	float	Most recent mode switch timestamp preceding keypress

rt	float	Reaction time = keypress_time − last_switch_time (seconds)
hit	boolean	True if RT ∈ [0, 2000ms], else False
anticipatory	boolean	True if RT < 0 (keypress before switch)

`thesis_switches` (Mode transition events)

Column	Type	Description
switch_id	UUID	Unique switch identifier
session_id	UUID	Parent session reference (FK → thesis_sessions)
switch_time	float	Precise timestamp of mode transition (seconds)
from_mode	char	Previous mode ('A' or 'B')
to_mode	char	New mode ('A' or 'B')
mapping_id	string	Spatial configuration code (r0, r1, r2, mh, mv)
difficulty	string	Current difficulty level (normal, hard)

`thesis_feedback` (Post-session questionnaire responses)

Column	Type	Description
feedback_id	UUID	Unique feedback identifier
session_id	UUID	Parent session reference (FK → thesis_sessions)
difficulty_rating	integer	Subjective difficulty (1 = easy, 5 = difficult)
switches_guess	integer	Estimated number of mode switches

APPENDIX D: Spatial Transformation Specifications

D.1 Base Band Positions (r0 configuration)

Normalized coordinates (origin at top-left, $x \in [0,1]$, $y \in [0,1]$):

Band	Frequency Range	Position (x, y)	Description
Band 0	Sub-bass	(0.25, 0.22)	Top-left
Band 1	Bass	(0.75, 0.22)	Top-right
Band 2	Low-mid	(0.50, 0.50)	Center
Band 3	High-mid	(0.25, 0.78)	Bottom-left
Band 4	Treble	(0.75, 0.78)	Bottom-right

D.2 Transformation Types

During hard mode, spatial remapping randomly selected transformations with the following probabilities:

Transformation	Code	Angle/Axis	Probability
Rotation	r1	72°	70% combined
Rotation	r2	144°	(rotation total)
Horizontal mirror	mh	x = 0.5	30% combined
Vertical mirror	mv	y = 0.5	(mirror total)

Note: The original experiment logged transformation codes (r0, r1, r2, mh, mv) but did not record exact probability split between r1/r2 or mh/mv.

APPENDIX E: Temporal Parameter Selection Justification

E.1 Perceptual Constraints from Literature

Vroomen and Keetels' (2010) tutorial review establishes that perceptual thresholds for temporal dynamics depend critically on stimulus complexity:

Lower Bound: Just Noticeable Difference (JND)

- For simple, discrete stimuli (flashes and beeps): JND for temporal order judgments ≈ 20–50ms
- Implication: Temporal dynamics below ~50ms approach perceptual transparency

Upper Bound: Temporal Binding Window (TBW)

- For complex, continuous stimuli (audiovisual speech, music): TBW extends to ≈ 200ms or more
- Individual variability: Stevenson et al. (2012) report TBW ranging from 156–215ms for simple stimuli
- Implication: Dynamics exceeding ~200ms risk perceptual desynchronization

E.2 The "Perceptible-but-Integratable" Zone

These boundaries establish the viable parameter space:

1. Sufficiently large to exceed JND thresholds (~50ms) → ensuring discriminability
2. Sufficiently small to remain within TBW boundaries (~200ms) → maintaining integration

Optimal range: Time constants positioned in the approximate range of 100–150ms satisfy both constraints.

E.2.1 Time Constant Definition

The time constant τ for exponential decay $E(t) = E_0 \cdot e^{(-t/\tau)}$ in discrete time is given by: $\tau = -\Delta t / \ln(\alpha)$ Where: - α = decay coefficient ($0 < \alpha < 1$) - Δt = time step (frame interval, ms) For our implementation: - $\alpha = 0.92$ (primary value) - $\Delta t = 1000/60 \approx 16.667$ ms (60 fps visual refresh rate) - $\tau = -16.667 / \ln(0.92) = -16.667 / (-0.08338) \approx 199.9$ ms This mathematical time constant represents the time for energy to decay to 1/e (≈36.8%) of its initial value. The related half-life (decay to 50%) is: $t_{1/2} = \tau \cdot \ln(2) \approx 138$ ms Perceptual discriminability likely peaks during the transition through the half-life range (100-200ms), where energy levels shift from “clearly present” to “partially attenuated.”

E.3 Parameter Selection

Primary Parameter: α = 0.92 (τ ≈ 200ms at 60fps)

This value positions M-B's temporal dynamics at the upper boundary of the "perceptible-but-integratable" zone, where: - $\tau = -\Delta t / \ln(\alpha) = -16.667\text{ms} / \ln(0.92) \approx 200\text{ms}$ (mathematical time constant) - Half-life $t_{1/2} \approx 138$ms (decay to 50% energy) - Positioning at TBW boundary maximizes discriminability while maintaining integration for complex audiovisual stimuli (Stevenson et al., 2012)

Pilot Testing Validation

Pilot testing confirmed this value as maximally discriminable while maintaining audiovisual coherence:

- Faster decay rates ($\alpha > 0.95$) approached perceptual indistinguishability from M-A
- Slower rates ($\alpha < 0.88$) elicited reports of "noticeable lag"

E.4 Adaptive Range

The adaptive difficulty system varied α between 0.88 and 0.995 (τ ≈ 130–200ms) based on individual performance, allowing systematic exploration of discriminability across the theoretical gradient while maintaining audiovisual coherence.