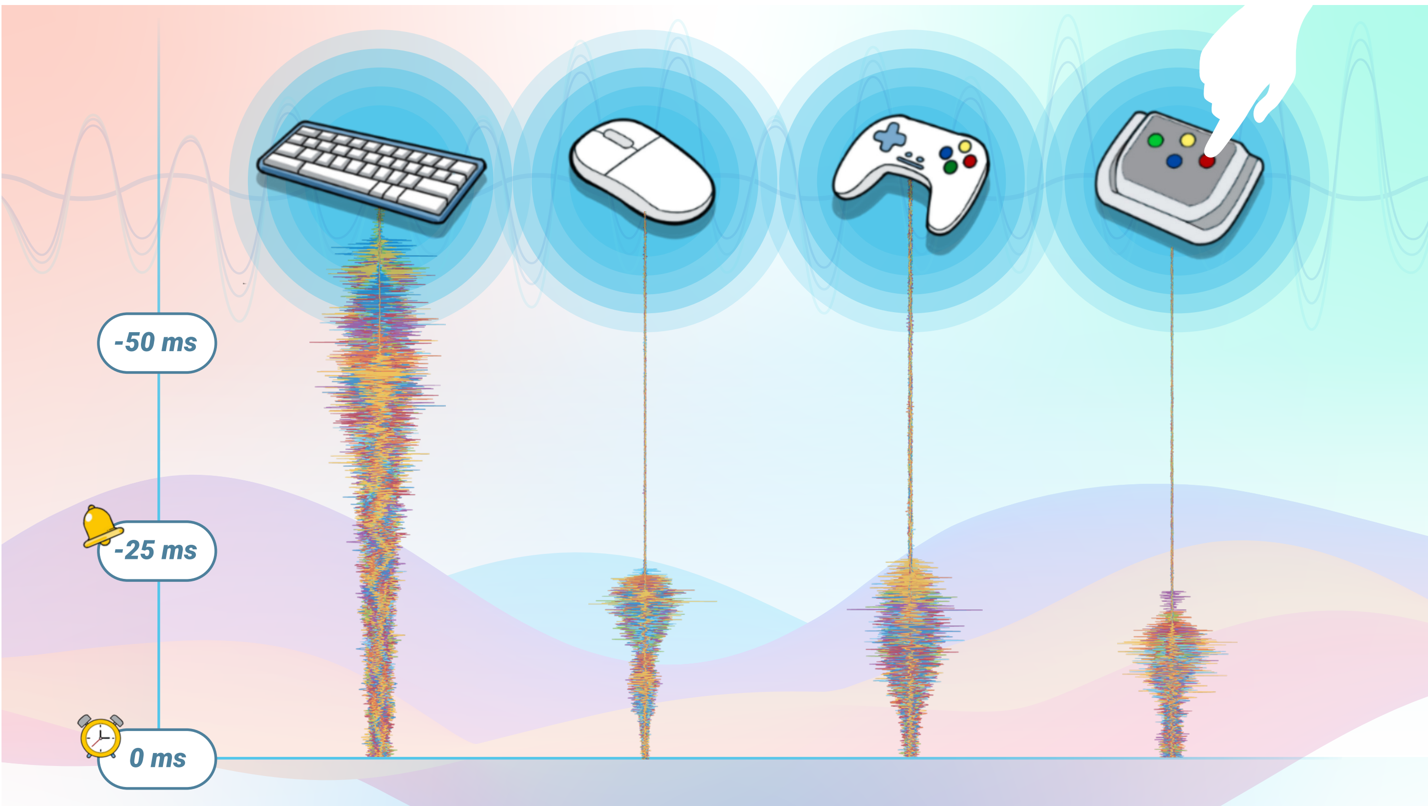
PsyRTDevice: Open-source toolkit for sub-millisecond evaluation of reaction time device accuracy

Ning Tang,[1,3] Aisu Li,[1,3] Bowen Chen,[1] Yiyuan Qiao,[1] Zhicheng Lin,[2,*] and Yang Zhang[1,*]
*Correspondence: zhichenglin@gmail.com (Z.L.); yzhangpsy@suda.edu.cn (Y.Z.)

Received: September 24, 2025; Accepted: November 6, 2025; Published Online: November 7, 2025; https://doi.org/10.59717/j.xinn-inform.2025.100017

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- We introduce an open-source toolkit for fast, convenient assessment of reaction time device precision.
- PsyRTDevice uses sound-wave technology to achieve sub-millisecond accuracy within 10 minutes of testing.
- Evaluation of 20 devices reveals substantial variability in response delays (7.58–47.46 ms) and precision.
- Low-precision devices yield longer absolute RTs but do not significantly affect cognitive effect magnitudes.
- PsyRTDevice enables transparent reporting of device accuracy, enhancing experimental rigor and replicability.

PsyRTDevice: Open-source toolkit for sub-millisecond evaluation of reaction time device accuracy

Ning Tang,[1,3] Aisu Li,[1,3] Bowen Chen,[1] Yiyuan Qiao,[1] Zhicheng Lin,[2,*] and Yang Zhang[1,*]
[1]Department of Psychology, Soochow University, Suzhou 215000, China
[2]Department of Psychology, Yonsei University, Seoul 03722, South Korea
[3]These authors contributed equally
*Correspondence: zhichenglin@gmail.com (Z.L.); yzhangpsy@suda.edu.cn (Y.Z.)
Received: September 24, 2025; Accepted: November 6, 2025; Published Online: November 7, 2025; https://doi.org/10.59717/j.xinn-inform.2025.100017

Citation: Tang N., Li A., Chen B., et al. (2025). PsyRTDevice: Open-source toolkit for sub-millisecond evaluation of reaction time device accuracy. *The Innovation Informatics* 1:100017.

Reaction time (RT) measurement—the interval between stimulus onset and response—provides a crucial window into cognitive processing speed and neural efficiency. Although widely used across cognitive science, psychology, neuroscience, and behavioral science, systematic validation of RT measurement devices remains scarce due to high testing costs and the lack of accessible toolkits. Here, we introduce PsyRTDevice, an open-source toolkit employing sound-wave technology to assess RT measurement latency and variability with sub-millisecond precision and within 10 minutes. Our evaluation of 20 popular wired and wireless response devices revealed substantial variability in response latency and precision. Comparing cognitive task performance between the most and least precise devices across three classic cognitive tasks showed that the lower-precision device consistently yielded longer RT measurements, but differences in device precision did not significantly affect cognitive effects based on RT differences (i.e., between conditions). Hierarchical drift diffusion modelling further revealed that these device-related differences primarily impacted non-decision time. PsyRTDevice thus offers researchers and educators an accessible toolkit to quantify device response latency and variability, supporting informed device selection and enhancing experimental rigor in RT-based research.

INTRODUCTION

Reaction time (RT) as a window into mental speed and neural efficiency

RT measurement stands as one of the most fundamental and enduring methodologies in psychological science and allied fields. Dating back to the 19th century, pioneering researchers such as Hermann von Helmholtz, who measured nerve conduction speeds, and Franciscus Donders, who developed methods to isolate distinct cognitive processing stages through RT differences, established RT as a staple metric for understanding human cognition.

In modern cognitive psychology, RT serves as an essential tool for mapping mental operations across domains including attention, decision-making, and memory retrieval. Classic paradigms such as the Stroop task and experimental frameworks like Hick's Law have demonstrated how RT can quantify processing speed and reveal how factors such as complexity, interference, or cognitive load affect performance. In sports science, precise RT measurement differentiates elite performers and guides training practices. Likewise, human factors research applies RT knowledge to design ergonomic interfaces, with implications for safety and efficiency in domains ranging from driving to aviation. In neuroscience, RT functions as a critical link between behavioral outcomes and underlying neural mechanisms, with models like the drift-diffusion framework connecting decision-making processes with neural activity patterns. This makes RT a valuable biomarker of both brain health and cognitive performance. Indeed, clinically, RT measurements are essential for detecting cognitive impairment, monitoring disease progression, evaluating recovery, and assessing treatment effectiveness.

RT measurement has also played a crucial role in understanding individual differences in cognitive abilities. Since Francis Galton's pioneering work in the 1880s, researchers have investigated relationships between RT and intelligence, demonstrating that individuals with higher psychometric intelligence tend to exhibit somewhat faster RTs, particularly on choice RT tasks. This relationship is thought to reflect central nervous system efficiency, with more consistent RTs and smaller increases in RT as tasks become more complex both correlating more strongly with cognitive abilities. In aging research, the systematic slowing of RTs across the lifespan has been fundamental to theories like Salthouse's Processing-Speed Theory, which proposes that decreased processing speed underlies many cognitive declines associated with aging.

Precise measurement of RT is fundamental across numerous research domains. In aging studies, accurate absolute RT values establish essential normative benchmarks across age groups and help detect subtle changes that may indicate early cognitive decline. Neuroscience research using EEG and MEG requires precise RT timing for response-locked analyses, where neural activity must be precisely aligned with response execution. These analyses depend on exact timing synchronization between behavioral responses and neural recordings to accurately identify components such as the lateralized readiness potential (LRP). In cognitive modeling, particularly drift-diffusion models, absolute timing accuracy directly influences parameter estimation for components like non-decision time, affecting theoretical interpretations about processing efficiency versus response execution. Consequently, the precision of RT recording devices plays a critical role in experimental validity and replicability in research areas requiring fine-grained temporal precision.

From a statistical perspective, measurement errors from input devices directly increase experimental variance, subsequently reducing statistical power and the likelihood of detecting small but genuine cognitive effects. By reducing error variance, high-precision measurement devices can increase statistical power and facilitate more reliable detection and replication of experimental findings. For tasks where subject variance is substantial, the additional error variance from RT measurement may be mathematically acceptable with modest increases in trial counts. However, for paradigms measuring intervals between nearly simultaneous responses or those requiring detection of small effect sizes, the choice of measurement device becomes paramount, as error variance can necessitate substantial increases in participant numbers or trial counts to maintain statistical power. Hence, the choice and calibration of RT measurement instruments directly impact experimental rigor and the ability to detect subtle cognitive phenomena.[1]

Evaluating RT measurement precision

The critical role of RT accuracy has motivated several approaches for systematic device evaluation, as summarized in Table 1. Early efforts by Myors[2] used simple graphical methods to assess device precision. Rather than comprehensively measuring overall response time error, this approach focused specifically on identifying errors within the scanning cycle, including those stemming from program clock and system processing intervals. The method visualized temporal precision through graphical techniques, based on the principle that high-precision devices produce smooth curves, while low-precision devices display stepped patterns.

Later methods achieved more complete response time error measurement but often required costly equipment. Plant et al.[3] employed a signal generator and custom-built device (VSCAR, Visual Stimulus Capture and Response) to simulate participant responses, calculating device precision by measuring the difference between signal emission time and computer-recorded time. Their study of 10 different mouse models revealed significant variations in both mean delay and standard deviation across devices. However, the custom-built device's high cost limited accessibility for many

Table 1. Summary of methods for evaluating response device accuracy.

Study	Measurement scope	Equipment cost	Operational difficulty	Potential device damage
Myors[2]	Partial (scanning cycle only)	Moderate	Moderate	None
Plant et al.[3]	Comprehensive	High	Low	None
Shimizu[4]	Comprehensive	Low	High	None
Neath et al.[5]	Comprehensive	Low	High	None
Bockes et al.[6]	Comprehensive	Low	High	Low (soldering required)
Zhang & Zhang[7]	Comprehensive	Low	High	High (potential cable damage)

Table 2. Device variability in RT research in two top psychology journals.

Device	*Journal of Experimental Psychology: General*, 2024	*Psychological Science*, 2024
	Percentage of RT studies	Percentage of RT studies
Response Box	6.45%	0%
Keyboard	83.87%	76.92%
Mouse	9.68%	7.69%
Gamepad	3.23%	0%
Digitizing Tablet	3.23%	0%
3D Motion Tracking	3.23%	7.69%
Unreported	0%	7.69%

Note: Our analysis of the RT research in two top psychology journal reveals that while keyboards are the predominant input device (83.87% and 76.92%), researchers also frequently employ mouse devices (9.68% and 7.69%) and various other devices such as gamepad (3.23%) and response box (6.45%). Perhaps most concerning is that 7.69% of published RT studies in *Psychological Science* failed to report any information about the specific input device used for RT measurement. This variability and occasional lack of methodological transparency poses challenges for cross-study comparability and replication efforts.

psychology researchers.

Subsequent researchers developed more affordable approaches, though these often sacrificed operational simplicity. Shimizu[4] constructed a custom hardware setup featuring a spring-loaded pen that simultaneously activated buttons on both a gamepad and a keyboard, with signals from both transmitted to the CPU. Using the gamepad-measured time as the reference, the absolute response delay of the keyboard was calculated as the difference between keyboard-recorded time and gamepad time. This method revealed that absolute response delays across different keyboard models ranged from 11 ms to 73 ms, demonstrating considerable variability. However, the approach required precise construction of a specialized spring-loaded pen device.

Similarly, Neath et al.[5] used a photodiode to detect screen brightness changes, which triggered a solenoid to press keyboard keys. Their comparison of different Apple keyboards showed that white keyboards had shorter average absolute response delays than aluminum keyboards, but response standard deviations remained similar across models. Bockes et al.[6] employed optocouplers soldered to input device buttons to record absolute response delays across six different devices, finding that the Logitech RX250 mouse demonstrated superior time precision based on both absolute response delay and response standard deviation. However, this method required soldering optocouplers directly to the devices.

Some researchers attempted to balance simplicity with cost-effectiveness, though sometimes with potential risk to equipment. Zhang and Zhang[7] introduced a method requiring no custom hardware or materials, using E-Prime to test overall absolute response delay. Their approach presented stimuli S1 and S2 sequentially, with a pulse signal sent simultaneously with S2 to simulate a participant's response to S1. The device's delay was calculated by subtracting the interval between S1 and S2 presentations from the recorded reaction time. Testing three reaction devices, they found the IBM-Lenovo optical mouse had the shortest absolute response delay and highest precision. However, this method required custom cables connecting the computer's parallel port to the mouse, creating potential risk of device damage.

The current study

The availability of abundant input devices creates flexibility in experimental design and implementation, often guided by practical considerations based on availability (using readily accessible devices) and convenience (selecting hardware compatible with established experimental procedures). Consequently, substantial variability has emerged, as highlighted by our review of research published in two top psychology journals (see Table 2). Critically, current methods for evaluating device response latency often require specialized equipment, complex setups, or risk damaging the hardware. These barriers have hindered systematic reporting of device accuracy, potentially affecting experimental replicability and interpretation of timing-critical research.

Here we develop and validate a comprehensive solution, the PsyRTDevice toolkit, which offers a sound-wave-based measurement approach with integrated analysis capabilities. This toolkit addresses key challenges in previous methodologies by providing a convenient, accurate method for measuring absolute response delays while eliminating risks of hardware damage. Through this resource, researchers can systematically quantify and transparently report the temporal accuracy of input devices, enabling informed decisions aligned with experimental precision requirements.

Using PsyRTDevice, we present our evaluations of 20 different input devices, quantifying their precision in terms of response delay and variability. Based on these measurements, we identify devices with the highest and lowest precision for use in subsequent experiments investigating the impact of device precision using three established cognitive tasks.

MATERIALS AND METHODS

Design principles

We developed PsyRTDevice to meet two critical requirements. The first concerns measurement precision: achieving sub-millisecond accuracy necessitates a sampling rate of at least 10,000 Hz to adequately capture timing differences between physical key presses and their digital registration. The second requirement focuses on accessibility: the solution must rely on standard computer components without requiring specialized hardware or modifications.

Modern computer sound cards fulfill both requirements, typically offering sampling rates of 48,000 Hz or higher—well above the minimum threshold for precise temporal measurement. By leveraging existing hardware, this approach eliminates the need for specialized equipment while maintaining high measurement accuracy.

Specifically, PsyRTDevice employs a sound-wave-based measurement approach that captures the temporal relationship between physical key actuation and digital registration. The testing procedure consists of three phases: (1) initialization, where users press the spacebar to begin a 3-second audio recording session; (2) response capture, where users press the target input device when prompted by the "Press key" display; and (3) dual measurement, which simultaneously records both the acoustic signal produced by the

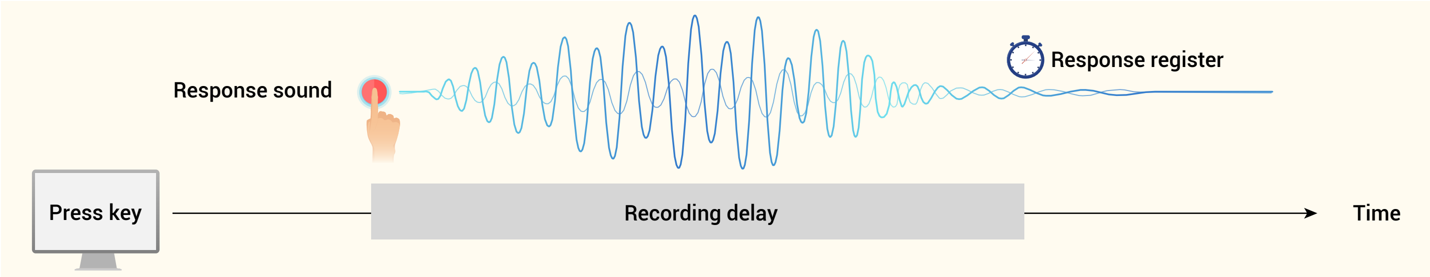


Figure 1. Illustration of the PsyRTDevice measurement method based on sound waves The procedure captures the temporal delay between a physical key press and its registration by the response device. When a key is pressed, the system simultaneously records the acoustic sound produced by the physical keystroke ("Response sound") and the moment when the response is registered by the measurement device ("Response register"). The time difference between these two events ("Recording delay") quantifies the device's response latency with sub-millisecond precision, allowing for accurate assessment of reaction time device performance. The physical keypress time ("Response sound") is identified by the moment the acoustic signal first reaches 40% of its peak amplitude.

physical key press (via microphone) and the timestamp when the response is registered by the computer. This process is repeated across 50 trials to establish robust measurement statistics. The toolkit then calculates the recording delay—the temporal difference between the acoustic event (representing physical actuation) and the computer-recorded response time (see Figure 1).

Signal processing

To ensure accurate measurement, PsyRTDevice implements several critical analysis steps. The process begins with noise filtering, where a high-pass filter (cutoff at 100 Hz) removes low-frequency background noise from the acoustic signal, as implemented through the *butter.m* and *filtfilt.m* functions in MATLAB.[8] This is followed by time-window extraction, in which a temporal window spanning from 70 ms before to 10 ms after the computer-recorded timestamp ("Response register" in Figure 1; the dashed line in Figure 2) is isolated for detailed analysis. Both time parameters (70 ms and 10 ms) can be adjusted as needed.

Within this extracted window, peak detection identifies the maximum amplitude as the signal peak. The toolkit then applies threshold crossing detection, defining the actual keypress time as the first moment when the signal amplitude reaches 40% of peak value—a threshold that can be adjusted. The 40% threshold was chosen empirically, as it reliably captures the initial onset of acoustic energy produced by key activation, minimizing variability from background noise and subsequent signal fluctuations. The interval between this acoustic event timestamp and the computer-recorded timestamp represents the recording delay. To establish reliability, 150 trials were conducted for each device. The mean delay and standard deviation across all the trials help quantify both central tendency and variability in device response times.

Code implementation

The PsyRTDevice toolkit consists of three primary MATLAB components. The first, *responseDevTest.m*, handles the measurement process and data collection. The second component, *extraEpochData.m*, cleans acoustic noise and preprocesses raw data into manageable time windows. The third component, *getRespDevPrecision.m*, performs statistical analysis and visualization. This streamlined implementation allows the complete testing procedure for a single device (50 measurements) to be completed in less than 10 minutes. The modular design also enables researchers to customize specific aspects of the analysis. The toolkit is available at https://github.com/yangzhangpsy/PsyRTDevice.[9]

Device evaluation

The experimental program was developed using PsyBuilder 0.1[10] and Psychtoolbox 3.0.18.13.[11,12] It was run on standard consumer-grade hardware: a Lenovo Legion R720 laptop featuring an Intel Core i7-7700HQ processor, an NVIDIA GTX1050Ti graphics card, and running Windows 10. Visual stimuli were displayed on an LG IPS monitor (model LP156WF6-SPK3; 15.6 inches; resolution: 1920 × 1080 pixels; refresh rate: 60 Hz). All external input devices were connected via the computer's USB 3.1 Gen 1 (Type-A) interface.

Testing was conducted in a controlled, quiet environment. A set of 20 external input devices (or their operating modes) was evaluated, comprising five gamepad models, six multi-mode mouse models, and seven keyboard models (Table 3). This diverse selection enabled robust assessment of latency characteristics across various common response devices typically employed in behavioral and psychological research.

Evaluation results

Testing with PsyRTDevice yielded comprehensive device precision profiles for all 20 input devices/modes. By overlaying audio signals from all 150 trials per device, we generated distribution maps revealing characteristic response delay patterns (Figure 2). Our measurements revealed substantial differences in both latency (mean absolute response delay) and variability (standard deviation) across device types.

For latency, the Logitech F310 gamepad operating in D-mode demonstrated superior performance with the shortest mean response delay (7.58 ms), while the Dell L100 keyboard exhibited the poorest performance with the longest delay (47.46 ms)—a difference of 39.88 ms. Regarding measurement consistency, the Logitech F710 gamepad (D-mode) showed the highest stability with the smallest standard deviation (1.82 ms), whereas the Dell L100 keyboard again performed poorest with the largest standard deviation (7.54 ms).

Device precision followed a clear pattern across input categories, with gamepads demonstrating the highest overall precision (shortest delays and smallest variability), mice showing intermediate precision, and keyboards exhibiting the lowest overall precision (longest delays and greatest variability). These findings provide empirical benchmarks that can inform the selection of appropriate input devices for RT-based studies requiring varying levels of temporal precision.

RESULTS

Impact of device precision on cognitive effects

After establishing device precision measurements, we conducted three experiments to determine how input device characteristics affect RT measurements and cognitive effects. We compared the highest-precision device (Logitech F310 gamepad in D-mode) with the lowest-precision device (Dell L100 keyboard) in three classic cognitive effects: inhibition of return (IOR); Stroop effect; and endogenous attention cueing effect.

Our analysis combined behavioral measures with computational modeling using hierarchical drift diffusion modeling (HDDM).[13-16] HDDM decomposes RT distributions into psychologically meaningful parameters, allowing us to determine whether device precision affects overall RTs and whether these effects specifically impact cognitive processing or response execution:

Drift rate (v): Rate of evidence accumulation, reflecting processing efficiency

Boundary separation (a): Decision threshold, indicating response caution

Non-decision time (t): Time required for processes outside decision-making (stimulus encoding and motor response execution)

Methods

Participants. We determined sample size using G*Power,[17] based on detecting a medium effect size (η^2 = 0.06, f = 0.25) with α = 0.05 and power = 0.8, requiring a minimum of 34 participants per experiment. A total of 112

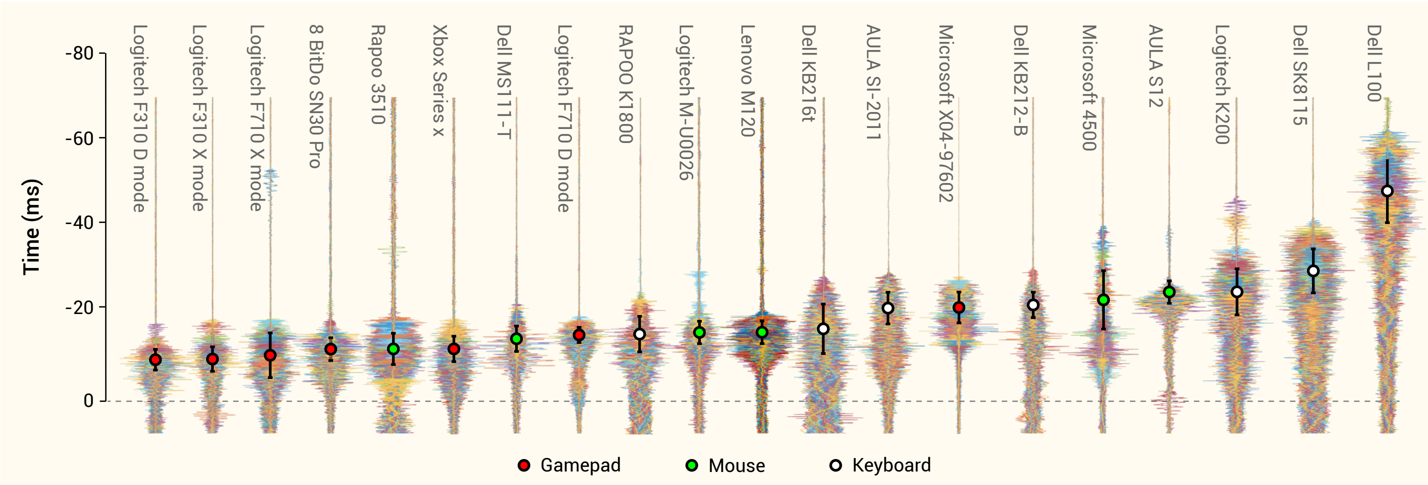


Figure 2. Recording delay distribution across 20 different reaction time devices measured with PsyRTDevice Each device's recording delay (ms) is represented by a violin plot showing the distribution of measurement latencies, with black dots indicating the mean value, error bars representing standard deviation, and colored lines depicting the overlay of 150 sound wave signals. Measurements are color-coded by input device type (yellow = gamepad, blue = mouse, white = keyboard). Devices are arranged in order of increasing mean recording delay from left to right, ranging from approximately 8 ms to over 50 ms. Notable variability is observed both between different device models and within individual devices (different modes). The dashed line indicates the internal response times recorded by the computer.

Table 3. Input devices tested using PsyRTDevice for latency evaluation.

Device type	Model	Wireless capability
Gamepad	Logitech F310 (X-mode)	No
	Logitech F310 (D-mode)	No
	Logitech F710 (X-mode)	Yes
	Logitech F710 (D-mode)	Yes
	8BitDo SN30 Pro	No
	Xbox Series X/S	No
	Microsoft X04-97602	No
Mouse	Rapoo 3510	Yes
	Dell MS111-T	No
	Logitech M-U0026	No
	Lenovo M120	No
	Microsoft 4500	No
	AULA S12	No
Keyboard	Rapoo K1800	Yes
	Dell KB216t	No
	AULA SI-2011	No
	Dell KB212-B	No
	Logitech K200	No
	Dell SK8115	No
	Dell L100	No

participants from Soochow University participated in the three experiments: IOR (n = 34; mean age 22.1 ± 2.3 years, 28 females); Stroop effect (n = 34; mean age 22.3 ± 2.4 years, 29 females); and cueing effect (n = 34; mean age 20.8 ± 2.0 years, 31 females). All participants were naive to the experimental purpose, had normal or corrected-to-normal vision, provided informed consent, and received compensation at 30 CNY per hour. The Academic Ethics Committee of Soochow University approved the study.

Equipment. The experiments were programmed using PsyBuilder[10] and MATLAB Psychtoolbox[11,12] and run on a computer with Ubuntu 18.04, an Intel Core i7-8700 processor, and an NVIDIA GTX-1060 graphics card. Stimuli appeared on a 27-inch ASUS PG279Q monitor (2560 × 1440 resolution, 120 Hz refresh rate, ULMB mode enabled).[18] Participants were seated 80 cm from the display in a dimly lit, quiet room with head position stabilized using a chinrest.

Design and procedure. The procedures for all three experiments are illustrated in Figure 3. Each experiment employed the following designs:

- IOR: A 2 (cue-target relationship: cued/uncued) × 2 (device type: gamepad/keyboard) within-subjects design with 60 trials per condition, totaling 240 trials across 10 blocks (5 per device type). Participants responded to target stimuli ("X" or "O") by pressing corresponding keys.
- Stroop: A 2 (Stroop consistency: consistent/inconsistent) × 2 (device type: gamepad/keyboard) mixed design with 48 trials per condition, totaling 192 trials across 4 blocks. Participants identified the color of target words by pressing corresponding keys.
- Attention cueing: A 2 (cue-target relationship: cued/uncued) × 2 (device type: gamepad/keyboard) within-subjects design with 80% valid cues. Each block contained 48 cued and 12 uncued trials. Participants completed 10 blocks (5 per device type), totaling 600 trials, responding to target stimuli ("X" or "O") with corresponding keys.

Experimental results. To assess device-related differences, we first confirmed the successful replication of the three cognitive effects: IOR (cued 493 ms; uncued 478 ms, $F(1, 33) = 32.21$, $p < .001$, $\eta_p^2 = .49$); Stroop effect (inconsistent 464 ms; consistent 447 ms, $F(1, 33) = 11.21$, $p < .01$, $\eta_p^2 = 0.25$); and attention cueing (uncued 497 ms; cued 441 ms, $F(1, 33) = 185.91$, $p < .001$, $\eta_p^2 = .85$).

Additionally, as Figure 4 shows, device type significantly influenced RTs, with the gamepad exhibiting shorter RTs than the keyboard: IOR (469 ms vs. 502 ms, $F(1, 33) = 10.45$, $p = .003$, $\eta_p^2 = .24$), Stroop (449 ms vs. 461 ms, $F(1, 33) = 2.13$, $p = .15$, $\eta_p^2 = .06$), and attention cueing (456 ms vs. 482 ms, $F(1, 33) = 18.23$, $p < .001$, $\eta_p^2 = .36$). HDDM analyses revealed that these differences were primarily attributable to shorter non-decision times with the gamepad (vs. the keyboard): IOR (242 ms vs. 270 ms, $F(1, 33) = 8.04$, $p = .008$, $\eta_p^2 = .20$), Stroop (234 ms vs. 252 ms, $F(1, 33) = 9.48$, $p = .004$, $\eta_p^2 = .22$), and attention cueing (243 ms vs. 276 ms, $F(1, 33) = 12.26$, $p = .001$, $\eta_p^2 = .27$).

Finally, no significant interactions emerged between device precision and cognitive effects, indicating that device precision did not systematically influence cognitive outcomes in either behavioral or HDDM analyses across all tested tasks (see Table 4).

DISCUSSION

The present study introduces PsyRTDevice, a novel sound-wave-based toolkit for evaluating RT device precision, and examines how device precision impacts both absolute RT measurements and relative cognitive effects. Our findings demonstrate that while device precision significantly affects absolute RT measurements, cognitive effects based on RT differences

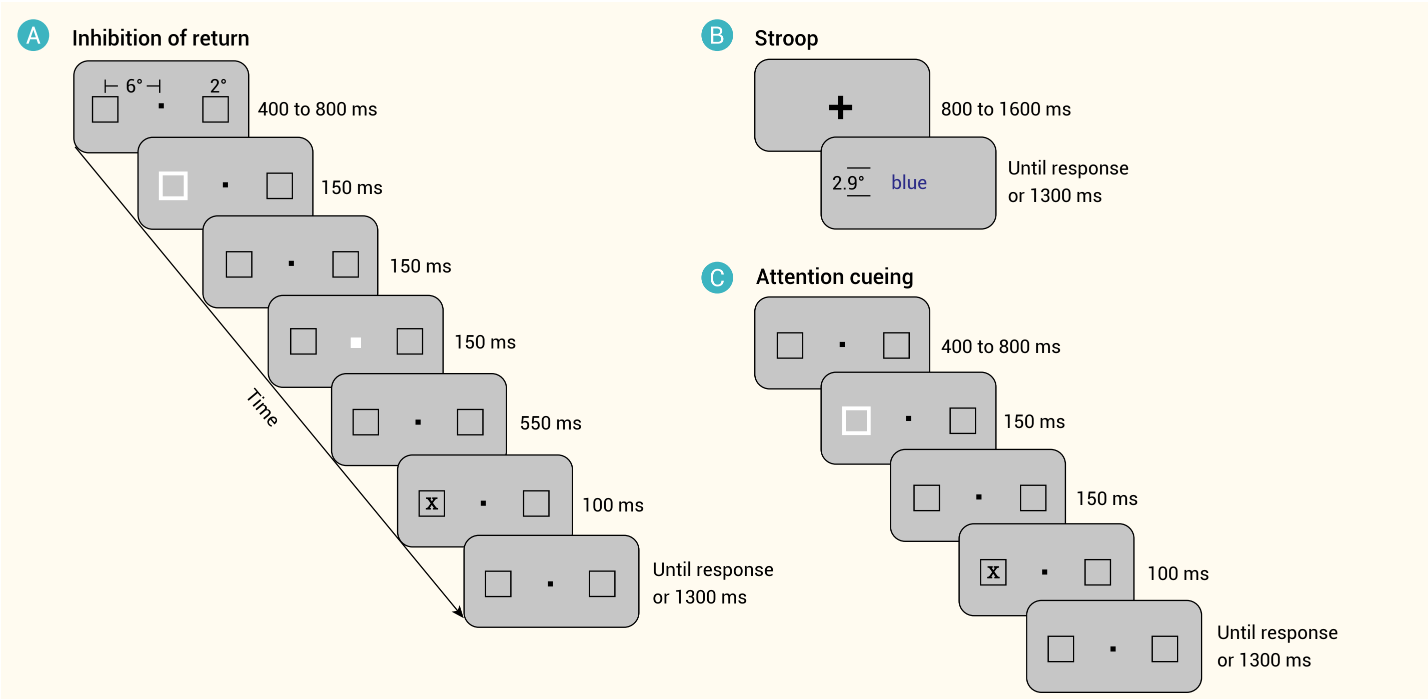


Figure 3. Procedures of the three cognitive tasks used to evaluate device performance (A) Inhibition of return task: A sequence of visual stimuli with centrally presented fixation point followed by a cue (white square), subsequent central fixation, and a target stimulus (X) requiring a response within 1300 ms. (B) Stroop task: Participants view a fixation cross followed by a color word (e.g., "blue") displayed in a congruent or incongruent color, requiring a response within 1300 ms. (C) Attention cueing task: A sequence beginning with a central fixation point, followed by a peripheral cue (white square), returning to central fixation, and a target (X) display requiring a response within 1300 ms.

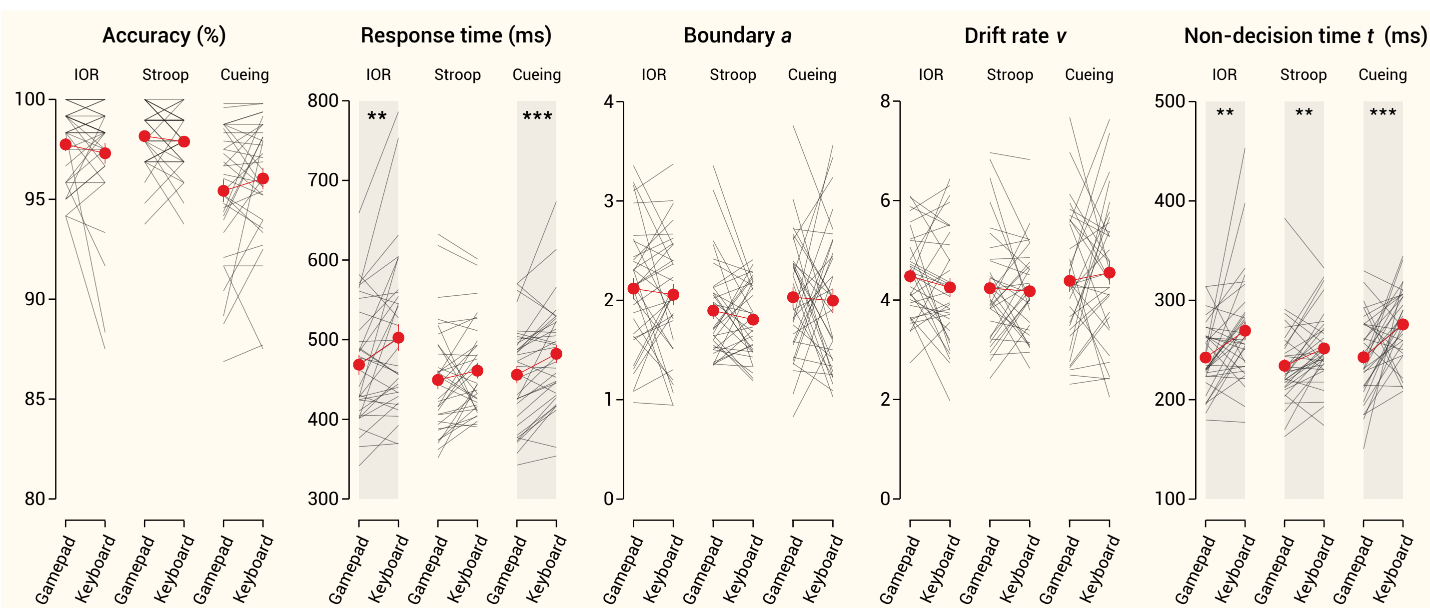


Figure 4. Comparison of behavioral performance and HDDM parameters between the most precise device (gamepad) and least precise device (keyboard) across three cognitive tasks There were five measurement categories: accuracy (%), response time (ms), boundary separation (*a*), drift rate (*v*), and non-decision time (*t*, ms). For each category, results are shown for three tasks: Inhibition of return (IOR), Stroop, and attention cueing. Gray lines connect individual participant data points between gamepad (left) and keyboard (right) conditions within each task. Red dots with error bars indicate the mean values with 95% confidence intervals. Significant differences between devices are indicated by asterisks ($*p < .05$, $**p < .01$, $***p < .001$), with shaded backgrounds highlighting these significant differences. Notably, response times and non-decision times show consistent significant differences across tasks, while accuracy, boundary separation, and drift rate show minimal device-dependent effects.

remain robust across devices with varying precision. These results have important implications for methodology selection, experimental design, and the interpretation of timing-sensitive research.

Methodological advancements in device precision testing

The PsyRTDevice toolkit addresses several limitations of previous approaches to measuring device precision. First, by leveraging standard audio hardware available in most computers, our method eliminates the need for specialized equipment while achieving sub-millisecond measurement accuracy. This approach reduces barriers to device evaluation compared to previous methods that required custom hardware,[3] complex setups,[5] or potentially risky device modifications.[7]

Second, the toolkit's streamlined implementation allows comprehensive precision evaluation of a single device (with 50 measurements) to be completed in just 10–15 minutes, making it practical for researchers to test multiple devices or routinely validate device precision before critical experiments. This efficiency addresses a significant barrier to widespread adoption of systematic precision testing in behavioral research.

Table 4. Statistics of interaction effects between device type and cognitive effect.

	Inhibition of return				Stroop				Attention cueing			
	RT	*a*	*v*	*t*	RT	*a*	*v*	*t*	RT	*a*	*v*	*t*
F	1.44	0.01	0.33	0.52	2.83	1.43	0.39	0.75	1.62	<0.01	0.51	0.53
p	0.24	0.97	0.57	0.48	0.10	0.24	0.54	0.39	0.21	0.95	0.48	0.47
η_p^2	0.04	0.00	0.01	0.02	0.08	0.04	0.01	0.02	0.05	<0.01	0.02	0.02

Third, unlike previous studies that typically focused on limited device types[4,5] or small samples of devices,[6] our evaluation of 20 diverse input devices provides a comprehensive comparison across device categories (gamepads, mice, and keyboards). This assessment revealed substantial variability in both average response delay (ranging from 7.58 ms to 47.46 ms) and measurement consistency (SDs ranging from 1.82 ms to 7.54 ms), highlighting the importance of informed device selection for timing-critical research.

While dedicated response boxes remain the gold standard for precision,[19] they are often impractical due to limited button configurations, higher costs, and reduced accessibility. The PsyRTDevice toolkit enables researchers to systematically evaluate more accessible alternatives like keyboards, mice, and gamepads, facilitating informed decisions based on specific experimental requirements rather than relying on assumptions about device precision.

Device precision on absolute RT measurements

Our experimental comparisons between the highest-precision device (Logitech F310 gamepad in D-mode) and lowest-precision device (Dell L100 keyboard) consistently demonstrated that device precision significantly affects absolute RT measurements. Across all three cognitive paradigms, the high-precision gamepad recorded shorter RTs than the keyboard: IOR (469 ms vs. 502 ms), Stroop (449 ms vs. 461 ms), and attention cueing (456 ms vs. 482 ms).

The HDDM analyses revealed that these device-related differences were primarily attributable to non-decision time parameters, which encompass both stimulus encoding and response execution processes. Given that stimulus parameters were identical across device conditions,[20,21] these differences likely stem from variations in response execution stages rather than perceptual encoding.[22,23] This suggests that the observed RT differences predominantly reflect inherent device response delays rather than cognitive processing variations.

These findings have important implications for research contexts where absolute RT values are critical, as alluded to in the introduction. For example, in studies involving electrophysiological measures like the LRP, which typically emerge approximately 200 ms before a behavioral response,[24-26] using high-latency devices could potentially distort the temporal relationship between neural activity and behavioral responses. When a device introduces a substantial delay (e.g., 50 ms), it complicates the attribution of neural components to specific cognitive processes, potentially obscuring the intrinsic relationship between response preparation and execution. Similarly, studies of visual perception phenomena like binocular rivalry[27,28] or precise RT modeling[3,13] may require careful consideration of device precision to ensure measurement validity.

Device precision on relative cognitive measurements

Our study shows that device precision did not significantly influence the magnitude of IOR, Stroop interference, or endogenous attention, indicating that the relative differences between experimental conditions remained stable across devices despite absolute RT differences.

This finding is noteworthy because it suggests that the relationship between device precision and cognitive effects is more complex than a simple linear accumulation of variance. If measurement error from the device simply added to participant response variability in a linear fashion, we would expect lower-precision devices to reduce statistical power for detecting cognitive effects.[1] However, our results suggest a non-linear interaction between device response variability and cognitive measurement variability.

One potential explanation is that human response variability is substantially larger than device variability, rendering the latter's contribution relatively negligible for between-condition comparisons. Within-subject or within-session designs may inherently control for device-related variance, as systematic device delays affect all conditions similarly, preserving relative differences even when absolute measurements differ.

These findings have important practical implications for experimental design and device selection. For studies primarily concerned with relative differences between conditions (e.g., many cognitive psychology experiments), device precision may be less critical than previously assumed, provided that the same device is used consistently across conditions being compared. However, researchers should still exercise caution when comparing results across studies using different input devices, as absolute RT values may differ substantially.

Limitations and future directions

Several limitations of the current study warrant consideration. First, while our device sample was diverse, it necessarily represents only a subset of available input devices. Future research should expand testing to include additional device categories (e.g., touchscreens, response boxes) and newer models as they become available.

Second, our experimental paradigms, while representative of common cognitive tasks, do not exhaust the full range of RT paradigms used in behavioral research. Tasks requiring extremely precise timing (e.g., millisecond-level discrimination) or those with smaller effect sizes might show different sensitivity to device precision.

Third, our methodology relies on acoustic signals from physical key presses, which may be less applicable to devices with minimal acoustic feedback (e.g., capacitive touchscreens). Future developments could incorporate alternative measurement approaches for such devices.

Future research directions include: (1) developing standardized device precision benchmarks to facilitate informed selection based on experimental requirements; (2) investigating how device precision interacts with special populations (e.g., older adults, clinical groups) who may show greater response variability; (3) exploring potential calibration methods to correct for device-specific delays in post-processing; and (4) examining how device precision affects more complex behavioral measures like sequential effects or RT distributions.

CONCLUSION

The PsyRTDevice toolkit provides an accessible, efficient, and precise method for evaluating RT device accuracy, addressing a long-standing methodological challenge in behavioral research. Our findings demonstrate that while device precision significantly affects absolute RT measurements, cognitive effects based on RT differences remain robust across devices with varying precision. This distinction has important implications for experimental design, interpretation, and cross-study comparisons.

We recommend that researchers: (1) systematically evaluate and report device precision, especially for studies where absolute RT values are critical; (2) maintain consistent device usage within experiments, particularly when making direct comparisons between conditions; and (3) consider device precision characteristics when interpreting results across studies using different input devices.

By facilitating more transparent reporting of device characteristics and a better understanding of how these characteristics influence behavioral measurements, the present work contributes to improving methodological rigor and replicability in RT-based research across psychological and behavioral sciences.

REFERENCES

1. Segalowitz S.J. and Graves R.E. (1990). Suitability of the IBM XT, AT, and PS/2 keyboard, mouse, and game port as response devices in reaction time paradigms. *Behav. Res. Methods Instrum. Comput.* 22:283–289. DOI:10.3758/BF03209817
2. Myors B. (1998). A simple graphical technique for assessing timer accuracy of computer systems. *Behav. Res. Methods Instrum. Comput.* 30:454–456. DOI:10.3758/bf03200679
3. Plant R.R., Hammond N. and Whitehouse T. (2003). How choice of mouse may affect response timing in psychological studies. *Behav. Res. Methods Instrum. Comput.* 35:276–284. DOI:10.3758/BF03202553
4. Shimizu H. (2002). Measuring keyboard response delays by comparing keyboard and joystick inputs. *Behav. Res. Methods Instrum. Comput.* 34:250–256. DOI:10.3758/BF03195452
5. Neath I., Earle A., Hallett D., et al. (2011). Response time accuracy in Apple Macintosh computers. *Behav. Res. Methods* 43:353–362. DOI:10.3758/s13428-011-0069-9
6. Bockes F., Wimmer R. and Schmid A. (2018). LagBox-Measuring the latency of USB-connected input devices. *CHI EA* 18:1-6. DOI:10.1145/3170427.3188632
7. Zhang Y. and Zhang M. (2010). Measure the time accuracy of response device: An accuracy yet easy-to-implement method. *J. Northeast Norm. Univ. Nat. Sci. Ed.* 42:66–70. https://qikan.cqvip.com/Qikan/Article/Detail?id=35396226
8. Van Drongelen W. (2007). Signal processing for neuroscientists (Academic Press). DOI:10.1016/B978-0-12-370867-0.X5000-1
9. Tang N., Li A.S., Chen B., et al. (2025). PsyRTDevice code. https://github.com/yangzhangpsy/PsyRTDevice.
10. Lin Z., Yang Z., Feng C., et al. (2022). PsyBuilder: An open-source, cross-platform graphical experiment builder for Psychtoolbox with built-in performance optimization. *Adv. Meth. Pract. Psychol. Sci.* 5:25152459211070573. DOI:10.1177/25152459211070573
11. Brainard D.H. (1997). The psychophysics toolbox. *Spat. Vis.* 10:433–436. DOI:10.1163/156856897x00357
12. Pelli D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* 10:437–442. DOI:10.1163/156856897x00366
13. Thapar A., Ratcliff R. and McKoon G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychol. Aging* 18:415–429. DOI:10.1037/0882-7974.18.3.415
14. Ratcliff R., Thapar A. and McKoon G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Percept. Psychophys.* 65:523–535. DOI:10.3758/BF03194580
15. Ratcliff R., Gomez P. and McKoon G. (2004). A diffusion model account of the lexical decision task. *Psychol. Rev.* 111:159–182. DOI:10.1037/0033-295X.111.1.159
16. Pan W., Geng H., Zhang L., et al. (2025). dockerHDDM: A user-friendly environment for Bayesian hierarchical drift-diffusion modeling. *Adv. Meth. Pract. Psychol. Sci.* 8:25152459241298700. DOI:10.1177/25152459241298700
17. Faul F., Erdfelder E., Buchner A., et al. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* 41:1149-1160. DOI:10.3758/BRM.41.4.1149
18. Zhang G.L., Li A.S., Miao C.G., et al. (2018). A consumer-grade LCD monitor for precise visual stimulation. *Behav. Res. Methods* 50:1496–1502. DOI:10.3758/s13428-018-1018-7
19. Li X., Liang Z., Kleiner M., et al. (2010). RTbox: A device for highly accurate response time measurements. *Behav. Res. Methods* 42:212–225. DOI:10.3758/BRM.42.1.212
20. Weindel G., Gajdos T., Burle B., et al. (2021). The decisive role of non-decision time for interpreting the parameters of decision making models. *PsyArXiv* preprint. DOI:10.31234/osf.io/gewb3
21. Voss A., Rothermund K. and Voss J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Mem. Cogn.* 32:1206–1220. DOI:10.3758/BF03196893
22. Ratcliff R. and McKoon G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* 20:873–922. DOI:10.1162/neco.2008.12-06-420
23. Ratcliff R. (2006). Modeling response signal and response time data. *Cogn. Psychol.* 53:195–237. DOI:10.1016/j.cogpsych.2005.10.002
24. Miller J., Patterson T. and Ulrich R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology* 35:99–115. DOI:10.1111/1469-8986.3510099
25. Kojouharova P., Nagy B., Czigler I., et al. (2023). Mechanisms of spatial contextual cueing in younger and older adults. *Psychophysiology* 60:e14361. DOI:10.1111/psyp.14361
26. Jaśkowski P., Kurczewska M., Nowik A., et al. (2007). Locus of the intensity effect in simple reaction time tasks. *Percept. Psychophys.* 69:1334–1343. DOI:10.3758/bf03192950
27. Levelt W.J. (1965). On binocular rivalry. (Doctoral dissertation, Van Gorcum Assen). https://pure.mpg.de/rest/items/item_77195/component/file_2424565/content
28. Blake R. and Logothetis N.K. (2002). Visual competition. *Nat. Rev. Neurosci.* 3:13–21. DOI:10.1038/nrn701

FUNDING AND ACKNOWLEDGMENTS

This research was supported by Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project (2025ZD0215700), National Natural Science Foundation of China (32500932; 32171049), Social Science Foundation of Jiangsu Province (22JYB015), Natural Science Foundation of Jiangsu Province (BK20240791; BK20250796), and Jiangsu Funding Program for Excellent Postdoctoral Talent (2025ZB642). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

N.T. contributed to investigation, software, and writing (original draft, review, and editing). A.L. was responsible for funding acquisition, project administration, validation, and writing (review and editing). B.C. contributed to data curation, formal analysis, investigation, software, and visualization. Y.Q. participated in data curation, investigation, software, and visualization. Z.L. contributed to conceptualization, funding acquisition, supervision, resources, and writing (review and editing). Y.Z. contributed to conceptualization, formal analysis, funding acquisition, investigation, resources, supervision, and writing (original draft, review, and editing). All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

This study was approved by the institutional review board of the ethics committee of Soochow University (SUDA20240119H01) and was conducted in accordance with the Declaration of Helsinki. All study participants provided written informed consent.

DATA AND CODE AVAILABILITY

The code that supports the findings of this study is available at https://github.com/yangzhangpsy/PsyRTDevice.[9] Data are available from the corresponding author upon reasonable request.